



全国高等工科教育计算机类规划教材

数据库原理与应用

— Access

杨 茹 柳 青 主编

 **机械工业出版社**
CHINA MACHINE PRESS



全国高等工科教育计算机类规划教材

数据库原理与应用

Access

主 编 杨 茹 柳 青

副主编 刘宇阳 史文集

参 编 吕松涛 张爱文

于燕飞 刘添华

葛冬梅 陆 上

主 审 雷国华



机械工业出版社

本书共分为 10 章, 分别介绍了数据库原理基础知识、建立 Access 数据库和表的方法、查询、窗体、报表、数据访问页、宏、VBA 程序设计、Access 与其他系统的联系、Access 应用系统开发实例。本书可作为高等工
科院校数据库课程教材, 也可以作为高职高专学校的教材, 本书也可以作
为计算机爱好者的自学参考书。

为了方便教师教学, 本书配有电子教案, 请发邮件至 wangyx@mail.
machineinfo.gov.cn 邮箱索取。

图书在版编目 (CIP) 数据

数据库原理与应用——Access/杨茹, 柳青主编.

—北京: 机械工业出版社, 2006.8

全国高等工科教育计算机类规划教材

ISBN 7-111-19752-6

I. 数… II. ①杨…②柳… 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 (2006) 第 095342 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 王玉鑫 孔熹峻

责任编辑: 李学锋 版式设计: 张世琴 责任校对: 姚培新

封面设计: 姚毅 责任印制: 洪汉军

北京京丰印刷厂印刷

2006 年 8 月第 1 版·第 1 次印刷

184mm×260mm·17 印张·420 千字

0 001—4 000 册

定价: 26.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

本社购书热线电话 (010) 68326294

编辑热线电话 (010) 68354423

封面无防伪标均为盗版

前言

Access 2003 是 Microsoft 公司最新推出的 Office 2003 的组件之一,是目前非常流行、功能强大的桌面数据库管理系统。它可以有效地组织、管理和共享数据库的信息,并将数据库信息与 Web 结合在一起,使用 Access 2003 可通过直观的可视化操作完成大部分数据的管理工作,也可通过 VBA 编程程序代码处理较复杂的问题,同时也为 VB 等编程工具提供了很好的接口

主要内容: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、窗体、报表、数据访问页、宏、模块的应用等,为用户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。通过本书的学习,读者只要按照步骤循序渐进地操作,即可在最短的时间内学会使用 Access 2003 软件。

本教材特点如下:

- 1) 注重基础,讲究使用,从入门到精通,循序渐进,通俗易懂。
- 2) 大量配图,有助于理解教材内容。
- 3) 内容的编排与“二级考试新大纲”的知识系统完全一致,充分体现“大纲”的知识能力要求。
- 4) 每一章后面设计了适量的习题,主要是针对本章重点、难点进行训练,对掌握本章内容有非常重要的作用。
- 5) 最后面单独开辟一章应用实例,完整地向读者介绍了如何应用 Access 开发数据库应用系统,使学生初步掌握数据库应用系统开发全过程。

编写分工:本书由杨茹、柳青主编,雷国华任主审。其中第 1~2 章由史文集编写、第 3~4 章由刘宇阳编写,第 5~6 章由吕松涛编写,第 7 章由柳青编写,第 8 章由杨茹编写,第 9 章由于燕飞编写,第 10 章由张爱文编写,刘添华、葛冬梅、陆上编写了其中部分内容。全书由杨茹、柳青规划和统稿。

由于时间仓促,加上水平有限,书中难免有不足之处,敬请广大读者批评指正。

编者

目 录

前言	2.4.1 使用表设计器创建一张表	41
	2.4.2 使用表向导创建一张表	44
	2.4.3 通过输入数据创建表	46
第 1 章 数据库基础知识	2.5 表布局设计的基本操作	47
1.1 数据库系统概述	2.5.1 调整行高和列宽	47
1.1.1 数据库常用术语	2.5.2 改变列的顺序	48
1.1.2 数据管理的发展历程	2.5.3 显示与隐藏列	49
1.1.3 数据库系统的特点	2.5.4 冻结列	50
1.1.4 数据库系统的三级模式结构	2.5.5 删除行或列	50
1.1.5 数据库系统的组成	2.6 特殊字段类型的编辑	51
1.1.6 数据库技术的研究领域和 三个发展阶段	2.6.1 日期/时间	51
1.2 关系数据库的基本概念	2.6.2 超级链接	51
1.2.1 数据模型	2.6.3 OLE 对象	51
1.2.2 关系的数学定义	2.7 对象的各种操作	51
1.2.3 关系数据库	2.7.1 复制表	51
1.2.4 关系的完整性	2.7.2 删除表	53
1.2.5 关系运算	2.7.3 重命名表	53
1.3 关系数据库标准语言 SQL	2.7.4 修改数据表结构	53
1.3.1 SQL 的特点	2.7.5 查找与替换记录	54
1.3.2 SQL 语言的基本概念	复习思考题	55
1.3.3 SQL 语言的概述	第 3 章 查询	57
1.4 Access 概述	3.1 查询的基础知识	57
1.4.1 Access 系统的基本特点	3.1.1 查询的类型	57
1.4.2 Access 系统的基本对象	3.1.2 查询的功能	58
复习思考题	3.1.3 查询的工作原理	58
第 2 章 建立 Access 数据库和表	3.2 使用向导创建查询	59
2.1 数据库设计的一般方法	3.2.1 使用简单查询向导创建 单表查询	59
2.2 创建 Access 数据库的方法	3.2.2 使用交叉表查询向导创建查询	61
2.2.1 Access 数据库文件	3.2.3 使用查找重复项查询向 导创建查询	63
2.2.2 创建空 Access 数据库	3.2.4 使用查找不匹配项查询向 导创建查询	65
2.2.3 利用 Access 数据库向导创建 Access 数据库	3.3 使用设计视图创建查询	66
2.3 表的基本概念	3.4 查询中的计算	68
2.3.1 Access 表的组成	3.4.1 字段表达式	68
2.3.2 Access 中的数据	3.4.2 建立字段表达式	70
2.3.3 四种视图		
2.4 表的创建		

