



教育部考试中心

授权独家出版

全国计算机等级考试

四级教程

——数据库工程师

(2008年版)



高等教育出版社
Higher Education Press

全国计算机等级考试

四级教程

——数据库工程师(2008年版)

教育部考试中心



高等教育出版社

内容提要

《全国计算机等级考试四级教程——数据库工程师(2008年版)》是根据教育部考试中心制定的《全国计算机等级考试四级数据库工程师考试大纲(2007年版)》的要求编写的。

这本教材涉及的考点是在三级考试的基础上进行的,全书按照数据库工程师考试大纲设定的六部分内容:数据库应用系统分析及规划、数据库设计及实现、数据库存储技术、并发控制技术、数据库管理与维护、数据库技术的发展及新技术,分为20章组织和论述数据库工程师考试涉及的相关知识和内容。

本书内容丰富翔实,理论和实践并重,实用性强,不仅可作为数据库工程师考试的标准教材,而且可用作高校计算机及相关专业的教材,也可作为实用的数据库技术培训教材和技术参考书。

图书在版编目(CIP)数据

全国计算机等级考试四级教程.数据库工程师:2008年版/
教育部考试中心. —北京:高等教育出版社,2007.9
ISBN 978-7-04-022294-4

I.全… II.教… III.①电子计算机—水平考试—教材
②数据库系统—水平考试—教材 IV.TP3

中国版本图书馆CIP数据核字(2007)第132542号

策划编辑 何新权 责任编辑 柳秀丽 封面设计 张志奇 责任印制 韩刚

出版发行	高等教育出版社	购书热线	010-58581118
社 址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landaco.com
经 销	蓝色畅想图书发行有限公司		http://www.landaco.com.cn
印 刷	北京民族印刷厂	畅想教育	http://www.widedu.com
开 本	787×1092 1/16	版 次	2007年9月第1版
印 张	24	印 次	2007年9月第1次印刷
字 数	570 000	定 价	38.20元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 22294-00

大力推行全国计算机等级考试 为发展知识经济、信息产业和培养计算机专门人才作出贡献 (代序)

中国科学院院士 北京大学信息与工程科学学部主任
全国计算机等级考试委员会主任委员
杨芙清

当今,人类正在步入一个以智力资源的占有和配置,知识生产、分配和使用为最重要因素的知识经济时代,也就是小平同志提出的“科学技术是第一生产力”的时代。科教是经济发展的基础,知识是人类创新的源泉。基础研究的科学发现、应用研究的原理探索和开发研究的技术发明,三者之间的联系愈来愈紧密,转换周期日趋缩短。世界各国的竞争已成为以经济为基础、以科技(特别是高科技)为先导的综合国力的竞争。

在高科技中,信息科学技术是知识高度密集、学科高度综合,具有科学与技术融合特征的学科。它直接渗透到经济、文化和社会的各个领域,迅速改变着人们的观念、生活和社会的结构,是当代发展知识经济的支柱之一。

在信息科学技术中,微电子是基础,计算机硬件及通信设施是载体,计算机软件是核心。软件是人类知识的固化,是知识经济的基本表征,软件已成为信息时代的新型“物理设施”。人类抽象的经验、知识正逐步由软件予以精确地体现。在信息时代,软件是信息化的核心,国民经济和国防建设、社会发展、人民生活都离不开软件,软件无处不在。软件产业是增长最快的朝阳产业,是具有高额附加值、高投入/高产出、无污染、低能耗的绿色产业。软件产业的发展将推动知识经济的进程,促进从注重量的增长向注重质的提高的方向发展,是典型的知识型产业。软件产业是关系到国家经济安全和文化安全,体现国家综合实力,决定 21 世纪国际竞争地位的战略性的产业。

为了适应知识经济发展的需要,大力推动信息产业的发展,需要在全民中普及计算机的基本知识,广开渠道,培养和造就一批又一批能熟练运用计算机和软件技术的各行各业的专门人才。

1994 年,原国家教委(现教育部)推出了全国计算机等级考试,它是一种重视应试人员对计算机和软件的实际掌握能力的考试。它不限制报考人员的学历背景,任何年龄段的人员都可以报考。这就为培养各行各业计算机的应用人才开辟了一条广阔的道路。

1994 年是推出计算机等级考试的第一年,当年参加考试的有 1 万余人;而 2006 年,年报考人数已近 356 万人。截至 2007 年上半年,全国计算机等级考试共开考 25 次,考生人数累计达 2269 万,其中有 826 万人获得了不同级别的计算机等级证书。

