



数据库原理

Database Principle

马瑾 张晶 著



中国电力出版社
CHINA ELECTRIC POWER PRESS

数据库原理

Database Principle

马瑾 张晶 著



中国电力出版社
CHINA ELECTRIC POWER PRESS

内 容 提 要

本书以关系数据库系统为核心,全面介绍了数据库系统的基本原理。全书共八章,主要内容包括:数据库系统基本概念、关系数据库、关系数据库标准语言 SQL、关系数据库设计理论、数据库保护、数据库设计、数据库管理系统、数据库技术新进展。

本书概念清楚、重点突出,理论与实践结合紧密。重点章节配以丰富案例,有助于读者从实际应用的角度出发,联系所学理论,掌握所学内容。

本书可供软件设计与开发的工程技术人员进行数据库结构的设计和数据库应用系统开发时参考,也可作为高等院校计算机专业、信息管理与信息系统以及相关专业的本科生的教材。

图书在版编目(CIP)数据

数据库原理/马瑾,张晶著. —北京:中国电力出版社,
2016.6

ISBN 978-7-5123-9490-2

I. ①数… II. ①马… ②张… III. ①数据库系统
IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2016)第 171866 号

中国电力出版社出版、发行

(北京市东城区北京站西街 19 号 100005 <http://www.cepp.sgcc.com.cn>)

北京教图印刷有限公司印刷

各地新华书店经售

*

2016 年 6 月第一版 2016 年 6 月北京第一次印刷
710 毫米×980 毫米 16 开本 17.25 印张 267 千字
定价 38.00 元

敬告读者

本书封底贴有防伪标签,刮开涂层可查询真伪
本书如有印装质量问题,我社发行部负责退换

版权专有 翻印必究



前言

数据库技术是计算机科学中发展最快的领域之一，已经成为信息系统的基础和核心。20世纪60年代末，数据库技术是作为数据处理中的一门新技术发展起来的。目前，数据库技术已是计算机软件领域的一个重要分支，形成了较为完整的理论体系和实用技术。

随着数据库系统的推广，计算机应用已深入人类社会的各个领域，如当前的管理信息系统（MIS）、企业资源规划（ERP）、计算机集成制造系统（CIMS）、地理信息系统（IGIS）、决策支持系统（DDS）等都是以数据库技术为基础。目前，我国实施的国家信息化、“金”字工程、数字城市等都是以数据库为基础的大型计算机系统。数据库的建设规模和性能、数据库信息量的大小和使用水平已成为衡量一个国家信息化程度的重要标志。

本书将数据库基本原理、方法和应用技术相结合，以关系数据库系统为核心，在完整论述数据库系统的基本概念、基本原理的基础上，对 Microsoft SQL Server 进行讲解，使读者能熟练地使用现有的数据库管理系统和软件设计与开发工具，进行数据库结构的设计和数据库应用系统开发。本书的主要特点：内容丰富、系统全面；突出重点、注重总结；概念清晰、分析深入；例题丰富，实用性强；叙述深入浅出，语言流畅。

本书由山西财贸职业技术学院马瑾、张晶著，马瑾老师编写了第一～四章，张晶老师编写了第五～八章。

2016年6月



目 录



前言

第一章 概述	1
第一节 数据库系统基本概念	1
第二节 数据模型概述	8
第三节 数据库系统的组成与结构	26
第二章 关系数据库	36
第一节 关系数据库概述	36
第二节 关系数据结构	38
第三节 关系的完整性	43
第四节 关系代数	47
第三章 关系数据库标准语言 SQL	57
第一节 SQL 简介	57
第二节 数据定义	61
第三节 数据查询	67
第四节 数据更新	94
第五节 视图	98
第四章 关系数据库设计理论	109
第一节 数据依赖	109
第二节 范式	114
第三节 关系模式的规范化	125
第五章 数据库保护	131
第一节 安全性	131

第二节	完整性	137
第三节	并发控制	145
第四节	恢复	159
第五节	数据库复制和数据库镜像	167
第六章	数据库设计	170
第一节	数据库设计概述	170
第二节	需求分析	177
第三节	概念结构设计	182
第四节	逻辑结构设计	197
第五节	数据库的物理设计	201
第六节	数据库的实施和维护	206
第七章	数据库管理系统	210
第一节	DBMS 的基本功能	210
第二节	DBMS 的进程结构和多线索控制	211
第三节	DBMS 系统结构	219
第四节	语言处理	222
第五节	数据存取层	227
第六节	缓冲区管理	231
第七节	数据库物理组织	233
第八章	数据库技术新进展	236
第一节	数据库技术发展概述	236
第二节	数据模型及数据库的发展	236
第三节	数据库技术与其他相关技术相结合	245
第四节	面向应用领域的数据库新技术	260
参考文献	270