3.5 操作查询	72	5.3.1 报表的结构	116
3.5.1 更新查询	72	5.3.2 利用报表设计视图创建报表	116
3.5.2 生成表查询	73	5.3.3 报表的属性	118
3.5.3 追加查询	74	5.4 创建高级报表	119
3.5.4 删除查询	75	5.4.1 基于一般查询的报表	119
3.6 联接表	75	5.4.2 基于参数查询的报表	119
3.6.1 表间关系	76	5.4.3 子报表的创建	120
3.6.2 在关系窗口中创建关系	80	5.5 打印报表的设置	121
3.7 创建交叉表和参数查询	82	复习思考题	122
3.7.1 创建交叉表查询	82	第6章 数据访问页	125
3.7.2 创建参数查询	83	6.1 Access 数据访问页简介	125
3.8 SQL 查询	84	6.2 创建数据访问页	125
3.8.1 联合查询	84	6.2.1 自动创建数据页	125
3.8.2 传递查询	85	6.2.2 使用数据页向导创建数据 访问页	126
3.8.3 数据定义查询	85	6.2.3 利用已有的网页生成数据 访问页	128
3.8.4 子查询	85	6.2.4 手动创建数据访问页	128
复习思考题	86	6.2.5 修改已有的数据访问页	131
第4章 窗体	89	6.3 使用控件	131
4.1 有关窗体的几个基本概念	89	6.3.1 数据访问页中的图表	131
4.1.1 窗体类型	89	6.3.2 添加电子表格控件	132
4.1.2 窗体的表现形式	89	6.4 使用超级链接	133
4.1.3 窗体的三种视图及组成	90	6.4.1 在已有的文件或 Web 页上创 建超级链接	133
4.2 使用向导创建窗体	90	6.4.2 创建链接新建数据访问页的 超级链接	134
4.2.1 自动窗体的创建	90	6.4.3 创建当前数据库中的数据访 问页的超级链接	134
4.2.2 创建数据透视表自动窗体	91	6.4.4 创建发送电子邮件的超级链接	135
4.2.3 使用窗体向导创建窗体	92	复习思考题	135
4.3 自定义窗体	93	第7章 宏	137
4.3.1 使用设计视图简介	93	7.1 宏的概念	137
4.3.2 窗体基本控件及应用	96	7.2 宏的创建	138
4.4 子窗体的创建	101	7.2.1 利用设计视图创建宏	138
4.5 窗体中数据的操作	104	7.2.2 创建与设计宏	141
复习思考题	105	7.2.3 创建与设计宏组	142
第5章 报表	107	7.3 宏的执行条件	143
5.1 报表基础知识	107	7.4 宏的执行	144
5.1.1 报表的功能	107	7.5 常用宏操作	144
5.1.2 报表的类型	107	复习思考题	146
5.1.3 报表工具栏	108	第8章 VBA 程序设计	148
5.2 创建报表的方法	108		
5.2.1 自动创建报表	108		
5.2.2 使用向导创建报表	109		
5.2.3 使用图表创建报表	112		
5.2.4 使用标签向导设计报表	113		
5.3 报表设计视图的使用	115		

