

普通高校计算机主干课程辅导与提高丛书

盛定宇 编著

数据库原理 辅导与提高



清华大学出版社

普通高校计算机主干课程辅导与提高丛书

数据库原理辅导与提高

盛定宇 编著

清华大学出版社

北 京

内 容 简 介

“数据库原理”作为研究数据库技术的基本理论已经成为学习信息技术的重要专业课程。在高等教育的计算机科学与技术、电子信息科学与技术、信息管理与信息系统等有关专业,以及自考考试的计算机科学与技术、计算机信息管理、计算机网络等专业都开设了相关的课程。

全书分为数据库系统的基本概念、关系数据语言、关系数据标准语言 SQL、关系数据理论、数据库设计、数据库系统的管理与控制、数据库技术的发展 7 个专题,对基本概念和方法做了扼要的阐述,并有大量习题和测试题供读者练习。本书适合于大学本专科学生学习,也可以作为报考硕士研究生考生的复习参考书。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

数据库原理辅导与提高/盛定宇编著. —北京:清华大学出版社, 2004

(普通高校计算机主干课程辅导与提高丛书)

ISBN 7-302-08126-3

I. 数… II. 盛… III. 数据库系统—高等学校—教学参考资料 IV. TP311.13

中国版本图书馆 CIP 数据核字(2004)第 012632 号

出 版 者:清华大学出版社 地 址:北京清华大学学研大厦

<http://www.tup.com.cn> 邮 编:100084

社 总 机:010-62770175 客户服务:010-62776969

责任编辑:刘利民

封面设计:钱 诚

版式设计:杨 洋

印 装 者:北京鑫海金澳胶印有限公司

发 行 者:新华书店总店北京发行所

开 本:185×260 印张:17 字数:372 千字

版 次:2004 年 3 月第 1 版 2004 年 3 月第 1 次印刷

书 号:ISBN 7-302-08126-3/TP·5870

印 数:1~5000

定 价:22.00 元

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770175-3103 或(010)62795704

丛 书 序

随着我国国民经济持续、稳定、快速的发展,我国高等教育正在经历一个前所未有的发展阶段,招生规模不断扩大。在世界范围内,信息技术得到了空前发展和广泛应用,社会迫切需要培养大量的计算机和信息技术方面的人才。计算机和信息类专业已经成为我国当前高等教育中最需要、发展最迅速、招生规模最大的热门专业之一。

由于计算机和信息类专业招生规模扩大,高等院校师资力量显得不足,讲授课程只能由小班改为大班或合班上课,对每个学生的批改作业时间和辅导答疑时间也相应地减少,但学生所使用的教材没有适应这种变化,以前学习计算机专业教材就比较难,现在相对难度更大了。这就要求配合使用针对性强、便于自学的辅导教材,以帮助解决在学习过程中遇到的各类问题,从而提高学习效率。现在社会上计算机类的教材比较多,但用于学生自主学习的辅助教材却很少,各地师生要求尽快组织出版此类教材的呼声越来越高。因此,清华大学出版社根据学生的迫切需要,本着对学生高度负责的精神,组织了一批具有相关课程教学经验和编写教材经验并深受读者好评的高等院校专家编写了这套辅助教材。

本套丛书将分批陆续推出,首批推出6本。分别是《计算机组成原理辅导与提高》、《数据结构辅导与提高》、《操作系统辅导与提高》、《计算机网络辅导与提高》、《数据库原理辅导与提高》、《软件工程辅导与提高》。

本套丛书的编委会从有利于学生自主学习的需要出发,经过多次酝酿和研讨,最后形成统一认识,采用统一的编写格式,主要特点如下:

- 从每门课程的学科内容出发,针对一般教材中的重点和难点,分专题进行循序渐进地解剖和分析,并尽量通过恰如其分的事例进行说明,使读者既能理解和掌握重要的概念,又能从理论和实际应用的结合上加深认识。
- 为了加强训练和应用,巩固所学的知识,在每本教材的附录中分别给出了针对课程内容的练习题和参考解答。练习题题型丰富,包含选择、填空、运算、分析、设计等题型,对于较难的习题在解答中同时给出了分析过程。
- 为了自测本人的学习效果,在附录中给出了一套试卷,要求两小时内做完。
- 本套丛书的读者对象为学习相应课程的本、专科学生以及考研的学生,他们可以根据各自需要选取有关内容。

