

XIANDAI
SHUJUKU JISHU

现代数据库

赵正文 / 主编

技术



电子科技大学出版社

现代数据库技术

赵正文 主编



电子科技大学出版社

图书在版编目 (CIP) 数据

现代数据库技术 / 赵正文主编. —成都: 电子科技大学出版社, 2013. 1

ISBN 978-7-5647-1432-1

I. ①现… II. ①赵… III. ①数据库系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2012) 第 317717 号

内 容 提 要

本书概括介绍了数据库系统的基本原理、基本概念以及数据库技术的最新研究动态, 重点以进销存管理系统为例详细阐述了数据库设计的过程, 较详细地介绍了 Oracle、SQL Server 等主流数据库和 PowerDesigner 建模工具的相关知识和具体操作。

全书分为四部分: (1) 基础篇, 包括数据库系统概述、关系理论、关系查询语言三章。(2) 应用篇, 包括数据库应用概述、需求分析、数据库概念设计、数据库逻辑设计、数据库用户事务综合设计、数据库物理设计及实现六章。(3) 专题篇, 包括数据库技术的发展状况、XML 技术、主流 DBMS 三章。(4) 附录, 包括 UML 类图和 PowerDesigner。

本书可作为高等院校有关数据库课程的教材, 也可供从事数据库设计和开发人员参阅。

现代数据库技术

赵正文 主编

出 版: 电子科技大学出版社 (成都市一环路东一段 159 号电子信息产业大厦 邮编: 610051)
策划编辑: 周 岚
责任编辑: 周 岚
主 页: www.uestcp.com.cn
电子邮箱: uestcp@uestcp.com.cn
发 行: 新华书店经销
印 刷: 河北永清县晔盛亚胶印有限公司
成品尺寸: 170mm×240mm 印张 24.5 字数 493 千字
版 次: 2013 年 4 月第一版
印 次: 2013 年 4 月第一次印刷
书 号: ISBN 978-7-5647-1432-1
定 价: 75.00 元

■ 版权所有 侵权必究 ■

- ◆ 本社发行部电话: 028-83202463; 本社邮购电话: 028-83201495。
- ◆ 本书如有缺页、破损、装订错误, 请寄回印刷厂调换。

序

数据库技术经过了近 40 年的飞速发展，已经渗透到社会的各个领域，企业对精通数据库技术的应用型和技能型专业人才的需求也日益迫切。这对数据库技术类书籍提出了新的要求：内容应突出实用和应用，方便教师组织教学和学生自学。

赵正文博士等编著的《现代数据库技术》一书，正好填补了这个空白，为解决这个问题提出了新的思路，切合了时代的需求。通读《现代数据库技术》一书，让人耳目一新，它是一本以基础理论和基本原理为基础，突出实用和应用的好教材或参考书，表现在以下方面：

- 本书的作者有深厚的理论基础，长期在一线从事数据库技术的教学、科研和信息系统的开发管理工作，具有丰富的实践经验。
- 内容组织遵循了合理的逻辑关系，读者可在循序渐进的学习过程中掌握数据库的基础理论和应用技术。
- 突出了实用和应用特点，引入了主流 DBMS 和 CASE 工具的介绍，真正使得读者可以快速学以致用。
- 在基础理论和基本原理的描述上，内容全面、简洁明了且通俗易懂。
- 本书以大家比较熟悉的进销存管理系统为例贯穿全书，该实例能体现系统性和完整性，其广度、深度和复杂度满足了不同层次的读者需求。据此，读者通过学习可以对数据库的应用（特别是数据库设计）有全面的认识，加深对相关知识系统性的掌握。
- 本书难易度的把握运用恰当，特别适合初学者自学和教师组织教学。
- 本书的数据库设计部分很有特色，较多地采用了直观的图表表现形式，并辅以详细讲解，使读者学习起来更轻松、更易于理解。