8.1 VBA 编程的基本概念	148	的概念	210
8.1.1 面向对象的编程思想	148	9.1.2 Access 数据库对象的导入	212
8.1.2 模块	153	9.1.3 与 Excel 交换数据	213
8.1.3 过程	153	9.1.4 与文本文件交换数据	217
8.2 VBA 编程的基本知识	153	9.1.5 与 Visual FoxPro 交换数据	222
8.2.1 VBA 的数据类型	155	9.1.6 链接	223
8.2.2 常量	158	9.2 Access 数据库与 Web 页	225
8.2.3 变量	159	9.2.1 将 Access 表或查询导出为	
8.2.4 运算符	161	静态 Web 页	225
8.2.5 表达式	163	9.2.2 链接 Web 页	229
8.2.6 VBA 的常用语句	166	9.2.3 Access 与动态 Web 页	231
8.3 VBA 的编程界面	167	9.3 超链接	234
8.3.1 打开一个 VBA 的编程器	167	复习思考题	236
8.3.2 创建 VBA 的过程	175	第 10 章 Access 应用系统开发实例	237
8.3.3 转变已有的宏为 VBA 过程	179	10.1 开发的步骤与过程	237
8.4 基本控制结构	182	10.2 系统需求分析	237
8.4.1 顺序结构	183	10.2.1 数据需求分析	238
8.4.2 选择结构	183	10.2.2 功能需求分析	238
8.4.3 循环结构	187	10.3 数据库设计	239
8.5 过程	189	10.3.1 数据库逻辑设计	239
8.5.1 Sub 过程	189	10.3.2 数据库物理设计	240
8.5.2 自定义函数过程	192	10.4 详细设计	242
8.5.3 参数的传递	195	10.4.1 创建数据库设计	242
8.5.4 变量的作用域	195	10.4.2 设计数据库的表	242
8.5.5 控件及应用	197	10.4.3 创建表间关系	244
8.6 VBA 编程的调试方法及错误处理	201	10.4.4 创建查询	244
8.6.1 错误类型	201	10.4.5 设计窗体	246
8.6.2 调试与排错	202	10.4.6 设计报表	258
8.6.3 调试窗口	204	10.4.7 设计数据访问页	261
8.6.4 错误处理	205	10.4.8 设计自动启动宏	262
复习思考题	208	10.5 系统的测试与发布	263
第 9 章 Access 与其他系统的联系	210	10.5.1 系统的测试	263
9.1 数据的导入、链接和导出	210	10.5.2 系统的发布	263
9.1.1 数据的导入、链接和导出		参考文献	265

第1章 数据库基础知识

学习目标

1) 理解数据库、数据模型、数据库管理系统。

2) 了解数据管理的发展、数据库管理系统的组成。

3) 理解关系数据库、关系运算基本概念。

4) 了解数据模型的组成要素。

5) 掌握概念模型的表示方法：实体-联系模型（E-R图的表示方法）。

6) 识记常用的三种数据模型：层次模型、网状模型、关系模型。

7) 识记数据库系统的三级模式结构。

8) 识记 SQL 基本命令：查询命令、操作命令。

9) 了解 Access 系统的基本特点及基本对象。

数据库是数据管理的最新技术，计算机科学的重要分支。众多应用领域开始使用数据库来存储和处理信息，因此以如何存储、使用和管理数据为主要任务的数据库技术也得到了越来越广泛的应用，并成为计算机科学与技术、信息管理等专业的重要课程之一。

1.1 数据库系统概述

1.1.1 数据库常用术语

本节主要给出与数据库相关联的几个基本概念。

1. 数据 (Data)

数据即描述事物的符号记录，根据被描述事物的特点可以采用数字、文字、声音、语言等多种形式来描述。

2. 数据库 (DataBase, DB)

数据库可简单理解为存放数据的仓库。具体来说数据库是长期储存在计算机内、有组织、可共享的数据集合。

3. 数据库管理系统 (DataBase Management System, DBMS)

数据库管理系统是位于用户与操作系统之间的一层数据管理软件，主要功能包括：

(1) 数据定义功能 DBMS 提供的数据定义语言 (Data Definition Language, DDL)，可使用户实现对数据库中的数据对象的定义。

(2) 数据操纵功能 DBMS 提供的数据操纵语言 (Data Manipulation Language, DML)，可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 数据库在建立、运行和维护时由数据库管理系统统一管理、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 它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 and 性能监视、分析功能等。这些功能通常由一些实用程序完成。

4. 数据库系统 (DataBase System, DBS)

数据库系统是指在计算机中引入数据库后的系统, 由数据库、数据库管理系统、数据库应用系统、数据库管理员和用户五部分构成。

1.1.2 数据管理的发展历程

数据管理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。人们借助计算机进行数据处理是近三十年的事, 其应用已远远超出了最初的满足复杂的科学计算的要求。随着计算机硬件、软件技术的不断发展和计算机应用范围的不断扩大, 计算机的应用已经从科学研究部门扩展到了企业和行政部门。因此, 在计算机的三大主要应用领域 (即科学计算、数据处理和过程控制) 中, 数据处理迅速上升为计算机应用的主要方面。数据库技术就是作为数据处理的一门新技术而发展起来的。数据管理经历了人工管理、文件系统、数据库系统三个阶段。

1. 人工管理阶段

20 世纪 50 年代中期以前, 计算机主要用于科学计算。当时的硬件状况是, 外存只有纸带、卡片、磁带, 没有磁盘等直接存取的存储设备; 软件方面, 只有汇编语言而没有操作系统和高级语言, 更没有管理数据的软件, 数据处理的方式是批处理。这些决定了当时的数据管理只能依赖人工来进行。

特点:

