

3导 自考 3导丛书

财经·管理·教育

最新版

核心学案

自学考试新教材·计算机及应用专业(二)

数据库原理

教材依据/经济科学出版社《数据库原理》丁宝康/主编
组编/全国高等教育自学考试命题研究组



同步辅导同步过关



指定教材核心浓缩



预测试卷历年真题



航空工业出版社

应对自学考试课程大规模修订后新教材内容



导数·导学·导考

高等教育自学考试了导丛书

数据库原理

教材依据 / 经济科学出版社《数据库原理》主编 / 丁宝康
组 编 / 全国高等教育自学考试命题研究组

应对自考课程大规模修订后新教材内容

自学考试新教材

核学案

江苏工业学院图书馆
藏书章



航空工业出版社

图书在版编目(CIP)数据

数据库原理/自学考试命题研究组,《数据库原理》
编委会编. —北京:航空工业出版社,2005.5
(自学考试新教材核心学案. 计算机及应用专业. 第
2辑)
ISBN 7-80183-608-1

I. 数... II. ①自... ②数... III. 数据库系统—高
等教育—自学考试—自学参考资料 IV. TP311.13

中国版本图书馆 CIP 数据核字(2005)第 056050 号

数据库原理

Shujuku Yuanli

航空工业出版社出版发行

(北京市安定门外小关东里 14 号 100029)

发行部电话:010-84926529 010-64978486

北京市通县华龙印刷厂印刷

全国各地新华书店经售

2005 年 6 月第 1 版

2005 年 6 月第 1 次印刷

开本: 850 × 1168 1/32

印张: 55

字数: 2030 千字

(全 10 册) 定价: 140.00 元

简 介



张立勇 一个普通的农民孩子，清华大学打工8年，一直坚持刻苦自学，不仅80分以上通过四级、六级考试，托福考试630分，而且获得了北京大学本科文凭。2004年10月共青团中央向张立勇颁发了“中国青年学习成才奖”，他被誉为共青团中央树立的全国十大杰出学习青年之一。

张立勇的事迹被中央电视台“东方之子”“面对面”“新闻会客厅”等多个栏目采访报道，被北京电视台、中国教育电视台等电视媒体，新浪网、雅虎网等网络媒体，《人民日报》《中国青年报》《大学生》等报纸杂志，共100多家媒体采访报道，在社会上引起很大反响。被众多青年学子视为学习的榜样。

“**因**为我选择了这样一条自己的人生道路，所以我没有机会像大多数的学子那样，经历从学校到学校，顺利地接受高等教育的过程。我只能通过自学来圆我的大学梦。”

“**我**常常想，上帝会厚爱每一个人的，它会用不同的方式对你所付出的艰辛和努力给予补偿。但是，上帝只钟爱那些自助的人。如果你不努力，你不拼搏，所有的机会都会和你失之交臂。如果在这十年之中，我放弃了对人生理想和人生价值的追求，那么，当这一切机遇到来的时候，我又怎么可能把握住呢？”

“**大**家觉得我是一个榜样，但我个人并不这么想。社会把我放到这样的位置，充当这样的角色，能够影响一些人，这是最让我自豪的。”

----- **张立勇**



★ 编委会 ★

导教 · 导学 · 导考

★ 编委主任：程 琨 魏 莹

★ 编委名单：（按姓氏笔画排列）

万 鹏 刘 斌 刘海飞 刘 涛

闫树茂 宋玉珍 张 沁 张远盛

肖 果 邵桂英 崔海燕 程 琨

董金波 董 蕾 蒋 怡 魏 莹



★前言★

导学·导考·导练



“其实人的智力相差并不悬殊,可毅力的差距却使每个人拥有各自不同的前途。尤其是对于参加自考的人来说,毅力是非常重要的,当然还需要有得当的学习方法。”

“有很多人抱怨自考难以通过,然而正是这种严格的管理制度保证了自考毕业生的质量,使自考生获得了社会的认可和一致的好评。”

