

附：数据库原理自学考试大纲



数据库原理

组编 / 全国高等教育自学考试指导委员会
主编 / 丁宝康

计算机及应用专业(独立本科段)

全国高等教育自学考试指定教材

全国高等教育自学考试指定教材

学出版社

數
據
基
礎
理



全国高等教育自学考试指定教材

计算机及应用专业（独立本科段）

数 据 库 原 理

（附：数据库原理自学考试大纲）

全国高等教育自学考试指导委员会组编

丁宝康 主编

丁宝康 编著

李大学

经济科学出版社

图书在版编目 (CIP) 数据

数据库原理/丁宝康主编：全国高等教育自学考试指导委员会组编. —北京：经济科学出版社，2000.1

全国高等教育自学考试指定教材

ISBN 7-5058-2063-X

I . 数… II . ①丁…②全… III . 数据库系统-理论-高等教育-自学考试-教材 IV . TP311.131

中国版本图书馆 CIP 数据核字 (2000) 第 11487 号

数据 原 理

(附数据库原理自学考试大纲)

全国高等教育自学考试指导委员会组编

丁宝康 主编

丁宝康 李大学 编著

经济科学出版社出版

社址：北京海淀区万泉河路 66 号 邮编：100086

网址：www.esp.com.cn

电子邮件：esp@public2.east.net.cn

北京市鑫鑫印刷厂印刷

787×1092 16 开 15.50 印张 280000 字

2000 年 3 月第一版 2000 年 7 月第三次印刷

印数：18001~23000 册

ISBN 7-5058-2063-X/G·443 定价：20.40 元

(图书出现印装问题，请与当地教材供应部门调换)

(版权所有 翻印必究)

责任编辑：刁其武
责任校对：杨晓莹
版式设计：代小卫
技术编辑：董永亭

组 编 前 言

当您开始阅读本书时，人类已经迈入了 21 世纪。

这是一个变幻难测的世纪，这是一个催人奋进的时代。科学技术飞速发展，知识更替日新月异。希望、困惑、机遇、挑战，随时随地都有可能出现在每一个社会成员的生活之中。抓住机遇，寻求发展，迎接挑战，适应变化的制胜法宝就是学习——依靠自己学习、终生学习。

作为我国高等教育组成部分的自学考试，其职责就是在高等教育这个水平上倡导自学、鼓励自学、帮助自学、推动自学，为每一个自学者铺就成才之路。组织编写供读者学习的教材就是履行这个职责的重要环节。毫无疑问，这种教材应当适合自学，应当有利于学习者掌握、了解新知识、新信息，有利于学习者增强创新意识、培养实践能力、形成自学能力，也有利于学习者学以致用、解决实际工作中所遇到的问题。具有如此特点的书，我们虽然沿用了“教材”这个概念，但它与那种仅供教师讲、学生听，教师不讲，学生不懂，以“教”为中心的教科书相比，已经在内容安排、形式体例、行文风格等方面都大不相同了。希望读者对此有所了解，以便从一开始就树立起依靠自己学习的坚定信念，不断探索适合自己的学习方法，充分利用已有的知识基础和实际工作经验，最大限度地发挥自己的潜能，以达到学习的目标。

欢迎读者提出意见和建议。

祝每一位读者自学成功。

全国高等教育自学考试指导委员会

1999.5

编者的话

《数据库原理》是全国高等教育自学考试指导委员会组织的计算机及应用专业的教材。

数据库技术是计算机软件领域的一个重要分支，产生于 20 世纪 60 年代末。30 多年来，数据库技术得到迅速发展，并已形成较为完整的理论体系和一大批实用系统。随着数据库系统的推广使用，计算机应用已深入到社会各个角落。当今的管理信息系统（MIS）、办公信息系统（OIS）、计算机辅助设计与制造（CAD/CAM）、计算机集成制造系统（CIMS）、地理信息系统（GIS）等都以数据库技术为基础。20 世纪 90 年代初，我国已在邮电、计委、银行、电力、铁路、气象、民航、情报、公安、军事、航天和财税等行业装备了以数据库为基础的大型计算机系统。

