

“十二五”规划大学教材

数据库 原理与应用

S HUIJUKUYUANLI
YUYINGYONG

滕以芳 庞 颖 谢志伟◎主编



东北师范大学出版社
NORTHEAST NORMAL UNIVERSITY PRESS

数据库原理与应用

滕以芳 庞 颖 谢志伟 主编

东北师范大学出版社

长 春

图书在版编目 (CIP) 数据

数据库原理与应用 / 滕以芳, 庞颖, 谢志伟主编. —长春 :
东北师范大学出版社, 2012. 4
ISBN 978-7-5602-8242-8

I. ①数… II. ①滕… ②庞… ③谢… III. ①数据库
—高等学校—教材 IV. ①F234. 3

中国版本图书馆 CIP 数据核字 (2012) 第 072447 号

☐责任编辑: 王春彦 ☐封面设计: 魔弹文化
☐责任校对: 赵世鹏 ☐责任印制: 张允豪

东北师范大学出版社出版发行
长春净月经济开发区金宝街 118 号 (邮政编码: 130117)
电话: 010—82920765
传真: 010—82920765

网址: <http://www.nenup.com>

电子函件: sdcbs@mail.jl.cn

北京魔弹文化制版

北京市彩虹印刷有限责任公司印装

2012 年 4 月第 1 版 2012 年 4 月第 1 次印刷

幅面尺寸: 185 mm×260 mm 印张: 16.25 字数: 416 千字

定价: 29.90 元

前言



数据库技术从诞生到现在，在不到半个世纪的时间里，逐渐形成了坚实的理论基础、成熟的商业产品和广泛的应用领域，使得数据库成为一个研究者众多且被广泛关注的研究领域。

数据库技术是计算机科学中的一个非常重要的部分，作为数据管理最有效的手段，它极大地促进了计算机应用的发展。数据库技术是计算机系统与应用系统的核心技术，它是信息管理系统开发的重要基础。数据库技术以及数据库的应用也正以日新月异的速度发展，因此作为现代的大学生，特别是计算机专业的学生，学习和掌握数据库知识是非常必要的。

本书是面向计算机专业及其它专业的学生学习数据库知识而编写的大学教材。其特点是内容全面，既包括数据库的基础理论知识，又包括数据库的前端和后端的应用技术。SQL Server 是目前广泛应用的后台数据库管理系统，Visual Basic 也是使用非常普遍、方便的可视化编程环境。将这些内容结合在一本书中，可以使读者系统、全面地学习数据库系统的整体概念和应用技术。在介绍数据库理论时，本书特别加强了解决实际问题的内容，包括在数据库管理系统中对索引的管理方法以及如何构建提高数据查询效率的索引，如何编写带参数的存储过程以及如何自定义函数以实现复杂的数据查询功能等。在实现数据完整性约束方面，本书除了介绍常用的完整性约束方法之外，还介绍了实现复杂的数据完整性约束的方法——触发器。



本书在介绍数据库原理理论概念的基础上,将理论与实际相结合作为重点,力求使读者通过学习本书,能够完成系统的需求分析、数据库的概念设计、逻辑设计、物理设计、安全设计、数据库实现等设计开发全过程,最终实现一个较为完整的、能够反映应用需求的管理信息系统,以此来培养和提高读者的数据库开发能力和水平。

本书主编为鹤岗师范高等专科学校滕以芳,编写本书的第一章至第四章;鹤岗师范高等专科学校庞颖编写本书的第五章至第七章;黑龙江农垦职业学院谢志伟编写本书的第八章至第十章。

本书附有习题,以加强读者对所学知识的理解,及时掌握重点和难点。本书结合数据库设计的过程,在相应的章节中编写了分析设计题,以便读者掌握各个设计环节的基本方法、基本技能和设计要点。本书还大量增加了图表,以提高记忆力和理解能力。

本书的主要特点如下:

