

全国高等学校计算机教育研究会
课程与教材建设委员会
李大友 主编

组编

计算机等级考试辅导

(四级)

数据库系统原理

李爱中 周成云 马务充 编著



机械工业出版社

计 算 机 等 级 考 试 辅 导

(四 级)

数据库系统原理

全国高等学校计算机教育研究会
课 程 与 教 材 建 设 委 员 会 组 编

李大友 主编

李爱中 周成云 马务充 编著

机 械 工 业 出 版 社

本书以阐述关系数据库为主,先系统地讲述了数据库技术的基本原理和应用实践的主要内容,包括数据库的发展过程、关系数据库的基本概念和基本原理、SQL 语言、数据库的安全与保护、DBMS 的设计与实现、数据库应用系统的设计与实现,介绍了在微机上广泛使用的小型 DBMS 产品 FoxPro、在大中小及微机上广泛使用的大型 DBMS 产品 ORACLE 及面向对象方法和面向对象数据库;然后指出重点,同时对于难点进行了剖析,并附有大量例题及其答案,以及模拟试题。

本书内容强调直观易懂,便于自学,可作为高等学校计算机专业数据库课程的辅导材料,也可作为非计算机专业人员参加计算机等级考试(四级)的辅导材料。

图书在版编目(CIP)数据

数据库系统原理/李爱中等编著. —北京:机械工业出版社,1998.1

(计算机等级考试辅导:四级/李大友主编)

ISBN 7-111-05742-2

I. 数… II. 李… III. 数据库系统-理论 IV. TP311.13

中国版本图书馆 CIP 数据核字(97)第 29261 号

出 版 人:马九荣(北京市百万庄大街 22 号 邮政编码 100037)

责任编辑:何文军 版式设计:张世琴 责任校对:申春香

封面设计:姚学峰 责任印制:路琳

中国建筑工业出版社密云印刷厂印刷·新华书店北京发行所发行

1998 年 5 月第 1 版第 1 次印刷

787mm×1092mm^{1/16}·9.75 印张·225 千字

0 001—5 000 册

定价:16.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

前 言

数据库是计算机科学的重要分支，是信息系统的基础。数据库管理系统是管理数据和信息的有效工具。现在，不论大型计算机、中小型计算机、工作站，还是微型计算机，都配有数据库管理系统。关系数据库管理系统是当今数据库技术的主流。近年来，数据库技术在我国的应用蓬勃开展。为了配合国家教育委员会组织的计算机等级考试（四级），按其大纲要求编写了本书。

全书以关系数据库为核心，全面地概述了数据库的基本原理。全书共分 14 章。第 1 章为概述性引论，介绍了有关数据库的一些基本概念。第 2 章为数据库的基本存储结构。第 3 章介绍了关系数据库的基本概念。第 4 章介绍了标准的数据库语言——SQL，第 5 章介绍了关系数据库的规范化理论。第 6 章介绍了数据库的安全与保护。第 7 章介绍了数据库应用系统的设计。第 8 章介绍了数据库管理系统的设计与实现。第 9 章介绍了在微机上广泛使用的数据库管理系统 FoxPro。第 10 章介绍了在大中小型机上均广泛使用的大型数据库管理系统 ORACLE。第 11 章介绍了分布式数据库的基本原理。第 12 章介绍了 ORACLE 的分布式功能。最后，第 13 章，介绍了面向对象方法和面向对象数据库。

本书在编写过程中，首先回顾了数据库原理的主要内容，列出了重点内容，详细分析了难点；编拟了大量习题并附有答案；最后给出了四套模拟试题及其答案。

本书在编写过程中得到了北京工业大学李大友教授的指点和帮助，在此作者致以衷心的感谢。

本书第 1~3 章由李爱中编写，第 4~9 章由周成云编写，第 10~13 章由马务充编写，并由李爱中统编全书。

由于作者水平所限，书中可能存在不正确之处，恳请读者指正。

编 者

《计算机等级考试辅导》序言

当前,在世界范围内,一个以微电子技术、计算机技术和通信技术为先导的,以信息技术和信息产业为中心的信息革命方兴未艾。信息技术和信息产业的发展,对国民经济的发展、国家经济信息化起着举足轻重的作用,并已成为衡量一个国家发展水平的重要标志。因此,实现国家经济信息化,已成为世界各国所追求的共同目标。

为了使我国尽快实现国家经济信息化,赶上发达国家的水平,必须加速发展我国的信息技术和信息产业。其中最关键的环节就是人才的培养,尤其是计算机应用人才的培养。有了人才,才能迅速提高全社会的计算机应用水平,促进国家经济信息化水平的提高。因此,解决全民普及计算机知识,尽快提高全民族整体的计算机应用水平,已成为当务之急。各行各业、各层次人员,不论年龄与知识背景如何,都应掌握和应用计算机,解决其各自专业领域的计算机应用问题,为本职工作或专业服务,使其与国家经济信息化的需要相适当。

国家教委考试中心为适应这一形势发展的需要,使所培养的计算机应用人才的水平有一个公正的、客观的统一标准,推出了全国计算机等级考试。这一考试,根据应试者所具有的计算机应用能力水平的不同,划分为不同等级,分别进行考核。

全国计算机等级考试共分 为四级六类,其内容范围如下:

一级分为 A、B 两类,均面向文字处理和数据库应用系统操作人员。

一级 A 类要求掌握计算机基础知识、微机系统基本组成、操作系统功能和使用、字表处理软件的功能和使用、数据库应用系统的基本概念和操作。

一级 B 类要求掌握计算机基础知识、微机系统基本组成、DOS 操作系统基本知识及操作、文字处理软件 WPS 和数据库语言 FoxBASE 的操作。

二级面向使用高级语言进行程序设计的人员,要求掌握计算机基础知识、操作系统的功能和使用、数据库的基本概念及应用和具有使用一种高级语言(C 语言、PASCAL 语言、FORTRAN 语言、BASIC 语言或数据库语言)进行程序设计的能力。

三级分为 A、B 两类

三级 A 类面向测控领域的应用人员。要求掌握微机原理、汇编语言程序设计,微机接口技术、软件技术基础以及微机在测控领域的应用。

三级 B 类面向软件方面的应用人员。要求掌握计算机基础知识、数据结构与算法、操作系统、软件工程方法以及具有微机在管理信息系统或数值计算或计算机辅助设计方面的应用能力。

四级要求达到相当于大学计算机专业本科毕业生水平,具有计算机软件 and 硬件系统的设计开发能力。要求掌握计算机系统原理、计算机体系结构、计算机网络与通信、离散数学、数据结构与算法、操作系统、软件工程和数据库系统原理等方面的基础理论知识。

为推动全国计算机等级考试的健康发展,满足社会上对等级考试教材的迫切要求,全国高等学校计算机教育研究会课程与教材建议委员会组织了高等院校多年从事计算机教育的第一线专家教授,编写了《计算机等级考试教程》系列教材和《计算机等级考试辅导》系列丛

书，并得到国家教委考试中心和机械工业出版社的大力支持，使得这套教程和辅导能够及时与广大读者见面。

这套《计算机等级考试辅导》系列丛书是《计算机等级考试教程》系列教材的配套辅导材料。它针对计算机等级考试中的主要内容、重点、难点进行剖析，通过大量的例题分析和模拟试题，使读者能够正确掌握所学知识、技能，把握考试内容、方向，顺利通过考试。它既可以作为初学者自学《计算机等级考试教程》时的辅导材料，也可以作为学过该课程的考生考前复习、热身的教材，还可以作为各种培训班的培训材料。

由于计算机技术是一门迅速发展的学科及作者水平所限，这套教程和辅导肯定会有很多不足之处，衷心希望得到社会各界和广大读者的批评指正。

主编 李大友

目 录

《计算机等级考试辅导》序言

前 言

第 1 章 概述

1.1 数据管理的沿革	1
1.1.1 数据管理简述	1
1.1.2 数据库技术的发展	3
1.2 数据模型	3
1.2.1 数据模型的三要素	4
1.2.2 概念模型	5
1.2.3 三种主要的数据库模型	7
1.3 数据库系统的结构	10
1.3.1 数据库系统的三级模式结构	10
1.3.2 使用数据库的计算机系统结构	13
1.3.3 数据库管理系统	15
1.4 重点内容	16
1.5 难点剖析	18
1.6 习题及其答案	19
1.6.1 选择题	19
1.6.2 填空题	19
1.6.3 答案	20

第 2 章 数据库的存储结构

2.1 顺序组织	21
2.2 顺序带链的组织	21
2.3 带次关键字索引的顺序组织	21
2.4 多表组织	21
2.5 完全倒排组织	22
2.6 Hash 定址组织	22
2.7 联系的存储	22
2.8 重点内容	22
2.9 难点剖析	23
2.10 习题及其答案	23
2.10.1 选择题	23

2.10.2 填空题	23
2.10.3 答案	23

第 3 章 关系数据库

3.1 关系模型的基本概念	24
3.1.1 表格	24
3.1.2 表名、表头和表体	25
3.1.3 关系的数学定义	25
3.1.4 关系模型	27
3.2 关系数据库语言概述	29
3.3 关系代数	30
3.4 关系演算	31
3.4.1 元组关系演算	31
3.4.2 域关系演算	32
3.5 重点内容	33
3.6 难点剖析	34
3.7 习题及其答案	34
3.7.1 选择题	34
3.7.2 填空题	35
3.7.3 答案	35

第 4 章 关系数据库的标准语言——SQL

4.1 SQL 概述	36
4.2 SQL 的数据定义功能	36
4.2.1 基本表的定义与删除	37
4.2.2 索引的建立与删除	37
4.3 SQL 的数据操纵功能	38
4.3.1 SQL 查询语句	38
4.3.2 SQL 更新语句	38
4.4 视图	39
4.4.1 视图的定义与删除	39
4.4.2 视图的查询语句	40
4.4.3 视图的更新语句	40
4.4.4 视图的优点	40
4.5 SQL 的数据控制功能	40
4.6 嵌入式 SQL	41