在人类迈向 21 世纪知识经济的时候，信息变为经济发展的战略资源，信息技术已成为社会生产力中重要的组成部分。人们充分认识到，数据库是信息化社会中信息资源管理与开发利用的基础。对于一个国家，数据库的建设规模、使用水平已成为衡量该国信息化程度的重要标志。因此，数据库是计算机技术中一门重要的课程。

本课程的直接先修课程为“程序设计语言”和“数据结构”，后继课程为“软件工程”。本课程与“软件工程”、“计算机网络”等课程一起将成为学生毕业实习的基础。

本书较为详细地介绍了数据库系统的基本原理、方法和应用技术。全书共分 8 章，组织如下：

第 1 章介绍数据库系统的基本概念，包括数据模型、体系结构、全局结构等内容。

第 2 章介绍关系模型的运算理论：关系代数和关系演算。

第 3 章介绍关系数据库标准语言 SQL 的全貌。

第 4 章介绍关系数据库的模式设计理论，包括函数依赖、分解特性和范式等内容。

第 5 章介绍数据库应用系统设计的全过程，重点在概念设计和逻辑设计阶段。

第 6 章介绍数据库系统的保护技术，包括恢复、并发控制、完整性和安全性等内容。

第 7 章介绍分布式数据库系统的基本概念、体系结构、查询处理和客户/服务器结构的分布式系统。

第 8 章介绍具有面向对象特征的数据库系统：对象关系数据库和面向对象数据库。

本书第 4 章有关模式设计方面的内容理论性较强。学习时，读者可侧重于掌握其结论性内容及简单应用。

本书第 1、2、3、7、8 章由丁宝康撰写，第 4、5、6 章由李大学撰写。

中国纺织大学乐嘉锦教授、复旦大学周傲英教授和上海理工大学王槐珍教授审阅了全稿，并提出许多宝贵的意见，在此向他们致以衷心的感谢。

复旦大学施伯乐教授对全书的结构提出了指导性意见，楼荣生教授仔细校对了全书。上海交通大学陈敏逊教授为本书的出版进行了大量的工作。在此一并向他们表示感谢。

本书也可作为计算机专业和广大从事计算机工作的科技人员的自学参考书。
限于水平，书中欠妥之处，敬请广大读者和专家指正。

丁宝康 李大学

1999.9

于复旦大学计算机科学系

目 录

数据库原理

第 1 章 数据库概论	(1)
1.1 数据管理技术的发展	(1)
1.1.1 人工管理阶段	(1)
1.1.2 文件系统阶段	(2)
1.1.3 数据库阶段	(3)
1.2 数据描述	(5)
1.2.1 数据描述的三个领域	(5)
1.2.2 物理存储介质层次及数据描述	(7)
1.2.3 数据联系的描述	(8)
1.3 数据模型	(9)
1.3.1 数据模型的概念	(9)
1.3.2 实体联系模型	(10)
1.3.3 结构数据模型	(11)
1.3.4 面向对象模型	(15)
1.4 数据库的体系结构	(16)
1.4.1 三级模式结构	(16)
1.4.2 三级结构和两级映象	(17)
1.4.3 两级数据独立性	(18)
1.4.4 用户及用户界面	(19)
1.5 数据库管理系统	(19)
1.5.1 DBMS 的主要功能	(20)
1.5.2 DBMS 的组成	(21)
1.5.3 用户访问数据的过程	(22)
1.6 数据库系统	(23)
1.6.1 DBS 的组成	(23)
1.6.2 DBS 的全局结构	(24)
1.6.3 DBS 的效益	(26)
小结	(26)

