



中等职业学校
21世纪计算机规划教材



数据库应用技术

(Access 2003 版)

□ 李振辉 鞠光明 鞠仪静 编著

冶金工业出版社

内 容 简 介

本书内容包括数据库基础知识、Access 2003 入门、创建数据库和表、表的基本操作、数据的查询、窗体的设计、报表的设计、宏、数据访问页、数据的导入和导出、模块、应用程序开发实例、数据库管理及安全等。

本书内容详尽、重点突出、语言通俗易懂、操作简单明了,通过学习能使读者快速掌握 Access 2003 的各项功能和操作技巧,并可以利用它们解决实际问题。本书既可作为中等职业学校计算机基础课程的教材,也可作为培训学校的培训教材或计算机初学者的自学参考书。

图书在版编目(CIP)数据

数据库应用技术: Access 2003 版 / 李振辉等编著.

北京:冶金工业出版社, 2005.12

中等职业学校 21 世纪计算机规划教材

ISBN 7-5024-3870-X

I. 数... II. 李... III. 关系数据库—数据库管理系统, Access 2003—专业学校—教材 IV.TP311.138

中国版本图书馆 CIP 数据核字(2005)第 139391 号

出版人 曹胜利(北京沙滩嵩祝院北巷 39 号,邮编 100009)

责任编辑 戈兰

佛山市新粤中印刷有限公司印刷;冶金工业出版社发行;各地新华书店经销

2006 年 1 月第 1 版第 1 次印刷

787mm×1092mm 1/16; 13.5 印张; 308 千字; 208 页

20.00 元

冶金工业出版社发行部 电话:(010) 64044283 传真:(010) 64027893

冶金书店 地址:北京东四西大街 46 号(100711) 电话:(010) 65289081

(本社图书如有印装质量问题,本社发行部负责退换)

前言

一、关于本书

Access 是 Microsoft 公司推出的数据库管理软件, 最适合用来作为中、小规模数据量应用软件的底层数据库。它具有功能强大、可靠、高效的管理方式, 支持网络和多媒体技术, 简单易学, 便于开发等主要特点, 目前已经获得相当广泛的应用。

Access 本身就是一个采用面向对象技术开发的数据库系统软件, 它能很好地支持面向对象技术。Access 的数据库对象都具有属性和方法等面向对象技术的基本特征。随着数据库技术在各个领域的广泛应用, 学习和掌握 Access 已成为广大计算机使用者的迫切要求。

在 Access 2003 版本中, Access 使用了表格来实现数据的采集、维护、分析和传播等功能, 此外, 它还提供了关系型数据库所要求的相当丰富的数据运算和数据汇总能力。

为满足各行业计算机用户及在校学生学习和应用 Access 2003 的需要, 编者根据多年的教学、科研和实际应用 Access 数据库管理软件的经验, 结合 Access 2003 的特点, 编写了本书, 希望本书的出版能有助于推动和普及 Access 数据库管理软件的应用。

二、本书内容结构

全书共分 13 章, 其具体结构如下:

第 1 章: 数据库基础知识。主要介绍了数据库技术的发展、数据库的体系结构以及关系数据库等内容。

第 2 章: Access 2003 入门。主要介绍了 Access 2003 的安装与卸载、启动与退出、Access 2003 的工作界面、Access 2003 的数据库文件和表等内容。

第 3 章: 创建数据库和表。主要介绍了创建数据库、在 Access 2003 中创建表等内容。

第 4 章: 表的基本操作。主要介绍了表的结构和字段编辑、记录的编辑、表的关系、设置数据表格式等内容。

第 5 章: 数据的查询。主要介绍了创建查询、设置查询、查询操作、高级查询等内容。

第 6 章: 窗体的设计。主要介绍了创建窗体、窗体控件、编辑窗体等内容。

第 7 章: 报表的设计。主要介绍了创建报表、报表控件、报表编辑、子报表、报表打印以及标签设计等内容。

第 8 章: 宏。主要介绍了创建宏和宏组、执行和调试宏、常用的宏命令以及保存宏等内容。

第 9 章: 数据访问页。主要介绍了创建数据访问页、编辑数据访问页、访问数据访问页、优化数据访问页的性能等内容。

第 10 章: 数据的导入和导出。主要介绍了导入数据、导出数据、链接 Access 表以及与 Word 和 Excel 的相关操作等内容。

第 11 章: 模块。主要介绍了模块的概念、创建、编辑以及 Visual Basic 编辑器窗口等内容。

第 12 章: 应用程序开发实例。主要介绍了一个应用程序设计实例。本章用一个具体

的应用程序的设计过程，详细介绍了从系统分析到数据库设计的完整过程。

第 13 章：数据库管理及安全。主要介绍了数据库的备份与还原、压缩与修复、加密与解密以及数据库的管理等内容。

建议本书课时为 78 课时，其中上机操作至少 32 课时。

三、本书特点

本书从易学性和实用性入手，主要有以下特点：

- (1) 本书语言简炼、层次分明、结构完整、图文并茂，便于读者理解和掌握。
- (2) 书中每章都配练习题，便于读者巩固所学知识。

四、本书适用对象

本书由暨南大学的李振辉、鞠光明、鞠仪静编写。

本书的作者是计算机教学第一线的教师，在长期的教学实践过程中积累了丰富的教学实战经验。他们根据中等职业技术学校学生的认知规律和特点，总结归纳出一套理论联系实际的教学方法，在讲解各知识结构和技巧的同时，将重要的知识点融于实例中，这样，既便于教学又利于自学，为培养读者的灵活应用能力和创造能力奠定了基础。本书既可作为中等职业学校计算机基础课程的理想教材，也可作为培训学校的培训教材或计算机初学者的自学参考书。

