

医药数据库应用 基础教程

郑先容 蒋世忠 张 英 齐惠颖 副主编

清华大学出版社

全国高等医药院校计算机与信息技术规划教材

医药数据库应用 基础教程

郭永青 周 怡 主编
郑先容 蒋世忠 张 英 齐惠颖 副主编

清华大学出版社
北京

内 容 简 介

本书从医药学生的角度出发,从数据的集成到数据库的建立,从简单的数据查询到编制应用程序访问数据库进行数据处理,结合医药学中相关数据库应用实例,循序渐进地介绍了数据需求分析、创建数据库、查询数据和处理数据的实际应用过程,学习者可以根据每章节所学习的知识,完成各个实验,以达到掌握数据库知识的目的和提高解决问题的能力。

本书适合作为医药院校的计算机课程教材,也可作为其他相关专业、相关课程的教材。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

医药数据库应用基础教程 / 郭永青, 周怡主编. —北京: 清华大学出版社, 2008.6
(全国高等医药院校计算机与信息技术规划教材)

ISBN 978-7-302-17359-5

I. 医… II. ①郭…②周… III. 数据库管理系统-应用-医药学-高等学校-教材
IV. R319

中国版本图书馆 CIP 数据核字 (2008) 第 052778 号

责任编辑: 索 梅

责任校对: 梁 毅

责任印制: 杨 艳

出版发行: 清华大学出版社

地 址: 北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编: 100084

社 总 机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者: 北京嘉实印刷有限公司

装 订 者: 北京市密云县京文制本装订厂

经 销: 全国新华书店

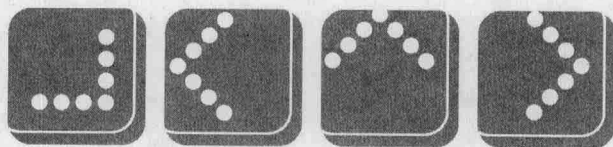
开 本: 185×260 印 张: 12.5 字 数: 296 千字

版 次: 2008 年 6 月第 1 版 印 次: 2008 年 6 月第 1 次印刷

印 数: 1~5000

定 价: 20.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话: (010)62770177 转 3103 产品编号: 028653-01



全国高等医药院校计算机与信息技术规划教材

编写委员会

主任

童隆正（首都医科大学）

副主任

王世伟（中国医科大学）

周 怡（广东药学院）

郭永青（北京大学）

委员

（按姓氏笔画排序）

刘 燕（中山大学）

孙 铭（首都医科大学）

齐惠颖（北京大学）

张志常（中国医科大学）

李 宁（首都医科大学）

李祥生（山西医科大学）

杨 进（广东药学院）

陈 卉（首都医科大学）

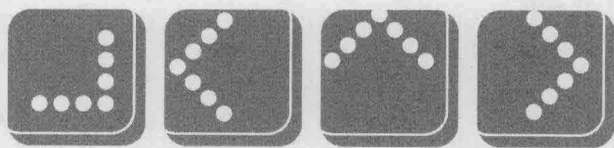
周 毅（中山大学）

赵小龙（新疆医科大学）

晏峻峰（湖南中医药大学）

韩 滨（大连医科大学）

韩绛青（复旦大学）



· 全国高等医药院校计算机与信息技术规划教材

序

长期以来我国高等医学教育主要以培养五年制应用型的专业人才为主,随着社会经济的发展,国家对高素质医学人才的需求日益增加,目前越来越多的院校开始招收七年制医学专业学生。开展长学制教育,优化课程体系,全面提升学生综合素质和创新实践能力,是培养高层次医学人才的新模式。

七年制医学高等教育不同于传统的研究生教育,也不同于本科生教育,没有现成的模式可循,需要我们在实践中不断探索。目前的计算机教材主要涵盖了研究生、本科生和高职高专生等三个层次,没有专门针对七年制医学生的教材。在这种形式下,全国普通高等院校计算机基础教育研究会医学专业委员会认为,七年制医学专业急需一套新的教材。为此,专业委员会根据国家对七年制医学专业人才培养的要求,参照国外经费教材并结合教师们多年的教学实践经验,组织全国多所医学院校的教师,编写此套“全国高等医药院校计算机与信息技术规划教材”。