- 1) 数据不进行保存
- 2) 没有专门的数据管理软件
- 3) 数据面向应用
- 4) 只有程序的概念

2. 文件系统阶段

20 世纪 50 年代后期至 60 年代中期, 硬件中出现了磁盘、磁鼓等直接存储设备; 软件方面, 操作系统中有了专门的数据管理软件, 称为文件系统, 数据按其内容、结构和用途分成若干个命名的文件, 用户可通过操作系统对文件进行打开、读、写、关闭等操作; 处理方式上不仅有了批处理, 还能够联机适时处理。

特点:

- 1) 数据可长期保存。
- 2) 由文件系统管理数据。
- 3) 数据共享性差, 冗余度大。
- 4) 数据独立性差。
- 5) 数据安全性差。

3. 数据库系统阶段

20 世纪 60 年代以来, 计算机用来管理数据的规模越来越大, 应用范围也越来越广泛, 硬件有了大容量磁盘, 在多用户、多应用共享数据的需求推动下出现了统一管理数据的专门

软件系统——数据库管理系统。

在文件系统中，应用程序直接存储数据的文件，而数据库系统中，应用程序是通过数据库管理系统来访问数据，数据库管理系统实际上是应用程序和存储数据的数据库之间的接口。从文件系统到数据库系统，是数据管理发展历程中的一个飞跃。

1.1.3 数据库系统的特点

与人工管理和文件系统相比，数据库系统的特点主要有以下几个方面。

1. 数据结构化

传统文件的最简单形式是等长同格式的记录集合。例如，一个学生人事记录文件，每个记录都有如图 1-1 所示的记录格式。

学号	姓名	性别	出生年月	系别	籍贯	简历
----	----	----	------	----	----	----

图 1-1 记录格式

针对不同的学生，其中前六项数据基本等长，而后一项数据信息量大小变化较大，采用等长的记录格式存储数据，必须按信息量最多的学生记录来存储，因而会浪费大量的存储空间。而在其他记录文件里，每个文件的数据项也包括了学号、姓名、性别，这又引发了大量的重复存储。

在数据库系统中，从整体角度来组织数据，数据不再针对某一应用，而是面向全组织，实现了整体数据的结构化，把文件系统中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，在描述数据时，不仅要描述数据本身，还要描述数据之间的联系。可按图 1-2 来组织该信息管理系统。

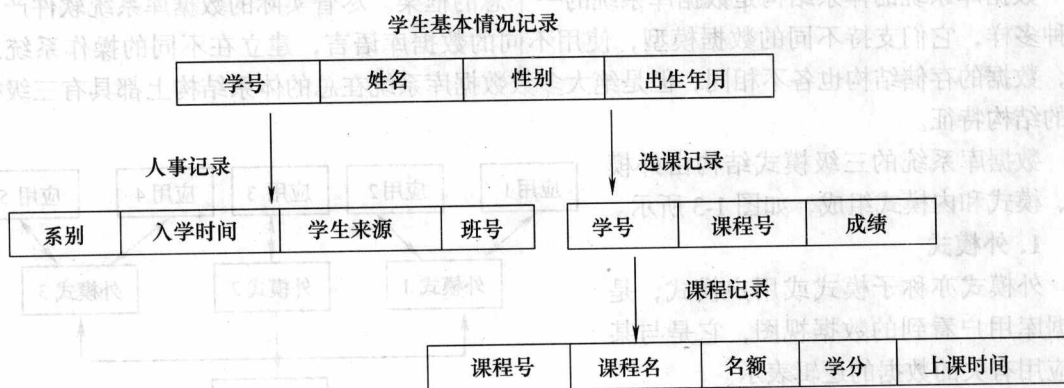


图 1-2 结构化的学生数据组织

2. 数据共享性高，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，数据可被多个用户、应用共享使用。例如可供人事部门了解学生的人事情况，教务部门了解学生的选课情况，而且各部门没有重复存储，减少了数据的冗余。

这种结构化的数据库系统弹性大，很容易增加一个新的应用，随时扩充以适应各种用户的需求。

3. 数据独立性高

数据的独立性包括数据的物理独立性和逻辑独立性。

物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的。即数据在磁盘上的数据库中存储是由 DBMS 管理的, 应用程序要处理的只是数据的逻辑结构, 当数据的物理存储改变了, 应用程序不用改变。

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, 当数据的逻辑结构改变时, 用户程序也可以不变。

4. 数据由 DBMS 统一管理和控制

DBMS 不仅要有基本的管理功能, 还要有如下的控制功能。

(1) 数据的安全性 (Security) 保护 数据的安全性是指保护数据, 防止不合法的使用造成数据的破坏和丢失。例如对不同的用户设置不同的使用权限, 使储户只能查询不能修改。

(2) 数据的完整性 (Integrity) 检查 数据的完整性是指数据的正确性、有效性和相容性。完整性检查可将数据控制在有效的范围内、例如学生成绩不能小于 0 分, 性别只能是男或女等。

(3) 并发 (Concurrency) 控制 当多个用户同时修改、存取数据库中的数据时, 必须对多用户的并发操作加以协调和控制。例如在售票系统中某次列车车票只有一张, 这时有多名乘客在不同终端上同时发出定票请求, 必须采取措施, 只能让一名乘客得到这张票。

