

21 世纪 高等学校本科系列教材

总主编 吴中福

大型数据库技术及应用

(33)

王 越 刘加伶 李 梁 编著



13-43

重庆大学出版社

清华大学出版社

大型数据库技术及应用

第二版

王 珊 陈明勇 李 强 编著



f62

TP311.13-43
12386

大型数据库技术及应用

王 越 刘加伶 李 梁 编著



A0954742

重庆大学出版社

内容简介

本书以 Oracle8i For windows NT 为蓝本,深入浅出地介绍了数据库基础知识,Oracle8i 系统管理数据库设计以及如何编写 Oracle8i 应用程序。

全书从介绍 Oracle8i 的基本概念及结构入手,讨论问题的求解方法、通过使用工具归纳使用 Oracle8i 的方法与技巧。

本书适合本科及专科计算机科学与技术、信息管理与信息系统的数据库设计课程的教材,同时也适合初学者,数据库管理员和数据库应用开发人员使用。

图书在版编目(CIP)数据

大型数据库技术及应用/王越,刘加伶,李梁编著. 重庆:
重庆大学出版社,2001.7

计算机科学与技术本科系列教材

ISBN 7-5624-2319-9

I. 大... II. ①王... ②刘... ③李... 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(2001)第 25006 号

大型数据库技术及应用

王 越 刘加伶 李 梁 编著

责任编辑 曾令维 周 立

*

重庆大学出版社出版发行

新华书店经销

重庆大学建大印刷厂印刷

*

开本:787×1092 1/16 印张:22 字数:549 千

2001 年 8 月第 1 版 2001 年 8 月第 1 次印刷

印数:1—6000

ISBN 7-5624-2319-9/TP·285 定价:30.00 元

前言

Oracle8i 数据库一直处于数据库技术的领先者的地位,这主要归功于其强大的功能。从技术的角度讲,它把数据库从基本的关系模型引入到对象关系模型,人们可以从传统的关系模型转到对象模型的应用上,可以使开发的系统更加有效和快捷,Oracle8i 支持单机到大型机。对于管理信息系统专业和计算机科学与技术专业的学生来说,掌握大型数据库的应用,把握当今数据库发展的潮流,既是对数据库原理课程的补充,也是对学生动手能力的一个锻炼,是学生走向工作岗位不可缺少的一门技术。

全书共分 12 章,第 1 章绪论,主要介绍数据库的基本概念及数据库系统的基本原理。第 2 章 Oracle8i 的体系结构,主要介绍 Oracle8i 的数据库结构和其基本运行机制。第 3 章 Oracle8i 操作初步,主要介绍 Oracle8i 数据库的安装和基本配置。第 4 章,设计与创建数据库,主要介绍 Oracle8i 数据库设计的基本方法以及设计步骤。第 5 章,表和表空间操作主要介绍 Oracle8i 表空间管理,表管理等方面的知识。第 6 章,Oracle8i 的对象管理,主要介绍 Oracle8i 索引操作,视图操作以及用户对象管理的操作。第 7 章,Oracle8i 数据保护,主要介绍 Oracle8i 数据库中的事务的概念,并发控制,系统控制,会话控制,日志、回滚段和控制文件管理等内容。第 8 章,Oracle8i 安全与审计,主要介绍 Oracle8i 数据库的用户管理,系统级权限管理,对象级权限管理,角色,审计等概念。第 9 章,备份与恢复,主要介绍 Oracle8i 数据库的恢复原理,脱机备份与恢复,逻辑备份与恢复,热备份与恢复等概念。第 10 章 PL/SQL 程序设计,主要介绍 PL/SQL 的控制结构,过程和函数的程序设计,包程序设计,触发器程序设计,异常处理。第