由于编者水平有限，编写时间仓促，书中如有疏漏和不足之处，敬请读者批评指正，联系方法如下：

电子邮箱：service@cnbook.net

网址：www.cnbook.net

本书电子教案、源代码及习题参考答案可在该网站免费下载。此外，该网站还有一些其他相关书籍的介绍，可供读者选购参考。

编 者
2005 年 9 月

目 录

第1章 数据库基础知识.....1	2.5.2 查询..... 19
1.1 数据库概述.....1	2.5.3 窗体..... 20
1.1.1 数据库的概念.....1	2.5.4 报表..... 21
1.1.2 数据库的特点.....1	2.5.5 页..... 22
1.1.3 数据库的术语.....2	2.5.6 宏..... 23
1.2 数据库技术的发展.....3	2.5.7 模块..... 24
1.2.1 数据库技术的萌芽.....3	小结..... 24
1.2.2 数据库技术的发展.....4	综合练习二..... 24
1.2.3 数据库技术的成熟.....4	一、选择题..... 24
1.3 数据库的体系结构.....4	二、填空题..... 25
1.4 关系数据库.....5	三、思考题..... 25
1.4.1 数据库系统模型和结构.....5	四、上机操作..... 25
1.4.2 关系的定义.....8	第3章 创建数据库和表..... 26
1.4.3 关系数据库.....9	3.1 创建数据库..... 26
小结.....9	3.1.1 数据库设计的基本步骤..... 26
综合练习一.....10	3.1.2 构建 E-R 关系数据库模型..... 27
一、选择题.....10	3.1.3 创建和打开数据库..... 33
二、填空题.....10	3.2 在 Access 2003 中创建表..... 34
三、思考题.....10	3.2.1 使用设计器创建表..... 35
第2章 Access 2003 入门..... 11	3.2.2 使用表向导创建表..... 40
2.1 Access 2003 概述..... 11	3.2.3 通过输入数据创建表..... 43
2.2 Access 2003 的安装与卸载..... 11	小结..... 44
2.2.1 按照安装程序的向导操作..... 11	综合练习三..... 45
2.2.2 Office 2003 的在线更新..... 12	一、选择题..... 45
2.2.3 Access 2003 的修复与卸载..... 12	二、填空题..... 45
2.3 Access 2003 的启动与退出..... 14	三、思考题..... 45
2.3.1 Access 2003 的启动..... 14	四、上机操作..... 45
2.3.2 Access 2003 的退出..... 15	第4章 表的基本操作..... 46
2.4 Access 2003 的工作界面..... 16	4.1 表的结构和字段编辑..... 46
2.4.1 Access 2003 的菜单栏..... 16	4.1.1 行高、列宽、列的顺序..... 46
2.4.2 Access 2003 的工具栏..... 17	4.1.2 显示、隐藏、冻结列..... 47
2.4.3 Access 2003 的帮助界面..... 17	4.1.3 字段属性..... 49
2.5 Access 2003 的数据库文件和表..... 18	4.1.4 字段的基本操作..... 49
2.5.1 表..... 18	4.2 记录的编辑..... 52

4.2.1 添加记录	52	5.4.2 生成表查询	77
4.2.2 修改记录	52	5.4.3 追加查询	78
4.2.3 查找与替换记录	52	5.4.4 删除查询	79
4.2.4 记录排序	53	5.5 高级查询	80
4.2.5 筛选记录	53	5.5.1 创建交叉表查询	80
4.3 表的关系	53	5.5.2 创建参数查询	81
4.3.1 建立表间关系	53	5.5.3 SQL 查询	82
4.3.2 修改表间关系	54	小结	84
4.4 设置数据表格式	55	综合练习五	84
4.4.1 设计视图	55	一、选择题	84
4.4.2 数据表视图	56	二、填空题	84
4.4.3 数据透视表视图	56	三、思考题	85
4.4.4 数据透视图视图	56	四、上机操作	85
小结	57	第 6 章 窗体的设计	86
综合练习四	57	6.1 窗体概述	86
一、选择题	57	6.1.1 窗体类型	86
二、填空题	57	6.1.2 窗体的表现形式	87
三、思考题	57	6.1.3 窗体的构成	88
四、上机操作	58	6.2 创建窗体	88
第 5 章 数据的查询	59	6.2.1 创建数据透视表窗体	89
5.1 查询概述	59	6.2.2 用窗体向导创建窗体	89
5.1.1 查询的类型	59	6.2.3 使用图表向导创建	90
5.1.2 查询的功能	60	6.2.4 使用数据透视表创建窗体	91
5.1.3 查询的工作原理	60	6.2.5 使用设计视图创建窗体	92
5.2 创建查询	61	6.3 窗体控件	92
5.2.1 使用简单查询向导创建 单表查询	61	6.3.1 标签	92
5.2.2 使用交叉表查询向导创建查询	63	6.3.2 文本框	93
5.2.3 使用查找重复项查询向导 创建查询	65	6.3.3 组合框与列表框	94
5.2.4 使用查找不匹配项查询向导 创建查询	66	6.3.4 选项组	95
5.2.5 使用设计视图创建查询	67	6.3.5 命令按钮	96
5.3 设置查询	68	6.3.6 子窗体	98
5.3.1 查询中的计算	68	6.3.7 日历	99
5.3.2 联接表	72	6.3.8 选项卡	100
5.4 查询操作	76	6.4 编辑窗体	101
5.4.1 更新查询	76	6.4.1 窗体设计视图	101
		6.4.2 设置窗体的数据源	102
		6.4.3 定义窗体的外观	103
		6.4.4 在窗体中添加和修改数据	104