(4) 数据库恢复 (Recovery) 当数据库系统出现硬件、软件的故障, 操作员失误以及被人破坏时, DBMS 必须具有将数据库恢复到某一已知的正确状态的能力。

1.1.4 数据库系统的三级模式结构

数据库系统的体系结构是数据库系统的一个总的框架。尽管实际的数据库系统软件产品多种多样, 它们支持不同的数据模型, 使用不同的数据库语言, 建立在不同的操作系统之上, 数据的存储结构也各不相同, 但是绝大多数数据库系统在总的体系结构上都具有三级模式的结构特征。

数据库系统的三级模式结构由外模式、模式和内模式组成, 如图 1-3 所示。

1. 外模式

外模式亦称子模式或用户模式, 是数据库用户看到的数据视图, 它是与某一应用有关的数据的逻辑表示。

外模式通常是模式的子集, 它是各个用户的数据视图, 而不同的用户因其需求、看待数据的方式、对数据的要求的不同, 使用的程序设计语言也可以不同, 因此不同用户的外模式描述是不同的。即使对模式中同一数据, 在外模式中的结构、类型、长度、保密级别等都可以不同。

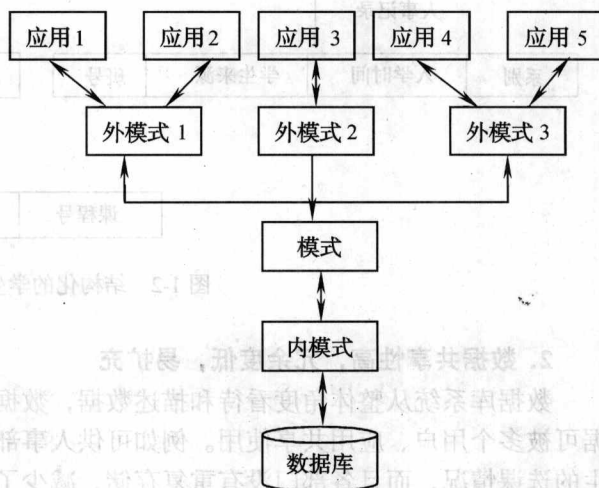


图 1-3 数据库系统的三级模式结构

数据库系统提供外模式描述语言（外模式 DDL）描述用户数据视图。用外模式 DDL 写出的一个用户数据视图的逻辑定义的全部语句称为此用户的外模式。外模式 DDL 和用户选用的程序设计语言具有相容的语法。

2. 模式

模式，亦称为逻辑模式，是数据库中全体数据的逻辑结构和特性的描述，是所有用户的公共数据视图。

模式通常以某一种数据类型为基础，它不仅仅是数据逻辑结构的定义，而且要定义与数据相关的安全性、完整性要求；不仅要定义数据记录内部的结构，而且要定义这些数据项之间的联系，以及不同记录之间的联系。

数据库系统提供模式描述语言（模式 DDL）来严格地表示这些内容。用模式 DDL 写出的一个数据库逻辑定义的全部语句，称为该数据库的模式。模式是对数据库结构的一种描述，而不是数据库本身，它是装配数据的一个框架。

3. 内模式

内模式是全体数据库数据的内部表示或者低层描述，用来定义数据的存储方式和物理结构。例如，记录是顺序存储还是按照 B+ 树结构存储，或是按 Hash 方法存储；索引的组织方式是什么；数据是否压缩存储，是否加密，以及数据的存储记录结构的规定等。

内模式通常用内模式数据描述语言（内模式 DDL 亦称存储模式 DDL）来描述和定义。

1.1.5 数据库系统的组成

一个数据库系统至少由以下四部分组成，如图 1-4 所示。

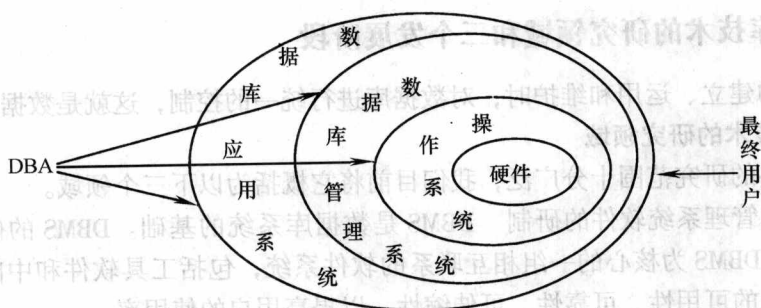


图 1-4 数据库系统组成示意图

1. 计算机系统

计算机系统包括硬件系统和软件系统。

(1) 硬件 硬件包括中央处理器（CPU）、内存储器、外部存储设备和其他外设，硬件是数据库赖以存在的物理设备。另外对于数据库系统来讲，它对硬件还有些特殊要求，例如需要足够大的内存来存放操作系统、DBMS 核心模块、数据库的数据缓冲区和应用程序；由于数据库数据庞大，因此需要足够大的磁盘等直接存取设备存储数据，需要足够的磁带作数据库的备份；此外还要求系统具有较高的通道能力，以提高数据传送率。