4.6.1 不用游标的 DML 语句	42	6.3.1 基本概念	65
4.6.2 使用游标的 DML 语句	42	6.3.2 封锁	67
4.7 SQL 的事务处理功能	43	6.3.3 活锁与死锁	68
4.7.1 事务处理的概述	43	6.3.4 可串行性	69
4.7.2 SQL 语言的事务 处理语句	44	6.3.5 两段锁协议	70
4.8 重点内容	44	6.4 恢复	70
4.9 难点剖析	46	6.4.1 故障的种类	71
4.10 习题及其答案	46	6.4.2 转储和恢复	71
4.10.1 选择题	46	6.4.3 日志文件	72
4.10.2 填空题	46	6.5 重点内容	73
4.10.3 答案	47	6.6 难点剖析	74
第 5 章 关系数据库的规范化理论		6.7 习题及其答案	74
5.1 问题的提出背景	48	6.7.1 习题	74
5.2 规范化理论	50	6.7.2 答案	74
5.2.1 函数依赖	50	第 7 章 数据库应用系统的设计	
5.2.2 关键字	51	7.1 数据库应用系统的设计 概述	75
5.2.3 范式	51	7.1.1 数据库和信息系统	75
5.2.4 2NF	52	7.1.2 数据库应用系统设计 的特点	75
5.2.5 3NF	53	7.1.3 数据库应用系统的设计 方法简述	76
5.2.6 BCNF	53	7.1.4 数据库应用系统的设 计步骤	76
5.2.7 多值依赖	54	7.2 需求分析	76
5.2.8 4NF	57	7.3 概念结构设计	77
5.3 模式的分解	58	7.3.1 数据抽象与局部视图 设计	78
5.3.1 模式分解的三个定义	58	7.3.2 视图的集成	79
5.3.2 分解的无损连接性与保持函 数依赖性	59	7.4 逻辑结构设计	80
5.4 查询优化	60	7.4.1 E-R 图向关系数据模型 的转换	80
5.4.1 查询优化概述	60	7.4.2 规范化理论的应用	81
5.4.2 优化的一般策略	60	7.5 数据库的物理设计	82
5.5 重点内容	61	7.6 数据库应用系统的实施 和维护	83
5.6 难点剖析	62	7.6.1 数据库数据的载入和应用程 序的开发	83
5.7 习题及其答案	62	7.6.2 数据库应用系统的试	
5.7.1 填空题	62		
5.7.2 答案	62		
第 6 章 数据库的安全与保护			
6.1 安全性	63		
6.2 完整性	64		
6.3 并发控制	65		

运行	83	9.2.4 表的互斥和共享	103
7.6.3 数据库应用系统的运行 和维护	84	9.2.5 Browse 窗口	103
7.7 重点内容	84	9.3 FoxPro2.5 的安装、启动 及退出	104
7.8 难点剖析	85	9.4 使用 FoxPro 数据库	104
7.9 习题及其答案	85	9.4.1 建立数据库	104
7.9.1 习题	85	9.4.2 RQBE 窗口	104
7.9.2 答案	86	9.4.3 报表的设计与制作	105
第8章 数据库管理系统的设计与实现		9.5 创建应用程序	105
8.1 DBMS 概述	87	9.5.1 建立数据库	105
8.1.1 DBMS 的目标	87	9.5.2 菜单生成器	105
8.1.2 DBMS 的基本功能	87	9.5.3 屏幕生成器	105
8.1.3 DBMS 与操作系统	88	9.5.4 项目管理器	105
8.1.4 DBMS 和第四代应用开发语言 环境	88	9.6 关于 FoxPro 的评述	105
8.2 DBMS 的系统结构	89	9.7 习题及其答案	106
8.2.1 DBMS 程序模块的组成	89	9.7.1 习题	106
8.2.2 DBMS 的层次结构	91	9.7.2 答案	107
8.3 语言处理	91	第10章 ORACLE 数据库管理系统介绍	
8.3.1 语言翻译处理层的任务和工作 步骤	91	10.1 ORACLE 的结构与 原理	108
8.3.2 解释方法	93	10.1.1 ORACLE 的总体结构	108
8.3.3 预编译方法	93	10.1.2 ORACLE 的逻辑结构	110
8.4 数据存取层	93	10.1.3 用户数据库对象	111
8.4.1 存取层的系统结构	94	10.1.4 数据字典	112
8.4.2 存取层的功能子系统	94	10.2 ORACLE 的 SQL、PL/SQL 与 SQL * PLUS	113
8.5 缓冲区管理	95	10.3 数据库系统的管理和 维护	114
8.6 数据库的物理组织	96	10.3.1 数据库的安装	114
8.7 重点内容	99	10.3.2 启停数据库	115
8.8 难点剖析	100	10.3.3 数据库的管理	116
8.9 习题及其答案	100	10.3.4 表空间的管理	117
8.9.1 习题	100	10.3.5 安全管理	117
8.9.2 答案	100	10.3.6 完整性约束	118
第9章 FoxPro 数据库管理系统介绍		10.3.7 ORACLE 的实用程序	118
9.1 FoxPro 简介	101	10.4 ORACLE 产品介绍	118
9.2 FoxPro 的基本原理	102	10.5 习题及其答案	119
9.2.1 FoxPro 的数据类型	102	10.5.1 习题	119
9.2.2 工作区	102	10.5.2 答案	119
9.2.3 View 窗口	103		