在十余所院校的积极参与和大力支持下,首批出版4本教材,包括《医学计算机与信息技术实用教程》及实验指导,《医学数据库实用教程》及实验指导。今后将陆续出版医学信息系统、医学影像技术和医学多媒体等相关内容的教材。

本套教材面向医学实际应用,突出了医学专业的特点。在写作过程中,力求做到概念准确、语言清晰、易学易用、通俗简明。本套教材舍弃了烦琐的理论说明,强调操作技能的训练,采用了项目驱动的方式,在介绍必备知识的基础上,引导读者学习必备的知识并完成各种技能的训练。

由于我国的七年制医学教育刚刚兴起,许多问题有待深入讨论,全国普通高等院校计算机基础教育研究会医学专业委员会为了满足许多院校的急需,编写了此套教材,以期抛砖引玉。教材中的不足之处,敬请读者不吝指正。

编写委员会

前言

随着信息技术在医药学领域的广泛应用（如医院的医药信息管理系统、基于 Web 的公共卫生信息系统以及医学生物信息数据库检索等），信息技术已逐渐成为医药学生参加工作后的工作环境和工具。因而，数据库技术也成为了医药学相关专业学生学习和研究的热点。以往我们向学生介绍的数据库知识常限定在单机版的桌面数据库，这往往会限制学生对网络环境数据库的了解和应用。在这本教材中，我们选择了 SQL Server 大型关系数据库管理系统作为学习数据库的平台，以让医学学生更直接参与和体验网络数据库的创建、查询等应用。

最近教育部成立的“医药类计算机基础教学分指导委员会”也特别指出了学习和掌握数据库知识的重要性。我们尝试着用医药相关的数据，对从组织建立数据结构到查询和处理数据进行深入浅出的讲解和实例列举，使读者逐步地了解、掌握和应用数据库知识。

本书分为 6 章。第 1 章介绍了信息、数据、数据库技术、数据库系统的基本概念和数据库技术的发展；第 2 章介绍了数据库工作环境和数据库创建、修改及删除等基本操作；第 3 章主要介绍了 SQL 结构化查询语句的核心部分——查询功能；第 4 章介绍了数据库视图和索引的基本概念；第 5 章通过实现电子处方功能的示例，说明如何使用 Visual Basic 6.0 和 SQL Server 2000 实现数据库技术应用程序的开发；第 6 章简单介绍了生物信息学及目前国内外常用的生物信息数据库，并用前几章所学的知识，示范性地介绍了建立二次数据库的实例。

本书由周怡、郑先容、赵文光、高理文编写第 1 章；李志文、郑先容编写第 2 章；张英、郑先容编写第 3 章；陶伟业、郑先容编写第 4 章；蒋世忠、周怡编写第 5 章；齐惠颖、王焱、陈卉、郭永青编写第 6 章；最后由郑先荣、周怡、郭永青统稿。在整个编写过程中，参编人员经过多次讨论和反复斟酌，力求以联系实际的方法，引导学习者理解和掌握数据库概念，提高分析问题和解决问题的能力。

由于编写时间较为仓促以及作者水平有限，书中疏漏和不足之处恳请广大读者提出宝贵意见和建议。

编 者

2008 年 4 月

读者意见反馈

亲爱的读者：

感谢您一直以来对清华版计算机教材的支持和爱护。为了今后为您提供更优秀的教材，请您抽出宝贵的时间来填写下面的意见反馈表，以便我们更好地对本教材做进一步改进。同时如果您在使用本教材的过程中遇到了什么问题，或者有什么好的建议，也请您来信告诉我们。