1. 全面介绍数据库技术的发展、基本原理与基本概念。
2. 应用学习目标、小结强调概念的理解和分析方法之间的关系,以此来概括该章的知识体系结构。
3. 系统地介绍了数据库系统的开发流程和方法。
4. 完整地介绍了中小型数据库管理系统 Access,以及适用于网络环境下进行数据库应用开发的 Microsoft SQL Server 的功能及具体的开发方法。
5. 应用两个具有代表性的案例把书中的数据库原理与开发技术进行串联,以此来介绍应用程序开发的全过程。
6. 本书以通俗易懂的语言、图片和图表,由浅入深地介绍了数据库的基本概念与操作、创建表、表的编辑操作、查询应用等,为读者轻松掌握数据库应用系统的使用开发提供了一个入门的捷径,并且在后面单独开辟一章应用实例,完整地向读者介绍了 SQL Serve 数据库管理系统平台及如何应用 Access 开发数据库应用系统,其中的实例接近生活易于接受。

本书是作者多年从事数据库教学的经验和体会的总结。在本书的编写过程中参考了国内外同行的研究成果,在此一并致谢。

编者



目 录 CONTENTS

第一章 绪 论	1
1.1 数据管理技术的产生和发展	2
1.2 数据库的基本概念	6
1.3 数据模型	8
1.4 数据库系统的结构	18
1.5 数据库管理系统	20
第二章 关系数据库	25
2.1 关系模型概述	26
2.2 关系定义和性质	27
2.3 关系模型的完整性约束	31
2.4 关系代数	34
第三章 关系数据库标准语言 SQL	42
3.1 SQL 概述	43
3.2 数据定义	45
3.3 查询	49
3.4 数据更新	61
3.5 视图	64
3.6 数据的控制	67
3.7 嵌入式 SQL	68
第四章 规范化理论	73
4.1 问题的提出	74
4.2 函数依赖	76
4.3 规范化	83
第五章 数据库设计	97
5.1 数据库设计概述	98

5.2 需求分析	103
5.3 概念结构设计	107
5.4 逻辑结构设计	113
5.5 物理结构设计	116
5.6 数据库的实施和维护	119
第六章 数据库编程	122
6.1 嵌入式 SQL	123
6.2 存储过程	132
6.3 ODBC 编程	137
第七章 数据库完整性和安全性	149
7.1 数据库完整性	150
7.2 数据库安全性	157
7.3 SQL Server 的完整性控制和安全控制	166
第八章 数据库恢复和并发控制	173
8.1 事务	174
8.2 数据库恢复技术	175
8.3 并发控制	182
第九章 数据库新技术	193
9.1 数据库技术发展概述	194
9.2 数据仓库与数据挖掘	196
9.3 数据库技术与其他相关技术的结合	205
9.4 数据库建设中应注意的问题	218
第十章 综合应用编程示例	219
10.1 学生信息管理系统	220
10.2 酒店库存管理系统	233
参考文献	252

第一章 绪论



本章概述

数据库的基本概念,并通过对数据管理技术发展的三个阶段的介绍,阐述了数据库技术产生和发展的背景,也说明了数据库系统的优点。

数据模型是数据库系统的核心和基础。本章介绍了组成数据模型的三个要素和概念模型。概念模型也称信息模型,用于信息世界的建模,E-R模型是这类模型的典型代表,E-R模型简单、清晰,应用十分广泛。

数据模型的发展经历了非关系模型(层次模型、网状模型)、关系模型,正在走向面向对象模型。

数据库系统中,数据具有三级模式结构的特点,由外模式、模式、内模式以及外模式/模式映像、模式/内模式映像组成。三级模式结构使数据库中的数据具有较高的逻辑独立性和物理独立性。一个数据库系统中,只有一个模式,一个内模式,但有多个外模式。因此,模式/内模式映像是唯一的,而每一个外模式都有自己的外模式/模式映像。

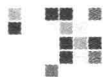
最后介绍了数据库管理系统的工作模式、功能与组成。

学习这一章应把注意力放在掌握基本概念和基本知识方面,为进一步学习下面章节打好基础。



学习目标

1. 了解数据、信息、数据处理、数据管理的概念和联系。
2. 掌握数据库、数据库管理系统、数据库系统的内容和含义。
3. 了解数据模型的组成要素、分类和组成要素。
4. 掌握数据库系统的三级模式结构。