相信本书的出版，能对广大教师、初学者和需要提高数据库技术的专业人员提供有益的帮助。在此，祝贺赵正文博士，祝贺《现代数据库技术》的出版。谨以此为序。

International Progress on WAMTIP 主编
International Journal of WMIP 副主编
国际小波分析应用研究中心主任
国家科学技术奖励评审委员



前 言

数据库技术从 20 世纪 60 年代产生至今，已经在商业流通、工业生产、金融、通信、文化教育、科学研究以及国防等领域得到了广泛的应用，并形成了较为完整的理论与技术体系，已成为计算机科学技术中的一个重要分支。

对于绝大多数学生和数据库应用系统开发人员来说，学习和掌握数据库技术的主要目的不是去研究和开发商品化的数据库管理系统 (DBMS)，而是为了运用现有的 DBMS 和相关数据库技术，解决实际工作中的各类计算机应用问题。

因此，本书融入了作者多年的教学和科研实践以及信息系统开发设计的丰富经验，立足于数据库技术的基本原理和基本概念，简化了过于深奥的理论知识，以较大篇幅详细地阐述了数据库的应用，突出了实用和应用两大特点。

全书分为四部分：

- 第一部分：基础篇。概括介绍了数据库系统的基本原理、基本概念、关系理论以及关系查询语言 SQL 等基础知识，使读者对数据库概貌和 SQL 的使用有个基本了解。
- 第二部分：应用篇。以进销存管理系统为实例，完整地讲解了数据库系统生存期的整个过程，从需求获取、需求分析、概念设计、逻辑设计到物理设计及实现。使读者在学习完本部分内容后对数据库应用有一个整体概念和系统性的认识，初步具备设计一个中小型应用系统数据库的能力。
- 第三部分：专题篇。主要介绍了面向对象数据库、分布式数据库、数据仓库、数据挖掘和联机分析处理技术等数据库技术的研究发展以及 XML 的基本概念和简单应用，并详细地介绍了 SQL Server、Oracle、DB2、Sybase、My SQL 等主流 DBMS，使读者对具体 DBMS 有一个初步的系统性的感性认识，为 DBMS 的选择提供帮助，也可作为一本简明手册来使用。
- 第四部分：附录。重点介绍了 PowerDesigner 在数据库建模上的具体操作。

本书由赵正文拟定目录提纲和编写要点并最后统稿，参与编写的人员有陈朝岳、侯艳、张惠茹、雷涛和钱广，蒋俊参与了后期的修改工作。

在编写过程中，我们力求精益求精，但难免存在一些不足之处，真诚希望同行专家和读者对本书提出宝贵的意见和建议。

编 者

目 录

第一部分 基础篇

第 1 章 数据库系统概述	2
1.1 数据库技术的发展	2
1.2 数据库的基本概念	5
1.3 数据抽象	6
1.4 数据库管理系统	15
1.5 数据库系统	33
1.6 小结	36
第 2 章 关系理论	37
2.1 关系模型概述	37
2.2 关系代数	38
2.3 关系演算	44
2.4 关系数据库设计	47
2.5 小结	55
第 3 章 关系查询语言	56
3.1 SQL 简述	56
3.2 数据定义	58
3.3 数据更新	61
3.4 数据查询	65
3.5 数据控制	80
3.6 视图	82
3.7 嵌入式 SQL	84
3.8 数据库查询优化	86
3.9 Oracle 和 SQL Server 对标准 SQL 的扩充	91
3.10 小结	94