11 章, Oracle8i WEB 程序设计, 主要介绍制作 WEB 页的基本方法, HTML 概述, 为 WEB 页增加色调, HTML 基本标记应用, HTML 高级标记应用。Oracle8i WEB 出版助手, WEB 出版助手操作初步, 修改、删除、生成和预览 WEB 页, 使用模板生成 WEB 页。第 12 章, 大型数据库应用实例, 主要介绍了医院管理系统数据库设计实例, 商业连锁店管理系统数据库设计实例, 财务管理系统数据库设计实例。

全书由王越审定, 本书第 1 章, 第 2 章, 第 3 章, 第 4 章, 第 5 章, 第 6 章, 第 7 章, 第 11 章由王越编写, 第 10 章由徐帆编写, 第 8 章, 第 9 章由刘加伶编写, 第 12 章由李梁编写。另外, 王仙, 钟静, 秦臻, 王俭林对完成本书的内容输入及资料收集给予了巨大的帮助, 在这里表示由衷的谢意。

目 录

第1章 绪 论	1
1.1 引言	1
1.2 数据模型	5
1.3 数据库系统的结构	15
1.4 数据库管理系统	19
小 结	23
第2章 Oracle8i 的体系结构	24
2.1 Oracle8i 数据库结构	24
2.2 Oracle8i 基本结构	27
2.3 Oracle8i 系统结构	30
小 结	35
第3章 Oracle8i 操作初步	36
3.1 Oracle8i FOR WINDOWS NT 安装与配置	36
3.2 Oracle8i 工具操作初步	45
3.3 SQL 概述	55
3.4 SQL * PLUS	57
小 结	69
第4章 设计与创建数据库	70
4.1 数据库设计概述	70
4.2 创建 Oracle8i 数据库	75
小 结	84
第5章 表和表空间操作	85
5.1 表空间管理	85
5.2 表管理	89
5.3 借助模式管理器进行表操作	102
小 结	103
第6章 操作 Oracle8i 的对象	104
6.1 索引操作	104
6.2 聚簇操作	109

6.3	视图操作	112
6.4	其他数据库对象操作	114
6.5	用模式管理器操作对象	118
	小 结	132
第 7 章	Oracle8i 数据保护	134
7.1	事务控制	134
7.2	并发控制	137
7.3	会话控制	138
7.4	系统控制	141
7.5	日志、回滚段和控制文件管理	144
	小 结	149
第 8 章	Oracle8i 安全与审计	150
8.1	用户管理	150
8.2	系统级权限管理	151
8.3	对象级权限管理	153
8.4	角色	155
8.5	配置文件	158
8.6	使用安全管理器	160
8.7	审计	169
	小 结	171
第 9 章	备份与恢复	172
9.1	数据库恢复综述	172
9.2	Oracle8i 数据库备份	174
9.3	Oracle8i 数据库恢复	177
9.4	数据库的恢复和实现	182
	小 结	187
第 10 章	PL/SQL 程序设计	188
10.1	PL/SQL 概述	189
10.2	PL/SQL 数据类型	193
10.3	PL/SQL 变量和常量	204
10.4	PL/SQL 表达式	209
10.5	PL/SQL 注释	210
10.6	PL/SQL 控制结构	212
10.7	PL/SQL 例外处理	225
10.8	PL/SQL 游标处理	232
10.9	过程和函数的程序设计	242
10.10	PL/SQL 包	255
10.11	数据库触发器	259
	小 结	265

第 11 章	Oracle8i WEB 程序设计	266
11.1	制作 WEB 页的基本方法	266
11.2	Oracle8i WEB 出版助手	275
小 结		283
第 12 章	大型数据库应用实例	284
12.1	信息系统应用中数据库设计概述	284
12.2	会计信息系统的数据库设计	301
12.3	医院管理系统数据库设计实例	317
参考文献		343

第 1 章

绪 论

数据库技术产生于 20 世纪 60 年代中期,是数据管理的最新技术,是计算机科学的重要分支,它的出现极大地促进了计算机应用向各行各业的渗透。本章将介绍数据库的有关概念以及为什么要发展数据库技术,从中可以看出数据库技术的重要性。