6.4.5 数据的查找和排序	105	综合练习七	128
6.4.6 数据的筛选	105	一、选择题	128
小结	107	二、填空题	128
综合练习六	107	三、思考题	129
一、选择题	107	四、上机操作	129
二、填空题	107	第 8 章 宏	130
三、思考题	107	8.1 宏的概念	130
四、上机操作	107	8.2 创建宏和宏组	131
第 7 章 报表的设计	108	8.2.1 利用设计视图创建宏	131
7.1 报表概述	108	8.2.2 创建与设计宏	134
7.1.1 报表的功能	108	8.2.3 创建与设计宏组	135
7.1.2 报表的类型	108	8.3 执行和调试宏	137
7.1.3 报表工具栏	110	8.3.1 执行宏	137
7.2 创建报表	111	8.3.2 调试宏	138
7.2.1 自动创建报表	111	8.4 常用的宏命令	138
7.2.2 使用向导创建报表	112	8.5 保存宏	141
7.2.3 使用图表向导创建报表	114	小结	141
7.2.4 使用标签向导设计报表	116	综合练习八	141
7.2.5 使用设计视图创建报表	118	一、选择题	141
7.3 报表控件	121	二、填空题	142
7.3.1 标签	121	三、思考题	142
7.3.2 文本框	121	四、上机操作	142
7.3.3 图像	122	第 9 章 数据访问页	143
7.3.4 子报表	122	9.1 数据访问页概述	143
7.4 报表编辑	123	9.2 创建数据访问页	143
7.4.1 设置报表属性	123	9.2.1 自动创建数据访问页	143
7.4.2 编辑报表的结构	123	9.2.2 利用数据页向导创建数据 访问页	145
7.4.3 对报表中的数据排序	124	9.2.3 利用现有的 Web 页创建数据 访问页	147
7.4.4 定义查询数据源	125	9.2.4 将 Access 对象转换为数据 访问页	148
7.5 子报表	125	9.2.5 利用设计视图创建数据访问页 ..	149
7.5.1 在已有报表中创建子报表	125	9.3 编辑数据访问页	152
7.5.2 把已有报表作为子报表	126	9.3.1 在设计视图中编辑	152
7.6 报表打印	126	9.3.2 利用脚本编辑器编辑	153
7.6.1 预览报表	126	9.4 访问数据访问页	155
7.6.2 页面设置	126		
7.6.3 打印报表	127		
7.7 标签设计	127		
小结	128		

9.5 优化数据访问页的性能	155	11.3.4 VBA 编程的调试方法	176
小结	156	小结	180
综合练习九	156	综合练习十一	180
一、选择题	156	一、选择题	180
二、填空题	156	二、填空题	181
三、思考题	156	三、思考题	181
四、上机操作	157	四、上机操作	181
第 10 章 数据的导入和导出	158	第 12 章 应用程序开发实例	182
10.1 导入数据	158	12.1 系统分析	182
10.1.1 XML 概述	158	12.1.1 需求分析	182
10.1.2 导入外部数据	158	12.1.2 总体设计	182
10.2 导出数据	160	12.2 数据库设计	184
10.2.1 导出数据的简单应用	160	12.2.1 系统界面设计	185
10.2.2 导出数据的其他应用	161	12.2.2 模块设计	192
10.3 链接 Access 表	163	12.2.3 系统菜单设计	196
10.4 与 Word 和 Excel 的相关操作	163	小结	200
10.4.1 用 Word 发布 Access 表	163	综合练习十二	200
10.4.2 导入或链接 Excel 电子表格	164	一、选择题	200
小结	164	二、填空题	201
综合练习十	164	三、思考题	201
一、选择题	164	四、上机操作	201
二、填空题	165	第 13 章 数据库管理及安全	202
三、思考题	165	13.1 数据库的安全性	202
四、上机操作	165	13.1.1 数据库的备份与还原	202
第 11 章 模块	166	13.1.2 数据库的压缩与修复	202
11.1 模块概述	166	13.1.3 数据库的加密与解密	203
11.1.1 类模块	166	13.2 数据库管理	205
11.1.2 标准模块	166	13.2.1 用户账号管理	205
11.2 创建模块	167	13.2.2 生成 MDE 文件	206
11.2.1 创建类模块	167	小结	206
11.2.2 宏转换为模块	168	综合练习十三	206
11.2.3 创建模块实例	168	一、选择题	206
11.3 Visual Basic 编辑器窗口	169	二、填空题	207
11.3.1 VBA 编程的基本概念	169	三、思考题	207
11.3.2 打开一个 VBA 的编辑器	169	四、上机操作	207
11.3.3 创建 VBA 的过程	171	参考文献	208