习题 1	(27)
第 2 章 关系模型.....	(28)
2.1 关系模型的基本概念	(28)
2.1.1 二维表格的基本术语	(28)
2.1.2 关系模式、关系子模式和存储模式	(29)
2.1.3 关系模型的三类完整性规则	(31)
2.1.4 关系模型的形式定义	(32)
2.2 关系代数	(32)
2.2.1 关系代数的五个基本操作	(33)
2.2.2 关系代数的四个组合操作	(34)
2.2.3 关系代数表达式及其应用实例	(36)
2.2.4 扩充的关系代数操作	(38)
2.3 关系演算	(39)
2.3.1 元组关系演算	(39)
2.3.2 域关系演算	(42)
2.3.3 关系运算的安全性和等价性	(43)
2.4 查询优化	(44)
2.4.1 关系代数表达式的优化问题	(44)
2.4.2 关系代数表达式的等价变换规则	(45)
2.4.3 优化的策略	(47)
2.4.4 关系代数表达式的优化算法	(48)
小结	(51)
习题 2	(51)
第 3 章 关系数据库 SQL 语言	(54)
3.1 SQL 概述	(54)
3.1.1 SQL 发展历程	(54)
3.1.2 SQL 数据库的体系结构	(55)
3.1.3 SQL 的组成	(55)
3.2 SQL 的数据定义	(56)
3.2.1 SQL 模式的创建和撤消	(56)
3.2.2 SQL 提供的基本数据类型	(56)
3.2.3 基本表的创建、修改和撤消	(57)
3.2.4 视图的创建和撤消	(59)
3.2.5 索引的创建和撤消	(60)
3.3 SQL 的数据查询	(61)
3.3.1 SELECT 语句的基本句法	(61)
3.3.2 SELECT 语句完整的句法	(65)
3.3.3 SELECT 语句中的限定	(66)
3.4 SQL 的数据更新	(71)

3.4.1 数据插入	(71)
3.4.2 数据删除	(72)
3.4.3 数据修改	(73)
3.4.4 对视图的更新操作	(74)
3.5 嵌入式 SQL	(75)
3.5.1 SQL 语言的运行环境	(75)
3.5.2 嵌入式 SQL 的使用规定	(76)
3.5.3 嵌入式 SQL 的使用技术	(77)
小结	(80)
习题 3	(81)
第 4 章 关系数据库的模式设计	(83)
4.1 关系模式的设计问题	(83)
4.2 函数依赖	(85)
4.2.1 函数依赖的定义	(85)
4.2.2 函数依赖的逻辑蕴涵	(86)
4.2.3 键	(87)
4.2.4 函数依赖的推理规则	(88)
4.2.5 函数依赖推理规则的完备性	(89)
4.2.6 闭包的计算	(90)
4.2.7 函数依赖集的等价和覆盖	(92)
4.3 关系模式的分解特性	(94)
4.3.1 模式分解中存在的问题	(94)
4.3.2 无损联接	(96)
4.3.3 无损联接的测试	(96)
4.3.4 保持函数依赖的分解	(101)
4.4 关系模式的范式	(102)
4.4.1 第一范式	(102)
4.4.2 第二范式	(103)
4.4.3 第三范式	(104)
4.4.4 BCNF	(105)
4.4.5 分解成 BCNF 模式集的算法	(106)
4.4.6 分解成 3NF 模式集	(109)
4.4.7 模式设计方法的原则	(111)
4.4.8 多值依赖	(112)
4.4.9 第四范式	(114)
小结	(114)
习题 4	(115)
第 5 章 数据库设计	(117)
5.1 数据库设计概述	(117)

5.1.1 软件生存期	(117)
5.1.2 数据库系统生存期	(118)
5.1.3 数据库设计过程的输入和输出	(119)
5.1.4 数据库设计方法学	(120)
5.1.5 数据库设计的步骤	(122)
5.2 规划	(122)
5.3 需求分析	(123)
5.3.1 需求描述与分析	(123)
5.3.2 需求分析阶段的输入和输出	(124)
5.3.3 需求分析的步骤	(124)
5.3.4 数据字典	(127)
5.4 概念设计	(129)
5.4.1 概念设计的必要性	(129)
5.4.2 概念模型	(130)
5.4.3 概念设计的主要步骤	(130)
5.4.4 数据抽象	(131)
5.4.5 基本 ER 模型的扩展	(133)
5.4.6 ER 模型的操作	(133)
5.4.7 采用 ER 方法的数据库概念设计	(135)
5.5 逻辑设计	(141)
5.5.1 逻辑设计环境	(141)
5.5.2 逻辑设计的步骤	(142)
5.5.3 ER 模型向关系模型的转换	(143)
5.5.4 关系数据库的逻辑设计	(144)
5.6 物理设计	(146)
5.6.1 物理设计的步骤	(146)
5.6.2 物理设计的环境	(147)
5.6.3 物理设计的性能	(148)
5.7 实现与维护	(150)
5.7.1 数据库的实现	(150)
5.7.2 其它设计工作	(151)
5.7.3 运行与维护	(151)
小结	(152)
习题 5	(152)
第 6 章 数据库保护	(153)
6.1 数据库的恢复	(153)
6.1.1 事务的概念	(153)
6.1.2 事务的性质	(153)
6.1.3 故障类型和恢复方法	(155)