近年来,随着计算机技术与网络技术的飞速发展,大大改变了人们的生活面貌,促进了社会的发展。信息技术在各个领域的深度融合推动了数据库技术在各个领域的广泛应用。从小型事务处理到大型信息系统,从联机事务处理到联机分析处理,从一般企业管理到计算机辅助设计与制造、电子政务、电子商务、地理信息系统等,数据库系统已经渗透到人们工作和生活的方方面面。因此,数据库技术已成为当今计算机信息系统的核心技术,是计算机技术和应用发展的基础。经过 40 多年的发展,它已形成了较为完整的理论体系和实用技术。

本章首先回顾了数据管理技术的发展过程,然后介绍数据库的基本概念和数据模型,最后阐述数据库系统的体系结构。

1.1 数据管理技术的产生和发展

数据库技术是应数据管理任务的需要而产生的。数据的处理是指对各种数据进行收集、存储、加工和传播的一系列活动的总和。数据管理则是指对数据进行分类、组织、编码、存储。检索和维护,它是数据处理的中心问题。

研制计算机的初衷是进行复杂的科学计算,随着计算机技术的发展和数据处理的需求,需要对大量的数据进行处理来得到有价值的信息,但由于数据量大,数据结构复杂,数据之间有复杂的逻辑联系,所以需要有一个通用、高效而又使用方便的管理软件,以便把数据有效地管理起来,这样,数据管理技术也就相应产生了。

随着计算机硬件和软件的发展,数据管理技术大致经历了三个阶段:人工管理阶段、文件系统阶段和数据库系统阶段。

1.1.1 人工管理阶段

20 世纪 50 年代中期以前,计算机技术相对落后,计算机主要用于科学计算。硬件方面,计算机的外存只有磁带、卡片、纸带,没有磁盘等直接存取的存储设备,存储量非常小。软件方面,只有汇编语言,没有操作系统和管理数据的软件。数据处理的方式是批处理,也即机器一次处理一批数据,直到运算完成为止,然后才能进行另外一批数据的处理,中间不能被打断,原因是当时的外存如磁带、卡片等只能顺序输入。

这个阶段数据管理的特点如下:

(1)数据不共享。数据是面向应用程序的,一组数据只能对应一个应用程序。多个应用程序若涉及某些相同的数据,也必须各自定义,无法相互利用,互相参照。数据不但高度冗余,而且不能共享。

(2)数据不保存。在需要计算时,利用卡片、纸带等将数据输入,经过运算得到运算结果,数据处理的过程就结束了。由于当时计算机主要用于科学计算,一般不需要将数据长期保存,只是在计算某一课题时将数据输入,用完就撤走。不仅对用户数据如此处置,对系统软件有时也是这样。

(3)数据不具有独立性。数据是程序的组成部分,即程序和数据是一个不可分割的整体,数据和程序同时提供给计算机运算使用。对数据进行管理,就像现在的操作系统以目

录、文件的形式管理数据。程序员不仅要知道数据的逻辑结构,也要规定数据的物理结构,程序员对存储结构、存取方法及输入/输出的格式有绝对的控制权,要修改数据必须修改程序。例如,要对 100 组数据进行同样的运算,就要给计算机输入 100 个独立的程序,因为数据无法独立存在。



图 1.1 人工管理阶段程序与数据之间的关系

(4)没有专用的软件对数据进行管理。数据需要由应用程序自己管理,没有相应的软件系统负责数据的管理工作。每个应用程序不仅要规定数据的逻辑结构,而且要设计物理结构,包括存储结构、存取方法、输入方式等。因此,程序员负担很重。在人工管理阶段,程序与数据之间的关系如图 1.1 所示。

1.1.2 文件系统阶段

数据管理发展到文件系统阶段,是在 20 世纪 50 年代后期到 60 年代中期。此时的计算机不仅用于科学计算,还大量应用于数据管理在硬件方面,有了磁盘、磁鼓等直接存储设备;在软件方面出现了高级语言和操作系统,且操作系统中有了专门管理数据的软件,一般称之为文件系统。从处理方式上讲,不仅有了文件批处理,而且能够联机实时处理,联机实时处理是指在需要的时候随时从存储设备中查询、修改或更新,因为操作系统的文件管理功能提供了这种可能。

用文件系统管理数据的特点如下:

(1)文件组织多样化。文件有顺序文件、链接文件、索引文件等,因而对文件的记录可顺序访问,也可随机访问,更便于存储和查找数据。但文件之间相互独立、缺乏联系。数据之间的联系要通过程序去构造。