欢迎广大读者对本丛书提出宝贵意见,电子邮件地址为 xuxk@crtvu.edu.cn。

丛书主编 徐孝凯

2003年12月

“普通高校计算机主干课程辅导与提高丛书”

组织与编著委员会

主 编 徐孝凯

成 员 (排列无先后)

任爱华 (北京航空航天大学)

宿红毅 (北京理工大学)

盛定宇 (首都经贸大学)

顾一禾 (南京理工大学)

王小铭 (华南师范大学)

刘世峰 (北京交通大学)

何 军 (中国人民大学)

郑 岩 (北京邮电大学)

顾 问 许卓群 (北京大学)

侯炳辉 (清华大学)

前 言

随着信息技术（包括计算机技术、通信技术和网络技术等的飞速发展，信息系统渗透到社会的各个领域，作为其核心和基础的数据库技术也得到了越来越广泛的应用。数据库的建设规模、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。

“数据库原理”作为研究数据库技术的基本理论已经成为学习信息技术的重要专业课程。在高等教育的计算机科学与技术、电子信息科学与技术、信息管理与信息系统等有关专业，以及自学考试的应用计算机科学与技术、计算机信息管理、计算机网络等专业都开设了相关的课程。

“数据库原理”所涉及的内容，不仅是这些相关专业必备的知识基础，也是从事信息产业工作人员的必备知识与技能，同时也是进一步深入研究数据管理与应用技术的出发点。

现在有很多关于数据库原理的教材和书籍，都很有特点，本书试图帮助读者对这一领域的知识建立一个比较系统的体系，掌握这一领域中必备的知识和方法。全书分为 7 个专题：数据库系统的基本概念、关系数据语言、关系数据标准语言 SQL、关系数据理论、数据库设计、数据库系统的管理与控制、数据库技术的发展。本书对所涉及的概念作了概要的叙述，重点内容通过实际例题加深对概念的理解，在一些更深入研究的问题上，提出主要思路。

在书后的附录中还配置了大量的习题，供读者通过做习题理解概念与方法，建议读者完成习题后再核对答案。本书针对大学专科、本科学生学习和报考硕士研究生的需要，内容深浅度有不同层次，供读者选择。

在附录中列出了 3 类自测试题，其中基础试题适合于大学专科或参加全国计算机等级考试二级学员的自我测试；综合测试题适合于大学本科或参加全国计算机等级考试三级（数据库技术）学员的自我测试；高级测试题适合于报考相关专业硕士研究生的学员的自我测试。在书后的参考文献中，还列出了相关知识所涉及的书籍。

本书是相关知识的总结，选用了相关书籍与教材中的内容与习题，虽然主要例题和习题是笔者在教学中编制的，但笔者确实得到很多启发和收益，因此向这些作者表示深切的感谢。

数据库技术涉及的领域很广，新的技术与方法也在不断出现。因本人水平有限，其中难免出现错误和不完善之处，真诚希望广大专家和读者给予批评指正，笔者将非常感谢。

编 者

2003 年 5 月于北京

目 录

专题 1 数据库系统的基本概念	1
1.1 本课程的目标与内容	1
1.1.1 数据库系统的概念与特点	1
1.1.2 数据库系统的发展	4
1.1.3 数据库学科的研究领域	5
1.1.4 学习本课程重点需要领会的问题	6
1.2 数据模型	6
1.2.1 数据模型的基本概念	6
1.2.2 关系数据模型的特点	9
专题 2 关系数据语言	11
2.1 关系数据模型的定义和性质	12
2.2 关系数据语言之一——关系代数	16
2.3 关系数据语言之二——关系演算	27
2.3.1 元组关系演算	27
2.3.2 域关系演算	33
专题 3 关系数据标准语言 SQL	38
3.1 数据库与表的定义	39
3.1.1 数据库定义	39
3.1.2 基本表结构的定义	39
3.1.3 数据完整性定义	40
3.1.4 索引的建立与删除	44
3.2 数据查询	45
3.2.1 简单查询	46
3.2.2 连接查询	48
3.2.3 嵌套查询	49
3.3 视图的定义与应用	53
3.3.1 建立视图	53
3.3.2 视图的应用	56
3.4 嵌入式 SQL	59
3.4.1 嵌入式 SQL 的基本原理和使用规则	59