第二部分 应用篇

第 4 章 数据库应用概述	97
4.1 数据库在应用系统中的重要地位	97
4.2 数据库设计	98
4.3 实例介绍	101

4.4 小 结	102
第 5 章 需求分析	103
5.1 需求分析的任务	103
5.2 需求分析的过程	104
5.3 数据流图和数据字典	105
5.4 获取需求的重要方法——调研	107
5.5 需求分析实例	108
5.6 小结	121
第 6 章 数据库概念设计	123
6.1 概念设计的任务	123
6.2 E-R 模型	123
6.3 概念设计的方法和步骤	129
6.4 概念设计实例	133
6.5 小结	147
第 7 章 数据库逻辑设计	148
7.1 逻辑设计的基本概念	148
7.2 逻辑设计的任务	150
7.3 逻辑设计的步骤	150
7.4 逻辑设计实例	153
7.5 小结	157
第 8 章 数据库用户事务综合设计	158
8.1 概述	158
8.2 用户事务综合设计实例	165
8.3 创建数据库设计文档	178
8.4 生成 SQL 脚本	181
8.5 小结	188
第 9 章 数据库物理设计及实现	190
9.1 概述	190
9.2 创建 ERM 系统数据库	191
9.3 小结	199

第三部分 专题篇

第 10 章 数据库技术的研究发展	201
10.1 概述	201
10.2 面向对象数据库系统	204

10.3	分布式数据库	217
10.4	数据仓库	235
10.5	联机分析处理技术	245
10.6	数据挖掘	253
10.7	其他数据库技术	259
第 11 章 XML 技术		277
11.1	XML 概述	277
11.2	XML 文档	280
11.3	在 .NET 中使用 XML	289
11.4	小结	293
第 12 章 主流 DBMS		294
12.1	SQL Server	294
12.2	Oracle	321
12.3	DB2	343
12.4	Informix	348
12.5	Sybase	352
12.6	MySQL	357
12.7	Access	362
12.8	小结	364
附录 A UML 类图		368
附录 B PowerDesigner		369
B.1	PowerDesigner 概述	369
B.2	建立概念数据模型	371
B.3	建立物理数据模型	375
B.4	创建报告文档	380
B.5	小结	383

第一部分

基础篇

第 1 章 数据库系统概述

数据库技术是计算机科学技术中发展最快、应用最广泛的领域之一，它是计算机信息系统与应用程序的核心技术和重要基础。自从 20 世纪 50 年代中期计算机的应用由科学计算扩展到数据处理，数据库技术的重要性也日益被人们所认识。时至今日，数据库技术已经形成了较为完善的理论体系和实用技术，成为计算机科学的一个重要分支。

数据库管理系统（Database Management System, DBMS）由一个互相关联的数据的集合和一组用以访问这些数据的程序组成。这个数据集合通常称为数据库，其中包含了关于某个企业组织的信息。DBMS 的基本目标是要提供一个能方便、有效地存取数据库信息的环境。在大多数组织里，信息是非常重要的，因而计算机科学家提出并丰富和发展了大量的用于有效管理数据的概念和技术。

1.1 数据库技术的发展

随着使用计算机进行数据管理的技术的不断发展，数据库技术经历了人工管理、文件系统、数据库系统和 Internet 时代四个发展阶段。

1.1.1 人工管理阶段

在 20 世纪 50 年代中期以前，计算机的外部设备只有磁带机、卡片机和纸带穿孔机等，而没有直接可以存取的磁盘设备，计算机主要用于科学计算，数据处理采取批处理的方式。

人工管理数据具有如下特点：

- 数据不保存

由于当时数据主要用于科学计算，数据管理规模小，加上计算机硬件、软件条件有限，一般不需要将数据长期保存，计算完毕后就将从数据撤走。

- 没有专门的软件系统对数据进行管理

由于没有专门的软件系统对数据进行管理，因此应用程序设计者不但要规定数据的逻辑结构，而且还要设计物理结构，即要设计数据的存储结构、存取方法和输入输出方法等。

- 数据不共享

数据面向应用程序，一组数据只能对应一个程序，造成程序间的大量冗余。

■ 数据不独立

数据和程序之间不具有独立性，如果数据的存储结构变化，则程序需要做相应的修改。

1.1.2 文件系统阶段

从 20 世纪 50 年代中期到 60 年代中期，计算机不仅用于科学计算，同时也开始用于信息处理，硬件方面有了很大改进，出现了磁盘、磁鼓等直接存储设备。软件方面出现了高级语言和操作系统，且操作系统中出现了专门的数据管理软件。