地址：北京市海淀区双清路学研大厦 A 座 602 室 计算机与信息分社营销室 收

邮编：100084

电子邮件：jsjic@tup.tsinghua.edu.cn

电话：010-62770175-4608/4409

邮购电话：010-62786544

教材名称：医药数据库应用基础教程

ISBN：978-7-302-17359-5

个人资料

姓名：_____ 年龄：_____ 所在院校/专业：_____

文化程度：_____ 通信地址：_____

联系电话：_____ 电子信箱：_____

您使用本书是作为：☐指定教材 ☐选用教材 ☐辅导教材 ☐自学教材

您对本书封面设计的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议：_____

您对本书印刷质量的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议：_____

您对本书的总体满意度：

从语言质量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

从科技含量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

本书最令您满意的是：

☐指导明确 ☐内容充实 ☐讲解详尽 ☐实例丰富

您认为本书在哪些地方应进行修改？（可附页）

您希望本书在哪些方面进行改进？（可附页）

电子教案支持

敬爱的教师：

为了配合本课程的教学需要，本教材配有配套的电子教案（素材），有需求的教师可以与我们联系，我们将向使用本教材进行教学的教师免费赠送电子教案（素材），希望有助于教学活动的开展。相关信息请拨打电话 010-62776969 或发送电子邮件至 jsjic@tup.tsinghua.edu.cn 咨询，也可以到清华大学出版社主页（<http://www.tup.com.cn> 或 <http://www.tup.tsinghua.edu.cn>）上查询。

目 录

第1章 数据库概述

1

1.1	数据库基本概念	1
1.1.1	信息与数据	1
1.1.2	数据管理方式的发展	2
1.2	数据模型	5
1.2.1	概念层数据模型	5
1.2.2	组织层数据模型	8
1.3	数据库系统的三层模式结构和二级映像	13
1.3.1	关系数据库的三级模式	13
1.3.2	数据库系统的二级映像	14
1.4	数据库系统的组成	15
1.4.1	数据库	15
1.4.2	数据库管理系统	15
1.4.3	数据库系统	16
1.5	数据库设计的基本步骤	17
1.5.1	需求分析阶段	17
1.5.2	数据库概念设计	18
1.5.3	数据库逻辑设计	18
1.5.4	数据库物理设计	19
1.6	数据库技术的发展动向	20
1.6.1	当前主流数据库—— 关系数据库	20
1.6.2	面向对象数据库	20
1.6.3	分布式数据库	21
1.6.4	移动数据库	22
1.7	小结	24
	习题1	24

第2章 数据库的建立与基本操作 27

2.1 创建数据库	27
2.1.1 数据库的建立	27
2.1.2 数据库的删除	35
2.1.3 数据库的备份	38
2.1.4 数据库的还原	44
2.1.5 分离数据库与附加数据库	45
2.2 数据表的创建与删除	47
2.2.1 数据类型	47
2.2.2 数据表结构的创建	48
2.2.3 删除数据表	54
2.3 表的基本操作	54
2.3.1 表结构的修改	54
2.3.2 表数据的插入	56
2.3.3 表数据的更新	59
2.3.4 表数据的删除	60
2.4 小结	61
习题2	61

第3章 数据查询 65

3.1 查询概述	65
3.2 准则表达式	67
3.3 用 SELECT 语句实现单表查询	67
3.3.1 选择表中的若干列	69
3.3.2 选择表中的若干记录	71
3.3.3 对查询结果排序	77
3.3.4 使用集函数	78
3.4 使用其他子句或关键字查询数据	79
3.4.1 带有 IN 谓词的子查询	80
3.4.2 带有比较运算符的子查询	82
3.4.3 带有 ANY 或 ALL 谓词 的子查询	83
3.4.4 带有 EXISTS 谓词的子查询	85
3.5 连接查询	87
3.6 小结	90
习题3	90

第4章 视图与索引 94

4.1 创建视图	94
4.1.1 使用 SQL Server 向导和企业 管理器创建视图	95

4.1.2	使用 Transact-SQL 语句 创建视图	101
4.2	管理视图	105
4.2.1	修改视图	106
4.2.2	通过视图修改数据	108
4.2.3	删除视图	109
4.3	索引	110
4.3.1	创建索引	112
4.3.2	删除索引	116
4.3.3	索引的管理	117
4.4	小结	119
	习题 4	119
第 5 章	数据库在医药领域中的应用示例	124
5.1	Visual Basic 数据库访问技术概述	124
5.1.1	ADO Data 控件	125
5.1.2	Visual Basic 的数据显示 控件	127
5.2	医院信息系统	128
5.3	电子处方功能组成	129
5.4	电子处方数据库设计与数据源 创建	130
5.4.1	电子处方数据库设计	130
5.4.2	创建 ODBC 数据源	131
5.5	电子处方界面设计与功能实现	137
5.5.1	电子处方界面与功能演示	137
5.5.2	电子处方界面及代码设计	140
5.6	小结	148
	习题 5	148
第 6 章	生物信息数据库	152
6.1	生物信息学	152
6.1.1	生物信息学的定义	152
6.1.2	生物信息学的研究领域	152
6.1.3	生物信息学中的计算机 技术	153
6.2	医学生物信息学	156
6.2.1	医学生物信息学概述	156
6.2.2	医学生物信息学的应用 领域与研究内容	156
6.3	生物信息数据库	158