1.1 引 言

1.1.1 数据、数据库、数据库系统、数据库管理系统

数据、数据库、数据库系统、数据库管理系统是与数据库技术密切相关的 4 个基本概念。

(1) 数据(data)

说起数据,人们首先想到的是数字。其实数字只是最简单的一种数据。数据的种类很多,在日常生活中数据无所不在,如文字、图形、图像、声音。

为了认识世界,交流信息,人们需要描述事物。数据实际上是事物的符号记录。在日常生活中人们直接用自然语言(如汉语)描述事物。在计算机中,为了存储和处理这些事物,就要抽出对这些事物感兴趣的特征组成一个记录来描述。例如,在学生档案中,如果人们最感兴趣的是学生的姓名、性别、出生年月、入学时间,那么可以这样描述:

李明,男,1972,江苏人,1990 年考入计算机系,而不了解其语义的人则无法理解其含义。可见,数据的形式本身并不能完全表达其内容,需要经过语义解释。

(2) 数据库(database,简称 DB)

收集并抽取出一个应用所需要的大量数据之后,应将其保存起来以供进一步加工处理和抽取有用信息。保存方法有很多种:人工保存、存放在文件里、存入在数据库里,其中数据库是存放数据的最佳场所,其原因将在 1.1.2 节中介绍。

所谓数据库就是长期储存在计算机内、有组织的、可共享的数据集合。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,并可为各种用户共享。

(3) 数据库管理系统(database management system, 简称 DBMS)

收集并抽取出一个应用所需要的大量数据之后,如何科学地组织这些数据并将其存储在数据库中,又如何高效地处理这些数据呢?完成这个任务的是一个软件系统——数据库管理系统。

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。

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、多用户对数据的并发使用及故障发生后的系统恢复。

(4) 数据库系统(database system, 简称 DBS)

数据库系统是指在计算机系统中引入数据库后的系统构成,一般由数据库、数据库管理系统(及其开发工具)、应用系统、数据库管理员和用户构成。应当指出的是,数据库的建立、使用和维护等工作只靠一个 DBMS 远远不够,还要有专门的人员来完成,这些人称为数据库管理员(database administrator, 简称 DBA)。

在不引起混淆的情况下人们常常把数据库系统简称为数据库。

1.1.2 数据库技术的产生与发展

数据库技术是应数据管理任务的需要而产生的。

数据管理是指对数据如何进行分类、组织、编码、储存、检索和维护,它是数据处理的中心问题。随着计算机硬件和软件的发展,数据管理经历了人工管理、文件系统和数据库系统 3 个发展阶段。

(1) 人工管理阶段

在 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;软件状况是,没有操作系统,没有管理数据的软件,数据处理方式是批处理。

人工管理数据具有以下特点:

①数据不保存。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。

②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。应用程序中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。

③数据不共享。数据是面向应用的,一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时,由于必须各自定义,无法互相利用、互相参照,因此程序与程序之间有大量冗余数据。

④数据不具有独立性。数据的逻辑结构或物理结构发生变化后,必须对应用程序做相应的修改,这就进一步加重了程序员的负担。

(2) 文件系统阶段

20 世纪 50 年代后期到 60 年代中期,计算机的应用范围逐渐扩大,计算机不仅用于科学计算,而且还大量用于管理。这时硬件上已有了磁盘、磁鼓等直接存取存储设备;软件方面,操

作系统中已经有了专门的数据管理软件,一般称为文件系统,处理方式不仅有了文件批处理,而且能够联机实时处理。

用文件系统管理数据具有如下特点:

①数据可以长期保存。由于计算机大量用于数据处理,数据需要长期保留在外存上,反复进行查询、修改、插入和删除等操作。

②由专门的软件即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,使应用程序与数据之间有了一定的独立性,程序员可以不必过多地考虑物理细节,将精力集中于算法。而且数据在存储上的改变不一定反映在程序上,大大节省了维护程序的工作量。