第 11 章 分布式数据库

11.1 分布式数据库的产生与发展	120
11.2 分布式数据库的分类	120
11.3 关于分布式数据库定义的讨论	122
11.4 分布式数据库与单一数据库的比较	122
11.5 分布式数据库管理系统 (DDBMS)	124
11.6 计算机网络概述	124
11.6.1 网络结构分类	125
11.6.2 协议	129
11.6.3 ISO/OSI 参考系统结构	129
11.7 习题及其答案	130
11.7.1 习题	130
11.7.2 答案	130

第 12 章 分布式 ORACLE 系统简介

12.1 分布式体系结构的 ORACLE	131
12.2 网络 ORACLE 的连接过程	131
12.3 网络 ORACLE 的使用	131
12.4 分布式查询	132
12.5 远程数据传输	134
12.6 应用程序中的分布式查询问题	135
12.7 习题及其答案	135
12.7.1 习题	135

12.7.2 答案	135
-----------------	-----

第 13 章 面向对象方法与面向对象数据库

13.1 面向对象技术的形成与发展	136
13.2 面向对象的方法学简介	136
13.2.1 程序设计方法学	136
13.2.2 面向对象的程序设计方法	137
13.3 面向对象的方法的基本概念与特征	138
13.3.1 对象	138
13.3.2 消息和方法	138
13.3.3 类和类的层次	138
13.3.4 继承性	139
13.3.5 封装性	139
13.3.6 多态性	139
13.3.7 动态聚束	140
13.4 面向对象的系统分析方法	140
13.5 面向对象设计和面向对象实现	140
13.6 面向对象数据库	141
13.7 习题及其答案	141
13.7.1 习题	141
13.7.2 答案	141

第 14 章 模拟试题

14.1 试题	142
14.2 答案	144

第1章 概 述

这一章将从数据管理技术的发展过程出发,介绍什么是数据库,什么是数据库系统以及一些相关的基本概念。本章的目的是使读者从宏观上了解数据库系统的总体概貌,并为学习后面的各章打下基础。

1.1 数据管理的沿革

1.1.1 数据管理简述

数据管理是数据处理的中心问题。数据管理是指对数据进行组织、分类、编码、存储、检索和维护等操作。对计算机而言,数据(Data)已不仅仅限于数值,数据随着应用的发展,数据本身的含义已变得极为广泛,包括图象、声音等。严格地说,数据是指所有通过编码能输入到计算机中并被计算机程序所处理的符号。文字也是数据。随着计算机硬件和软件的发展,数据管理经历了下列三个发展阶段:

- (1) 人工管理阶段 50年代中期以前。
- (2) 文件系统阶段 50年代后期至60年代中期。
- (3) 数据库系统阶段 从60年代后期开始至今。

数据库(DataBase——DB),是指有组织地、动态地存储在外存上的相互关联着的数据的集合。数据库具有下列特点:

(1) 面向组织的复杂的数据结构 举一个例子来说明这个问题。例如一个职工的人事记录文件,每一个记录记载一个职工的全部人事信息。若采用传统的等长格式记录的文件形式,记录格式如图1-1-1(注意,已简化)所示。

部 门	职 工 号	姓 名	性 别	年 龄	籍 贯	社 会 关 系
--------	-------------	--------	--------	--------	--------	------------------

图 1-1-1 记录格式

在社会关系一栏,各个职工所具有的信息量变化很大。若以等长记录形式存储,只能按信息量最多时设定该栏所需的存储空间,这样当栏内信量少而职工人数很多时,势必造成大量地浪费存储空间。因此,改用不等长记录或主记录与明细记录相结合的方法就形成了。在人事记录中规定,除社会关系一栏外,其它任意一栏对所有职工均是等长的。把这些等长的记录存放在一个主文件之中,而把社会关系也以等长多个记录存放在明细文件中。这样,职工的人事信息分为两个文件,其记录结构如图1-1-2所示。

部 门	职 工 号	姓 名	性 别	年 龄	籍 贯
--------	-------------	--------	--------	--------	--------

主文件结构

职 工 号	姓 名	关 系	状 况
-------------	--------	--------	--------

明细文件结构

图 1-1-2 人事数据库文件结构

这样每个职工在主文件中有一个记录,在明细文件中有若干个记录。注意,在明细文件中有一栏职工号,用于区分明细文件中的记录属于哪个职工。

(2) 数据冗余少, 容易扩充 由于在设计数据库一开始就从整体的观点来看待和描述数据。数据不再面向某一具体的应用, 而是面向整个系统。这样可以大大减小数据冗余, 节约存储空间, 减少数据存取时间, 同时也可以避免数据之间的不一致性, 保持数据的完整性。对数据库中的数据可以以灵活的方式加以应用。可以取数据库整体的各种子集用于不同的应用系统中。而且当应用需求变化时, 只需选取不同的子集便可满足新的需求。这样, 数据库具有容易扩充的特点。

(3) 数据和程序的独立性 数据库系统提供了两种映象功能。一个是数据的存储结构与逻辑结构之间的映象或转换功能; 另一个是数据的总体逻辑结构与某类应用所需的局部逻辑结构之间的映象功能。第一种映象使得当数据的存储结构 (或物理结构) 改变时, 数据的逻辑结构可以不变, 因此不必修改应用程序。这就是数据和程序之间的物理独立性。