概 述

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，它的出现极大促进了计算机应用向各行各业的渗透。从 20 世纪 50 年代开始，数据处理迅速发展为计算机应用的主要方面，在计算机的三大应用（科学计算、数据处理与过程控制）中所占比例为 70% 左右。20 世纪 60 年代末，数据库技术开始发展起来。如今，数据库技术在各行业已经得到了广泛应用，成为存储、使用和更新信息资源的主要手段，产生了巨大的经济效益和社会效益。因此，学习和研究数据库，不断推广数据库，不断利用现有数据库技术为社会经济发展服务，是每个计算机工作者必须做的工作。

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，从中不难看出数据库技术的重要性。

第一节 数据库系统基本概念

在系统介绍数据库的基本概念之前，先介绍一些数据库最常用的术语和基本概念。

一、数据、数据库、数据库管理系统、数据库系统

数据、数据库、数据库管理系统和数据库系统是数据库技术密切相关的四个基本概念。

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处不在；文字、图像、声音、学生的档案



记录、货物的运输情况……这些都是数据。

为了认识世界、交流信息，人们需要描述事物。数据实际上是描述事物的符号记录。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、籍贯、所在系别、入学时间，那么可以这样描述：

（李明，男，1972，江苏，计算机系别，1990）

数据与其语义是不可分的。对于上面一条学生记录，了解其语义的人会得到如下消息：李明是个大学生，1972年出生，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（Data Base, DB）

数据库是存放数据的仓库。只不过这个仓库是在计算机存储设备上，而且数据是按一定的格式存放的。

人们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，进一步抽取有用信息。在科学技术飞速发展的今天，人们的视野越来越广，数据量急剧增加。过去人们把数据存放在文件柜里，现在人们借助计算机和数据库技术科学地存储和管理大量的复杂的数据，以便能方便而充分地利用这些宝贵的信息资源。

严格地讲，数据库是长期储存在计算机内、有组织的、可共享的大量数据的集合。数据库中的数据按一定的数据模型组织、描述和存储，具有较小的冗长度（Redundancy）、较高的数据独立性（Data Independence）和易扩展性，并可为各种用户共享。

概括地讲，数据库数据具有永久存储、有组织和可共享三个基本特点。

3. 数据库管理系统（Data Base Management System, 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发生故障后的系统恢复。

4. 数据库系统 (Data Base System, DBS)

数据库系统是指在计算机系统中引入数据库后的系统,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,这些人被称为数据库管理员 (Data Base Administrator, DBA)。

在不引起混淆的情况下常常把数据库系统简称为数据库。

数据库系统可以用图 1-1 表示。数据库系统在整个计算机系统中的地位如图 1-2 所示。

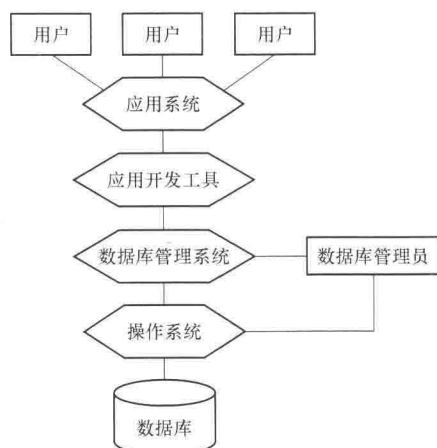


图 1-1 数据库系统

图 1-2 所示。

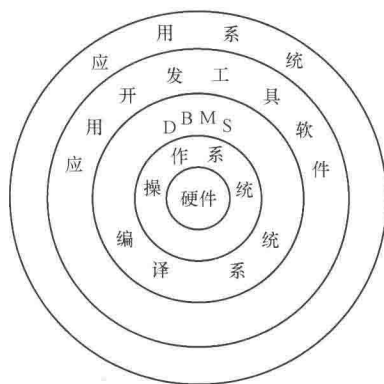


图 1-2 数据库系统在计算机系统中的地位

二、数据库技术的发展历程

数据库技术是应数据管理任务的需要而产生的。

数据管理则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。而数据的处理是指对各种数据进行收集、存储、加工和传播等一系列活动的总和。

随着计算机的普及以及电子计算机软件和硬件技术的发展,数据处理的速



度和规模是手工方式和机械方式无可比拟的。随着数据处理量的增长,产生了数据库技术。数据管理技术的发展与计算机硬件、软件和计算机的应用范围有着密切的关系,使数据处理跨入了一个崭新的阶段。采用数据库技术管理数据就是把相关的数据集中存放在一个或多个数据库文件里,用户通过数据库管理软件来使用数据库中的数据。