——一名从自考获得本科学历后又考上硕士生直到博士生的成功者的自述

参加自学考试,除了需要具备以上成功者所提到的毅力和方法外,还应该了解自考的每门课程都采用我们通常所说的“过关”考试——只要通过课程的一次性考试,就可拿到课程的学分,通过某专业要求课程的全部考试,也就会顺利获得这个专业的自考毕业证。然而,一分之差也会导致参考课程过关失败,有些考生难免多次重考才能修完规定课程。因此,在本书的编写过程中,编委们反复研讨自学考试的特点,努力寻求帮助自考生的有效途径。本书是多位学者、专家,历时数年的产物,具有以下优点。



掌握核心内容,了解命题动态,注重知识系统化

了解命题精神,是自学考试的核心,是达到专业标准的关键。自学考试的课程命题以课程自学考试大纲为依据,以最新指定教材为范围。本书紧紧贴住每一门课程的考试大纲和指定教材,用【考纲要求提示】、【知识结构图示】、【核心内容速记】、【同步精华题解】、【典型例题解析】等多个栏目解剖教材内容,是一套脉络清晰的速成讲义,可以使考生在厚厚的教材中抓住重点,对教材的系统学习有极强的指导作用。同时,对于临考考生,它又可以成为离开教材仍能独立使用的贴身笔记。《核心学案》摒弃了一些辅导书的题海战术,引导考生重视教材的学习。那么怎样去自学才能弄懂教材并将厚书读“薄”呢?抓住重点才是关键。《核心学案》用清晰的思路,帮助考生将教材知识系统化,使考生在答卷时知识系统、逻辑清晰、胸有成竹。



依据权威资料,重视最新信息,紧跟时代脉搏

参加高等教育自学考试的考生,常常会感到市面上的辅导资料甚至教材都有



滞后性。全国高教自考办也认可这一事实,并采取了一些有效措施,比如在发布考试大纲和指定教材的基础上又组编了《全国高等教育自学考试活页丛书》等补充学习材料,并明文规定增补内容纳入统一命题范围,要占卷面 5~10 分。同时高教自考办还加快了教材的修订频率。面对这种情况,原有的一些辅导资料的严重滞后和内容缺陷也是必然的。本套《核心学案》则高度重视这一现象,在依据考试大纲和指定教材时,选用高教自考办的最新修订本(2004 年起自考课程已在做大规模修订),并将活页丛书等内容融会贯通其中,有的科目还特意增加了【最新内容补充】以引起考生重视。另外,本套书还吸收了许多自考强化班的授课精华,目的是帮助考生了解最新考试动态。我们还将开通网上自考辅导随时更新有关内容和提供特色售后服务,欢迎点击 www.study-book.com.cn。

三

做到讲练结合,力求精讲精练,提高辅导命中率

本套书配有【同步精华题解】和综合演练题,是在对考纲、教材归纳总结后选编的一些经典同步练习题。这些练习题的题型与考试题型完全一致,使考生能够迅速掌握答题方法与同步要点。另外,本书的编者还依据各科内容,遴选考点,在对历年实考真题做详细分析的基础上精编了《命题预测试卷》。这些试卷不仅题型题量完全与真考试卷保持一致,而且力求覆盖考试大纲的各科重点。考生如果在学习《核心学案》的基础上再认真研习《命题预测试卷》,既可熟悉题型、了解试卷难易度,又可将其作为自测、练习之用,找出差距,查漏补缺。因此,在《核心学案》的首印首发优惠活动中,为了帮助考生用好的学习方法提高应试过关率,我们特意将《命题预测试卷》作为《核心学案》的赠品送给每个考生。这样,本书即成为真正具有命中率的辅导用书。

总之,面对数千万的自考考生,我们是抱着高度的责任感来完成这项使命的。我们的目的是:减轻考生的学习负担;我们口号是:用最短的时间使考生自考过关!因为工作量的巨大和考期的压力,也许我们遗留了某些不足,欢迎读者批评指正。来函可致: reader@study-book.com.cn,我们将高度重视,以求完善。