事实说明,鼓励社会各阶层的人士通过各种途径掌握计算机应用技术,并运用等级考试对他们的才干予以认真的、有权威性的认证,是一种人才培养的有效途径,是比较符合我国具体情况的。等级考试也为用人单位录用和考核人员提供了一种测评手段。从有关公司对等级考试所作



的社会抽样调查结果看,不论是管理人员还是应试人员,对该项考试的内容和形式都给予了充分的肯定。

计算机等级考试所取得的良好效果,也同全国各有关单位专家们在等级考试的大纲编写、试题设计、阅卷评分及效果分析等多项工作中所付出的大量心血和辛勤劳动密切相关,他们为这项工作的顺利开展作出了重要的贡献。

计算机与软件技术是一项日新月异的高新技术。计算机等级考试大纲有必要根据计算机与软件技术在近年的新发展,进行适当的修正,从而使等级考试更能反映当前计算机与软件技术的应用实际,使培养计算机应用人才的基础工作更健康地向前发展。

从面临知识经济的机遇与挑战这样一个社会大环境的背景出发,考察全国计算机等级考试,就会看到,这一举措是符合知识经济和信息产业的发展方向的,是值得大力推行的。

我们相信,在 21 世纪知识经济和信息产业加快发展的形势下,在教育部考试中心的精心组织领导下,在全国各有关专家们的大力配合下,全国计算机等级考试一定会以更新的面貌出现,从而为我国培养计算机应用专门人才的宏大事业作出更多的贡献。

2007 年 7 月

前言

本教程是根据教育部考试中心制定的《全国计算机等级考试四级数据库工程师考试大纲(2007年版)》的要求编写的。

本教程涉及的考点是在三级考试的基础上进行的,全书按照《全国计算机等级考试四级数据库工程师考试大纲(2007年版)》的要求设定六部分内容,即数据库应用系统分析及规划、数据库设计及实现、数据库存储技术、并发控制技术、数据库管理与维护、数据库技术的发展及新技术。全书内容分为20章,论述了数据库工程师考试涉及的相关知识和内容。第1章引论介绍了数据库系统的一些基本概念、数据模型、数据库的系统结构、数据库管理系统,使读者对数据库技术和方法有一个总体的了解;第2章数据库应用系统的生命周期从整体上介绍了数据库应用系统的开发方法及过程;第3章需求分析及功能建模方法详细讨论了需求分析的概念和任务,获取需求的过程以及两种较为流行的需求建模方法DFD和IDEFO方法;第4章数据库概念设计及数据建模主要介绍了数据库概念设计阶段的工作,数据库概念设计做什么,怎么做以及概念设计方法:ER方法和IDEFIX方法;第5章关系数据库逻辑设计主要介绍了数据库逻辑设计如何做,以及相关概念、理论、技术和方法;第6章存储技术与数据库物理设计主要讨论了数据库文件的组织和结构形式、索引技术以及数据库物理设计的内容和步骤;第7章数据库应用系统功能设计首先介绍了软件体系结构和软件设计方面的基础知识,然后按照DBAS生命周期模型,依次讨论了数据库应用程序总体设计、事务概要设计、应用程序概要设计、事务详细设计和应用程序详细设计的步骤和设计内容;第8章关系数据库操作语言SQL主要介绍了关系数据库操作语言SQL;第9章事务调度与并发控制主要介绍了事务的基本概念、性质、事务调度,并发控制技术和死锁的处理方法;第10章数据库的实施、运行和维护主要介绍了数据库实施的基本步骤,讨论了数据库运行和维护中涉及的相关问题,包括监控分析、空间管理、参数调整和查询优化;第11章故障管理主要介绍了数据库恢复的一般步骤,数据库中常见的故障及恢复策略;第12章至第15章结合考试用数据库系统,讨论了SQL Server 2000数据库的安装及环境设置、Transact-SQL语言、SQL Server数据库对象的创建及维护以及SQL Server数据库的安全控制模型及安全管理;第16章、第17章结合考试指定的编程环境,介绍了VB 6.0开发环境与数据访问接口,以及VB数据库编程的内容;第18章至第20章主要介绍了数据库技术的新进展。

本书在第2章就介绍了数据库应用系统的开发方法,并给出应用案例及需求,以后各章相关知识点的介绍都以这个应用案例展开,包括建模、数据库设计、案例开发及实现。

为便于考生自学,本书各章后的习题参考答案也一并给出,考虑到篇幅,将参考答案放在中国教育考试网网站上,供考生自由下载,网站地址是://www.eduexam.com.cn。

本书由清华大学计算机系汤荷美主编,参加编写的人员有:何玉洁(第8章、第12章、第13章、第14章、第15章、第16章、第17章),叶文(第1章、第2章、第6章、第7章、第9章),林友芳(第10章、第11章、第18章、第19章、第20章),汤荷美(第3章、第4章、第5章)。清华大学计算机系周立柱教授对教程进行了审阅。