③数据共享性差。在文件系统中,一个文件基本上对应于一个应用程序,即文件仍然是面向应用的。当不同的应用程序具有部分相同的数据时,也必须建立各自的文件,而不能共享相同的数据,因此数据的冗余度大,浪费存储空间。同时由于相同数据的重复存储、各自管理,给数据的修改和维护带来了困难,容易造成数据的不一致性。

④数据独立性低。文件系统中的文件是为某一特定应用服务的,文件的逻辑结构对该应用程序来说是优化的,因此要想对现有的数据再加一些新的应用会很困难,系统不容易扩充。一旦数据的逻辑结构改变,必须修改应用程序,修改文件结构的定义。而应用程序的改变,例如,应用程序改用不同的高级语言等,也将引起文件的数据结构的改变。因此数据与程序之间仍缺乏独立性。可见,文件系统仍然是不具有弹性的无结构的数据集合,即文件之间是孤立的,不能反映现实世界事物之间的内在联系。

(3) 数据库系统阶段

20世纪60年代后期以来,计算机用于管理的规模更为庞大,应用越来越广泛,数据急剧增长,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。这时硬件已有大容量磁盘,硬件价格下降,软件价格上升,为编制和维护系统软件及应用程序所需的成本相对增加,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。在这种背景下,以文件系统作为数据管理手段已经不能满足应用的需求,于是为解决多用户、多应用共享数据的需求,使数据尽可能多地为应用服务,就出现了数据库技术,出现了统一管理数据的专门软件系统——数据库管理系统。

用数据库系统来管理数据具有如下特点:

1) 数据结构化 数据结构化是数据库与文件系统的根本区别。

在文件系统中,相互独立文件的记录内部是独立的。传统文件的最简单形式是等长同格式的记录集合。文件系统尽管其记录内部已经有了某些结构,但记录之间没有联系。数据库系统实现整体数据的结构化,这是数据库的主要特征之一,也是数据系统与文件系统的本质区别。

在数据库系统中,不仅数据是结构化的,而且存取数据的方式也很灵活,可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。而在文件系统中,数据的最小存取单位是记录,粒度不能细到数据项。

2) 数据的共享性好,冗余度低 数据的共享程度直接关系到数据的冗余度。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。这样既可以大大减少数据冗余,节约存储空间,又能够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拷贝的值不一样。采用人工管理或文件系统管理时,由于数据被重复存储,当不同的应用和修改不同的拷贝时就易造成数据的不一致。

3) 数据的独立性高 数据库系统提供了两方面的映像功能,从而使数据既具有物理独立性,又有逻辑独立性。

数据库系统的一个映像功能是数据的总体逻辑结构与某类应用所涉及的局部逻辑结构之间的映像或转换功能。这一映像功能保证了数据的总体逻辑结构改变时,通过对映像的相应改变可以保持数据的局部逻辑结构不变,由于应用程序是依据数据的局部逻辑结构编写的,所以应用程序不必修改。这就是数据与程序的逻辑独立性,简称数据的逻辑独立性。

数据库系统的另一个映像功能是数据的存储结构与逻辑结构之间的映像或转换功能。这一映像功能保证了当数据的存储结构(或物理结构)改变时,通过对映像的相应改变可以保持数据的逻辑结构不变,从而应用程序也不必改变。这就是数据与程序的物理独立性,简称数据的物理独立性。

数据与程序之间的独立性,可以把数据的定义和描述从应用程序中分离出去。另外,由于数据的存取由 DBMS 管理,用户不必考虑存取路径等细节,从而简化了应用程序的编制,大大减少了应用程序的维护和修改。