用文件系统管理数据的优缺点如下：

(1) 优点

■ 数据可以长期保存

在文件系统中，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。

■ 由文件系统管理数据

有专门的计算机软件提供数据存取、查询、修改和管理功能，文件系统把数据组织成相互独立的数据文件，即可以按文件名访问数据。

■ 数据文件已经多样化

文件系统中的数据文件不仅有索引文件、链接文件、直接存储文件等多种形式，而且还可以使用倒排文件进行多码检索。

(2) 缺点

■ 数据冗余度大

文件系统采用面向应用的设计思想，系统中的数据文件都是与应用程序相对应的。当不同的应用程序所需要的数据有部分相同时，也必须建立各自的文件，而不能共享相同的数据。

■ 数据独立性差

文件系统中的数据文件是为某一特定应用服务的，数据文件是孤立的，不能反映现实世界事物之间的内在联系。同时由于数据与应用程序间缺乏独立性，使得应用系统不容易扩充。

1.1.3 数据库系统阶段

进入 20 世纪 60 年代，计算机软件、硬件技术得到了飞速发展。1969 年美国 IBM 公司研制开发的层次型信息管理系统（Information Management System, IMS 系统）、美国数据系统语言协会（Conference on Data System Language, CODASYL）发布的数据库任务组（Data Base Task Group, DBTG）关于网状数据库的报告以及 1970 年 IBM 公司的研究员 E.F.Codd 在发表的论文“大型共享数据库数据的关系模型”中提

出的“关系模型”是数据库技术发展史上具有里程碑意义的重大事件，这些研究成果大大促进了数据管理技术的发展和應用。

数据库系统具有如下特点：

■ 数据高度结构化

在数据库系统中，使用规范的数据模型表示数据结构，数据不再针对某一项应用，而是面对系统的整体，应用程序可通过 DBMS 访问数据库中的所有数据。

■ 较少的数据冗余，共享性高

较少的数据冗余，可以使数据统一，便于数据维护，并可节约存储空间，易于数据的存储、管理和查询。数据库中的数据是高度共享的：同一用户可以就不同的程序访问同一数据；多用户多程序可以同时访问同一数据。

■ 数据与应用程序相互独立

在数据库系统中用户只需要关注数据库名称、数据文件名称和文件中的属性名称等概念，而不用过多考虑数据的物理存储方面的问题。这样一来，数据的定义和描述就可从应用程序中分离出来，数据与应用程序之间具有独立性。对于系统具体处理数据的存取等技术问题，已专门开发出数据管理的系统软件，即数据库管理系统，从而简化应用程序的编写，减少开销。

■ 通过 DBMS 进行数据安全性和完整性控制

数据库管理系统 DBMS 可以有效地防止数据库中的数据被非法使用或修改。对于完整性控制，DBMS 提供了数据完整性定义方法和进行数据完整性检验的功能，数据控制机制还包括数据的并发控制和数据恢复，排除由于数据共享所造成的数据不完整和系统运行错误问题，保证数据在受到破坏时能够及时使数据库恢复到正确状态。

1.1.4 Internet 时代

自 20 世纪 80 年代以来，以分布式数据库和面向对象数据库技术为代表，使数据管理技术进入了高级数据库阶段。20 世纪 90 年代以来，涌现了一大批新一代数据库，如数据仓库、OLAP 分析、数据挖掘、数字图书馆、电子出版物、电子商务、远程教育、Web 上的数据管理和信息检索等。新一代数据库的应用大大地激发了数据库技术的研究和开发，出现了一大批具有 Internet 时代特征的数据库技术和相应的数据库管理系统，如 Web 信息检索技术与系统、Web 数据集成和共享技术与系统、数据流技术与系统、电子商务和电子政务技术与系统、数据图书馆技术与系统、安全数据库技术与系统等，使数据库的应用发展到了一个新的阶段。