6.1.4 恢复的基本原则和实现方法	(156)
6.1.5 运行记录优先原则	(157)
6.1.6 SQL 中的恢复操作	(157)
6.2 数据库的并发控制	(158)
6.2.1 数据库的并发操作带来的问题	(158)
6.2.2 排它型封锁	(160)
6.2.3 活锁与死锁	(162)
6.2.4 共享型封锁	(163)
6.2.5 两段封锁法	(165)
6.3 数据库的完整性	(167)
6.3.1 完整性子系统	(167)
6.3.2 完整性规则	(167)
6.3.3 SQL 中的完整性约束	(168)
6.4 数据库的安全性	(172)
6.4.1 安全性级别	(172)
6.4.2 权限	(173)
6.4.3 权限的转授与回收	(173)
6.4.4 SQL 中的安全性控制	(174)
6.4.5 数据加密法	(176)
6.4.6 自然环境的安全性	(177)
小结	(178)
习题 6	(178)
第 7 章 分布式数据库系统	(180)
7.1 分布式数据库系统概述	(180)
7.1.1 集中式系统与分布式系统	(180)
7.1.2 分布式数据库系统的定义	(181)
7.1.3 分布式数据库系统的透明性	(182)
7.1.4 分布式数据库系统的优缺点	(183)
7.1.5 分布式数据库系统的分类	(184)
7.2 分布式数据库系统的体系结构	(184)
7.2.1 分布式数据存储	(184)
7.2.2 分布式数据库系统的体系结构	(186)
7.2.3 分布透明性	(188)
7.2.4 分布式数据库管理系统的功能及组成	(189)
7.2.5 分布式数据库系统中存在的问题	(190)
7.3 分布式查询处理	(190)
7.3.1 查询处理的传输代价	(191)
7.3.2 基于半联接的查询优化策略	(192)
7.3.3 基于联接的查询优化策略	(194)

7.4 客户/服务器结构的分布式系统	(194)
7.4.1 客户/服务器式 DBS	(195)
7.4.2 客户/服务器结构的分布式 DBS	(195)
小结	(197)
习题 7	(198)
第 8 章 具有面向对象特征的数据库系统	(199)
8.1 对象联系图	(200)
8.1.1 从关系到嵌套关系、复合对象	(200)
8.1.2 对象联系图	(201)
8.1.3 数据的泛化/细化	(202)
8.2 对象关系数据库	(203)
8.2.1 ORDB 的定义语言	(203)
8.2.2 ORDB 的查询语言	(206)
8.3 面向对象数据库	(207)
8.3.1 面向对象数据模型的基本概念	(208)
8.3.2 持久化程序设计语言	(210)
8.3.3 持久化 C++ 系统	(212)
小结	(214)
习题 8	(214)
参考文献	(216)

数据库原理自学考试大纲

出版前言	(219)
一、课程性质与设置目的	(221)
二、课程内容与考核目标	(222)
第 1 章 数据库概论	(222)
第 2 章 关系模型	(223)
第 3 章 关系数据库 SQL 语言	(224)
第 4 章 关系数据库的模式设计	(225)
第 5 章 数据库设计	(226)
第 6 章 数据库保护	(227)
第 7 章 分布式数据库系统	(228)
第 8 章 具有面向对象特征的数据库系统	(229)
实践环节	(230)
三、有关说明与实施要求	(232)
附录 题型举例	(234)
后记	(235)