(2) 软件 软件方面必须有支持 DBMS 运行的操作系统以及 DBMS 本身；具有与数据库接口的高级语言及其编译系统；以及各种实用程序等必要的软件，它们构成数据库系统必不

可少的软件环境。

2. 数据库管理系统 (DBMS)

DBMS 是管理数据库的系统软件, 它实现数据库系统的各种功能。

3. 数据库

它是一个特定组织的各项应用相关的全部数据的集合。通常由两大部分组成: 一部分是相关应用所需要的工作数据的集合, 称作物理数据库, 它是数据库的主体; 另一部分是关于各级数据结构的描述数据, 称作描述数据库, 通常由一个数据词典系统管理。

4. 人员

人员主要是指数据库管理员 (DBA)、系统分析员、数据库设计人员、应用程序员和最终用户 (End User)。

其中 DBA 是一组熟悉计算机数据处理业务, 负责设计和维护数据库的技术人员, 在数据库系统中, 数据是面向整个用户集体的, 这样在设计数据库时, 既要考虑数据之间的固有联系, 又要顾及整个用户系统使用数据的方式和系统运行的效率, 因而要由专业人员对用户需求进行广泛的深入细致的反复调查、归纳和总结; 即使在系统建成以后, 由于系统相当复杂和新的应用不断出现, 也必须由专业人员进行管理, 并在使用中对数据库不断进行调整、修改和扩充, 甚至重新组织。因此, 这些专业人员在数据库的整个生命周期中处于十分重要的地位, 人们常把他们视为数据库系统的一个必要组成部分, 主要承担建立、维护和管理数据库系统的任务。

最终用户通过应用系统的用户接口使用数据库。常用接口方式有浏览器、菜单、表格、图形、报表等。

1.1.6 数据库技术的研究领域和三个发展阶段

在数据库的建立、运用和维护时, 对数据库进行统一的控制, 这就是数据库技术。

1. 数据库技术的研究领域

数据库学科的研究范围十分广泛, 我们目前将它概括为以下三个领域。

(1) 数据库管理系统软件的研制 DBMS 是数据库系统的基础。DBMS 的研制包括研制 DBMS 本身及以 DBMS 为核心的一组相互联系的软件系统, 包括工具软件和中间件。研制的目标是提高系统的可用性、可靠性、可伸缩性, 以提高用户的使用率。

(2) 数据库设计 数据库设计的主要任务是在 DBMS 支持下, 为某一部门或组织设计一个结构合理、使用方便、效率较高的数据库及其应用系统。其主要的研究方向是数据库设计方法学和设计工具, 包括数据库设计方法、设计工具和设计理论的研究、数据模型和数据建模的研究、计算机辅助数据库设计方法及其软件系统的研究、数据库设计规范和标准的研究。

(3) 数据库理论 数据库理论的研究主要集中在关系的规范化理论、关系数据理论等。近年来又出现数据库逻辑演绎和逻辑推理、数据库中的知识发现 (Knowledge Discovery from Database, KDD)、并行算法等新的理论研究方向。

2. 数据库技术的三个发展阶段

短短三十年数据库已从第一代数据库系统——层次和网状数据库系统, 第二代数据库系统——关系数据库系统, 发展到第三代以面向对象模型为主要特征的新一代数据库系统。

数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透,互相结合,成为当前数据库技术发展的主要特征。从20世纪80年代以来,数据库技术在商业领域的巨大成功刺激了其他领域对数据库技术需求的迅速增长。另一方面在应用中提出的一些新的数据管理的需求也直接推动了数据库技术的研究与发展,尤其是面向对象数据库系统(Object Oriented Database System, OODBS)的研究与发展。新一代数据库技术的研究和发展导致了众多不同于第一、二代数据库的系统诞生,构成了当今数据库系统的大家族。这些新的数据库系统无论它是基于扩展关系数据模型的、还是OO模型的;是分布式、客户/服务器或混合式体系结构的;是在SMP还是在MPP并行机上运行的并行数据库系统;是用于某一领域(如工程、统计、GIS)的工程数据库、统计数据库、空间数据库等,我们都可以广泛地称之为新一代数据库系统。新一代数据库系统的发展呈现了百花齐放的局面。

特点:

1) 面向对象的方法和技术对数据库发展的影响最为深远。

2) 数据库技术与多学科技术的有机结合。

3) 面向应用领域的数据库技术的研究。

经过多年的研究和讨论,对第三代数据库系统的基本特征已有了共识:

1) 第三代数据库系统应支持数据管理、对象管理和知识管理。

2) 第三代数据库系统必须保持或继承第二代数据库系统的技术。

3) 第三代数据库系统必须对其他系统开放。

1.2 关系数据库的基本概念

1.2.1 数据模型

数据库中存储的是数据,这些数据反映了现实世界中有意义、有价值的信息,它不仅反映数据本身的内容,而且反映数据之间的联系。那么如何抽象表示、处理现实世界中的数据和信息呢?这就需要使用数据模型这个工具。数据模型是数据库中用于提供信息表示和操作手段的形式框架,它是我们将现实世界转换为数据世界的桥梁。