一般情况下, 数据库系统中某一类应用所使用的是数据全体 (即数据库) 的子集, 而且各类应用对同一数据的使用要求也不一定相同。为此, 数据库系统通常提供局部数据结构的说明 (或定义) 功能。当局部数据结构改变时, 系统自动地提供对这些映象的改变 (上述第二种映象功能)。当总体逻辑结构改变时而局部逻辑结构不变的情况下, 程序员根据此局部逻辑结构所编写的程序就可以不必改变了。这就是数据和程序之间的逻辑独立性。

正是数据和程序的上述两种独立性, 便可以把数据的定义和描述, 从应用程序中分离出来。另外, 数据的存取由数据库管理软件完成, 用户不必考虑存取数据的细节, 从而简化了应用程序的编写, 大大地减少了应用程序调试和维护时的工作量。

(4) 统一的数据控制功能 数据库是系统中所有用户可共享的资源。在计算机科学中, 共享是并发 (concurrency) 的, 即多个用户同时使用数据库。为此, 系统必须提供下列三个方面的数据控制功能:

- 数据的安全性 (security) 数据的安全性是指保护数据以防止不合法的使用, 而造成数据的泄密和破坏。这就需要采取一些安全保密措施。例如, 系统使用了检查口令、定义用户保密级或数据库存取权限等机制。当用户对数据进行操作时, 系统自动地检查用户能否执行此项操作。只有通过检查后才执行所允许的操作。

- 数据的完整性 (integrity) 数据的完整性是指数据的正确性、有效性和相容性 (或一致性)。系统提供必要的功能, 以保证数据库中的数据在输入和使用过程中始终符合原来的完整性定义。例如: 月份是 1 至 12 之间的正整数、职工不能无性别、职工号码是唯一的。在一般情况下, 一个职工不能在两个以上的部门工作。

另外, 当计算机硬件或软件发生故障而破坏了数据的完整性, 或者用户对数据库中的数据进行了错误的操作时, 系统能够进行应急的善后处理, 把数据库恢复到正确的状态或损失最小的状态。

- 并发控制 (concurrent control) 当多个用户的并发进程同时存取或修改数据库时, 可能会发生互相干扰而造成得到错误的操作结果或死锁。因此, 为了确保数据完整性和避免死锁, 必须对多用户的并发操作加以控制和协调。

具有上述特征的数据库系统 (DataBase System——DBS) 主要由数据库、数据库管理系统 (DataBase Management System——DBMS) 和一些应用程序 (application program) 组成, 如图 1-1-3 所示。

数据库是存储在外存设备上的通用化的综合性数据集合。DBMS 是一个数据管理软件系

统,对数据库的建立、使用和维护进行统一的控制。应用程序是针对某项应用而开发的程序,它完成用户对数据操作。应用程序通过调用 DBMS 来实施具体的操作。DBMS 规定了应用程序的语法和语义。

1.1.2 数据库技术的发展

虽然目前难以考证“数据库”一词源于何处,但众所公认,C. W. Bachman 是数据库领域的早期开拓者之一。早在 1962 年,他就设计了 IDS (Integrated Data Store) 系统。该系统于 1963 年投入运行。IDS 的设计基础是用一个灵活的模式把不同类型的记录链接在一起,其目的在于更有效地使用直接存取设备上的数据。IDS 可以认为是数据库系统的雏形,对数据库的早期研究工作影响很大。为此, Bachman 于 1973 年获得了 ACM 颁发的图灵奖(计算机科学工作者所能得到最高级别的大奖)。

从 60 年代末期到 70 年代初期出现了下面三个重要的事件,奠定了数据库技术的坚实的理论基础,标志着数据库技术的日益成熟。这三个事件是:

(1) 数据系统语言协商会 (Conference On Data System Language——CODASYL) 下属的数据库任务组 (Data Base Task Group——DBTG) 对数据库方法进行了系统的研究,总结并形成了若干报告(即 DBTG 报告)。DBTG 报告所提出的方法是基于网状结构的。DBTG 报告奠定了数据库网状模型的基础。属于网状模型的知名数据库管理系统有: Univac 公司的 DMS1100 系统(1971 年), Honeywell 公司的 IDS/2 系统(1975 年)等等。今天世界上运转的数据库仍有一些是基于网状模型的,但数量已经很少了。

(2) 1969 年 IBM 公司开发了数据库管理系统的商品化软件 IMS (Information Management System)。IMS 的数据模型是层次结构的。

(3) 1970 年 IBM 公司 San Jose 研究所的 E. F. Codd 提出了数据库的关系模型,开创了数据库关系方法和关系数据库理论的研究。关系方法是当今数据库的主流,由于 E. F. Codd 在关系数据库方面的杰出工作,他于 1981 年获得了 ACM 图灵奖。1981 年 IBM 公司宣布了关系数据库管理系统的商品软件 SQL/DS。与此同时,加州大学伯克利分校研制了 INGRES 关系数据库管理系统并加以商品化。从此,关系方法从实验室走向了社会。故此,在计算机科学领域,把 70 年代称为数据库的时代。在 80 年代,几乎所有的新开发的数据库管理系统都是基于关系方法的。微机上的关系数据库管理系统的功能越来越强,性能越来越好,大大地促进了数据库的应用。现在数据库已遍及整个社会的各个领域。数据库是计算机应用的技术基础。没有数据库,就没有现在遍及整个社会的计算机应用。