第1章 数据库概论

从20世纪50年代中期开始，计算机的应用由科学部门逐步扩展到企业、行政部门。至60年代，数据处理成为计算机的主要应用。数据库技术作为数据管理技术，是计算机软件领域的一个重要分支，产生于60年代末。现已形成相当规模的理论体系和实用技术。本章先回顾数据管理技术的发展过程，然后介绍数据库中主要的概念，使读者对数据库技术的概貌有所了解。

1.1 数据管理技术的发展

计算机的数据处理应用，首先要把大量的信息以数据形式存放在存储器中。存储器的容量、存储速率直接影响到数据管理技术的发展。1956年生产的第一台计算机磁盘容量仅为5M字节（1M为 10^6 ），而现在已达10000M字节（表1.1）。目前光盘已经广泛使用，容量已达数百G字节（1G为 10^9 ）。存储器的发展，为数据库技术提供了良好的物质基础。

表1.1 磁盘容量的发展概况

时间(年)	1956	1965	1971	1978	1981	1985	1995
容量(MB/轴)	5	30	100	600	1200	5000	10000

使用计算机以后，数据处理的速度和规模，无论是相对于手工方式，还是机械方式，都是无可比拟的。在数据处理中，通常计算是比较简单的，而数据的管理比较复杂。数据管理是指数据的收集、整理、组织、存储、维护、检索、传送等操作，这部分操作是数据处理业务的基本环节，而且是任何数据处理业务中必不可少的共有部分。数据管理技术的优劣，将直接影响数据处理的效率。

数据管理技术的发展，与硬件（主要是外存）、软件、计算机应用的范围有密切的联系。数据管理技术的发展经过三个阶段：人工管理阶段、文件系统阶段和数据库阶段。

1.1.1 人工管理阶段（20世纪50年代中期以前）

这一阶段的计算机主要用于科技计算。外存只有磁带、卡片和纸带等，软件只有汇编语言，尚无数据管理方面的软件。数据处理方式是批处理。这个时期数据管理的特点如下：

- (1) 数据不保存在机器中。计算机主要用于计算，算题时将数据输入，计算后将结果数据输出。随着计算任务的完成，数据空间与程序空间一起被释放。
- (2) 没有专用的软件对数据进行管理。每个应用程序要包括存储结构、存取方法、输入

输出方式等。程序中的存取子程序随着存储结构的改变而改变，因而程序与数据不具有独立性。存储结构改变时，应用程序必须改变。

(3) 只有程序 (program) 的概念，没有文件 (file) 的概念。数据的组织方式必须由程序员自行设计与安排。

(4) 数据面向应用。即一组数据对应一个程序。

1.1.2 文件系统阶段 (20世纪50年代后期至60年代中期)

这一阶段，计算机不仅用于科技计算，还用于信息管理。随着数据量的增加，数据的存储、检索和维护问题成为紧迫的需要，数据结构和数据管理软件迅速发展起来。此时，外存已有磁盘、磁鼓等直接存取存储设备。软件领域出现了高级语言和操作系统。操作系统中的文件系统是专门管理外存的数据管理软件。处理的方式有批处理，也有联机实时处理。这一阶段数据管理的特点如下：

(1) 数据可长期保存在外存的磁盘上。由于计算机应用转向信息管理，因此对文件要进行大量的查询、修改和插入等操作。

(2) 数据的逻辑结构与物理结构有了区别。程序与数据之间具有“设备独立性”，即程序只需用文件名就可与数据打交道，不必关心数据的物理位置，由操作系统的文件系统提供存取方法 (读/写)。

(3) 文件组织已呈现多样化，有索引文件、链接文件和散列文件等。但文件之间相互独立，缺乏联系。数据间的联系要通过程序去构造。

(4) 数据不再属于某个特定的程序，可以重复使用。但文件结构的设计仍然是基于特定的用途，程序基于特定的存储结构和存取方法，因此程序与数据结构之间的依赖关系并未根本改变。

在文件系统阶段，由于具有设备独立性，因此改变存储设备，不必改变应用程序。但这只是初级的数据管理，还未能彻底体现用户观点下的数据逻辑结构独立于数据在外存的物理结构的要求。在数据物理结构修改时，仍然需要修改用户的应用程序。