3.4.2 嵌入式 SQL 的使用技术	60
专题 4 关系数据理论	64
4.1 关系数据模型的构造	64
4.1.1 问题的提出	64
4.1.2 函数依赖的概念和性质	66
4.1.3 关系模型的范式	67
4.1.4 多值依赖与 4NF	72
4.1.5 关系模式规范化的步骤	74
4.2 数据模型分析	76
4.2.1 函数依赖集合的规范化	76
4.2.2 关系模型的分析	79
4.2.3 关系模式分解的规则	80
4.2.4 连接依赖与 5NF	85
4.2.5 多值依赖公理系统概述	87
4.3 数据规范化理论证明问题	88
专题 5 数据库设计	92
5.1 数据库设计的方法与步骤	92
5.1.1 数据库设计的特点与步骤	92
5.1.2 需求分析	96
5.2 概念结构设计	100
5.2.1 概念结构设计的方法	100
5.2.2 概念模型的设计方法	104
5.2.3 概念结构模型的形成	108
5.3 逻辑结构设计	111
5.3.1 数据库逻辑模型的产生	112
5.3.2 设计用户模式	113
5.3.3 数据库逻辑设计实例分析	114
5.4 数据库物理设计及实施、运行与维护	120
5.4.1 数据库物理设计	120
5.4.2 数据库的实施、运行与维护	121
专题 6 数据库系统的管理与控制	125
6.1 数据库管理系统	125
6.1.1 数据库管理系统的功能	125
6.1.2 数据库管理系统的组成	126
6.1.3 应用系统对数据库管理系统的要求与选择	129

6.2 数据库系统的安全性与完整性	131
6.2.1 数据库的安全性	131
6.2.2 数据库的完整性	136
6.3 并发控制	141
6.3.1 基本概念	142
6.3.2 封锁机制	143
6.3.3 活锁和死锁	145
6.3.4 可串行性与两段锁协议	146
6.4 数据库恢复技术	148
6.4.1 故障的种类	149
6.4.2 数据转储与恢复	150
6.4.3 日志文件及其应用	151
专题 7 数据库技术的发展	154
7.1 面向对象数据库系统	154
7.1.1 面向对象程序设计方法	155
7.1.2 面向对象数据模型	156
7.1.3 面向对象数据库语言	159
7.1.4 面向对象数据库的模式演进	160
7.1.5 对象-关系数据库	161
7.2 分布式数据库系统	163
7.2.1 分布式数据库系统的特点与目标	163
7.2.2 分布式数据库系统的体系结构	165
7.3 网络环境下的数据库体系	167
7.3.1 客户机/服务器系统	167
7.3.2 开放式数据库的互连技术 (ODBC)	170
7.4 数据仓库与数据挖掘	174
7.4.1 数据仓库	174
7.4.2 数据仓库设计实例	181
7.4.3 数据挖掘技术概述	185
附录 A 综合练习题	188
附录 B 综合练习题参考答案	219
附录 C 自测试卷	248
附录 D 参考书目	258

专题 1 数据库系统的基本概念

随着信息技术（包括计算机技术、通信技术和网络技术等）的飞速发展，信息系统渗透到社会的各个领域，作为其核心和基础的数据库技术也得到了越来越广泛的应用。数据库的建设规模、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。

“数据库原理”作为研究数据库技术的基本理论已经成为学习信息技术的重要专业课程。在高等教育的计算机科学与技术、电子信息科学与技术、信息管理与信息系统等相关专业，以及自学考试的计算机科学与技术、计算机信息管理、计算机网络等专业都开设了相关的课程。

“数据库原理”所涉及的内容，不仅是这些相关专业必备的知识基础，也是从事信息产业工作人员的必备知识与技能，同时也是进一步深入研究数据管理与应用技术的出发点。

1.1 本课程的目标与内容

本课程从软件工程的角度研究数据管理技术，属于软件基础课程，而不是一般程序设计语言的课程。新的软件工具在不断地更新和改进，这种更新和改进是建立在一定的理论基础之上的，并且随着这些更新与改进，理论也在不断地提高与完善。本课程主要包含以下几个方面的问题。