第 1 章 数据库基础知识

教学目标

随着计算机技术的飞速发展及其应用领域的不断扩大，特别是随着计算机网络和 Internet 的不断发展，基于计算机网络和数据库技术的信息管理系统、应用系统得到了突飞猛进的发展。因此，数据库技术的基础知识和基本技能已成为计算机及相关专业的必修内容。

- (1) 数据库概述。
- (2) 数据库技术的发展。
- (3) 数据库的体系结构。
- (4) 关系数据库。

1.1 数据库概述

1.1.1 数据库的概念

数据库是为了满足某一部门中多个用户的多种应用的需要，按照一定的数据模型在计算机中组织、存储和使用相互联系的数据集合。数据库系统就是管理大量的、持久的、可靠的和共享的数据的工具。

例如，每个人都有很多朋友和熟人，为了保持联系，常常用一个笔记本将他们的姓名、地址、电话等信息都记录下来，以便联系。这个“通讯录”就是一个最原始的人工管理的“数据库”，每个人的姓名、地址、电话等信息就是这个数据库中的数据项。可以随意在笔记本这个“数据库”中添加新朋友的个人信息，也可以由于某个朋友的家庭地址或者其他变动而修改他的相关信息，如图 1-1 所示。

	公司/家庭	地址	城市	省/自治区/直辖市	邮政编码	国家
▶ 陈磊		学院路28号	北京	北京	100083	中国
完世		解放路110号	西安	陕西	710045	中国
凤铃		幸福大道230号	石家庄	河北	050031	中国
孔雀		临江大道123号	武汉	湖北	430034	中国
Buchanan	14 Garrett Hill		伦敦	伦敦		英国
※						

图 1-1

1.1.2 数据库的特点

从数据库的定义看，它有以下特性：

(1) 集成性。所谓集成，就是指把某特定应用环境中各种相关的数据以及数据之间的联系全部集中在一起，并且按照特定的结构形式存储至介质，这种联系也可以看成一种数据。也可以把这种数据库看成是由若干个性质的数据文件联合而形成的并且在文件之间局部或者全部消除了冗余的数据整体。

(2) 共享性。共享是指数据库中的数据可以让不同的用户共同使用。各用户可以使用不同的编程语言、以不同的访问方式同时访问同一个数据库。其实，共享性就是数据库

的继承性带来的必然结果。当然数据库会提供安全访问机制保证各用户能正确地访问数据。

(3) 大量性。数据库的一个特征就是拥有大量的数据。一般企业的数据库容量会高达数百 MB, 而如银行、证券公司这类信息量较大的部门, 其业务数据量会高达 GB 甚至 TB。因此数据库中存放的数据一般不能直接在内存中进行处理, 需要使用大容量而速度相对较低的外部存储设备。

(4) 持久性。数据库作为信息的存储工具, 里面的数据需要在一定时间内保持有效性。例如交易行的业务数据、公司企业的商业数据等, 这些资料往往需要保存几年、几十年甚至更长。这时候人们甚至会使用光盘等可靠性比一般磁盘更高的存储介质进行数据库的数据备份。

1.1.3 数据库的术语

要学习使用数据库, 首选应该掌握数据库中的各种术语。数据库系统是组织、存取和维护大量数据的信息管理应用系统。

数据库技术涉及的概念主要有:

(1) 实体: 客观存在并可相互区别的事物称为实体。实体可以是具体的人、事、物, 也可以是抽象的概念或联系。

(2) 属性: 实体所具有的某一特性称为属性。一个实体可以由若干个属性来刻画。例如, 学生实体可以由学号、姓名、性别、出生年份、系、入学时间等属性组成。

(3) 码: 惟一标识实体的属性集称为码。例如, 学号是学生实体的码。

(4) 域: 属性的取值范围称为该属性的域。例如, 学号的域为 8 位整数, 姓名的域为字符串集合。

(5) 实体型: 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地刻画同类实体, 称为实体型。

(6) 实体集: 同型实体的集合称为实体集。例如, 全体学生就是一个实体集。

(7) 联系: 在现实世界中, 事物内部以及事物之间是有联系的, 这些联系在信息世界中反映为实体(型)内部的联系和实体(型)之间的联系。实体内部的联系通常是指组成实体的各属性之间的联系。实体之间的联系通常是指不同实体集之间的联系。

(8) 计算机系统: 计算机系统指用于数据库管理的计算机硬件和软件系统。由于数据库系统的数据量一般都很大, 因此对硬件资源要求较高。

① 需要足够大的内存以存放和运行操作系统、数据库管理系统的核心模块和应用程序等。

② 要有大容量的磁盘直接存取数据库数据, 有足够的磁带(或软盘)作数据备份。

③ 系统应具有较高的通道能力, 以提高数据传送率。

④ 系统还应具有网络功能, 以实现数据资源的共享。

(9) 数据库: 数据库是数据库系统中存储的结构化的、逻辑相关的数据集合, 它是数据库系统操作的对象, 并为多种应用服务。数据库中的数据具有集中性和共享性。集中性是指把数据库看成性质不同的数据文件的集合, 数据冗余很小。共享性是指不同的用户使用不同的语言, 为了不同的应用目的可同时存取数据库中的数据。