在应用需求的推动下,在计算机硬件、软件发展的基础上,数据管理技术经历了人工管理、文件系统、数据库系统三个阶段。这三个阶段的特点见表 1-1。

表 1-1 数据管理三个阶段的特点

阶 段		人工管理阶段	文件系统阶段	数据库系统阶段
背景	应用背景	科学计算	科学计算、数据管理	大规模数据管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘,磁盘阵列
	软件背景	没有操作系统	有文件系统	有数据库管理系统
特点	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理
	数据的管理者	用户(程序员)	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用	现实世界(一个部门、企业、跨国组织等)
	数据的共享程度	无共享,冗余度较大	共享性差,冗余度大	共享性高,冗余度小
	数据的独立性	不独立,完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据结构化	无结构	记录内有结构、整体无结构	整体结构化,用数据模型描述
数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力	

三、数据库系统的特点

与人工管理和系统文件系统相比,数据库系统的特点主要有以下四个方面。

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。数据库设计的基础是数据模型。设计数据库时,要站在全局需要的角度抽象和组织数据;要完整、准确地描述数据自身和数据之间联系的情况;要建立适合整体需要的数据模型。数据



库系统是以数据库为基础的，各种应用程序都建立在数据库之上。这种特点决定它的设计特点，先设计数据库，再设计程序功能。

所谓“整体”结构化是指在数据库中的数据不再仅仅针对某一个应用，而是面向全组织；不仅数据内部是结构化的，而且整体是结构化的，数据之间是具有联系的。在数据库系统中实现了整体数据的结构化。也就是说，不仅要考虑某个应用的数据结构，还要考虑整个组织的数据结构。

2. 数据的共享性好，冗余度低

数据冗余度是指重复数据的多少。减少数据冗余度可以带来以下优点：数据量小，可以节约存储空间，使数据的存储、管理和查询都容易实现；数据冗余小，可以使数据统一，避免数据的不一致；便于数据维护，避免数据统计错误。

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。可以选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

数据库中的数据定义功能和数据管理功能是由 DBMS 提供的，所以数据对应用程序的依赖度大大降低，数据和程序之间具有较高的独立性。数据独立性高使得程序中不需要有关数据结构的储存方式的描述，从而减轻了程序设计的负担。而且当数据结构变化时，程序的维护也会比较容易。

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当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特性，简称数据的物理独特性。

数据与程序之间的独立性，使得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与修改。

4. 数据由 DBMS 统一管理和控制

由于对数据实行了统一管理，而且所管理是有结构的数据，因此在使用数据时可以有灵活的方式，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，只需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。

(1) 数据的安全性 (Security)。数据的安全性是指保护数据，防止不合法使用数据造成的数据的泄密和破坏，是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。

(2) 数据的完整性 (Integrity)。数据的完整是指数据的正确性、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。

(3) 并发 (Concurrency) 控制。当多个用户的并发进程同时存取、修改数据库时，可能会发生互相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。

(4) 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的功能，这就是数据库的恢复功能。

数据库管理阶段应用程序与数据之间的对应关系可用图 1-3 表示。

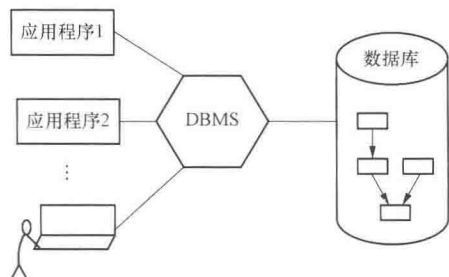


图 1-3 数据库系统阶段应用程序与数据之间的对应关系



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的数据集合。他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，在发生故障后进行恢复。

四、数据库技术的研究领域

虽然目前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地前进。概括地讲，当前数据库主要的研究范围有以下三个领域。

1. 数据库管理系统的研制

DBMS 是系统数据库的基础。DBMS 的研制包括研制 DBMS 本身以及 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产生率。

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、多媒体数据库系统等的兴起就是应这些新的需求和应用背景而产生的。

2. 数据库设计

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究都已成为新的研究方向。



第二节 数据模型概述

模型，特别是具体模型，人们并不陌生。一张地图、一组建筑设计沙盘、一架精致的航模飞机都是具体的模型。一眼望去，就会使人联想到真实生活中的事物。模型是对现实世界中某个对象特征的模拟和抽象。例如，航模飞机是对生活中飞机的一种模拟和抽象，它可以模拟飞机的起飞、飞行和降落，他抽象了飞机的基本特征——机头、机身、机翼、机尾。

数据模型也是一种模型，它是对实际数据特征的抽象。也就是说，数据模型是用来描述数据、组织数据和对数据进行操作的。

为了把具体事物抽象、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，首先把现实世界中的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，这一过程如图 1-4 所示。