限于水平及编写时间仓促,书中难免有欠妥之处,望读者提出宝贵意见,以便修订时更正。

作者

2007年6月



第1章 引论	(1)	2.3.2 业务规则需求分析	(21)
1.1 基本概念	(1)	2.3.3 性能需求分析	(22)
1.1.1 信息与数据	(1)	2.3.4 其他需求分析	(22)
1.1.2 数据库系统	(2)	2.4 系统设计	(23)
1.2 数据模型	(3)	2.4.1 概念设计	(23)
1.2.1 数据模型概念	(3)	2.4.2 逻辑设计	(24)
1.2.2 数据模型分类	(4)	2.4.3 物理设计	(25)
1.3 数据视图与模式结构	(6)	2.5 实现与部署	(26)
1.3.1 数据视图与数据抽象	(6)	2.6 运行管理与维护	(27)
1.3.2 三级模式结构	(6)	2.6.1 日常维护	(28)
1.3.3 二级映像与数据独立性	(8)	2.6.2 系统性能监控和分析	(28)
1.4 数据库系统体系结构	(8)	2.6.3 系统性能优化调整	(29)
1.5 数据库管理系统	(9)	2.6.4 系统升级	(30)
1.5.1 数据库管理系统的功能	(10)	2.7 应用案例	(30)
1.5.2 数据库系统的全局结构	(10)	2.7.1 需求描述	(30)
1.6 数据库技术的发展与应用	(11)	2.7.2 数据对象	(31)
小结	(12)	小结	(34)
习题	(13)	习题	(35)
第2章 数据库应用系统生命周期	(14)	第3章 需求分析及功能建模方法	(36)
2.1 数据库应用系统生命周期	(14)	3.1 需求分析概述	(36)
2.1.1 软件工程与软件开发方法	(14)	3.1.1 需求分析概念	(36)
2.1.2 DBAS 软件组成	(15)	3.1.2 系统分析员的职能	(37)
2.1.3 DBAS 生命周期模型	(16)	3.1.3 需求获取的方法	(38)
2.2 规划与分析	(18)	3.1.4 需求分析过程	(39)
2.2.1 系统规划与定义	(18)	3.2 DFD 建模方法	(42)
2.2.2 可行性分析	(19)	3.2.1 DFD 方法的基本对象	(42)
2.2.3 项目规划	(20)	3.2.2 开发 DFD 图	(43)
2.3 需求分析	(20)	3.2.3 建模案例	(43)
2.3.1 数据与数据处理需求分析	(21)	3.2.4 数据字典	(45)

3.3 IDEF0 建模方法	(46)	6.1.2 文件组织	(96)
3.3.1 概述	(46)	6.2 文件结构与存取	(97)
3.3.2 IDEF0 方法	(46)	6.2.1 堆文件	(97)
3.3.3 建模过程及步骤	(49)	6.2.2 顺序文件	(98)
3.4 DFD 与 IDEF0 的比较	(52)	6.2.3 聚集文件	(99)
小结	(52)	6.2.4 索引文件	(100)
习题	(53)	6.2.5 散列文件	(100)
第4章 数据库概念设计及数据建模	(54)	6.3 索引技术	(100)
4.1 数据库概念设计概述	(54)	6.3.1 基本概念	(100)
4.1.1 数据库概念设计的任务	(54)	6.3.2 有序索引的分类及特点	(102)
4.1.2 概念设计过程	(55)	6.4 散列技术	(105)
4.2 数据建模方法	(55)	6.4.1 散列文件	(105)
4.3 ER 建模方法	(56)	6.4.2 散列函数	(107)
4.3.1 基本概念	(56)	6.4.3 桶溢出	(107)
4.3.2 ER 方法语法	(57)	6.5 数据字典	(108)
4.3.3 ER 方法建模实例	(58)	6.6 数据库物理设计	(108)
4.4 IDEF1X 建模方法	(60)	6.6.1 设计步骤和内容	(108)
4.4.1 IDEF1X 概述	(60)	6.6.2 数据库逻辑模式调整	(109)
4.4.2 IDEF1X 模型元素	(61)	6.6.3 DB 文件组织与存取设计	(111)
4.4.3 建模过程	(64)	6.6.4 数据分布设计	(113)
小结	(69)	6.6.5 安全模式设计	(115)
习题	(69)	6.6.6 确定系统配置	(116)
第5章 关系数据库逻辑设计	(71)	6.6.7 物理模式评估	(117)
5.1 概述	(71)	小结	(117)
5.2 基本概念	(71)	习题	(117)
5.2.1 关系模型	(71)	第7章 数据库应用系统功能设计	(119)
5.2.2 关系数据库	(75)	7.1 软件体系结构与设计过程	(119)
5.2.3 关系的完整性	(75)	7.1.1 软件体系结构	(119)
5.3 关系数据库设计理论	(76)	7.1.2 软件设计过程	(120)
5.3.1 问题的提出	(76)	7.2 DBAS 总体设计	(121)
5.3.2 函数依赖	(77)	7.2.1 系统总体设计	(121)
5.3.3 规范化设计方法	(81)	7.2.2 软件总体设计	(123)
5.4 数据库模式设计	(89)	7.2.3 客户/服务器体系结构	(123)
5.4.1 初始关系模式的设计	(89)	7.3 概要设计	(125)
5.4.2 优化关系模式	(91)	7.3.1 数据库事务概要设计	(125)
5.4.3 数据完整性设计	(93)	7.3.2 应用软件概要设计	(127)
5.4.4 安全模式和外模式的设计	(93)	7.4 详细设计	(127)
小结	(93)	7.4.1 数据库事务详细设计	(128)
习题	(94)	7.4.2 应用软件详细设计	(128)
第6章 存储技术与数据库物理设计	(95)	7.5 人机界面设计	(129)
6.1 文件组织	(95)	小结	(131)
6.1.1 数据库的物理结构	(95)	习题	(131)