4) 数据由 DBMS 统一管理和控制 由于对数据实行了统一管理,而且所管理的是有结构的数据,因此在使用数据时有很灵活的方式,可以取整体数据的各种合理子集用于不同的应用系统,而且当应用需求改变或增加时,只要重新选取不同子集或者加上一小部分数据,便可以有更多的用途,满足新的要求,因此使数据库系统弹性大,易于扩充。

除了管理功能以外,为了适应数据共享的环境,DBMS 还必须提供以下几方面的数据控制功能。

- 数据的安全性(security)

数据的安全性是指数据保护,防止不合法使用数据造成数据的泄密和破坏,使每个用户只能按规定对某些数据以某些方式进行访问和处理。

- 数据的完整性(integrity)

数据的完整性指数据的正确性、有效性和相容性。即将数据控制在有效的范围内,或要求数据之间满足一定的关系。

- 并发(concurrency)控制

当多用户的并发进程同时存取、修改数据库时,可能会发生相互干扰而得到错误的结果,并使得数据库的完整性遭到破坏,因此必须对多用户的并发操作加以控制和协调。

- 数据库恢复(recovery)

计算机系统的硬件故障、软件故障、操作员的失误以及故意的破坏也会影响数据库中数据的正确性,甚至造成数据库部分或全部数据的丢失。DBMS 必须具有将数据库从错误状态恢复到某一已知的正确状态(也称为完整状态或一致状态)的功能,这就是数据库的恢复功能。

综上所述,数据库是长期存储在计算机内有组织的大量的共享数据集合。它可以供各种用户共享,具有最小冗余度和较高的数据独立性。DBMS 在数据库建立、运用和维护时对数据库进行统一控制,以保证数据的完整性、安全性,并在多用户同时使用数据库时进行并发控制,在发生故障后对系统进行恢复。

数据库系统的出现使信息系统的研制从以加工数据的程序为中心转向围绕共享的数据库

来进行。这样既便于数据的集中管理,又有利于应用程序的研制和维护,提高了数据的利用率和相容性,提高了决策的可靠性。

数据库技术从20世纪60年代中期产生到今天仅仅30年的历史,但其发展速度之快,使用范围之广是其他技术所不及的。60年代末出现了第一代数据库——网状数据库、层次数据库,70年代出现了第二代数据库——关系数据库。目前关系数据库系统已逐渐淘汰了网状数据库和层次数据库,成为当今最为流行的商用数据库系统。而80年代出现的以面向对象模型为主要特征的数据库系统又向关系数据库系统挑战。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、并行计算机技术等互相渗透,互相结合,成为当前数据库技术发展的主要特征。

1.1.3 数据库技术的研究领域

目前虽然已有了一些比较成熟的数据库技术,但随着计算机硬件的发展和应用范围的扩大,数据库技术也需要不断向前发展,概括地讲,当前数据库学科的主要研究范围有以下3个领域:

(1) 数据库管理系统软件的研制

数据库管理系统 DBMS 是数据库系统的基础。DBMS 的研制包括研制 DBMS 本身以及以 DBMS 为核心的一组相互联系的软件系统。研制的目标是扩大功能、提高性能和提高用户的生产率。

随着数据库应用领域的不断扩大,许多新的应用领域如自动控制、计算机辅助设计等,要求数据库能够处理与传统数据类型不同的新的数据类型,例如,声音、图像等非格式化数据,面向对象的数据库系统、扩展的数据库系统、多媒体数据库系统的兴起就是应这些新的需求和应用背景而产生的。

(2) 数据操作

数据库设计的主要任务是在 DBMS 的支持下,按照应用的要求,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。其中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等。

(3) 数据库理论

数据库理论的研究主要集中于关系的规范化理论、关系数据理论等。近年来,随着人工智能与数据库理论的结合以及并行计算机的发展,数据库逻辑演绎和知识推理、并行算法等理论研究,以及演绎数据库系统、知识库系统和数据仓库的研制都已成为新的研究方向。