(2)程序与数据之间有一定独立性。由文件系统进行数据管理,程序和数据之间由软件提供的存取方法进行转换,数据存储发生变化不一定影响程序的运行,既可大大节省维护的工作量,又可减轻程序员的负担。

(3)数据可长期保存。数据可以长期保留在外存上反复处理,即可以经常对数据文件进行查询、修改、插入和删除等操作。

(4)由文件系统对数据的存取进行管理。文件系统把数据组织成相互独立的数据文件,利用“按名访问,按记录存取”的管理技术,对文件进行修改、插入和删除等操作。因此,程序员只与文件名打交道,不必明确数据的物理存储结构,大大减轻了程序员的负担。

与人工管理阶段相比,文件系统阶段对数据的管理有了很大的进步,但一些根本性问题仍没有彻底解决,主要表现在以下三方面:

(1)数据联系弱。不同的数据文件之间相互独立,缺乏联系。虽然某些数据之间存在着紧密的联系,但是由于实现的复杂性,很少在系统中提供数据之间的紧密联系。

(2)数据独立性差。数据和程序相互依赖,一旦改变数据的逻辑结构,必须修改相应的应用程序。而应用程序发生变化,如改用另一种程序设计语言来编写程序,有时也需修改数据结构。因此,数据和程序之间缺乏独立性。

(3)数据冗余度大,一致性差。在文件系统中,一个文件基本上对应于一个应用程序。如果多个应用程序的数据完全相同,可以以文件为单位实现共享。当不同的应用程序具有部分相同的数据时,仍需建立各自的独立数据文件,而不能共享相同的数据。因此,数据的冗余度大,空间浪费严重。同时由于相同数据的重复存储、各自管理,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。

在文件系统阶段,程序与数据之间的关系如图 1.2 所示。

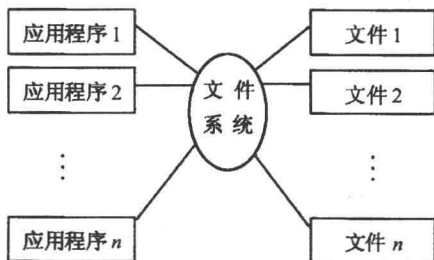


图 1.2 文件系统阶段程序与数据之间的关系

1.1.3 数据库系统阶段

20 世纪 60 年代后期,计算机用于管理的规模更为庞大,应用越来越广泛,数据量急剧增长,同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。硬件已有了大容量的磁盘,硬件价格下降。在处理方式上,联机实时处理要求更多,并开始提出和考虑分布处理。在这样的背景下,以文件系统作为数据管理手段已经不能满足应用的需求,为解决多用户、多应用共享数据的需求,出现了统一管理数据的专门软件系统——数据库系统。

从文件系统到数据库系统,是数据管理技术的一个飞跃。数据库系统管理数据的特点如下:

(1)数据共享性高、冗余小,易扩充。数据库系统从全局角度看待和描述数据,数据不再面向某个应用程序而是面向整个系统,因此数据可以被多个用户、多个应用共享使用。这样便减小了不必要的数据冗余,节约存储空间,同时也避免了数据之间的不相容性与不一致性。

由于数据面向整个系统,是有结构的数据,不仅可被多个应用共享使用,而且容易增加新的应用,这就使得数据库系统弹性大,易于扩充,可以适应各种用户的要求。

(2)数据独立性高。数据独立性是指数据库中的数据与应用程序之间依赖性弱,独立性强,包括数据的物理独立性和数据的逻辑独立性。

所谓物理独立性就是指用户的应用程序与存储在磁盘数据库中的数据是相互独立的。或者说,数据在数据库中怎样存储是由数据库系统管理的,应用程序不需要了解,应用程序要处理的只是数据的逻辑结构,即使数据的物理存储改变了,应用程序也不用改变。

所谓逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。或者说,数据的逻辑结构改变了,应用程序也可以不变。

由于数据与程序的独立,把数据的定义从程序中分离出去,加上数据的存取又由数据库系统负责,从而简化了应用程序,大大减少了应用程序的维护和修改的工作量。

(3)数据结构化。数据结构化是数据库系统与文件系统的根本区别。在文件系统阶段,只考虑了同一文件记录内部数据项之间的联系,而不同文件的记录之间是没有联系的,也就是说,从整体上看数据是无结构的。在数据库中,实现了整体数据的结构化,把文