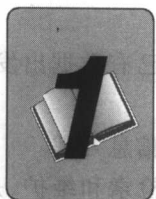
第 8 章 关系数据库操作语言 SQL	(132)	10.2 数据库的运行和维护	(176)
8.1 SQL 支持的数据类型	(132)	10.3 监控分析	(177)
8.1.1 数值型	(132)	10.4 空间管理	(177)
8.1.2 字符串型	(133)	10.5 参数调整	(179)
8.1.3 日期时间类型	(134)	10.6 查询优化	(180)
8.1.4 货币类型	(134)	小结	(182)
8.2 定义和维护关系表	(134)	习题	(182)
8.2.1 关系表的定义与删除	(135)	第 11 章 故障管理	(183)
8.2.2 修改表结构	(137)	11.1 事务	(183)
8.3 数据操作语言	(137)	11.2 故障的种类及解决方法	(184)
8.3.1 数据查询	(138)	11.2.1 事务内部故障	(184)
8.3.2 数据修改	(151)	11.2.2 系统故障	(184)
8.4 索引	(153)	11.2.3 介质故障	(185)
8.5 视图	(154)	11.2.4 计算机病毒故障	(185)
8.5.1 定义视图	(155)	11.3 数据库恢复技术概述	(185)
8.5.2 删除视图	(156)	11.4 数据转储	(186)
8.5.3 视图的作用	(157)	11.5 登记日志文件	(187)
小结	(157)	11.5.1 日志文件的格式和内容	(187)
习题	(157)	11.5.2 日志文件的作用	(188)
第 9 章 事务调度与并发控制	(160)	11.5.3 登记日志文件的原则	(190)
9.1 事务与事务调度	(160)	11.6 具有检查点的恢复技术	(190)
9.1.1 事务的概念	(160)	11.6.1 检查点的作用	(190)
9.1.2 事务的特性	(161)	11.6.2 检查点的引入	(191)
9.1.3 事务调度	(162)	11.6.3 恢复的步骤	(191)
9.1.4 可串行化调度	(164)	11.7 数据库镜像	(192)
9.2 基于锁的并发控制技术	(166)	11.7.1 数据库镜像的引入	(192)
9.2.1 锁的概念	(166)	11.7.2 数据库镜像简介	(192)
9.2.2 加锁协议	(167)	11.7.3 数据库镜像分类	(192)
9.2.3 两阶段锁协议	(167)	11.7.4 工作方式	(193)
9.2.4 锁粒度	(168)	11.8 RAID 的恢复技术	(194)
9.3 死锁处理	(168)	小结	(196)
9.3.1 死锁预防	(169)	习题	(196)
9.3.2 死锁检测与恢复	(170)	第 12 章 SQL Server 2000 数据库管理	
9.4 活锁处理	(170)	系统	(197)
小结	(171)	12.1 SQL Server 2000 概述	(197)
习题	(172)	12.2 SQL Server 2000 的安装	(198)
第 10 章 数据库的实施、运行和维护	(173)	12.2.1 安装前的准备	(198)
10.1 数据库的实施	(173)	12.2.2 安装及安装选项	(199)
10.1.1 定义数据库结构	(173)	12.2.3 测试安装	(203)
10.1.2 数据装载	(173)	12.3 SQL Server 2000 常用工具简介	(203)
10.1.3 编写与调试应用程序	(174)	12.3.1 企业管理器	(203)
10.1.4 数据库的试运行	(174)	12.3.2 查询分析器	(204)