6.4 序列数据库	160
6.4.1 三大核酸数据库	160
6.4.2 蛋白质序列数据库	170
6.5 基因组数据库和结构数据库	171
6.5.1 基因组数据库	171
6.5.2 结构数据库	172
6.6 国际、国内常用医学文献数据库	173
6.6.1 国际常用医学文献数据库	173
6.6.2 国内常用医学文献数据库	175
6.7 检索和提交数据	177
6.7.1 GenBank 数据检索	177
6.7.2 向 GenBank 提交序列数据	178
6.8 应用举例	178
6.8.1 背景介绍	178
6.8.2 数据与方法	179
6.8.3 数据库设计与实现	179
6.8.4 数据库查询	182
6.9 国内外重要的生物信息中心	182
6.10 小结	183
习题6	183
参考文献	184

第1章

数据库概述

近年来,随着计算机技术与网络技术的飞速发展,信息技术在医药学领域的深度融合推动了数据库技术在医药学领域的广泛应用。例如,医院及其医疗机构对病人诊疗的相关数据、医院物资管理的数据、核算分析数据、决策分析统计数据等都是医药领域管理层十分关注的数据库资源。医药相关部门需要对这些数据进行收集、处理和发布,基于这些数据库之上的药品数据库、疾病数据库、新药数据库、生物数据库等具有医药特色的数据库的开发与应用已经越来越普遍。

医药领域存在大量的数据和数据处理的需求,因而数据库技术也成为了医药学相关专业学生学习和研究的热点。

本章首先介绍了数据库的基本概念与数据模型,然后阐述了数据库系统的组成,并分别对数据库技术中的数据库、数据库管理系统、数据库系统三个文字相近而容易混淆的概念进行了较为详细的辨析。随后,对数据库技术的发展动向做了相应的探讨,分别介绍了不同类型的几种数据库——关系数据库、面向对象数据库、分布式数据库以及移动数据库。

1.1 数据库基本概念

在人们的日常生活和社会生产中都有大量的数据产生,例如在医疗机构及其各部门,病人治疗的相关数据、医院物资管理的数据、核算分析数据、决策分析统计数据等都是被人们关注的数据库。数据成为一种需要被管理和加工的非常重要的资源。如何实现对数据科学地进行收集、整理、存储、加工、传输是人们长期以来十分关注的问题。数据处理就是指对原始数据进行上述活动的技术。数据处理的目的是从大量的数据库中获得所需的资料,提取有用的数据库成分作为指挥生产、优化管理、补充知识的决策依据。数据库就是为了实现高效率的数据库处理和数据的合理存储,它有利于数据相对于处理程序的独立性和数据库的共享,并保证数据库的完整性和安全性。本节在系统地介绍数据库的基本概念之前,先介绍信息与数据库的概念与联系。

1.1.1 信息与数据库

信息与数据库是数据库领域中两个基本的概念。信息与数据库既相互关联,又相互区别。数据库是信息的载体;信息则是数据库的内涵,是对数据库的语义解释。

1. 信息 (information)

信息是指人脑对现实世界事务的存在方式或运动状态的反应。信息源于物质和能量,

一切事物,包括自然界和人类都产生信息。信息是可以感知和存储的,并且可以加工、传递和再生。电子计算机是信息处理领域中最先进的工具之一,人类对收集到的信息使用计算机进行处理。信息需要被加工、处理后才能交流和使用,人们往往用数据去记载、描述和传播信息。同一个信息,其表现和传播方式可以是多种多样的。例如,对发现新型流感病毒这个信息,可以登报说明、网上发布,也可以通过电视等媒体进行传播,这几种表现和传播方式,都是对同一个信息的不同反映(表现形式)方式。

2. 数据(data)