1.2 数据库的基本概念

1.2.1 数 据

数据可以看做是给定的事实，并且可以从中推出新的事实。狭义的数据是指能够进行加、减、乘、除以及各种统计计算的数值，如 1、2、3 等；广义的数据是指描述客观事物和性质的一种抽象的符号表示，它可以是数值、字符、图形、图像和声音等。数据集中的每一个个体称为数据元素，它是数据的基本单位。

计算机领域的数据是指能够进行数据化处理而进入计算机的广义数据。它是描述客观事物的数、字符以及所有能输入计算机并被计算机处理的符号的集合。

人们通过解释、推理、归纳、分析和综合等方法，从数据中所获得的有一定意义和价值且具有特定形式的内容称为信息。数据是信息存在的一种形式，只有通过解释和处理的数据才能成为有用的信息。

1.2.2 数据库的基本术语

数据库、数据库管理系统和数据库系统是数据库技术中最基本的术语。

(1) 数据库

数据库 (Database, DB) 是指长期储存在计算机内的、有组织并可共享的数据集合。数据库不是独立存在的，它是基于计算机系统的，能持久性地存放数据的“仓库”。数据库中的数据按一定的数据模型组织、描述和储存，具有较小冗余度，数据对程序的独立性以及由数据库管理系统统一管理和控制等特点。数据库是数据库系统的组成部分。

(2) 数据库管理系统

数据库管理系统是对数据库进行管理的软件，它位于用户与操作系统之间。它的功能包括对数据库的建立、更新、查询、统计、显示、打印和各种数据控制。数据库管理系统是数据库系统的一个重要组成部分。

(3) 数据库系统

数据库系统 (Database System, DBS) 是指计算机系统在引入了数据库后的系统，它由数据库、数据库管理系统、数据库管理员 (Database Administrator, DBA)、数据库应用程序和用户组成。数据库管理员是对数据库进行规划、设计、协调、维护和管理人员。

1.3 数据抽象

数据库系统为用户提供了一个数据的抽象视图，使用户不必了解数据库文件的存储结构、存储位置、存取方法等细节就可以存取数据库。模型是对事物、对象、过程等客观系统中感兴趣的内容的模拟和抽象表达，是理解系统的思维工具。数据模型也是一种模型，它是对现实世界数据特征的抽象。现有的数据库系统都是基于某种数据模型的。

数据模型应满足三个方面的要求：能比较真实地模拟现实世界；容易为人们所理解；便于在计算机上实现。

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。①概念模型：也称为信息模型，它是按用户的观点来对数据和信息建模，主要用于数据库设计。②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，主要用于 DBMS 的实现。数据模型是数据库系统的核心和基础。

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。即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 1-1 所示。

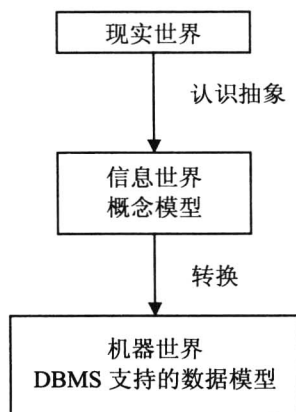


图 1-1 现实世界中客观对象的抽象过程

1.3.1 概念模型

概念模型实际上是现实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。

一、概念模型的基本概念

（1）实体（Entity）

客观存在并可相互区别的事物称为实体。实体可以是具体的事物，也可以是抽

象的概念或联系。如学生、老师、课程、产品就是具体的事物，而学生的选课、教师的授课、产品的订货也可以看成是实体，只是它们是抽象的实体。

(2) 属性 (Attribute)