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一个综合,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的具体事物,所以人们必须事先把事物转换成能够处理的数据。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中的数据和信息。通俗地讲数据模型就是现实世界的模拟。

数据模型应满足三方面的要求：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。在数据库系统中针对不同的使用对象和应用目的，采用不同的数据模型。

不同的数据模型实际上是提供给我们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。第一类模型是概念模型，也称信息模型，它是按用户的观点对数据和信息建模。另一类模型是数据模型，主要包括网状模型、层次模型、关系模型等，它是按计算机系统的观点对数据建模。

本节首先介绍数据模型的共性——数据模型的组成要素，然后分别介绍两类不同的数据模型——概念模型和数据模型。

1.2.1 数据模型的要素

一般地讲，任何一种数据模型都是严格定义的概念的集合。这些概念必须能够精确地描述系统的静态特性、动态特性和完整性约束条件。因此数据模型通常都是由数据结构、数据操作和完整性约束 3 个要素组成。

(1) 数据结构

数据结构用于描述系统的静态特性。

数据结构是所研究的对象类型(object type)的集合。这些对象是数据库的组成成分，它们包括两类，一类是与数据类型、内容、性质有关的对象，例如网状模型中的数据项、记录，关系模型中的域、属性、关系等；一类是与数据之间联系有关的对象，例如网状模型中的系型(set type)。

数据结构是刻画一个数据模型性质最重要的方面。因此在数据库系统中，人们通常按照其数据结构的类型来命名数据模型。例如，层次结构、网状结构和关系结构中的数据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

数据操作用于描述系统的动态特性。

数据操作是指对数据库中各种对象(型)的实例(值)允许执行的操作的集合，包括操作及有关的操作规则。数据库主要有检索和更新(包括插入、删除、修改)两大类操作。数据模型必须定义这些操作的确切含义、操作符号、操作规则(如优先级)以及实现操作的语言。

(3) 数据的约束条件

数据的约束条件是一组完整性规则的集合。完整性规则是给定的数据模型中数据及其联系所具有的制约和储存规则，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本数据模型必须遵守的基本的通用的完整性约束条件。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。

此外，数据模型还应该提供定义完整性约束条件的机制，以反映具体应用所涉及的数据必须遵守的特定的语义约束条件。例如，在学校的数据库中规定大学生年龄不得超过 29 岁，硕士研究生不得超过 38 岁，学生累计成绩不得有 3 门以上不及格等。

1.2.2 概念模型

数据模型是数据库系统的核心和基础。各种机器上实现的 BDMS 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抽象、组织为某一 DBMS 支持的数据模型,人们常常首先将现实世界抽象为信息世界,然后将信息世界转换为机器世界。也就是说,首先把现实世界中客观对象抽象为某一种信息结构,这种信息结构并不依赖于具体的计算机系统,不是某一个 DBMS 支持的数据模型,而是概念级的模型;然后再把概念模型转换为计算机上某一 DBMS 支持的数据模型。不难看出,概念模型实际上是现实世界到机器世界的一个中间层次。

由于概念模型用于信息世界的建模,是现实世界到信息世界的第一层抽象,是用户与数据库设计人员之间进行交流的语言,因此概念模型一方面应该具有较强的语义表达能力,能够方便、直接地表达应用中的各种语义知识;另一方面它还应该简单、清晰、易于用户理解。

(1) 信息世界中的基本概念

信息世界涉及的概念主要有:

1) 实体(entity) 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,例如,一个职工、一个学生、一个部门、一门课、学生的一次选课、部门的一次订货、老师与系的工作关系(即某位老师在某系工作)等都是实体。

2) 属性(attribute) 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(983602, 张山, 男, 1976, 计算机系, 1994)。这些属性组合起来表征了一个学生。

3) 码(key) 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(domain) 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,年龄的域为小于 35 的整数,性别的域为(男,女)。