数据是对客观事物的符号表示,是用于表示客观事物未经加工的原始素材,如图形符号、数字及字母等。或者说,数据是通过物理观察得来的事实和概念,是关于现实世界中的地方、事件、其他对象或概念的描述。在计算机科学中,数据则是指所有能输入到计算机并被计算机程序处理的符号的介质的总称。

数据是从一系列的观察和测量中得到的,并以数字或符号的形式来描述,计算机可以很方便地对数据进行处理。数据是对客观现象的表示形式。例如,在药房的药品记录中,如果人们最感兴趣的是药品的编号、名称、药品类型、药品价格等,可以这样描述:(2030,甘草片,中成药,盒,2.00)。如果单纯给出这条记录,可能难以让人理解其中的含义,但是了解这个记录每一项所含意义的人,会通过这条记录,获得这样的信息:编号为2030这种药,是甘草片,它是一种中成药,每盒价格是2元。这种对事物描述的符号记录就是数据。数据有一定的格式,例如,药品编号栏最多允许4个字节,药品类型栏允许4个汉字等。

人们通过解释、归纳、分析和综合等方法,从数据所获得的有意义的内容称为信息。因此,数据是信息存在的一种形式,只有通过解释或处理才能成为有用的信息。

3. 数据处理(data processing)

数据处理实际指的是利用计算机对各种类型的数据进行收集、存储、分类、计算、加工、检索及传输的全过程。这一过程主要是由人来对数据进行有效的组织,并把数据输入到计算机中。上述数据处理的全过程可以分为两个层次的操作:一是数据的收集、存储、分类、检索及传输和维护等,称为基本操作,这些基本操作环节构成为数据处理流程;二是加工、计算和输出等,它们随管理对象的不同,操作要求也有所不同,这些操作被称为应用操作。

1.1.2 数据管理方式的发展

计算机诞生于1946年。最初,计算机主要用于科学研究和工程技术领域的数值计算。但随着社会生产力和文明的不断发展,信息在人类社会活动中起着越来越重要的作用。与此同时,越来越大的信息量使得人们迫切需要能够快速处理大量信息的工具和手段。20世纪50年代初期,人们开始用计算机进行数据处理。几十年来,数据处理技术随着计算机软硬件的发展而不断地发展,它大致经历了以下的三个阶段。

1. 人工管理阶段

人工管理阶段是计算机用于数据处理的初级阶段。在该阶段,应用程序中除了要明确数据的逻辑结构外,还要考虑数据在计算机中如何存储和组织,并为数据分配空间及决定存取方法。应用程序完全依赖于特定数据,数据结构的改变以及数据存取物理地址或存储

设备的变化,都会要求修改相应的应用程序。也就是说,应用程序和数据一一对应,数据和处理它的应用程序融为一个整体,如图 1.1 所示。由于数据的物理组织是由应用程序员根据应用的要求设计的,因此很难实现多个应用程序共享数据资源,造成数据的大量重复。这一时期数据的处理主要是手工性质的,具有下述特点:

(1) 数据与程序不具有独立性

一组数据对应一组程序,这就使得程序依赖于数据。如果数据的类型、格式或者数据量、存取方法、输入输出方式等改变了,程序必须做相应的修改。

(2) 数据不宜长期保存

由于数据是面向应用程序的,在一个程序中定义的数据,无法被其他程序利用,因此程序与程序之间存在大量的冗余(重复)数据,占用资源。

(3) 系统中没有对数据进行管理的软件

数据管理任务(包括存储结构、存储方法、输入输出方法等)完全由计算机所执行的任务决定,其存储形式和存储位置均无规律可循。这就给应用程序设计人员增加了很大的负担。

2. 文件系统管理阶段

人工管理阶段数据处理的缺点显而易见,即数据独立性差、冗余度高等,从而造成数据的处理效率低,维护困难,数据分散。20 世纪 50 年代后期到 60 年代中期,随着计算机软、硬件技术的发展,在操作系统引入了文件管理系统后,对上述问题有了较大的改进。数据文件可以按文件名引用,应用程序通过文件管理系统与数据文件发生联系,数据的物理结构和逻辑结构间实现了转换,从而提高了数据的物理独立性,如图 1.2 所示。在文件系统中,还提供了多种文件组织形式,如顺序文件组织、索引文件组织和直接存取文件组织等。