第一章 数据库概论

考纲要求提示	(1)
知识结构图示	(1)
核心内容速记	(2)
同步精华题解	(14)



第二章 关系模型

考纲要求提示	(18)
知识结构图示	(18)
核心内容速记	(18)
同步精华题解	(27)



第三章 关系数据库 SQL 语言

考纲要求提示	(31)
知识结构图示	(31)
核心内容速记	(31)
同步精华题解	(41)



第四章 关系数据库的模式设计

考纲要求提示	(45)
知识结构图示	(45)
核心内容速记	(46)
同步精华题解	(55)



第五章 数据库设计

考纲要求提示	(59)
知识结构图示	(59)
核心内容速记	(60)
同步精华题解	(73)



第六章 数据库保护

考纲要求提示	(78)
知识结构图示	(78)
核心内容速记	(79)
同步精华题解	(85)



第七章 分布式数据库系统

考纲要求提示	(89)
知识结构图示	(89)
核心内容速记	(89)
同步精华题解	(94)



第八章 具有面向对象特征的数据库系统

考纲要求提示	(98)
知识结构图示	(98)
核心内容速记	(98)
同步精华题解	(101)



综合演练题

(105)



综合演练题参考答案

(109)



第一章 数据库概论



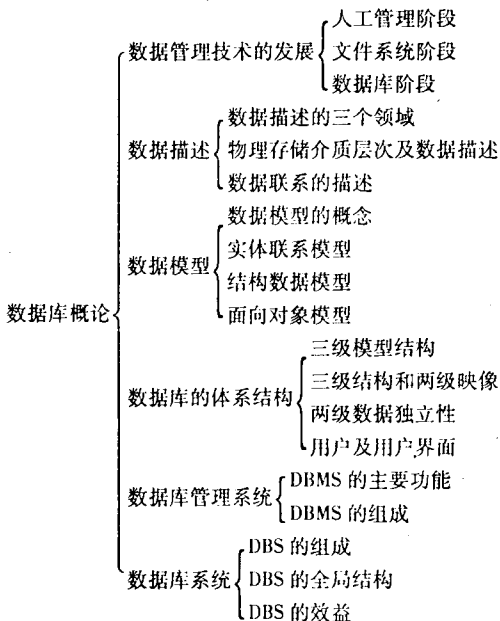
考纲要求提示

1. 了解数据管理技术的发展阶段,数据描述的术语,数据模型的概念,数据库的体系结构,数据库管理系统的功能及组成,数据库系统的组成及全局结构;

2. 本章的重点是实体间联系,数据模型,数据库体系结构,数据库系统的全局结构。



知识结构图示



核心内容速记

一、数据管理技术的发展

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。

(一)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。这个时期数据管理的特点如下:

- (1)数据不保存在机器中。
- (2)没有专用的软件对数据进行处理。
- (3)只有程序(program)的概念,没有文件(file)的概念。
- (4)数据面向应用,即一组数据对应一个程序。

(二)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。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。
- (2)数据的逻辑结构与物理结构有了区别。
- (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。
- (4)数据不再属于某个特定的程序,可以重复使用。

文件系统阶段是数据管理技术发展中的一个重要阶段。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)。
- (2)数据不一致性(inconsistency)。
- (3)数据联系弱(poor data relationship)。

(三)数据库阶段(20世纪60年代末开始)

1. 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功能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。

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。

2. 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。

(2)有较高的数据独立性。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。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。

(3)数据库系统为用户提供方便的用户接口,用户可以使用查询语言或终端命令操作数据库,也可以用程序方式(用 COBOL、C 一类高级语言和数据库语言编制的程序)操作数据库。