12.4 创建与管理数据库	(205)	14.4 管理数据库用户	(240)
12.4.1 系统数据库	(206)	14.4.1 建立数据库用户	(241)
12.4.2 SQL Server 数据库的构成	(206)	14.4.2 删除数据库用户	(242)
12.4.3 创建数据库	(207)	14.5 管理权限	(242)
12.4.4 删除数据库	(211)	14.5.1 SQL Server 权限种类	(242)
12.5 Transact - SQL 简介	(212)	14.5.2 权限的管理	(242)
12.5.1 Transact - SQL 语言基础知识	(212)	14.6 角色	(245)
12.5.2 流程控制语句	(213)	14.6.1 固定的服务器角色	(246)
12.6 数据传输	(214)	14.6.2 固定的数据库角色	(247)
12.6.1 DTS 功能概述	(215)	14.6.3 用户自定义的角色	(249)
12.6.2 利用 DTS 向导实现数据传输	(215)	小结	(251)
小结	(219)	习题	(251)
习题	(219)	上机练习	(252)
上机练习	(219)	第 15 章 备份和恢复数据库	(253)
第 13 章 数据库对象	(221)	15.1 备份数据库	(253)
13.1 存储过程	(221)	15.1.1 概述	(253)
13.1.1 存储过程基本概念	(221)	15.1.2 备份设备	(253)
13.1.2 创建和执行存储过程	(222)	15.1.3 创建备份设备	(254)
13.2 用户自定义函数	(225)	15.1.4 备份类型	(255)
13.2.1 基本概念	(225)	15.1.5 备份策略	(256)
13.2.2 创建和调用标量函数	(225)	15.1.6 实现备份	(257)
13.2.3 创建和调用内嵌表值函数	(226)	15.1.7 备份媒体集	(261)
13.2.4 创建和调用多语句表值函数	(227)	15.2 恢复数据库	(262)
13.3 触发器	(229)	15.2.1 恢复前的准备及恢复顺序	(262)
13.3.1 触发器基本概念	(229)	15.2.2 实现恢复	(263)
13.3.2 创建触发器	(229)	小结	(267)
13.4 查看、修改及删除对象	(232)	习题	(267)
13.4.1 查看对象	(232)	上机练习	(268)
13.4.2 修改对象	(232)	第 16 章 VB 开发环境与数据访问接口	(269)
13.4.3 删除对象	(233)	16.1 Visual Basic 6.0 简介	(269)
小结	(234)	16.1.1 Visual Basic 6.0 的特点	(269)
习题	(234)	16.1.2 Visual Basic 6.0 环境需求	(270)
第 14 章 安全管理	(236)	16.2 Visual Basic 程序设计基础	(270)
14.1 安全控制	(236)	16.2.1 Visual Basic 集成开发环境	(270)
14.1.1 安全控制模型	(236)	16.2.2 Visual Basic 程序设计的 基本步骤	(271)
14.1.2 数据库权限的种类及用户的 分类	(237)	16.3 数据绑定控件	(274)
14.2 SQL Server 的安全控制	(237)	16.3.1 数据绑定控件概述	(274)
14.3 管理 SQL Server 登录账户	(238)	16.3.2 DataGrid 控件的主要属性	(274)
14.3.1 系统内置的登录账户	(238)	16.3.3 DataList 和 DataCombo 控件	(275)
14.3.2 建立登录账户	(238)	16.4 数据访问接口	(276)
14.3.3 删除登录账户	(240)	16.4.1 ODBC	(276)