(10) 数据库管理系统：数据库管理系统 (DataBase Management System, DBMS) 是负责数据库管理和维护的软件系统。它通常由三个部分组成：数据定义语言及其翻译程序、数据操纵语言及其编译 (或解释) 程序、数据库管理例程序。

DBMS 是数据库系统的核心软件，学习使用数据库，通常要学习某个 DBMS 的使用方法。在关系数据库中有许多 DBMS 系统，例如：dBase、FoxBase、FoxPro、ORACLE、UNIFY、CLIPPER、Ingres 和 DB2 等等。

(11) 用户：用户是使用数据库的人。数据库系统中主要有数据库管理员、系统分析员、数据库设计人员、应用程序员和最终用户。

① 数据库管理员。

数据库管理员是指全面负责数据库系统正常运转的高级人员。主要负责：决定数据库中的信息内容和结构，决定数据库的存储结构和存取策略，定义数据的安全性要求和完整性约束条件，监控数据库的使用、运行和数据库的改进、重组、重构。

② 系统分析员。

系统分析员负责应用系统的需求分析和规范说明，负责和最终用户及数据库管理员相结合，确定系统的硬件和软件配置，并参与数据库系统的概要设计。

③ 数据库设计人员。

数据库设计人员负责数据库中数据的确定、数据库各级模式的设计。设计人员应参加用户需求调查和系统分析。

④ 用户 (即最终用户)。

最终用户是指并没有掌握太多计算机知识的工程人员和管理人员，通过数据库系统提供的用户接口 (如：浏览器、命令语言、表格操作、菜单、报表等交互式手段) 使用数据库。

1.2 数据库技术的发展

从第一代电子计算机的出现到现在，计算机已经有几十年的发展历史。20 世纪 60 年代以前，计算机主要还是用于科学计算，这是“计算机”这个名词的由来。随后信息时代来临，各行各业中所出现的数据越来越大并且越来越复杂，人们对计算机的应用提出了新的要求。老式的数据采集、存储和处理等数据加工手段已经远远不能满足人们的需要。计算机数据库技术就在这样的时代背景下产生了。

1.2.1 数据库技术的萌芽

“Data Base” (数据库) 一词首先被美国系统发展公司在 20 世纪 60 年代为美国海军基地研制数据库时使用。1968 年，国际商用机器公司 IBM 在数据库管理系统方面取得了重大的突破，率先研制成功了集成数据存储系统 (IMS——Information Management Systems, 信息管理系统)，它可以运行多个程序共享同一个数据库，是属于层次数据模型系统。但是该系统的设计是完全面向程序员的，所以可操作性很差。

1969 年，CODASYL (Conference On Data System Language, 数据系统语言会议) 组织的下属组织 DBTG (Data Base Task Group) 提出了网状数据库模型，使数据库系统开始走向规范化和标准化。在那个时期，CODASYL 组织做了大量的数据库技术的规范和推广工作，为数据库的成熟作出了贡献，所以一般认为数据库技术诞生于 20 世纪 60 年代末。

1.2.2 数据库技术的发展

20 世纪 70 年代, IBM 属下的 San Jose 研究所提出了关系数据库模型的概念, 开创了数据库的关系方法和关系规范化的理论。这个关系模型的提出以关系的数学理论为基础, 具有严谨的数学理论支持, 也继承了数学理论的完美和结构上的简单等优点。这个关系数据库理论的提出者 E.F.Code 还因此获得了计算机科学的最高奖项——ACM 图灵奖。

随后, 众多学者开始致力于关系方法的研究, 一些专用的或者实验性质的数据库系统陆续出现。1971 年, 美国数据系统语言协会正式提出了三级抽象模式, 即外模式、内模式和概念模式, 成功解决了数据独立性的问题。1974 年, IBM 公司的关系数据库管理系统 SyStemR 作为一个成功的软件产品投放市场。1978 年, 新奥尔良在其发表的 DBDWD 报告中把数据库的设计过程分为需求分析、信息分析与定义、逻辑设计和物理设计四个阶段。

1.2.3 数据库技术的成熟

1984 年, David Marer 所著的《关系数据库理论》一书的出版, 正式标志着数据库理论的成熟。尽管目前已经出现了面向对象的数据库理论, 但关系数据库理论仍然是应用最广泛的理论。本书介绍的 Access 数据库也是基于关系数据库理论的。

1.3 数据库的体系结构

美国 ANSI/X3/SPARC 的管理系统小组将数据库结构从逻辑上分为三个层次: 外层、概念层和内层; 它们分别对应三个抽象模式: 外模式、概念模式和内模式。这种划分称为数据库的三级结构。