件系统中的简单的记录结构变成了记录和记录之间的联系所构成的结构化数据。在描述数据的时候,不仅要描述数据本身,还要描述数据之间的联系。

(4)有统一的数据控制功能。数据库为多个用户和应用程序所共享,对数据的存取往往是并发的,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的同一个数据,为确保数据库数据的正确有效和数据库系统的有效运行,数据库系统提供下述四方面的数据控制功能。

①数据的完整性(integrity)控制。系统通过设置一些完整性规则以确保数据的正确性、有效性和相容性称做数据的完整性。完整性控制将数据控制在有效的范围内,或保证数据之间满足一定的关系。其中,有效性是指数据是否在其定义的有效范围内,如月份只能用1~12之间的正整数表示。正确性是指数据的合法性,如年龄属于数值型数据,只能含0,1,...,9,不能含字母或特殊符号。相容性是指表示同一事实的两个数据应相同,否则就不相容,如一个人不能有两个性别。

②并发(concurrency)控制。多用户同时存取或修改数据时,可能会发生相互干扰而提供给用户不正确的数据,并使数据库的完整性受到破坏,因此必须对多用户的并发操作加以控制和协调。

③数据的安全性(security)控制。保护数据以防止不合法使用数据造成数据的泄露和破坏,保证数据的安全和机密,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称做数据的安全性。例如,系统提供口令检查或其他手段来验证用户身份,防止非法用户使用系统;也可以对数据的存取权限进行限制,只有通过权限检查后才能执行相应的操作。

④数据恢复(recovery)。计算机系统出现各种故障是很正常的,数据库中的数据被破坏、丢失也是可能的。当数据库被破坏或数据不可靠时,系统应有能力将数据库从错误状态恢复到最近某一时刻的正确状态。

在数据库系统阶段,程序与数据之间的关系如图1.3所示。

从文件系统发展到数据库系统是信息处理领域的一个重大变化。在文件系统阶段,人们对信息处理方式关心的中心问题是应用系统功能的设计,程序设计处于主导地位,数据只起着服从程序设计需要的作用。在数据库系统管理阶段,人们关心的问题是数据结构的设计,它是整个数据库的应用的核心,而数据库应用的设计则处于以既定的数据结构为基础的外围地位。

从数据库技术的发展过程和演变趋势来看,数据库系统本身也在不断发展,从最初的层次数据库系统、网状数据库系统,向关系型数据库系统、关系对象数据库系统、对象数据库系统发展。本书重点介绍关系数据库系统。

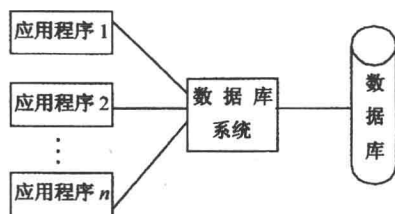


图 1.3 数据库系统阶段程序与数据之间的关系



1.2 数据库的基本概念

下面介绍一些数据库常用的术语和基本概念。

1.2.1 数据

所谓数据,就是数据库中存储的基本对象。它有多种表现形式,在大多数人的头脑中,数据就是数字,其实数字只是最简单的一种数据,这是数据的一种传统和狭义的理解。广义地理解,数据的种类很多,文字、图形、图像、声频、视频等,这些都可以作为数据,经过数字化后存入计算机。可以对数据做如下定义:数据是指描述事物的符号记录,这些符号可以是文字、图形、图像、声频、视频等。

数据的含义称为数据的语义,数据与其语义是不可分的。例如,学生档案表中有一个记录的描述如下:

(081637121,黎明,男,1987—5—11,计算机系)

这里的学生记录就是数据。对于这条记录,了解其含义后将得到如下信息:学号 081637121,姓名黎明,男,出生于 1987 年 5 月 11 日,在计算机系读书;而不了解含义的人则无法得到有用信息。可见,数据的形式本身还不能完全表达其内容,需要经过语义解释,数据与其语义是不可分的。