16.4.2	OLE DB	(279)	第 19 章 分布式数据库、对象数据库和并行数据库	(331)
16.4.3	ADO	(280)	19.1 分布式数据库	(331)
16.4.4	JDBC	(280)	19.1.1 分布式数据库系统概述	(331)
小结		(282)	19.1.2 分布式数据库目标与数据分布策略	(331)
习题		(282)	19.1.3 分布式数据库系统的体系结构	(333)
第 17 章 VB 数据库应用编程		(283)	19.1.4 分布式数据库的相关技术	(334)
17.1	ADO 数据控件	(283)	19.2 对象数据库	(335)
17.1.1	ADO 数据控件的主要属性和方法	(284)	19.2.1 面向对象数据库的基本概念	(335)
17.1.2	RecordSet 对象的主要属性和方法	(287)	19.2.2 面向对象技术与数据库技术	(336)
17.2	ADO 数据控件使用示例	(289)	19.2.3 对象-关系数据库	(338)
17.2.1	示例 1	(289)	19.3 并行数据库	(339)
17.2.2	示例 2	(290)	19.3.1 并行数据库概述	(339)
17.3	ADO 对象	(296)	19.3.2 并行数据库系统结构	(340)
17.3.1	Connection 对象	(297)	19.3.3 数据划分与并行算法	(341)
17.3.2	Command 对象	(298)	小结	(343)
17.3.3	RecordSet 对象	(299)	习题	(344)
17.3.4	Field 对象	(303)	第 20 章 数据仓库与数据挖掘	(345)
17.3.5	Parameter 对象	(304)	20.1 决策支持系统的发展	(345)
17.3.6	Error 对象	(305)	20.1.1 决策支持系统及其演化	(346)
17.3.7	使用 ADO 对象模型访问数据库	(306)	20.1.2 基于数据仓库的决策支持系统	(347)
小结		(312)	20.2 数据仓库技术概述	(348)
习题		(312)	20.2.1 数据仓库的概念与特性	(348)
上机练习		(313)	20.2.2 数据仓库的体系结构与环境	(350)
第 18 章 统一建模语言 UML		(315)	20.2.3 数据仓库的数据组织	(350)
18.1	UML 简介	(315)	20.2.4 元数据	(352)
18.2	UML 静态建模机制	(316)	20.2.5 操作型数据存储(ODS)	(352)
18.2.1	用例模型和用例图	(316)	20.3 设计与建造数据仓库	(354)
18.2.2	类图和对象图	(318)	20.3.1 数据仓库设计的需求与方法	(354)
18.2.3	包和包图	(323)	20.3.2 数据仓库的数据模型	(355)
18.2.4	组件和组件图	(324)	20.3.3 数据仓库设计步骤	(356)
18.2.5	部署图	(324)	20.4 数据仓库的运行与维护	(356)
18.3	UML 动态建模机制	(325)	20.4.1 数据仓库数据的更新维护	(357)
18.3.1	顺序图	(325)	20.4.2 数据仓库监控与元数据管理	(358)
18.3.2	协作图	(326)	20.5 联机分析处理与多维数据模型	(359)
18.3.3	状态图	(327)	20.5.1 OLAP 简介	(359)
18.3.4	活动图	(328)	20.5.2 多维分析的基本概念	(359)
小结		(328)	20.5.3 多维分析的基本活动	(359)
习题		(329)	20.5.4 OLAP 的实现方式	(360)

20.6 数据挖掘技术	(361)	小结	(364)
20.6.1 数据挖掘步骤	(361)	习题	(364)
20.6.2 关联规则挖掘	(362)	附录 全国计算机等级考试四级数据库	
20.6.3 分类挖掘	(363)	工程师考试大纲(2007年版)	(366)
20.6.4 聚类挖掘	(363)	参考文献	(368)
20.6.5 时间序列分析	(364)		

第



章

引 论

现代社会是信息社会。随着计算机、通信、网络等现代信息技术的快速发展,对信息的处理和应用已深入到社会各行各业,信息系统为信息的处理与应用提供了技术基础。数据库技术产生于20世纪60年代,是信息系统的核心技术和重要基础,也是计算机科学与技术的一个重要分支。数据库技术的飞速发展极大地促进了基于计算机的数据管理技术在各领域的广泛应用。

数据库技术是计算机科学与技术学科知识领域的一个重要组成部分。2006年9月出版的“高等学校计算机科学与技术专业发展战略暨专业规范”将计算机科学与技术学科划分为计算机科学(CS)、计算机工程(CE)、软件工程(SE)和信息技术(IT)四个专业方向,每个专业方向的知识体系由知识领域、知识单元和知识点三个层次组成。在这四个专业方向上,与数据库相关的核心课程分别为:CS和CE专业中的“数据库系统原理”,分别涵盖了信息管理知识领域和数据库系统知识领域;SE专业中的“数据库”,涵盖了计算基础和软件建模与分析两个知识领域中部分知识单元;IT专业中的“数据库与信息管理”,涵盖了信息管理知识领域。

在系统介绍数据库技术之前,本章首先阐述信息、数据等基本概念和数据库系统基础知识,介绍数据管理技术的发展情况,为后续各章奠定理论基础。

1.1

基本概念

1.1.1 信息与数据

信息、物质和能量是组成客观世界并促进社会发展的三大基本要素。

信息依附于物质和能量,但又不同于物质和能量。没有信息就不能更好地利用物质和能量,人类利用信息和知识改造物质、创造新物质、提高能量利用效率、发现新的能量形式。

信息(Information)是现实世界事物的存在方式和运动状态的反映,是对事物之间相互联系、相互作用的描述。信息具有可感知、可存储、可加工、可传递和可再生的自然属性。信息又是社会各行各业不可缺少的资源,这反映了信息的社会属性。