1.2 数据模型

数据库存放的是某个企业、组织或部门的综合性数据。这些数据库不仅要反映数据本身的内容,而且更重要的是要反映数据之间的联系。那么,在数据库系统的形式化结构中如何抽象地表示和处理现实世界中的数据和信息呢?在数据库中,使用的是数据模型(Data Model)这个工具来对现实世界进行抽象的。数据模型是数据库系统中用于信息表示和提供操作手段的形式化工具。了解数据模型这个基本概念,是学习数据库的基础。

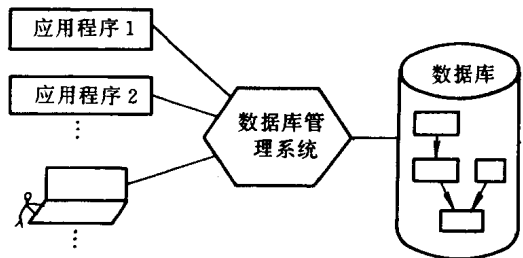


图 1-1-3 数据库系统的组成

不同的模型为数据和信息的模型化或形式化,提供了不同的工具。根据模型应用的目的的不同,可以把模型分成两类或两个层次:一个是概念模型(也称信息模型),另一个是数据模型(包括网状、层次和关系模型)。概念模型(Concept Model)或信息模型(Information Model)是按用户的观点来对数据和信息进行建模的工具,而数据模型则是按计算机系统的观点来对数据进行建模的工具。

概念模型用于直接对现实世界的信息建模。对于概念模型主要强调其语义的表达能力和使用的方便性,并尽可能直接地表达现实世界中的各种语义知识。概念模型应当是概念清晰、易于使用。概念模型是现实世界到信息世界的第一层抽象的结果,是用户和数据库设计人员进行交流的语言。数据模型用于对信息世界的信息进行第二层抽象。由于数据模型是面向机器世界的,一般要求数据模型有严格的语法和语义,另外加上一些限制或规定,以便于在机器上实现。故此,数据模型是通过一种严格规定了语法和语义的形式化语言来描述的,人们可以利用该语言来定义和操作数据库中的数据。

下面首先介绍数据模型的3个基本组成部分和概念模型。然后,在下一节介绍3种常用的数据模型。

1.2.1 数据模型的三要素

一般来讲,数据模型是严格定义的概念的集合。这些概念精确地描述系统的静态特性、动态特性和完整性约束条件。因此,数据模型通常由数据结构、数据操作和完整性约束条件三个部分组成。

(1) 数据结构 数据结构是所研究的对象类型(Object Type)的集合。这些对象和对象类型是数据库的组成成分。一般可分为两类:一类是与数据类型、内容和其它性质有关的对象;一类是与数据之间的联系有关的对象。前者如网状模型中的数据项和记录,关系模型中的域、属性和关系等。后者如网状模型中的关系模型(set type)。

在数据库领域中,通常按照数据结构的类型来命名数据模型,进而对数据库管理系统进行分类。如层次结构、网状结构和关系结构的数据模型分别称作为层次模型、网状模型和关系模型。相应地,数据库分别称作为层次数据库、网状数据库和关系数据库。

(2) 数据操作 数据操作是指对各种对象类型的实例(或值)所允许执行的操作的集合,包括操作及有关的操作规则。在数据库中,主要的操作有检索和更新(包括插入、删除、修改)两大类。数据模型定义了这些操作的语义、语法(即使用这些操作时所用的语言)。

数据结构是对系统静态特性的描述,而数据操作是对系统动态特性的描述。两者既有联系,又有区别。

(3) 数据的约束条件 数据的约束条件是完整性规则的集合。完整性规则是指在给定的数据模型中,数据及其联系所具有的制约条件和依存条件,用以限制符合数据模型的数据库的状态以及状态的变化,确保数据的正确性、有效性和一致性。例如,在关系模型中,任何关系必须满足实体完整性和参照完整性两个条件(在以后的章节里将详细讨论)。

另外,数据模型还应该提供定义完整性约束条件的机制,以反映须遵守的约束条件。例如,在学校的数据库中,规定本科生的年龄不得超过30岁,学习成绩累计不得超过3门不及格等。

数据模型的这三个方面的内容完整地刻画了一个数据模型。而模型结构是描述模型性质的最重要的方面。因此,在下面的两节中,为了使读者对模型有一个基本的了解,先重点介

绍模型的结构方面的内容。关于操作和约束将在后面的有关章节中讨论。

1.2.2 概念模型

数据模型是数据库系统的核心和基础。每个 DBMS 软件都是基于某种数据模型的。为了把现实世界中的具体事物或事物之间的联系表示成 DBMS 所支持的数据模型，人们首先必须将现实世界的事物及其之间的联系进行抽象，转换为信息世界的概念模型；然后将信息世界的概念模型转换为机器世界的数据模型。也就是说，首先把现实世界中的客观对象抽象成一种信息结构。这种信息结构并不依赖于具体的计算机系统和 DBMS。然后，再把概念模型转换为某一计算机系统上某一 DBMS 所支持的数据模型。因此，概念模型是从现实世界到机器世界的一个中间层次，如图 1-2-1 所示。