(4)系统提供四个方面的数据控制功能:数据库的恢复,并发控制,数据完整性和数据安全性,以保证数据库中数据是安全的、正确的和可靠的。

(5)对数据的操作不一定以记录为单位,也可以以数据项为单位,增加了系统的灵活性。

这个阶段的程序和数据间的联系可用图 1.1 来表示。

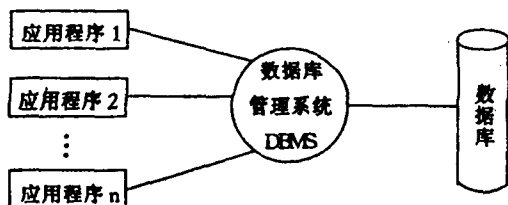


图 1.1 程序和数据的联系

3. 在数据库技术中有四个名词,分别是:

(1)数据库(Database, DB):DB 是统一管理的相关数据的集合。

(2)数据库管理系统(Database Management System, DBMS):DBMS 是位于用户与操作系统之间的一层数据管理软件,为用户或应用程序提供访问 DB 的方法,包括 DB 的建立、查询、更新及各种数据控制。DBMS 总是基于某种数据模型,可以分为层次型、网状型、关系型或面向对象型 DBMS。

(3)数据库系统(Database System, DBS):DBS 是实现有组织地、动态地存储大量关联数据,方便多用户访问的计算机软件、硬件和数据资源组成的系统,即采用了数据库技术的计算机系统。

(4)数据库技术:这是一门研究数据库的结构、存储、管理和使用的软件学科。

二、数据描述

(一)数据描述的三个领域

在数据处理中,数据描述将涉及不同的范畴。从事物的特性到计算机中的数据表示,经历了三个领域:现实世界、信息世界和机器世界。

1. 现实世界

存在于人们头脑之外的客观世界,称为现实世界(real world)。例如,仓库管理中涉及的货物管理,有货物的存放、货物的进出、货物的检查等,这里就可能有许多报表、图表、表格。

2. 信息世界

信息世界是现实世界在人们头脑中的反映,人们把它用文字和符号记载下来。在信息世界中,数据库技术用到下列一些术语:

(1)实体(entity):客观存在并且可以相互区别的事物称为实体。实体可以是具体的对象,例如,一个男学生、一辆汽车等。也可以是抽象的事件,例如,一次借书、一次足球比赛等。

(2)实体集(entity set):性质相同的同类实体的集合,称为实体集。例如,所有的男学生、全国足球锦标赛的所有比赛等。

(3)属性(attribute):实体有很多特性,每一个特性称为属性。每个属性有一个值域,其类型可以是整数型、实数型或字符串型。例如,学生有学号、姓名、年龄、性别等属性,相应值域为字符、字符串、整数和字符串型。

(4)实体标识符(identifier):能惟一标识每个实体的属性或属性集,称为实体标识符,或简称键。例如,学生的学号可以作为学生实体标识符。

3. 机器世界

信息世界的信息在机器世界中以数据形式存储。机器世界中数据描述的术语有以下四个:

(1)字段(field):标记实体属性的命名单位称为字段或数据项。它是以命名的最小信息单位,所以又称为数据元素或初等项。字段的命名往往和属性名相同。例如,学生有学号、姓名、年龄、性别等字段。

(2)记录(record):字段的有序集合称为记录。一般用一个记录描述一个实体,所以记录又可以定义为能完整地描述一个实体的字段集。例如,一个学生记录,由有序的字段集组成,即学号,姓名,年龄,性别等。

(3)文件(file):同一类记录的汇集称为文件。文件是描述实体集的,所以它又可以定义为描述一个实体集的所有记录的集合。例如,所有的学生记录组成了一个学生文件。

(4)关键码(key):能惟一标识文件中每个记录的字段或字段集,称为记录的关键码(简称“键”)。这个概念与实体标识符的概念相对应。例如,学生的学号可以作为学生记录的关键码。

机器世界和信息世界术语的对应关系如表 1.1 所示

表 1.1 术语的对应关系

信息世界	机器世界
实体·····	记录
属性·····	字段 (数据项)
实体集·····	文件
实体标识符·····	关键字

在数据库中,每个概念都有类型(type)和值(value)的区分。类型是概念的内涵,而值是概念的外延。

数据描述有两种形式:物理描述和逻辑描述。物理数据描述指数据在存储设备上的存储方式,物理数据是实际存放在存储设备上的数据。逻辑数据描述指程序员或用户用以操作的数据形式,是抽象的概念化数据。