在数据库技术中,我们用数据模型的概念来描述数据库的结构与语义,对现实世界进行抽象。数据模型(Data Model)是客观事物及其联系的数据描述,它具有描述数据和数据联系两方面功能。数据模型是数据库系统的核心问题,不同类型的数据库系统的主要区别就在于所支持的数据模型不同。

1. 数据模型的三个组成部分

数据模型有严格的形式化定义,以便在计算机系统中实现。数据模型应包含数据结构、数据操作和数据完整性约束三部分:

(1) 数据结构 数据结构是所研究的对象类型的集合,是对实体类型和实体间联系的表达和实现。由于数据模型是向用户提供的,因此,数据结构应该易于为用户所理解,并且还必须具有足够强的表示能力。

在数据库系统中,通常按照数据结构的类型来命名数据模型,如层次结构、网状结构和

关系结构的模型分别命名为层次模型、网状模型和关系模型。

(2) 数据操作 数据操作是指对数据库的检索和更新（包括插入、删除、修改）两类操作的实现。数据模型要定义这些操作的确切含义、操作符号、操作规则（如优先级别）以及实现操作的语言。

数据结构是对系统静态特性的描述，数据操作是对系统动态特性的描述。

(3) 数据完整性约束 数据完整性约束给出了数据及其联系应具有制约和依赖规则，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、有效和相容。数据模型应该反映和规定符合这种数据模型所必须遵守的完整性约束条件。另外，数据模型还应该提供定义完整性约束条件的机制，以反映某一部门的应用所涉及的数据必须遵守的语义约束条件。

一个数据模型由以上三部分组成，其中数据结构的区别是区分数据模型最本质的部分。

2. 概念模型

概念模型用于信息世界的建模，是数据库设计人员进行数据库设计的有利工具，也是数据库设计人员和用户之间进行交流的语言。

(1) 信息世界中的基本概念

1) 实体 (Entity)。实体是客观存在可相互区别的事物，实体可以是具体的人、事、物，如一名教师、一名学生、一门课程；也可以是抽象的概念或联系，如学生的一次选课、学生的基本情况等都是实体。

2) 属性 (Attribute)。实体所具有的某一特性称为属性。例如，要说明一个学生实体就可以抓住他的许多特性：姓名、学号、性别、出生年月等，这些属性组合起来可以完整的、全面的说明一个学生。例如，张楠，2004020312，男，1989，表明学生姓名是张楠的学生，学号是“2004020312”，是名男同学，1989年出生。

3) 联系。现实世界中，事物内部以及事物之间是有联系的，我们将它分为实体内部联系和实体之间的联系。实体内部联系是指组成实体各属性之间的联系。实体之间的联系是指不同实体集之间的联系，实体之间的联系可分为三类，如图 1-5 所示。

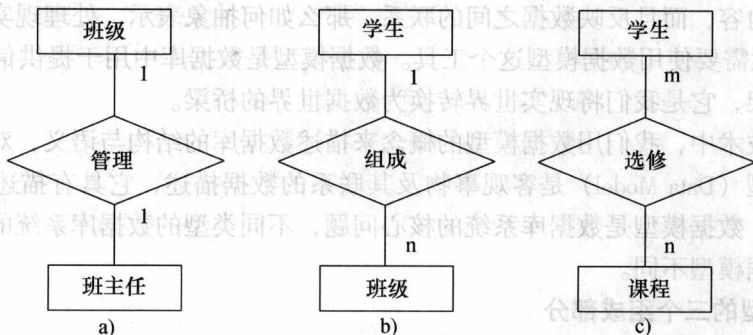


图 1-5 实体之间三类联系实例

a) 1:1 联系 b) 1:n 联系 c) m:n 联系

① 一对一联系 (1:1)。如果实体 A 中的每一个实体，实体集 B 中至多有一个（也可没有）实体与之联系，反之亦然，则实体 A 与实体 B 存在一对一联系，记为 1:1。

例如，学校里，一个班级只有一名班主任，而一名班主任只在一个班中任职，则班级与

班主任之间具有一对一联系。

②一对多联系 (1:n)。如果对于实体 A 中的每一个实体, 实体 B 中都有 n 个实体与之联系, 反之对于实体 B 中的每一个实体, 实体 A 中至多只有一个实体与之联系, 则实体 A 与实体 B 存在一对多联系。

例如, 一个班级中有若干名学生, 每个学生只在一个班级学习, 则班级与学生存在一对多联系。

③多对多联系 (m:n)

如果对于实体 A 中的每一个实体, 实体 B 中都有 n 个实体与之联系, 反之对于实体 B 中的每一个实体, 实体 A 中也有 m 个实体与之联系, 则实体 A 与实体 B 存在多对多联系。

例如, 一名学生可以选修多门课程, 一门课程可被多名学生选修, 则课程与学生之间具有多对多联系。

实际上, 一对一联系是一对多联系的特例, 而一对多联系又是多对多联系的特例。不论几个实体之间都存在以上三种关系。

3. 概念模型的表示方法

概念模型的表示方法很多, 其中最为著名的是实体-联系方法 (Entity-Relationship Approach)。该方法用 E-R 图来描述现实世界的概念模型, E-R 图也称为概念模型。