数据(Data)是描述现实世界事物的符号记录,是用物理符号记录下来的可以识别的信息。不同的物理符号体现出数据的不同表现形式,如数字、文字、图形、图像、声音等。数据的各种表

现形式可以经过数字化后存入计算机进行进一步处理。

信息与数据间存在固有联系。数据是信息的符号表示,或称为载体。信息则是数据的语义解释,是数据的内涵。信息以数据的形式表现出来,并为人们所理解接受。

数据处理(Data Processing)是指对数据进行分类、收集、组织、存储,进而从已有数据出发,抽取或推导出新的数据,这些新数据表示了新的信息。

信息与数据经常被不加严格区分地等同使用,因此数据处理也可以说成是信息处理。

数据管理(Data Management)是指对数据的分类、收集、组织、编码、存储、检索和维护,是数据处理业务的重要环节。因此,数据处理与数据管理紧密相联,数据管理技术的优劣将直接影响数据处理的效率。

数据处理与数据管理的区别在于数据管理侧重于如何对数据进行组织和存储,并根据用户需要提供对数据访问的支持。而数据处理除了具有数据管理功能外,还可以对通过数据管理得到的数据进行进一步深加工,从中获取新的数据和信息。

1.1.2 数据库系统

数据库(DB, DataBase)是长期存储在计算机内有组织的、大量的、共享的数据集合。

数据库中的数据按照一定的数据规范进行组织、描述和储存,可以供多个用户共享,数据具有合理的冗余度,数据间联系紧密同时又具有较高的数据独立性。

对于应用领域的各类数据,为了在数据库中科学合理地组织和存储这些数据,进而高效地访问和维护数据库中的大量数据,需要利用数据库管理系统这一计算机系统中重要的数据管理系统软件。

数据库管理系统(DBMS, DataBase Management System)是计算机系统中位于用户与操作系统之间的数据管理系统软件,是数据库系统的核心。

数据库系统(DBS, DataBase System)是指在计算机系统中引入数据库后的软硬件系统构成,如图1.1所示。DBS分为硬件、软件、用户三个层次:

(1) 计算机硬件平台。数据库系统存储的数据量很大,要求计算机硬件平台具有较快的CPU处理速度、足够大的内存和外存、较高的系统通信能力。

(2) 系统软件和应用软件,包括数据库管理系统DBMS、支持DBMS运行的操作系统、数据库应用开发工具和数据库应用程序。

(3) 用户,包括建立与维护数据库系统的数据库管理员(DBA, DataBase Administrator)、负责数据库系统设计与实现的系统分析员、数据库设计人员、应用程序员、使用数据库的最终用户(End User)等。

在不引起混淆或歧义的情况下,数据库系统经常被简称为数据库。

在许多实际数据库应用领域中,不仅需要大量应用数据进行组织、存储和访问,还需要对

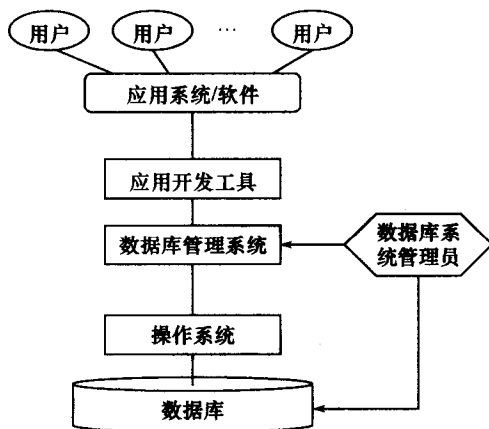


图 1.1 数据库系统组成

数据进行进一步的深加工处理,并且这种处理密切依赖于具体应用领域业务逻辑。由于数据库管理系统只是一种数据管理软件,它虽然提供了查询、插入、删除、更新等通用数据访问操作,但却无法提供各种面向具体应用领域的业务处理功能。

因此,在既需要数据管理又需要数据处理的应用场合,可以利用数据库管理系统对应用数据进行管理,利用应用软件(或应用系统)实现具体的应用业务处理流程。

图 1.1 给出的是一个广义的、由软硬件组成的数据库系统,它不仅通过 DBMS 软件具备了数据管理功能,还借助应用系统/软件具备了数据处理功能。

考虑到数据库这门学科主要从软件角度,研究各种通用的、与具体应用领域无关的数据管理技术,而不同数据库应用领域中数据处理需求又千差万别。下面给出一个狭义的数据库系统定义和建立在此概念基础上的数据库应用系统的定义。