如图 1-2 所示, 数据库的三级结构反映了三种看待数据库的角度。

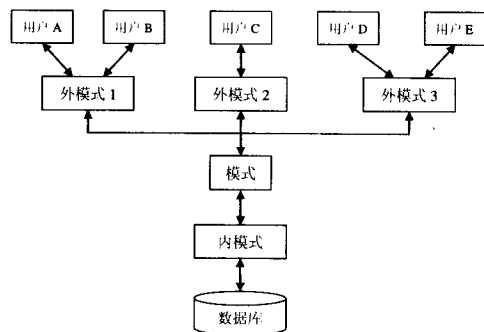


图 1-2

外层, 又称为外视图 (External View) 或者个别用户视图, 指的是离用户最近的一层, 它是数据库的用户层或称为用户级。用户视图也就是某个用户所看到的内容。对用户来说, 用户视图就是数据库, 而且一个用户往往只用到整个数据库的一部分。外模式也称子模式, 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。

概念层, 又称为概念视图 (Conceptual View) 或者用户共同视图, 是整个数据库信息内容的一种表示, 是所有外部视图的一个最小并级。概念视图用概念模式来定义, 概念模式又简称为模式, 是数据库中全体数据的逻辑结构和特征的描述, 包括了数据库中所用记

录类型的逻辑定义、权限检查以及概念模式和内模式之间的映射关系。

内层,又称为内视图 (Internal View) 或者存储视图,是最接近物理存储的一层。存储视图是数据库中最底一级的逻辑表达,它用内模式来描述。内模式是数据物理结构和存储方式的描述,是数据库内部的表示方式。例如,记录的存储方式是顺序存储、按照 B 树结构存储还是按 Hash 方法存储;数据是否压缩存储;数据是否加密;存储记录结构有何规定等。一个数据库只有一个内模式。

数据库的三级结构所具有的优点:

(1) 实现数据共享,减少数据冗余度。多个用户使用同一个概念模式派生的不同模式,减少了数据冗余,有利于用户间共享数据。

数据共享包括三个方面:

- ① 所有用户可以同时存取数据。
- ② 数据库可以为当前的用户服务,也可以为将来的新用户服务。
- ③ 可以使用多种语言完成与数据库的接口。

(2) 保证数据的独立性。

数据独立是指应用程序不必随数据存储结构的改变而改变。三级结构将外层和概念层分开,保证了数据的逻辑独立性。当数据库逻辑结构发生变化(如数据定义的修改、数据间联系的变更等)时,用户的应用程序不需要做任何修改。

数据独立提高了数据库系统的稳定性,因此提高了程序的效益。

(3) 有利于对数据的保护。

特定的用户只能看到属于自己的外视图,无法访问和修改其他数据库组成部分,因而缩小了程序错误传播的范围,保证了数据的正确性。

因此,数据库管理系统的核心任务是完成三级数据库之间的转换,把用户对数据库的操作从外层换到内层去执行。

1.4 关系数据库

关系数据库是建立在严密的数学基础之上的,它应用数学方法来处理数据库中的数据。

关系数据库是目前各类数据库中最重要、最流行的数据库,是目前使用最广泛的数据库系统。20 世纪 70 年代以后开发的数据库管理系统产品几乎都是基于关系的。

1.4.1 数据库系统模型和结构

早期的数据库系统都是在大型机上实现的大型数据库系统。20 世纪 80 年代初,随着微型机的普及和广泛使用,微型机数据库系统得到了很大的发展。

数据模型是数据库系统中一个关键概念,各种机器上实现的数据库系统都是基于某种数据模型的。因此,了解数据模型的基本概念是学习数据库的基础。

数据模型是指描述记录内的数据项间的联系和记录之间的联系的数据结构形式。它应满足三方面要求:能较真实地模拟现实世界;容易被人所理解;便于在计算机上实现。

在数据库的发展历史上,最常用的数据库模型有:层次模型 (Hierarchical Model)、网状模型 (Network Model) 和关系模型 (Relational Model)。下面分别介绍这三种模型。

1. 层次模型

层次模型是数据库系统中最早使用的一种模型。数据的层次模型用树型结构来表示各类实体的类型和实体之间的联系。在现实世界中许多实体之间的联系很自然的呈现出一种层次关系,例如:家族关系、行政机构、地理位置关系等。

在数据库中满足下述两个条件的“基本层次联系”的集合称为层次模型。

(1) 有且只有一个节点无双亲节点,这个节点就是根节点。

(2) 其他节点有且仅有一个双亲节点。

在层次模型中,每个节点表示一个记录类型(实体),记录之间的连线表示节点之间的联系。每个节点的上方的节点称为该节点的双亲节点,而其下方的节点称为该节点的子节点。没有子节点的节点称为叶节点。

层次模型的结构示意图如图 1-3 所示。

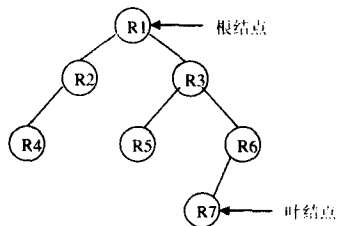


图 1-3

在图 1-3 中, R1 为根节点, 同时是 R2、R3 的双亲节点。R2、R3 是 R1 的子节点, 同时又是 R4、R5、R6 的双亲节点。R6 是 R7 的双亲节点, R7 是 R6 的子节点。R4、R5、R7 是叶节点。