实体所具有的某一特性称为属性。属性是描述实体或者联系的性质和特征的数据项。一个实体可以由若干个属性来刻画。如学生实体可以由学号、姓名、性别等属性组成。

(3) 码 (Key)

唯一标识实体的属性集称为码。如学号是学生实体的码。在有些实体集中可以有多个码。如学生实体中如果姓名不能没有重名，那么属性“姓名”也可以作为码。当一个实体集包括多个码时，通常要选定其中一个码作为主码 (Primary Key)，其他码就是候选码。

(4) 域 (Domain)

属性的取值范围称为该属性的域。如性别的域为 (男, 女)。

(5) 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指组成实体的各属性之间的联系。实体之间的联系通常指不同实体集之间的联系。两个实体集之间的联系可以分为三类：

- 一对一联系 (1:1)。如果实体集 A 中的每个实体在实体集 B 中至多有一个 (也可以没有) 实体与之关联，反之亦然，则称实体集 A 与实体集 B 具有一对一的关系。记为 1:1。如部门和经理 (假设一个部门只有一个经理)，班长和班级 (假设一个班只有一个班长) 都是一对一联系，如图 1-2 (a) 所示。
- 一对多联系 (1:n)。如果实体集 A 中的每一个实体在实体集 B 中有 n 个实体 ($n \geq 0$) 与之关联，而实体集 B 中每个实体在实体集 A 中最多只有一个实体与之关联，则称实体集 A 与实体集 B 是一对多联系。记为 1:n。如一个部门有若干个职工，而一个职工只在一个部门工作，则部门和职工之间就是一对多联系，如图 1-2 (b) 所示。
- 多对多联系 (m:n)。如果对于实体集 A 中的每个实体在实体集 B 中有 n 个实体 ($n \geq 0$) 与之关联，而对于实体集 B 中的每个实体在实体集 A 中也有 m 个实体 ($m \geq 0$) 与之关联，则称实体集 A 与实体集 B 是多对多联系。记为 m:n。如一个学生可以选修多门课程，一门课程也可以被多个学生选修，则学生和课程之间是多对多联系，如图 1-2 (c) 所示。

实际上，一对一联系是一对多联系的特例，一对多联系又是多对多联系的特例。可用图来表示两个实体集之间的这三类联系，如图 1-2 所示。

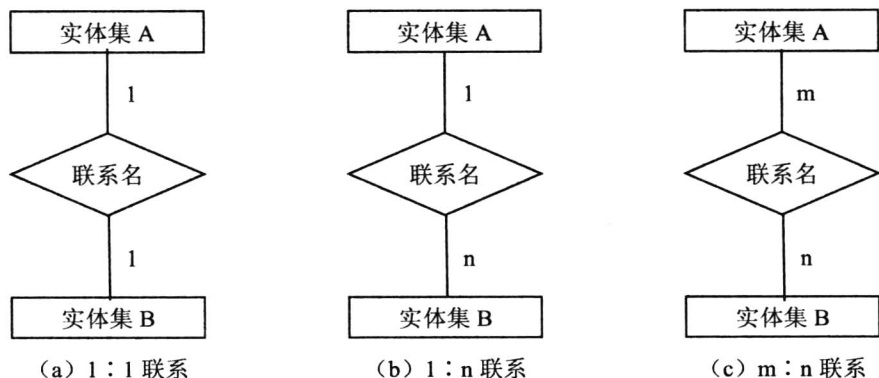


图 1-2 两个实体集之间的三类联系

二、概念模型的表示方法

概念模型是对信息世界的建模，概念模型应能够全面、准确地描述出信息世界中的基本概念。概念模型的表示方法很多，其中最为著名和使用最广泛的是 E-R 图法，即实体联系方法 (Entity Relationship Approach)。该方法用 E-R 图来描述现实世界的概念模型，提供了表示实体集、属性和联系的方法。E-R 图也称为 E-R 模型。在 E-R 图中：