5) 实体型(entity type) 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和刻画同类实体,称为实体型。例如,学生(学号,姓名,出生年份,系,入学时间)就是一个实体型。

6) 实体集(entity set) 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(relationship) 在现实世界中,事务内部以及事务之间是有联系的,这些联系在信息世界中反映为实体内部的联系和实体之间的联系。实体内部的联系通常是指组成实体的各属性之间的联系。两个实体型之间的联系可以分为 3 类:

• 一对一联系(1:1)

如果对于实体集 A 中的每一个实体,实体集 B 中至多有一个实体与之联系,反之亦然,则称为实体集 A 与实体集 B 具有一对一联系。记为 1:1。

例如,学校里面,一个班级只有一个正班长,而一个班长只在一个班中任职,则班级与班长之间具有一对一联系。

• 一对多联系(1:n)

如果对于实体集 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 A 中至多有一个实体与之联系,则称为实体集 A 与实体集 B 有一对多联系。记为 1:n。

例如,一个班级中有若干名同学,而每个学生只在一个班级中学习,则班级与学生之间是一对多的联系。

• 多对多联系(m:n)

如果对于实体集 A 中的每个实体,实体集 B 中有 n 个实体与之联系,反之,对于实体集 B

中的每一个实体,实体集 A 中也有 m 个实体与之联系,则称为实体集 A 与实体集 B 具有多对多联系。记为 $m:n$ 。

例如,一门课程同时有若干学生选修,而一个学生可以同时选修多门课程,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。

实体型之间的这种一对一、一对多、多对多联系不仅存在于两个实体型之间,也存在于两个以上的实体型之间。

若实体集 $E_1, E_2, \dots, E_n$ 存在联系,对于实体集 $E_i (i=1, 2, \dots, i-1, i, i+1, \dots, n)$ 中的给定实体,最多只和 E_i 中的一个实体相联系,则说 E_i 与 $E_1, E_2, \dots, E_{i-1}, E_{i+1}, \dots, E_n$ 之间的联系是一对多的。

例如,对于课程、教师与参考书 3 个实体型,如果一门课程可以有若干教师讲授,使用若干本参考书,而每一个教师只讲授一门课程,每一本参考书只供一门课程使用,则课程与教师、参考书之间的联系是一对多的。

同一实体集内的各实体之间也可以存在一对一、一对多、多对多的联系。例如,学生实体集内部具有领导与被领导的联系,即某一学生(班干部)“领导”若干名学生,而一个学生仅被另一个学生直接领导,因此这是一对多的联系。

(2) 概念模型的表示方法

概念模型是对信息世界建模,所以概念模型应该能够方便、准确地表示出上述信息世界中的常用概念。概念模型的表示方法很多,其中最为常用的是 P. P. S. Chen 于 1976 年提出的实体-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 图提供了表示实体型、属性和联系的方法。

- 实体型:用矩形表示,矩形框内写明实体名。
- 属性:用椭圆表示,并用无向边将其与相应的实体连接起来。
- 联系:用菱形表示,菱形框内写明联系名,并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的类型($1:1, 1:n$ 或 $m:n$)。

需要注意的是,联系本身也是一种实体型,也可以有属性。如果一个联系具有属性,则这些属性也要用无向边与该联系连接起来。

图 1.1 用 E-R 图描述了上面有关两个实体型之间的 3 类联系、3 个实体型之间的一对多联系和 1 个实体型内部的一对多联系的例子。

假设上面的 5 个实体型即学生、班级、课程、教师、参考书分别具有下列属性:

学生:学号、姓名、性别、年龄

班级:班级编号、所属专业系

课程:课程号、课程名、学分

教师:职工号、姓名、性别、年龄、职称

参考书:书号、书名、内容提要、价格

这 5 个实体的属性用 E-R 图表示,如图 1.2 所示。这 5 个实体之间的联系可以用 E-R 图表示,如图 1.3 所示。注意,选修和班级两个联系又都分别具有各自的属性。