1.2.2 数据库

数据库(Database)是按照数据结构来组织、存储和管理数据的仓库。换句话说,数据库是存放数据的仓库。只不过这个仓库是在计算机存储设备上,而且数据是按照一定的格式存放的。它产生于距今五十年前,随着信息技术和市场的发展,特别是 20 世纪 90 年代以后,数据管理不再仅仅是存储和管理数据,而转变成用户所需要的各种数据管理的方式。数据库有很多种类型,从最简单的存储有各种数据的表格到能够进行海量数据存储的大型数据库系统都在各个方面得到了广泛的应用。

在经济管理的日常工作中,常常需要把某些相关的数据放进这样的“仓库”,并根据管理的需要进行相应的处理。例如,企业或事业单位的人事部门常常要把本单位职工的基本情况(职工号、姓名、年龄、性别、籍贯、工资、简历等)存放在表中,这张表就可以看成一个数据库。有了这个“数据仓库”我们就可以根据需要随时查询某职工的基本情况,也可以查询工资在某个范围内的职工人数等等。这些工作如果都能在计算机上自动进行,那我们的人事管理就可以达到极高的水平。此外,在财务管理、仓库管理、生产管理中也需建立众多的这种“数据库”,使其可以利用计算机实现财务、仓库、生产的自动化管理。

在科学技术飞速发展的今天,人们的视野越来越广,数据量急剧增加。过去人们把数据存放在文件柜里,现在人们借助计算机和数据库技术科学地保存和管理大量复杂的数据,以便能方便而充分地利用这些宝贵的信息资源。

数据库是一个长期存储在计算机内、有组织的、可共享的、统一管理的数据集合。数据库中的数据按一定的数据模型组织、描述和储存,具有较小的冗余度、较高的数据独立

性和易扩展性,可以为各种用户共享。

数据的长期存储、有组织和可共享是数据库的三个基本特点。

1.2.3 数据库管理系统

操纵和管理数据库的软件系统就是数据库管理系统(DBMS),它由一组计算机程序构成,管理并控制数据资源的使用。在计算机软件系统的体系结构中,数据库管理系统位于用户和操作系统之间。

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。数据库管理系统使用户能方便地定义数据和操纵数据,并能够保证数据的安全性、完整性,以及多用户对数据的并发使用及发生故障后的系统恢复。

1.2.4 数据库系统

在计算机系统中引入数据库后的系统就叫数据库系统。数据库系统不但能够实现有组织地、动态地存储大量相关的数据,而且为数据处理和信息资源共享提供了便利条件。数据库系统主要由四部分组成:软件系统、硬件系统、数据库和用户。

(1)软件系统主要包括数据库管理系统及其开发工具、操作系统(OS)和应用系统等。数据库系统的核心软件是数据库管理系统 DBMS 和数据库应用系统 DBAS(Database Application System)。

(2)硬件系统是指运行数据库系统所需要的硬件设备。任何一个计算机系统都需要有中央处理器、存储器和输入/输出设备等硬件。一个数据库系统需要有足够快的处理器来处理数据,以便快速响应用户的数据处理和数据检索请求,同时还需要有足够容量的内存与外存来存储大量的数据。对于分布式数据库系统,还需要有网络通信设备的支持。

(3)数据库。在一个数据库系统中,可以根据需要创建多个数据库,并且数据库中的数据通常可以被多个用户所共享。

(4)用户是指使用和管理数据库的人,他们可以对数据库进行存储、维护和检索等操作。数据库系统中用户可分为三类:终端用户、应用程序员和数据库管理员。

终端用户主要是使用数据库的各级管理人员、工程技术人员、科研人员,一般为非计算机专业人员。应用程序员负责为终端用户设计和编制应用程序,以便终端用户对数据库进行存/取操作。

数据库管理员(Database Administrator, DBA)是指对数据库进行设计、维护和管理的专业人员。DBA 是数据库所属单位的代表。一个单位决定开发一个数据库系统时,首先就应确定 DBA 的人选。DBA 不仅应当熟悉系统软件,还应熟悉本单位的业务工作。DBA 应自始至终参加整个数据库系统的研制开发工作,开发成功后,DBA 将全面负责数据库系统的管理、维护和正常使用。其主要职责如下:

(1)定义数据的安全性和完整性,负责分配用户对数据库的使用权限和口令管理。

(2)参与数据库设计的全过程,决定数据库的结构和内容。

(3)监督控制数据库的使用和运行,改进和重新构造数据库系统。当数据库受到破坏时,应负责恢复数据库;当数据库的结构需要改变时,完成对数据结构的修改。



因此, DBA 不仅要有较高的技术水平, 还应具有了解和阐明管理要求的能力。特别对于大型数据库系统, DBA 极为重要。对于常见的微机数据库系统, 通常只有一个用户, 常常不设 DBA, DBA 的职责由应用程序员或终端用户代替。在不引起混淆的情况下, 数据库系统可以简称为数据库。数据库系统的组成如图 1.4 所示。

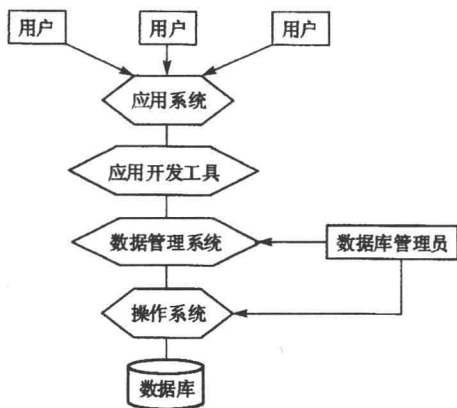


图 1.4 数据库系统的组成

1.3 数据模型

数据(data)是描述事物的符号记录。模型(Model)是现实世界的抽象。数据模型(Data Model)是数据特征的抽象, 是数据库管理的教学形式框架。数据库系统中用以提供信息表示和操作手段的形式构架。数据模型包括数据库数据的结构部分、数据库数据的操作部分和数据库数据的约束条件。由于计算机不可能直接处理现实世界中的具体事物, 所以人们必须事先把具体事物转换成计算机能够处理的数据。也就是首先要数字化, 把现实世界中具体的人、物、活动、概念用数据模型这个工具来抽象、表示和处理。通俗地讲, 数据模型是现实世界的模拟。现有的数据库系统都是基于某种数据模型的。

数据模型应满足三个方面的要求: 一是能比较真实地模拟现实世界; 二是容易为人所理解; 三是便于在计算机上实现。一种数据模型要很好地满足这三方面的要求, 在目前尚很困难。在数据库系统中应针对不同的使用对象和应用目的, 采用不同的数据模型。数据模型是数据库系统的核心和基础。

1.3.1 数据模型

数据模型是指数据库中数据的存贮和组织方式, 即如何表示实体以及实体之间的联系。数据模型是数据库系统的核心和基础, 了解数据库的数据模型特征, 可以帮助用户在建立和配置数据库时确定合理的系统应用结构, 在使用中灵活高效地发挥不同数据库系统的优势。

目前, 数据库领域最常用是数据模型主要有三种, 它们分别是层次模型(Hierarchical Model)、网状模型(Network Model)和关系模型(Relational Model)。随着数据库理论与

实践的不断发展,对象关系模型(Object Relational Model)、面向对象模型(Object Oriented Model)等正处于不断发展和完善之中。

在三种主要的逻辑数据模型中,层次模型和网状模型统称为格式化模型,也称为非关系模型。非关系模型的数据库系统在 20 世纪 70 年代至 80 年代初占据主导地位,现在已被关系模型的数据库系统所取代。但在欧美等国家,一些早期开发的应用系统仍在使用非关系模型的数据库系统。

20 世纪 80 年代以来,面向对象的方法和技术在计算机各个领域,包括程序设计语言、软件工程、信息系统设计、计算机硬件设计等各方面都产生了深远的影响,也促进了数据库中面向对象数据模型的研究和发展。许多关系数据库厂商为了支持面向对象数据模型,对关系模型做了扩展,从而产生了对象关系数据模型。

关系模型是目前使用最广泛的数据模型,占据数据库的主导地位。

在非关系模型中,实体用记录表示,实体的属性对应记录的数据项。实体之间的联系为记录之间的联系。

非关系模型中数据结构的单位是基本层次联系。所谓基本层次联系是指两个记录及它们之间的一对多(包括一对一)联系,基本层次联系如图 1.5 所示。图 1.16 中 R_i 位于联系 L_{ij} 的始点,称为双亲结点, R_j 位于联系 L_{ij} 的终点,称为子女结点。