从图 1-3 上可以看到层次模型像一棵倒立的树, 节点的双亲是惟一的。

层次模型适用于描述客观存在的事物中主次之分的结构关系, 具有层次分明、结构清晰的特点。它的缺点是层次模型只能反映记录类型间的一对多的关系, 不能反映多对多的关系。

如图 1-4 所示是一个教师学生层次数据库。在这个层次数据库中有四个记录类型。记录类型院系是根节点, 由系号、系名、系主任名三个字段组成。它有两个子节点: 教研室和学生。记录类型教研室是院系的子节点, 同时也是授课的双亲节点, 它由室号、室名、室主任名三个字段组成。记录类型学生由学号、姓名、成绩三个字段组成。记录类型授课由课号、班级、教师姓名三个字段组成。学生与授课是叶节点, 它们没有子节点。由院系到教研室、由教研室到授课、由院系到学生均是一对多的联系。

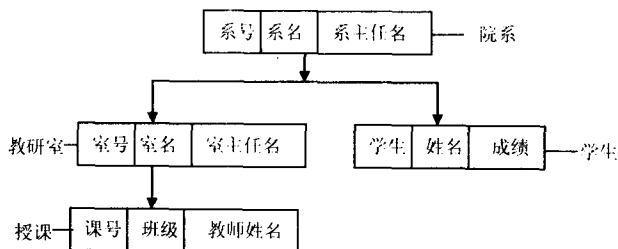


图 1-4

1969年,美国IBM公司研制的IMS系统是一个最典型的层次模型系统。

2. 网状模型

由于层次结构不能描述多对多的关系,因而产生了网状模型。网状模型就是在层次模型的基础上取消层次模型的限制,将树型结构变成图的结构。网状模型是以记录类型为节点的网络,反映的是现实世界中较为复杂的事物间的联系。

广义上,任意一个连通的基本层次联系的集合就是一个网状模型。在数据库中,把满足下列条件的基本层次联系集合称为网状模型:

- (1) 允许一个以上的节点无双亲节点。
- (2) 允许一个节点可以有多于一个的双亲节点。
- (3) 两个节点之间可以有两种或两种以上的关系。

图1-5为网状模型示意图。

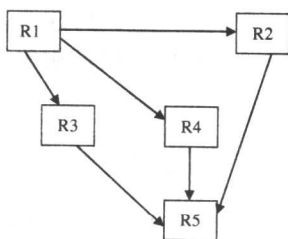


图 1-5

网状模型能够更为直接地描述现实世界,具有良好的性能,存取效率较高。但是,由于网状模型的结构比较复杂,而且随着应用环境的扩大,数据库的结构就变得越来越复杂,不利于用户使用。

CODASYL 系统(即 DBTG)是网状模型的典型代表,它表示了实体间的多种复杂联系。

3. 关系模型

关系模型是当前最重要的一种数据模型。20世纪80年代以来,几乎所有的数据库系统都是关系型的。关系模型是建立在严格的数学概念的基础上的。

从用户的角度看,关系模型的数据结构是一个二维表,它使用表格描述实体间的关系,由行和列组成。每个表格就是一个关系。下面以供应商表(如图1-6所示)为例,介绍关系模型中的一些术语。

■ 供应商 : 表						
	供应商ID	公司名称	联系人姓名	联系人职务	地址	城市
▶	1	佳佳乐	陈小姐	采购经理	西直门大街 110 号	北京
*	2	康富食品	黄小姐	订购主管	幸福大街 290 号	北京
+	3	妙生	胡先生	销售代表	南京路 23 号	上海
*	4	方全	王先生	市场经理	永定路 342 号	北京
*	5	日正	李先生	出口主管	体育场东街 34 号	北京
*	6	德昌	刘先生	市场代表	学院北路 67 号	北京
+	7	正一	方先生	市场经理	高邮路 115 号	上海

图 1-6

(1) 关系 (Relation): 一个关系也就是通常所说的一张表,如图1-6所示的供应商表。

(2) 元组 (Tuple): 表中的一行就是一个元组,或称为一条记录。

(3) 属性 (Attribute): 表中的一列即为一个属性,每个属性有一个名称即属性名。

例如,图1-6中的供应商表有六列,对应六个属性:供应商ID、公司名称、联系人姓名、

联系人职务、地址和城市。

(4) 关键字 (Key): 表中的某个属性组可以惟一地确定一个元组, 如图 1-6 中的供应商 ID, 可以惟一确定一个供应商, 即该关系中的关键字。

(5) 关系模式: 对关系的描述, 一般的表示方法为:

关系名 (属性 1, 属性 2, …… 属性 n)

例如, 图 1-6 中的关系可表示为:

供应商 (供应商 ID, 公司名称, 联系人姓名, 联系人职务, 地址, 城市)。

关系模型要求关系必须规范化, 即要求关系必须满足一定的规范条件。最基本的一条规范条件是: 关系的每一个分量必须是一个不可分的数据项, 也就是不允许表中有表。

关系模型所具有的优点是:

- (1) 关系模型是建立在严格的数学概念基础上的, 具有较强的理论根据。
- (2) 可表示一对一的关系, 也能表示一对多的关系, 还能表示多对多的关系。
- (3) 描述的一致性。无论实体还是实体之间都用关系来表示。
- (4) 概念简单, 操作方便, 数据独立性强。