1.1.1 数据库系统的概念与特点

在计算机存储设备中，数据是以字节为单位存储的。数据存储的基本单位是数据项（字段），进一步的组织是记录和（数据）文件。但数据库不是文件的简单堆积，而是加入了对数据的组织和管理。

1. 数据管理的意义

数据作为信息的载体，在计算机应用技术深入发展的今天，对数据处理的问题也有了更高的要求。在数据处理中，常常不仅需要应用复杂的数据模型和算法，而且还需要处理大量的数据。因此，数据组织和管理的问题就更加突出了。数据管理手段从人工管理到文件系统再到数据库系统，反映了对信息资源的重要性的认识和管理措施的提高。

2. 数据库系统的特点

(1) 数据结构化

在文件系统中,相互独立的文件记录内部结构的最简单形式是等长同格式记录的集合,数据库系统把文件系统中记录内部有结构的思想扩大到了多个记录型之间。从整体观点来看,不仅要考虑一个应用(程序)的数据结构,而且要考虑整个组织的数据结构问题。整个组织的数据结构化要求在描述数据时不仅描述数据本身,还要描述数据之间的联系。文件系统中尽管记录内部已有了某些结构,但记录之间是没有联系的,是孤立的。因此,数据的结构化是数据库主要特征之一,是数据库与文件系统的根本区别。

(2) 数据共享性高、冗余度小、易扩充

数据库从整体观点来看待和描述数据,数据不再是面向某一应用,而是面向整个系统,这可以大大减小数据的冗余度,既节约存储空间,减少存取时间;又可避免数据之间的不相容性和不一致性。

对数据库数据的应用可以有很灵活的方式,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。这就是弹性大、易扩充的特点。

(3) 数据独立性高

数据独立性是数据库领域的一个常用术语,它包括数据的物理独立性和数据的逻辑独立性。数据库系统提供了两方面的映像功能,一个是数据的存储结构与逻辑结构之间的映像或转换功能,另一个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。

第一种映像功能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,从而应用程序也不必改变。这就是数据和程序的物理独立性。简称数据的物理独立性。

数据库系统中某一类应用使用的数据通常是总体数据的子集,而且各类应用对同一数据的使用要求也不一定相同。数据库系统通常提供局部数据结构的说明功能。局部的数据结构可以按具体应用要求作一定的改变。系统提供对这些改变的映像和转换功能(即上面的第二种映像功能),使得当总体逻辑结构改变时,通过对映像的相应改变而保持局部逻辑结构不变。程序员是根据局部逻辑结构编写应用程序的,因而应用程序也就可以不必改变。这就是数据和程序的逻辑独立性,简称数据的逻辑独立性。

数据和程序的独立性,把数据的定义和描述从应用程序中分离出去。此外,数据的存取又由 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) 统一的数据管理和控制

数据库是系统中对用户的共享资源。计算机的共享一般是并发的,即多个用户同时存取数据库中的数据甚至可以同时存取数据库中的同一个数据。因此,数据库管理系统必须提供以下几个方面的数据控制功能。

■ 数据的安全性(Security)保护

数据的安全性,指保护数据以防止不合法的使用所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使用和处理。例如,系统用检查口令或其他手段来检查用户的身份,只有身份合格的用户才能进入数据库系统进行操作。提供用户密级和数据存取权限的定义机制,当用户对数据库执行操作时,系统自动检查用户能否执行这些操作。检查通过后才执行允许的操作。

■ 数据的完整性(Integrity)控制

数据的完整性,指数据的正确性,有效性和相容性。完整性检查提供必要的功能,保证数据库中的数据在输入、修改过程中始终符合原来的定义和规范,在有效的范围内,保证数据之间满足一定的关系。例如:月份是1~12之间的正整数;研究生性别是男或女;研究生年龄是大于15小于45的整数;研究生学号是惟一的;研究生所在的系、院必须是存在的有效的系、院等。

■ 数据库恢复(Recovery)