- 用长方形表示实体集，长方形内写实体集名。
- 用椭圆形表示实体集的属性，并用线段将其与相应的实体集联系起来。例如学生具有学号、姓名、性别、年龄、年级等属性，如图 1-3 (a) 所示。
- 用菱形表示实体集间的联系，菱形内写上联系名，并用线段分别与有关实体集连接起来，同时在线段旁标出联系的类型。如果联系具有属性，则该属性仍用椭圆形框表示，仍要用线段将属性与其连接起来。联系的属性必须在 E-R 图上标出，不能通过数据字典说明。如供应商、项目和零件之间存在有供应关系，如图 1-3 (b) 所示。客户和银行贷款及其他之间的联系，如图 1-3 (c) 所示。

1.3.2 数据模型

数据抽象是数据库系统的主要特点之一，具有很大的优越性。数据模型是实现数据抽象的主要工具。下面介绍数据模型的三要素和几种重要的数据模型。

一、数据模型的三要素

(1) 数据结构

数据结构是所研究的对象类型的集合，它是刻画一个数据模型性质最基本的方面。数据结构是对系统静态特征的描述。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的数据模型分别命名为层次模型、网状模型和关系模型。

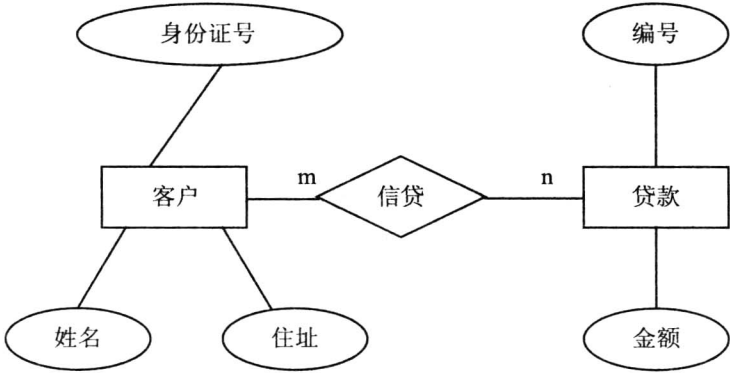
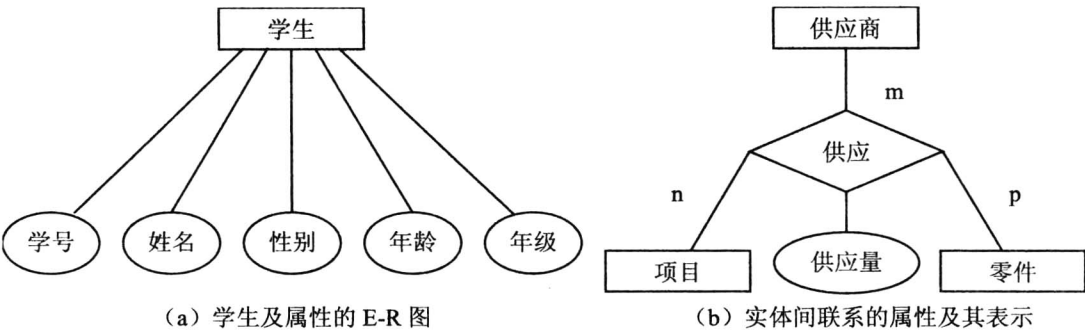


图 1-3 实体与属性以及实体间联系的属性的 E-R 图

(2) 数据操作

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，包括操作和相关的操作规则。数据库主要有检索和更新（插入、删除、修改）两大类操作。数据操作是对系统动态特性的描述。

(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依存规则，用于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、相容。

不同数据模型实际上是数据和信息建模的不同工具，根据模型应用的不同目的，可将数据模型分为 4 种类型：层次模型、网状模型、关系模型和面向对象模型。