当然, 关系数据模型也存在缺点, 其中最明显的缺点是, 由于存取路径对用户透明, 查询效率不如层次模型和网状模型。

由于关系模型具有较多的优点, 所以以关系模型为基础的数据库管理系统的开发发展很快, 其应用范围不断扩大。目前, 较为典型的关系型数据库管理系统有 SQL Server、Oracle 等。

1.4.2 关系的定义

关系也就是一张二维表, 指的是实体与实体之间或者实体内部之间一对一的性质。

关系具有以下特征:

1. 关系中不能有任意两条完全相同的记录

关系中实体与实体之间或实体内部之间是一一对应的联系, 因此不可以有完全相同的记录。

2. 关系中的记录是非排序的

关系中的记录顺序可以随意交换, 改变顺序后的关系性质不会发生变化。在实际的操作中, 用户常常需要对记录进行某种排序, 使其具有某种序列。

3. 关系中记录的字段是非排序的

记录的字段顺序不一样, 其代表的是同一个表。如图 1-7 和图 1-8 所示, 两图均表示同一个关系, 即都是供应商表, 只是表现形式有所不同。

■ 供应商 : 表			
	供应商ID	公司名称	联系人姓名
*	1	佳佳乐	陈小姐
*	2	康富食品	黄小姐
*	3	妙生	胡先生
*	4	为全	王先生
*	5	日正	李先生

图 1-7

■ 供应商 : 表			
	供应商ID	联系人姓名	联系人职务
*	1	佳佳乐	采购经理
*	2	康富食品	订购主管
*	3	妙生	销售代表
*	4	为全	市场经理
*	5	日正	出口主管

图 1-8

4. 字段名称不能相同

在同一个表中, 字段必须有不同的字段名, 否则无法区分; 在不同的表中, 字段名可

以相同。如图 1-9 所示的两个表格：地址表和客人表，在两个不同的表格中，都有名字和姓氏字段名，但在同一表格中，没有相同的字段名。

字段名称	数据类型
地址 ID	自动编号
名字	文本
姓氏	文本

字段名称	数据类型
客人 ID	自动编号
名字	文本
姓氏	文本
地址	文本

图 1-9

5. 字段不可再分

每个表中的字段都是惟一的，不可再分。也就是前面所说的不可以“表中有表”。

满足以上五个条件的关系才是规范化关系，反之则称为非规范化关系。当一个数据库不满足规范化条件时，在数据库操作和数据存储上将造成资源浪费，甚至无法正常使用数据库。

1.4.3 关系数据库

在关系模型中，用关系来表示实体以及实体间的联系。在一个给定的应用领域中，所有实体及实体之间联系的集合构成一个关系数据库。简单地说，关系数据库就是表的集合，结构化的数据存储在这些表中。但是关系数据库是一个整体，并不只是把不同的表随意的堆积在一起，表与表之间有着密切的联系，通常一个关系数据库中的表格和其他表格都有关系。例如：在某个公司的订单库中，有两张表如图 1-10 和图 1-11 所示。

订单ID	客户	雇员	订购日期	到货日期	发货日期	运营商
10248	山泰企业	赵军	1996-07-04	1996-08-01	1996-07-16	联邦货运
10249	东帝星	孙林	1996-07-05	1996-08-16	1996-07-10	急速快递
10250	实美	郑建杰	1996-07-08	1996-08-05	1996-07-12	统一包裹
10251	千图	李芳	1996-07-08	1996-08-05	1996-07-15	急速快递
10252	福星制衣厂股份有限公司	郑建杰	1996-07-09	1996-08-06	1996-07-11	统一包裹
10253	实美	李芳	1996-07-10	1996-07-24	1996-07-16	统一包裹

图 1-10

订单ID	产品	单价	数量	折扣
10248	猪肉	¥ 14.00	12	0%
10248	糯米	¥ 9.80	10	0%
10248	酸奶油	¥ 34.80	5	0%
10249	砂茶	¥ 18.60	9	0%
10249	猪肉干	¥ 42.40	40	0%
10250	虾子	¥ 7.70	10	0%
10250	猪肉干	¥ 42.40	35	15%
10250	海苔	¥ 16.80	15	15%

图 1-11

在上面的两张表格中，有相同的字段，即订单 ID。两张表格记录了订单的不同信息，但相互间有联系，由两张表格一起才能得到订单的全部信息。

关系数据库理论是建立在关系代数理论基础上的，借助数学工具形成了一整套数据库设计的理论和方法，因此关系数据库理论具有科学的严谨性和严密性。

小结

本章主要介绍了数据库的概念和特性，数据库技术的发展，数据库的体系结构和关系数据库等。

数据库系统是组织、存取和维护大量数据的管理应用系统，它一般由计算机系统、数据库、数据库管理系统和用户构成。数据库结构从逻辑上可以分为三个层次：外层、概念层和内层；它们三个分别对应着三个抽象模式：外模式、概念模式和内模式，被称为数据库的三级结构。

目前所使用的数据库管理系统产品几乎都是基于关系的，即关系数据库。关系也就是一张二维表；而关系数据库就是表的集合，人们将结构化的数据存储在这些表中。但是关