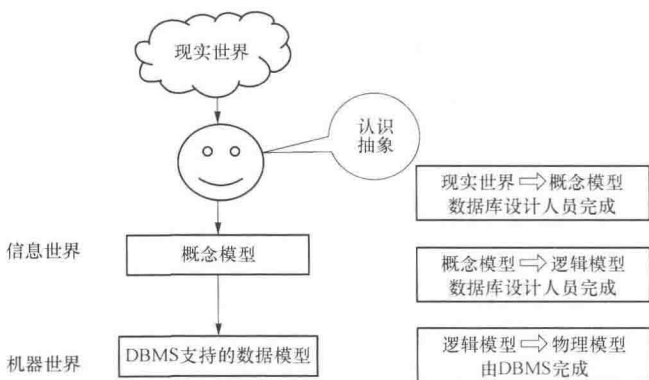


图 1-4 现实世界中客观对象的抽象过程

一、概念模型

概念模型是数据库系统的核心和基础。各种机器上实现的 DBMS 软件都是



基于某种数据模型的。为了实现世界中的具体事务抽象、组织为某一 DBMS 支持的数据模型，人们常常首先将现实世界的具体事务抽象化为信息世界，然后将信息世界转换为机器世界。也就是说，首先把现实世界中的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某一 DBMS 支持的数据模型。概念模型实际是实现世界到机器世界的一个中间层次。

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，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七个。

(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，一个工人、一个银行账号、一本书、公司间的一次交易等都是一个实体。

(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，学生实体可以用学号、姓名、年龄、性别、系别等来描述，一本书可以用书号、作者、出版社、日期、价格等属性来描述。

(3) 码 (Key)。唯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)。域是一组具有相同数据类型的值的集合。属性的取值范围来自某个域。例如，学号的域为 8 位整数，姓名的域为字符串集合，学生年龄的域为整数，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性和性质。用实体名及其属性名来抽象和刻画同类实体，称为实体型。例如，学生 (学号，姓名，性别，出生年月，所在院系，入学时间) 就是一个实体型。

(6) 实体集 (Entity Set)。同一类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; 实体之间的联系通常是指不同实体集之间的联系。

2. 两个实体型之间的联系

两个实体型之间的联系可以分为三种。

(1) 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 中至少有一个 (也可以没有) 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的联系。

(2) 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 具有一对多联系, 记为 1:n。

例如, 一个班中有若干名学生, 而每个学生只在一个班级中学习, 则班级与学生之间具有一对多的联系。

(3) 多对多联系 (m:n)。如果对于实体集 A 中的每一个实体, 实体集 B 中有 n ($n \geq 0$) 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 ($m \geq 0$) 个实体与之联系, 则称实体集 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个学生选修, 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。

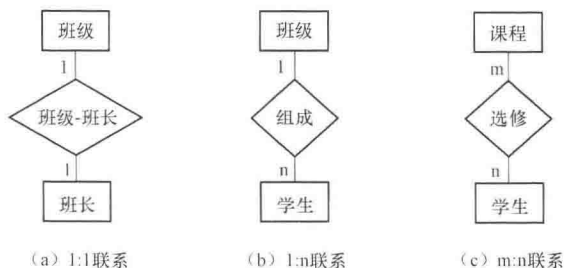


图 1-5 两个实体之间的三类联系

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。

可以用图形来表示两个实体型之间的这三类联系, 如图 1-5 所示。



3. 两个以上的实体型之间的联系

一般地，两个以上的实体型之间也存在着一对一、一对多、多对多联系。

若实体型 E_1, E_2, \dots, E_n 之间存在联系，对于实体型 $E_j (j=1, 2, \dots, i-1, \dots, n)$ 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。请读者给出实体型之间一对一、多对多联系的定义。

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个教师讲授，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，如图 1-6 (a) 所示。

又如，有 3 个实体型：供应商、项目、零件。一个供应商可以供给多个项目多种零件，而每一个项目可以使用多个供应商供应的零件，而每一种零件可由不同供应商供给。由此看出，供应商、项目、零件三者之间是多对多联系，如图 1-6 (b) 所示。要注意，3 个实体型之间多对多联系和 3 个实体型两两之间的多对多联系的语义是不同的。

4. 单个实体型内的联系

同一个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，职工实体型内部具有领导与被领导的联系，即某一职工（干部）“领导”若干名职工，而一个职工仅被另外一个职工直接领导，因此这是一对多的联系，如图 1-7 所示。

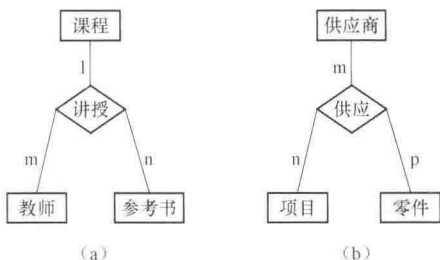


图 1-6 3 个实体之间的联系示例



图 1-7 单个实体型之间
一对多联系示例

5. 概念模型的一种表示方法：实体—联系方法

概念模型针对信息世界建模，所以概念模型应该能够方便、准确地表示出