在这一阶段,由于计算机硬件方面有了磁盘、磁鼓等存储设备,数据可以长期保存,并实现了以文件为单位的数据共享,可以对文件进行查询、修改、插入和删除操作。程序和数据之间由文件系统提供存取方法和进行转换,使应用程序与数据之间有了一定的独立性。



图 1.1 人工管理阶段应用程序与数据之间的对应关系

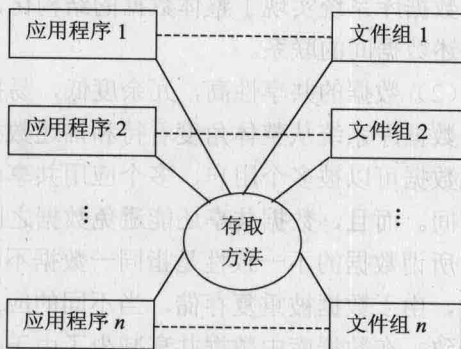


图 1.2 文件系统管理阶段应用程序与数据之间的对应关系

进入文件管理阶段以后,文件系统存在以下问题:

(1) 数据冗余度大

在文件系统中,应用程序与文件相对应,即文件面向应用。文件系统中未能实现以记录或数据项为单位的数据共享,数据仍然是分散的,是面向应用的,数据的共享必须以文件为单位,所以数据还存在大量的冗余。即使不同的应用程序具有大部分相同的数据,只要有不同的数据部分存在,应用程序也不能共用数据文件。这就造成了相同的数据可能因应用程序的不同具有多个数据备份,容易造成数据的不一致性,给数据的修改和维护带来困难。

(2) 数据完整性和独立性差

在文件系统中,一个或一组文件与具体的应用相对应,数据文件是面向应用程序的。文件的逻辑结构对它所服务的应用程序来说是优化的,因而数据的改变,势必会引起应用程序的改变;反之,应用程序的改变亦有可能引起数据文件的改变。文件与应用程序之间缺乏数据独立性。

此外,在文件系统管理阶段,应用程序和数据结构之间的相互依赖程度高,数据的完整性和安全性等也无法得到保证。

3. 数据库系统管理阶段

文件系统管理阶段处理数据存在着诸多不足,而计算机管理规模越来越大,数据量急剧增长。随着计算机工业的迅速发展,大容量和快速存取的磁盘设备开始进入市场,给数据库系统的研究提供了良好的物质基础。

数据库系统是在文件系统的基础上发展起来的新技术,它克服了文件系统的缺点,解决了冗余和数据依赖问题,提供了更广泛的数据共享,为应用程序提供了更高的独立性,保证了数据的完整性和安全性,并为用户提供了方便的用户接口。

与人工管理和文件系统阶段相比,数据库系统的特点主要有以下几个方面:

(1) 数据结构化

数据结构化是数据库与文件系统阶段的根本区别。在文件系统中,相互独立的文件的记录内部是有结构的,传统文件的最简单形式是等长同格式的记录集合。在数据库管理阶段,数据库系统实现了整体数据的结构化,所使用的数据模型不仅描述数据本身的特点,还描述数据间的联系。

(2) 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,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。数据共享可以大大减少数据冗余,节省存储空间。而且,数据共享还能避免数据之间的不相容性与不一致性。

所谓数据的不一致性是指同一数据不同备份的值不一样。采用人工管理或文件系统管理时,由于数据被重复存储,当不同的应用使用和修改不同的备份时就很容易造成数据的不一致。在数据库中数据共享减少了由于数据冗余造成的一致现象。

由于数据面向整个系统,是有结构的数据,不仅可以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。可以取整体数据的各种子集用于不同的应用系统,当应用需求改变或增加时,只要重新选取不

同的子集或者加上一部分数据便可以满足新的需求。

(3) 数据独立性高

数据独立性是数据库领域中一个常用术语,包括数据的物理独立性和数据的逻辑独立性。

物理独立性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。也就是说,数据在磁盘数据库中怎样存储是由数据库管理系统管理的,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。

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。也就是说,数据的逻辑结构改变了,应用程序也可以不变。

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管理系统负责,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。

(4) 数据由数据库管理系统统一管理和控制

数据库的共享是并发 (concurrency) 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个数据,而这些都是由数据库管理系统统一管理与控制,从而简化了应用程序。