在数据库系统中,逻辑数据与物理数据之间可以差别很大。数据管理软件的功能之一,就是要把逻辑数据转换成物理数据,以及把物理数据转换成逻辑数据。

(二)物理存储介质层次及数据描述

1. 物理存储介质层次

根据访问数据的速度、成本和可靠性,计算机系统的存储介质可分为以下六类:

(1)高速缓冲存储器(cache):cache是访问速度最快,也是最昂贵的存储器,容量小,由操作系统直接管理。

(2)主存储器(main memory):简称为主存或内存。机器指令可以直接对主存中的数据进行操作。最致命的一点是,在掉电或系统崩溃时,主存数据立即全部丢失。

(3)快擦写存储器(flash memory):又称为“电可擦可编程只读存储器”(即EEPROM),简称“快闪存”。快闪存存在掉电后能保持数据不丢失,操作速度略低于主存。

(4)磁盘存储器(magnetic-disk):磁盘是目前最流行的外部存储器,能长时间地联机存储数据。并能直接读取数据,被称为“直接存取存储器”。在掉电或系统崩溃后,数据不会丢失。

(5)光存储器(optical storage):目前流行的光存储器是“光盘只读存储器”(CD-ROM)。数据以光的形式存储在盘里,然后用一个激光器去读。

(6)磁带(tape storage):用磁带存储拷贝的数据或归档的数据。在存储器中,磁带价格最便宜,属于“顺序存取存储器”,每盘有5G字节。

存储介质组成了计算机系统的存储层次(见图1.2)。图1.2中最高一级

的高速缓存价格最昂贵,访问速度也最快。自上而下,每位(bit)数据的成本越来越低,但访问速度越来越慢。

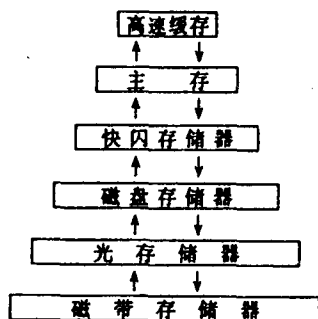


图 1.2 存储器设备层次

2. 物理存储中的数据描述

在存储器中用到下列数据描述的术语:

- (1) 位(bit, 比特): 一个二进制位称为“位”。一位只能取 0 或 1 状态。
- (2) 字节(byte): 8 个比特称为一个字节, 可以存放一个字符对应的 ASCII 码。
- (3) 字(word): 若干个字节组成一个字。一个字所含的二进制位数称为字长。各种计算机的字长是不一样的。
- (4) 块(block): 又称为物理块或物理记录。块是内存和外存交换信息的最小单位, 每块的大小通常为 $2^8 \sim 2^{12}$ 字节。
- (5) 桶(bucket): 外存的逻辑单位, 一个桶可以包含一个物理块或多个在空间上不一定连续的物理块。

(6) 卷(volume): 一台输入输出设备所能装载的全部有用信息, 称为“卷”。

(三) 数据联系描述

现实世界中, 事物是相互联系的。这种联系必然要在信息世界中有所反映, 即实体并不是孤立静止存在的。实体的联系有两类: 一类是实体内部的联系, 反映在数据上是同一记录内部各字段间的联系; 另一类是实体与实体之间的联系, 反映在数据上是记录之间的联系。