文件系统阶段是数据管理技术发展中的一个重要阶段。在这一阶段中，得到充分发展的各种数据结构和算法丰富了计算机科学，为数据管理技术的进一步发展打下了基础。

随着数据管理规模的扩大，数据量急剧增加，文件系统显露出三个缺陷：

(1) 数据冗余性 (redundancy)。由于文件之间缺乏联系，造成每个应用程序都有对应的文件，有可能同样的数据在多个文件中重复存储。

(2) 数据不一致性 (inconsistency)。这往往是由数据冗余造成的，在进行更新操作时，稍不谨慎，就可能使同样的数据在不同的文件中不一样。

(3) 数据联系弱 (poor data relationship)。这是由文件之间相互独立，缺乏联系造成的。

由于这些原因，促使人们研究新的数据管理技术，因此，在20世纪60年代末产生了数据库技术。

1.1.3 数据库阶段（20世纪60年代末开始）

随着数据管理规模一再扩大，数据量急剧增加，为了提高效率，开始时，人们只是对文件系统加以扩充，在应用文件中建立了许多辅助索引，形成倒排文件系统。但这并不能最终解决问题。在20世纪60年代末，磁盘技术取得重要进展，具有数百兆容量和快速存取的磁盘陆续进入市场，成本也不高，为数据库技术的产生提供了良好的物质条件。

20世纪60年代中期，在美国社会上出现的许多系统号称 database 或 databank，但还不能真正称为数据库系统，大多是倒排文件系统。数据管理技术进入数据库阶段的标志，是60年代末发生的三件大事：

- (1) 1968年，美国IBM公司推出层次模型的IMS系统；
- (2) 1969年10月，美国数据系统语言协会(CODASYL)的数据库任务组(DBTG)发表关于网状模型的DBTG报告(1971年通过)；
- (3) 1970年，美国IBM公司的E.F.Codd连续发表论文，提出关系模型，奠定了关系数据库的理论基础。

20世纪70年代以来，数据库技术得到迅速发展，开发了许多有效的产品并投入运行。数据库系统克服了文件系统的缺陷，提供了对数据更高级更有效的管理。概括起来，数据库阶段的管理方式具有以下特点：

(1) 采用复杂的数据模型表示数据结构。数据模型不仅描述数据本身的特点，还描述数据之间的联系。这种联系通过存取路径实现。通过所有存取路径表示自然的数据联系是数据库与传统文件的根本区别。数据不再面向特定的某个或多个应用，而是面向整个应用系统。数据冗余明显减少，实现了数据共享。

(2) 有较高的数据独立性。数据的逻辑结构与物理结构之间，差别可以很大。用户以简单的逻辑结构操作数据，而无需考虑数据的物理结构。数据库的结构分成用户的逻辑结构、整体逻辑结构和物理结构三级(图1.1)。用户(应用程序或终端用户)数据和外存中数据之间的转换由数据库管理系统实现。在物理结构改变时，尽量不影响整体逻辑结构、用户的逻辑结构以及应用程序，这样我们就认为数据库达到了物理数据独立性。在整体逻辑结构改变时，尽量不影响用户的逻辑结构以及应用程序，这样我们就认为数据库达到了逻辑数据独立性。

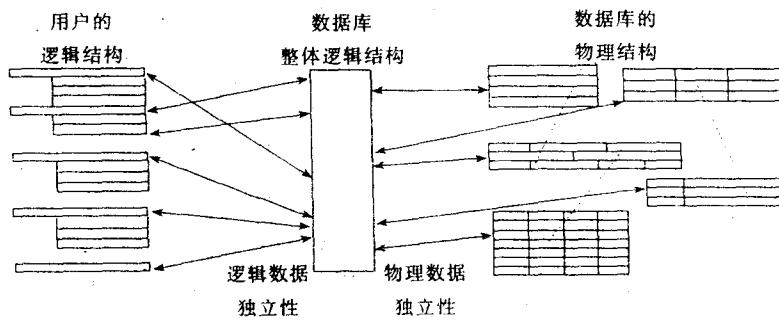


图1.1 数据库系统的结构