计算机系统的硬件、软件故障,操作员的失误以及人为的攻击和破坏,都会影响数据库中数据的正确性,甚至会造成数据库部分或全部数据的丢失。因此数据库管理系统必须能够进行应急处理,将数据库从错误状态恢复到某一已知的正确状态。

■ 并发控制(Concurrency)

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

(5) 数据的最小存取单位是数据项

数据库既可以存取数据库中某一个数据项或一组数据项,也可以存取一个记录或一组记录。

综上所述,数据库是长期存储在计算机内有大量的、组织的、共享的数据集合。它可以供各种用户共享且具有最小的冗余度和较高的数据与程序的独立性。由于多种程序并发地使用数据库,为了能及时有效地处理数据,并提高安全性和完整性,必须有一个软件系统——数据库管理系统DBMS(DataBase Management System),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,同时有多用户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。

3. 数据库的定义

有关数据库的定义,许多书籍有不同的叙述方式,本书从应用的角度定义如下。

数据库是长期存储在计算机存储设备内、有组织的、共享的数据集合。这些数据按一定的数据模型组织、描述、存储,具有较小冗余度、较高的数据独立性和易扩充性,并可各种用户共享。

4. 数据库系统的组成

使用数据库技术的计算机系统称为数据库系统,由如下几部分组成。

- (1) 数据库:包括实际存储的数据和对数据库的定义。
- (2) 硬件支持系统:包括计算机、内外存储器、输入输出设备和通信设备等。
- (3) 软件支持系统:包括操作系统、数据库管理系统以及应用开发系统。

(4) 人员：与数据库系统的设计、创建、使用、维护等工作相关的人员。包括

- ① 数据库管理员。
- ② 数据库设计人员。
- ③ 系统分析员。
- ⑤ 应用程序设计人员。
- ⑥ 各种系统用户和其他相关人员。

1.1.2 数据库系统的发展

数据库系统起源于 20 世纪 60 年代中期，其发展可以划分为三代。

第一代数据库系统，即层次数据库系统和网状数据库系统，主要支持层次和网状数据模型，其特点是支持三级抽象模式的体系结构；用指针来表示数据之间的联系；数据定义语言和数据操纵语言相对独立；数据库语言采用过程性语言。该数据库系统的发展过程如下：

1. 1964 年，美国通用电气公司的 Bachman 等人开发成功世界上第一个数据库管理系统 (DBMS) ——IDS (Integrated Data Store) 系统，奠定了网状数据库系统的基础。

2. 1969 年，美国 IBM 公司开发成功世界上第一个商品化 DBMS 产品——IMS (Information Management System) 系统，这是一个层次数据库系统。

3. 1969—1970 年，美国 CODASYL (Conference On Data System Language) 协会下属的 DBTG (Database Task Group) 对数据库方法进行了系统的研究，提出了 DBTG 报告，建立了以网状数据模型为基础的数据库系统概念、方法和技术。

第二代数据库系统，即关系数据库系统 (RDBMS)，主要支持关系数据模型，有严格的理论基础，概念简单、清晰，易于理解和使用。关系模型一经提出，便迅速发展，成为实用性最强的产品。该数据库系统的主要特点是：概念单一化，数据及其数据间的联系都用关系来表示；以关系代数为理论基础；数据独立性强；数据库语言采用说明性语言，简化了编程难度。该数据库系统的发展如下：

1. 1970 年，美国 IBM 公司 San Jose 研究实验室的研究员 E.F.Codd 发表了题为“大型共享数据库数据的关系模型”论文，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理论的研究，奠定了关系数据模型的理论基础。由于 E.F.Codd 的杰出工作，他于 1981 年获得了 ACM 图灵奖。

2. 1974 年，IBM 公司 San Jose 研究实验室在 IBM System/370 系列机上研制成功关系数据库实验系统 System R (1974—1979 年)。这是世界上最早的、功能强大的关系数据库系统。1981 年 IBM 公司又宣布了具有 System R 全部特征的新的数据库软件产品 SQL/DS 的问世。

3. 1980 年以后，RDBMS 的产品迅速推出，如 Oracle、Sybase、Informix、dBASE、Foxbase、FoxPro 等。

4. 1990 年以后，RDBMS 产品的版本不断更新，功能更加强大，支持分布式数据库和

客户/服务器数据库以及客户/浏览器/服务器数据库等,同时实现了开放式网络环境下异质数据库的互联操作,以及在整个企业/行业范围内的 OLTP(在线事务处理)应用支持。