1.2 数据模型

数据库是某个企业、组织或部门所涉及的数据的综合,它不仅要反映数据本身的内容,而且要反映数据之间的联系。由于计算机不可能直接处理现实世界中具体事物,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,在数据库中用数据模型这个工具来抽象、表示和处理现实世界的模拟。

不同的数据模型实际上是提供给用户模型化数据和信息的不同工具。根据模型应用的不同目的,可以将这些模型划分为两类,它们分属于两个不同的层次。第一类模型是概念层数据模型,习惯称它为概念模型,按照用户的观点来对数据和信息建模,主要用于数据库的设计。另一类模型是组织层数据模型,简称数据模型,主要分为关系数据库、层次数据库和网状数据库,它是按照计算机系统的观点对数据建模,主要用于数据库的实现。其中,关系型数据库是目前最重要的一种数据库。20 世纪 80 年代以来,计算机厂商推出的数据库管理系统 (database management system, DBMS) 大多都采用关系型数据库系统。

1.2.1 概念层数据模型

概念模型用于现实世界的建模,是现实世界到信息世界的第一层抽象,是数据库设计人员进行数据库设计的有力工具,也是数据库设计人员和用户之间进行交流的语言。概念模型面向用户,描述用户的需求,它不依赖于某一个 DBMS,但可以转换为计算机上某一 DBMS 支持的特定数据模型。概念模型的表示方法很多,其中最为著名、常用的是 P.P.S.Chen 于 1976 年提出的实体-联系方法。

1. 实体-联系模型

在实体-联系 (entity-relationship, E-R) 模型中,使用的与信息有关的概念主要有如下

一些:

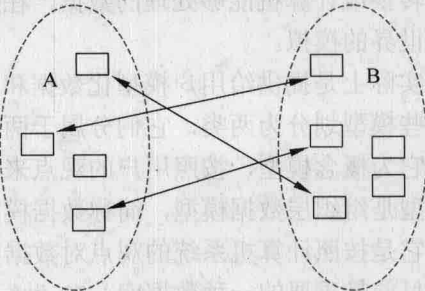
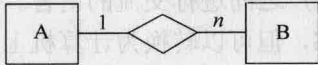
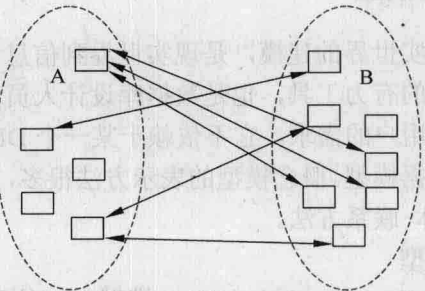
- (1) 实体 客观存在并可相互区别的事物称为实体,例如一位医生、一个病人。
- (2) 属性 实体具有的某一特性称为属性。一个实体可以用若干属性来刻画,例如(1001, 阿莫西林, 西药, 24s, 盒, 24.6, 1100, 2006-3-13, 2008-2-4) 这些属性组合起来表征了一种药品。
- (3) 实体集 同型实体的集合。例如,所有药品就是一个实体集,所有病人也是一个实体集。
- (4) 联系 不同实体集之间的联系。

两个实体之间的联系可以分为以下三类:

- 一对一联系 例如一个医院只有一个正院长,而一个院长也只有一个医院任职,则院长与医院之间是一对一联系。
- 一对多联系 例如一个医院有若干名入编医生,而每一入编医生只在一个医院任职,则医院与入编医生是一对多联系。
- 多对多联系 例如一名医生有多位病人,而一位病人可能也看多名医生,则医生与病人之间是多对多联系。

表 1.1 列出了概念模式的三种联系。

表 1.1 实体与实体间的三种联系

分 类	定 义	举 例
1:1(一对一联系) 	A 中任意实体至多对应 B 中的一个实体;反之, B 中的任意实体至多对应 A 中的一个实体 	观众与座位、 乘客与车票、 病人与病床、 学校与校长、 医院与正院长
1:n(一对多联系) 	A 中至少有一个实体对应 B 中的多个实体;反之, B 中的任意实体至多对应 A 中的一个实体 	城市与街道、 宿舍与学生、 父亲与子女、 班级与学生、 医院与医生