实体间联系有两种: 一种是同一实体集中各个实体之间的联系; 另一种是不同实体集的各个实体之间的联系。

两个不同实体集的实体间联系有以下三种情况。

1. 一对一联系

如果实体集 E1 中每个实体至多和实体集 E2 中一个实体有联系,反之亦然,那么实体集 E1 对 E2 的联系称为“一对一联系”,记为“1: 1”。

例如,飞机的座位和乘客之间、学校和校长之间都是 1: 1 联系,如图 1.3 所示。

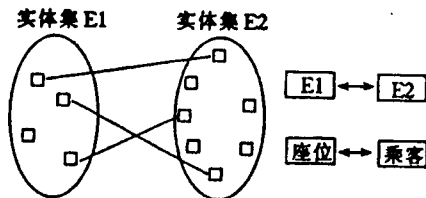


图 1.3 一对一联系

2. 一对多联系

如果实体集 E1 中每个实体与实体集 E2 中任意个(零个或多个)实体有联系,而 E2 中每个实体至多和 E1 中一个实体有联系,那么称 E1 对 E2 的联系是“一对多联系”,记为“1: N”。

例如,工厂里车间和工人之间、学校里系和学生之间都是 1: N 联系,如图 1.4 所示。

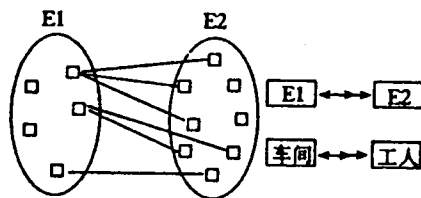


图 1.4 一对多联系

3. 多对多联系

如果实体集 E1 中每个实体与实体集 E2 中任意一个(零个或多个)实体有联系,反之亦然,那么称 E1 和 E2 的联系是“多对多联系”,记为“M: N”。

例如,学生与课程之间、商店和顾客之间都是 M: N 联系,如图 1.5 所示。上面三种联系是实体之间最基本的联系。

三、数据模型

(一) 数据模型的概念

模型是对现实世界的抽象。在数据库技术中,人们用模型的概念描述数据库的结构与语义,对现实世界进行抽象。表示实体类型及实体间联系的模式称为“数据模型”(data model)。

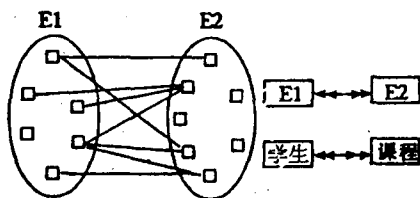


图 1.5 多对多联系

1. 数据模型的种类

目前广泛使用的数据模型可分为两种类型。

一种是独立于计算机系统的模型,完全不涉及信息在系统中的表示,只是用来描述某个特定组织所关心的信息结构,这类模型称为“概念数据模型”。这一类中著名的模型是“实体联系模型”。

另一种数据模型是直接面向数据库的逻辑结构,它是现实世界的第二层抽象。这类模型涉及到计算机系统和数据库管理系统,又称为“结构数据模型”。例如,层次、网状、关系、面向对象等模型。这类模型有严格的形式化定义,以便于在计算机系统中实现。

2. 结构数据模型的三个组成部分

结构数据模型应包含数据结构、数据操作和数据完整性约束三个部分。

数据结构是指对实体类型和实体间联系的表达和实现。

数据操作是指对数据库的检索和更新(包括插入、删除、修改)两类操作的实现。

数据完整性约束给出数据及其联系应具有的限制和依赖规则。

(二) 实体联系模型(ER 模型)

ER 模型是 P. P. Chen 于 1976 年提出的。这个模型是直接从现实世界中抽象出实体类型及实体间联系,然后用 ER 图表示的数据模型。设计 ER 图的方法称为 ER 方法。

ER 图是直观表示概念模型的工具。在 ER 图中有四个基本成分:

(1) 矩形框,表示实体类型(考虑问题的对象)。

(2) 菱形框,表示联系类型(实体间的联系)。

(3) 椭圆形框,表示实体类型和联系类型的属性。

相应的命名均记入各种框中。对于关键码的属性,在属性名下画一横线。

(4) 直线,联系类型与其涉及的实体类型之间以直线连接,并在直线端部标上联系的种类(1: 1, 1: N, M: N)。

ER 模型有两个明显的优点:接近于人的思维,容易理解;与计算机无关,