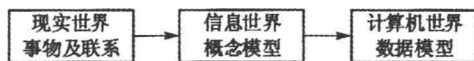


图 1.5 基本层次联系

1. 层次模型

层次数据模型是数据库系统中最早出现的数据模型。层次数据库系统的典型代表是 IBM 公司于 1968 年推出的第一个大型商用数据库管理系统——IMS(Information Management System)信息管理系统。

层次模型用树形结构表示实体及其之间的联系。树中每一个节点代表一个记录类型,树状结构表示实体型之间的联系。在一个层次模型中的限制条件是:有且仅有一个节点,无父节点,此节点为树的根;其他节点有且仅有一个父节点。该模型适合描述现实世界原本具有层次关系的系统,如一个系有若干专业和教研室,一个专业有若干班级,一个班级有若干学生,一个教研室有若干教师,其层次模型如图 1.6 所示。

层次模型的基本特点:

- (1)有且仅有一个节点无双亲结点,称其为根结点。
- (2)根以外的其他结点有且只有一个双亲结点。
- (3)记录之间的联系通过指针实现,查询效率高。

层次模型可以直接方便地表示一对一联系和一对多联系,但不能直接表示多对多联系。

层次模型的实际存储数据由链接指针来体现联系。在查找层次模型的任何一个记录值时,必须从根结点出发,按其路径查看才能找出它的全部意义。

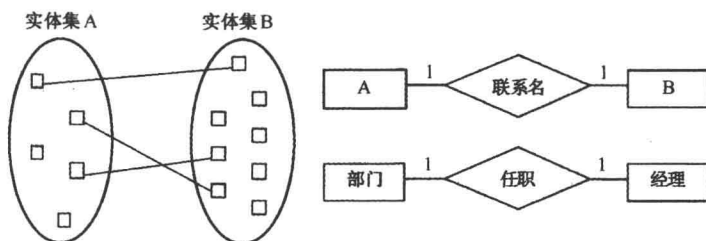


图 1.6 一个层次模型示例

2. 网状模型

网状模型(又称为网络模型)就是用网状结构表示实体及其之间联系的模型。也就是用网络结构表示实体类型及其实体之间联系的模型。顾名思义,一个事物和另外的几个都有联系……这样构成一张网状图。最具代表性的网状数据库管理系统是 1971 年 4 月 CODASYL(Conference On Data System Language, 数据系统语言协会)组织通过 DBTG(Database Task Group, 数据库任务组)报告(和其后的修改文件)规范的系统,习惯称之为 CODASYL 系统或 DBTG 系统。

网状模型的基本特点:

- (1) 允许一个以上结点无双亲。
- (2) 一个结点可以有多于一个的双亲。

网状模型反映了现实世界中实体间更为复杂的联系,可以直接表示多对多联系,但其中的结点间指针更加复杂,因而数据结构更加复杂。学生选课联系就是网状模型的典型例子,如图 1.7 所示。

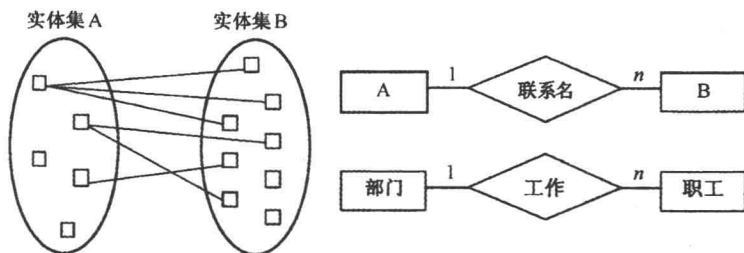


图 1.7 一个网状模型示例

3. 关系模型

关系实际上就是关系模式在某一时刻的状态或内容。也就是说,关系模式是型,关系是它的值。关系模式是静态的、稳定的,而关系是动态的、随时间不断变化的,因为关系操作在不断地更新着数据库中的数据。

关系模型是目前最常用的一种数据模型。关系数据库系统采用关系模型作为数据的组织方式。

1970 年,美国 IBM 公司的研究员 E. F. Codd 首次提出了数据系统的关系数据模型,标志着数据库系统新时代的来临,开创了数据库关系方法和关系数据理论的研究,为数据