现实世界的事物反映到人的大脑之中，然后人们开始认识这些事物，经过选择、命名、分类和组织等抽象工作之后形成概念模型，并进入到信息世界。

- 用户 (user) 关心的是现实世界中的事物、事物的属性及其相互关系。例如，用户可能关心他的顾客及其属性，如顾客地址、银行帐号等等。用户也关心自己的定货帐目，如谁订的货、订的什么和订多少等等。

- 系统分析员 (analyst) 同样也关心现实世界，但是系统分析员需要分析用户的信息需求。作为需求分析的结果，分析员必须以文档的形式对需求进行结构化的描述；这个文档就是信息模型。关于系统分析的方法，将在后继章节中介绍。下面先介绍一些基本的概念。

- 实体 (Entity) 实体是构成数据库的基本元素。实体是指一个存在的东西以区别这个东西所具有的属性和这个东西与其它东西的联系。实体可以是人，也可以是物；可以是实际对象，也可以是概念；可以是事物本身，也可以是指事物之间的联系。例如，一个学生，某部门的人数。

- 属性 (Attribute) 一个实体可以由若干个属性来刻画。属性是相对实体而言的，是实体所具有的特性。例如，学生这个实体具有学号、姓名、年龄、性别、系、年级等属性 (7803066, 张三, 19, 男, 计算机, 三年级)。这些属性值组合起来表征了一个学生实体 (的值)。

- 关键字 (Key) 能唯一地标识实体的属性的集合称为关键字 (或码)。例如，学号可以作为学生实体的关键字；若无重名现象发生，在同一个年级里，姓名和年级也可以作为学生实体的关键字。

- 域 (Domain) 属性的取值范围称作域。例如，学号的域为 7 位整数，年龄域为小于 35 的正整数，性别的域为 {男, 女}。域的名称叫作型，域中的元素叫作值或型的一次出现。

- 实体型 (Entity Type) 一类实体所具有的共同特征或属性的集合称为实体型。一般用实体名及其属性名来抽象地刻画一类实体的实体型。例如，学生 (学号, 姓名, 年龄, 系, 年级) 是一个实体型，其中学生是实体型的名，学号、姓名等为属性名。而 7803066 是属性型学号的一个值 (或一次出现)，(7803066, 张三, 19, 男, 计算机, 三年) 为实体型 (学号, 姓名, 年龄, 系, 年级) 或学生的一个值 (或一次出现)、名 (或型)，是用来刻画或命名一个集合的，而值是该集合中的一个元素。

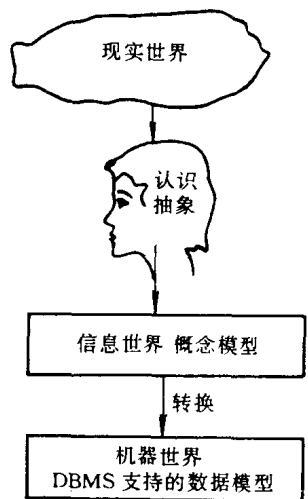


图 1-2-1 三个世界的划分及其关系

• **实体集 (Entity Set)** 同型实体的集合叫实体集。例如, 学生就是一个实体集。实体集的名即是实体型。对于学生和 (学号, 姓名, 年龄, 系, 年级) 均是实体型, 而学生是对实体型 (学号, 姓名, 年龄, 系, 年级) 所起的名称, 两者是指同一客观对象。但本科生和研究生可以为相同实体型, 而实体集不同。

• **联系 (Relationship)** 现实世界的事物之间是有联系的。一般存在两类联系: 一是实体内部的组成实体的属性之间的联系, 二是实体之间的联系。在考虑实体内部的联系时, 是把属性看作为实体。

一般来说, 两个实体之间的联系可分为三种:

(1) **一对一 (1:1) 联系** 若对于实体集 A 中的每一个实体, 实体集 B 中至多有唯一的一个实体与之联系, 反之亦然, 则称实体集 A 与实体集 B 具有一对一联系, 记作 1:1, 用图 1-2-2 表示。

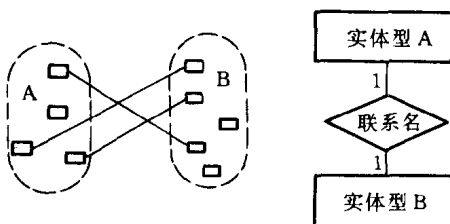


图 1-2-2 1:1 联系及其 E-R 图

(2) **一对多 (1:n) 联系** 若对于实体集 A 中的每个实体, 实体集 B 中有几个实体 ($n \geq 0$) 与之联系; 反之, 对于实体集 B 中的每一个实体, 实体集 A 中至多只有一个实体与之联系, 则称实体集 A 与实体集 B 有一对多联系, 记为 1:n, 用图 1-2-3 表示。