E-R 图提供了表示实体型、属性和联系的方法:

(1) 实体型 用矩形表示, 矩形框内写明实体名。

(2) 属性 用椭圆形表示, 椭圆形内写明属性名。

(3) 联系 用菱形表示, 菱形框内写明联系名, 用无向边分别与有关实体连接起来, 同时在无向边旁标上联系类型。

例 1-1 用 E-R 图表示实体及其联系 (见图 1-6)。

学生: 学号、姓名、性别、出生日期、系别。

课程: 课程号、课程名、学分、教师、上课时间、名额。

成绩: 学号、课程号、成绩。

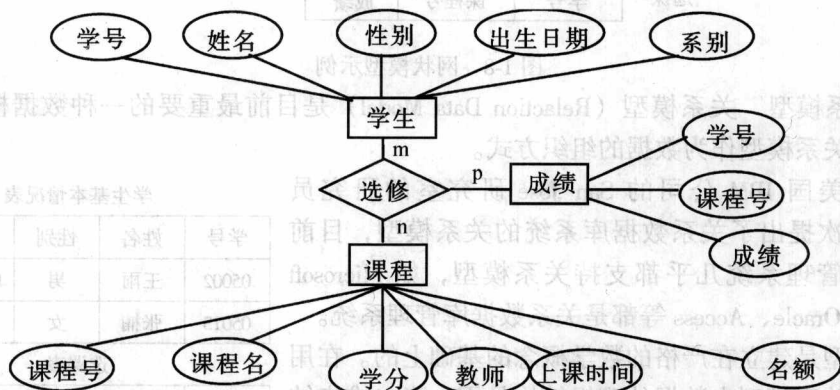


图 1-6 用 E-R 图表示实体及其联系

4. 常见数据模型

(1) 层次模型 层次模型的基本数据结构是层次结构, 也称树形结构。把满足以下两个

条件的基本层次联系的集合称为层次模型。

- 1) 有且只有一个结点没有双亲结点，这个结点称为根结点。
- 2) 根以外的其他结点有且只有一个双亲结点。

层次模型中每个结点表示一个实体类型，结点之间的连线表示实体间的联系。两个实体之间的联系可分为一对一、一对多和多对多三种联系，层次模型中表示的是实体间的一对多联系，使得层次数据库只能处理一对多的实体联系。层次模型中任何一个给定的记录值只有按其路径查看时才能看出它的全部意义。图 1-7 是一个采用层次模型的医院组织机构，对于医院这一实体，其下可以有门诊部、住院处、员工等多个子结点，而住院处又可设有病床、病人等子结点，它们都是一对多的联系。

(2) 网状模型 网状数据库采用网状模型作为数据的组织方式。在数据库中，把满足以下两个条件的基本层次联系的集合称为网状模型，图 1-8 为网状模型实例。

- 1) 已有一个以上的结点无父结点。
- 2) 至少有一个结点具有多于一个的父结点。

网状模型比层次模型更具有普遍性。它去掉了层次模型的两个限制，并允许两个结点之间有多种联系，可以更直接的描述现实世界，而层次模型实际上是网状模型的特例。层次模型与网状模型之间没有本质的区别，它们都用连线表示实体间的联系，它们都被称为格式化的数据模型。

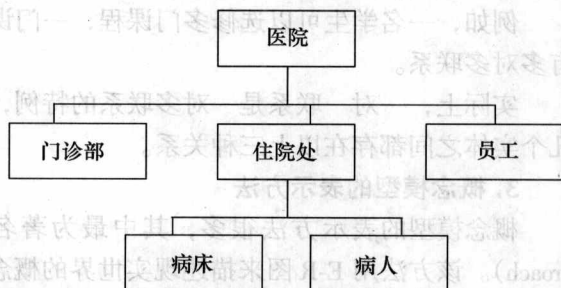


图 1-7 采用层次模型的医院组织机构



图 1-8 网状模型示例

(3) 关系模型 关系模型 (Relacion Data Model) 是目前最重要的一种数据模型。关系数据库采用关系模型作为数据的组织方式。

1970 年美国 IBM 公司的 San Jose 研究室的研究员 E.F.Codd 首次提出了关系数据库系统的关系模型，目前许多数据库管理系统几乎都支持关系模型，如 Microsoft SQL Server、Oracle、Access 等都是关系数据库管理系统。

关系模型是建立在严格的数学概念的基础上的。在用户观点下，关系模型中数据的逻辑结构是用一张二维表结构来表示实体及实体之间的联系，用关系（表）表示实体和实体之间的联系的模型称为关系数据模型，图 1-9 是关系模型实例。

学生基本情况表			
学号	姓名	性别	系别
05002	王雨	男	电子工程系
05013	张楠	女	人文系

选课表			
学号	课程名	学分	教师
05002	C 程序设计	2.5	孙华
05013	英语	3	李力

图 1-9 关系模型实例