第三代数据库系统,即面向对象的数据库系统,它基于扩展的关系数据模型或面向对象的数据模型,是尚未成熟的一代数据库系统,其主要特点是支持包括数据、对象和知识的管理;在保持和继承第二代数据库系统的技术基础上引入面向对象技术;对其他系统开放,具有良好的可移植性、可连接性、可扩展性和可互操作性。

第三代数据库系统的典型代表包括 Servio 公司的 Gem Stone、OWTOS 公司的 ONTOS、Object Design 公司的 Object Stone、Objectivity 公司的 Objectivity/DB、Versant Object Technology 公司的 Versant 等。它们都支持严格面向对象的数据模型。与此同时,大多数商品化的关系数据库系统也对支持的数据模型进行了扩充,发展成了对象关系数据库系统(ORDBMS)。

1.1.3 数据库学科的研究领域

数据库学科的研究范围十分广泛,可以概括为 3 个主要领域:

1. 数据库管理系统软件的研制

DBMS 是数据库系统的基础。研制 DBMS 的基本目标是扩大功能、提高性能和可用性,从而提高用户的生产率。研制以 DBMS 为核心的一组相互联系的软件系统已成为当前数据库软件产品的开发方向,这些在 DBMS 基础上运行的软件系统有数据通信(DC)软件、表格软件(Forms)、数据字典、报表书写、图形系统等。

由于数据库应用领域的不断扩大,数据库不仅广泛应用于管理,而且已开始应用到工程设计、图形图像和声音等多介质处理、自动控制和计算机辅助设计等新的应用领域。这些新的应用领域所处理的数据和管理领域中数据的格式有极大的区别,如声音、图像等,称之为非格式化的数据,处理的要求也不大相同。因而,研究这些新的应用领域中的数据库方法、技术是一个新的课题。它不仅涉及应用系统的设计方法,而且涉及数据库系统的模型、实现技术等各种新的问题。面向对象的数据库系统、扩展的数据库系统、多介质数据库等研究方向的兴起就是基于这些新的需求和应用背景而产生的。

2. 数据库设计

在数据库管理系统的支持下,按照应用要求为某一部门或组织设计一个结构良好、使用方便、效率较高的数据库及其应用系统,这是数据库设计的主要含义。在这一领域内,主要的研究课题是数据库设计方法学和设计工具的探索。它包括数据库设计方法、设计工具和理论的研究,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

数据库理论研究主要集中于关系的规范化理论及关系数据理论。近年来,随着人工智能与数据库的结合,数据库和逻辑、逻辑演绎和知识推理等理论研究,以及演绎数据库、知识库系统的研制都已成为新的研究方向。

1.1.4 学习本课程重点需要领会的问题

本书从如下几个方面探讨学习数据库应深入领会的问题。

1. 关系数据语言

包括：关系代数、元组关系演算和域关系演算。

2. 关系数据标准语言 SQL

包括：数据库与表的定义、数据库的连接查询、数据库的嵌套查询、库函数的应用、视图的定义与应用以及嵌入式 SQL 等。

3. 关系数据理论

包括：规范化理论基础、数据模型分析以及证明题方法。

4. 数据库设计

包括：概念模型的构造、逻辑模型的构造、数据库设计、运行与维护。

5. 数据库的控制与管理

包括：日志文件及其应用、并发控制与封锁技术。

6. 数据库技术的发展

包括：面向对象数据库系统、布式数据库系统、网络环境下的数据库体系、数据仓库与数据挖掘。

1.2 数据模型

数据模型是现实世界数据特征的抽象，构造成一些相关的数据组织的集合。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。

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，所以读者必须事先把具体事物转换成计算机能够处理的数据。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。

1.2.1 数据模型的基本概念

现有的数据库系统都是基于某种数据模型的。数据模型是数据库系统的数学形式框架，

是用来描述数据的一组概念和定义。它包括以下方面的内容：

- 数据的静态特征，它包括对数据结构和数据间联系的描述。
- 数据的动态特征，它是一组定义在数据上的操作，包括操作的含义、操作符、运算规则及其语言等。
- 数据的完整性约束，这是一组规则，数据库中的数据必须满足这组规则。