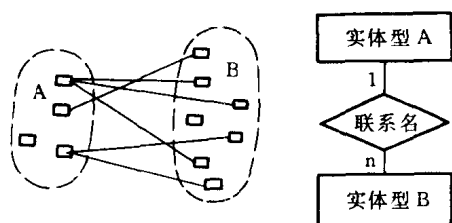


图 1-2-3 1:n 联系及其 E-R 图

相应地有多对一 ($n:1$) 联系 多对一联系, 从本质上说, 是一对多联系的逆转。其定义同 一对多联系类似, 不再赘述。

(3) **多对多 ($m:n$) 联系** 若对于实体集 A 中的每一个实体, 实体集 B 中有 n 个实体 ($n \geq 0$) 与之联系; 反之, 对于实体集 B 中的每一个实体, 实体集 A 中也有 m 个实体 ($m \geq 0$) 与之对应, 则称实体集 A 与实体集 B 具有多对多联系, 记作 $m:n$, 用图 1-2-4 表示。实质上, 多对多联系是任意一种联系。

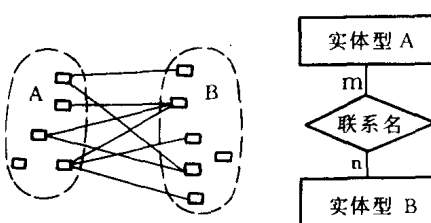


图 1-2-4 $m:n$ 联系及其 E-R 图

例如, 部门和经理之间有 1:1 联系; 一个部门有一个经理, 一个经理只在一个部门任职。部门和职工之间有 1:n 联系; 一个部门有多个职工, 而一个职工只在一个部门工作。项目和职工之间有 $m:n$ 联系; 一个项目有多个职工参加, 一个职工可以参加多个项目的工作。注意, 职工和部门之间有 $n:1$ 联系。

另外, 同一实体集内的各个实体间也可以有各种联系。例如职工实体中具有领导和被领导的联系。一个职工可领导若干名其他职工, 一个职工只被另一个职工领导, 因此这是一个 1:n 联系, 如图 1-2-5 所示。

概念模型的表示方法最常用的是实体—联系方法 (Entity-Relationship Approach), 简称 E—R 方法。该方法是由 P. P. S. Chen 在 1976 年提出的。E—R 方法用 E—R 图来描述某一组织的概念模型。在这里仅介绍 E—R 图的要点。有关如何认识、分析现实世界, 并从中抽取实体、联系及其建立概念模型的方法, 将在数据库应用系统设计一章中详细介绍。

在 E—R 图中:

(1) 长方形框表示实体集, 框内写上实体型的名称。

(2) 用椭圆框表示实体的属性, 并用有向边把实体框及其属性框连接起来。

(3) 用菱形框表示实体间的联系, 框内写上联系名, 用无向边把菱形框及其有关的实体框连接起来, 在旁边标明联系的种类。如果联系也具有属性, 则把属性框和菱形框也用无向边连接上。

例如, 学生实体具有学号、姓名、年龄、性别、系、年级六个属性; 用 E—R 图表示如图 1-2-6 所示。

例如, 在图 1-2-7 中, 可以使用供应量作为供应联系的属性, 表示某供应商供应给某个部门多少数量的某种零件。

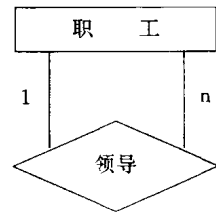


图 1-2-5 同一实体集内的联系

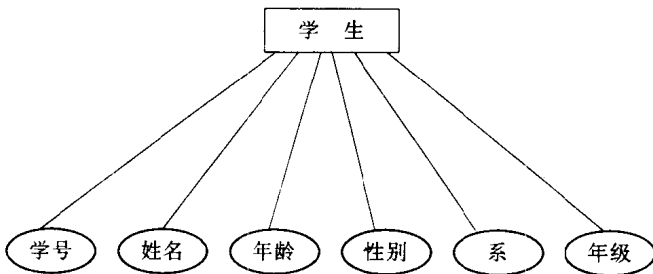


图 1-2-6 学生实体及其属性

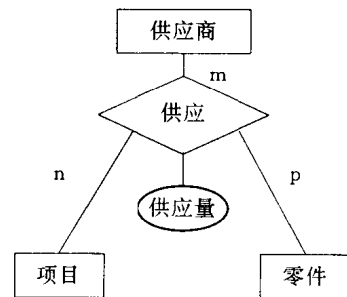


图 1-2-7 实体间联系及联系的属性

1.2.3 三种主要的数据库模型

实际 DBMS 所支持的数据库模型主要有三种:

- 层次模型 (Hierarchical Model)
- 网状模型 (Network Model)
- 关系模型 (Relational Model)

其中, 关系模型是当前 DBMS 所支持的数据库模型的主流。90 年代运行的 DBMS 几乎都是基于关系模型的。层次模型和网状模型统称为非关系模型。非关系模型的数据结构可以和图论中的图相对应, 比较直观, 但在理论上不完备, 实现效率较低, 故此目前很少用。本书只作概括的介绍。但是最近, 层次模型在研究面向对象的 DBMS 中已得到重视。有兴趣的读者可参考 C. J. 戴特的《数据库系统导论》或其它书籍中关于层次模型和网状模型的详细论述。

在关系模型中, 数据在用户的观点中 (或在用户视图中) 的逻辑结构是一张二维表 (Table)。下面我们以学生学籍表为例, 介绍关系模型中的一些主要术语, 如图 1-2-8 所示。

- 关系 (Relation), 对应于平常讲的一张表。
- 元组 (Tuple), 表中的一行。