数据模型应满足三方面要求：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模型化数据和信息的不同工具。根据模型应用的目的不同，可以将这些模型划分为两类，它们分别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。另一类模型是结构模型，主要包括网状模型、层次模型、关系模型和面向对象模型等，它是按计算机系统的观点对数据建模。

数据模型是数据库系统的核心和基础。

1. 数据模型三要素

(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，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，如网状模型中的系型（set type）。

数据结构是一个数据模型性质最重要的方面。因此，在数据库系统中，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。

(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存储规则，用以限定符合数据模型的数据库状态以及状态的变化，以保证数据的正确性、有效性和相容性。

数据模型应该反映和规定本数据模型必须遵守的通用的基本的完整性约束条件。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。

2. 主要数据模型

数据模型包括概念数据模型和结构数据模型，关于概念数据模型在后面数据库设计中将详细叙述，现在先介绍结构数据模型。

不同的数据模型具有不同的数据结构形式。目前最常用的数据模型有层次模型（Hierarchical model）、网状模型（Network model）、关系模型（Relational model）和面向对象模型（Object oriented model），其中层次模型和网状模型统称为非关系模型。非关系模型的数据库系统在 20 世纪 70 年代至 80 年代初非常流行，在数据库系统产品中占据了主导地位，现在已逐渐被关系模型的数据库系统所取代。

20 世纪 80 年代以来，面向对象的方法和技术在计算机各个领域，包括程序设计语言、软件工程、信息系统设计、计算机硬件设计等各方面都产生了深远的影响，也促进了数据库中面向对象数据模型的研究和发展。

（1）层次数据模型

层次模型是数据库系统中最早出现的数据模型，它用树形结构表示各类实体以及实体间的联系。现实世界中许多实体之间的联系本来就呈现出一种很自然的层次关系，如行政机构、家族关系等。层次模型数据库系统的典型代表是 IBM 公司的 IMS（Information Management Systems）数据库管理系统，这是一个曾经广泛使用的数据库管理系统。

（2）网状数据模型

在现实世界中实体型间的联系更多的是非层次关系，用层次模型表示非树形结构是很不直接的，而网状模型采用网状模型作为数据的组织方式，则可以克服这一弊病。网状数据模型的典型代表是 DBTG 系统，也称 CODASYL 系统。

层次数据模型和网状数据模型都是早期的数据库数据模型。数据库系统与文件系统的主要区别就在于不仅定义数据的存储而且还定义存储数据之间的联系，所谓“层次”和“网状”就是指这种联系的方式。

图1-1、图1-2表示的是这两种数据模型的实例。

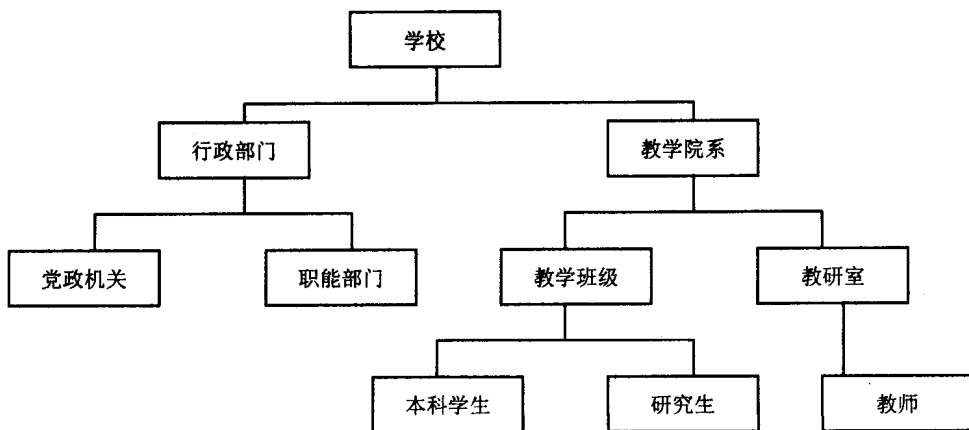


图 1-1 层次数据模型实例