数据库系统(狭义)是由数据库和数据库管理系统组成的软件系统。

数据库应用系统(DBAS, DataBase Application System)是由数据库、数据库管理系统、数据库应用程序组成的软件系统。

数据库技术是研究数据库的结构、存储、设计、管理和使用的一门计算机应用学科。

狭义上的数据库系统主要为用户提供数据存储和查询、插入、修改、删除、更新等数据管理功能。数据库应用系统则面向具体应用领域,提供了更为复杂的数据处理功能。

数据库技术与计算机科学与技术学科中许多其他研究内容都有密切联系。首先,数据库技术是在操作系统的文件系统基础上发展起来的,DBMS 需要操作系统的支持才能正常工作。在数据库系统中,数据库是以文件的形式存储在外设存储设备(如磁盘)上的。其次,数据库与数据结构间的关系也很密切。数据库技术不仅要用到数据结构中的链表、树、图等知识,数据库所采用的各类数据模型(如关系数据模型、面向对象数据模型)本身就属于复杂数据结构。再有,目前使用最多的主流数据库系统为关系数据库系统,其理论基础是关系数据模型,而关系数据模型是在离散数学集合论中“关系”这一基本概念基础上发展起来的。此外,当用户访问数据库时,数据库管理系统需要对用户提交的各类查询请求做查询处理和优化。这种查询处理和优化的原理和过程类似于计算机编译系统对用高级程序设计语言编写的程序进行的编译处理过程。最后,一些大型数据库系统或数据库应用系统属于复杂软件系统,开发这类软件系统既要遵循数据库系统的原理和方法,也要按照软件工程所定义的软件生命周期模型,采用系统化、工程化开发模式,以保证软件系统的开发质量和开发效率。

1.2 数据模型

1.2.1 数据模型概念

数据库系统建立在数据模型的基础上,数据模型是对现实世界数据特征的抽象。

一个具体数据库是面向某个应用领域的应用数据的集合,它反映了应用数据本身的内容和数据之间的联系。由于计算机无法直接处理现实世界中的具体事物,所以人们必须事先将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。

数据模型(Data Model)是数据库系统的形式框架,是用来描述数据的一组概念和定义,包括描述数据、数据联系、数据操作、数据语义以及数据一致性的概念工具。

数据模型是数据库系统的核心和基础。数据模型应满足三方面要求:能够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。目前还没有一种数据模型能够很好地同时满足这三方面的要求。在数据库系统中,可以针对不同的应用领域特征和使用对象,采用不同的数据模型。

数据模型定义了数据库系统中数据组织、存储和管理必须遵循的规范。这种规范精确地描述了系统的静态特征、动态特征和完整性约束条件。因此,数据模型通常可以看作由数据结构、数据操作和完整性约束3个要素组成。

1. 数据结构

数据结构用于描述系统的静态特征。它从语法角度表述了客观世界中数据对象本身的结构和数据对象间的关联关系,是刻画一个数据模型性质最重要的方面。

在数据库系统中,通常按照数据模型中数据结构的类型来区分、命名各种不同的数据模型。例如层次结构、网状结构、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征,是一组对数据库中各种数据对象允许执行的操作和操作规则组成的集合。

数据操作可以是检索、插入、删除和更新等。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)以及实现操作的数据库语言。

3. 数据完整性约束

数据完整性约束是一组完整性规则的集合。它定义了数据模型必须遵守的语义约束,也规定了根据数据模型所构建的数据库中数据内部及其数据相互间联系所必须满足的语义约束。

完整性约束是数据库系统必须遵守的约束,它限定了根据数据模型所构建的数据库的状态以及状态的变化,以便维护数据库中数据的正确性、有效性和相容性。

1.2.2 数据模型分类

模型是对现实世界的抽象。在数据库技术中,人们可以用数据模型这一概念来描述数据库的结构和语义,通过现实世界—信息世界—机器世界的抽象转换过程构建数据库系统,并根据数据模型所定义的规范去管理和使用数据库中的应用数据。

如图1.2所示,从数据建模角度,人们可以首先将现实世界中的客观数据对象抽象组织为信息世界中的某一种信息结构(这种信息结构并不依赖于具体的计算机系统或数据库管理系统DBMS,是一种概念数据模型),然后再将信息结构转换为机器世界中某一具体DBMS支持的数据模型,并存储在计算机系统中。

目前已经出现了多种数据模型,按照这些数据模型在数据建模和数据管理中的不同作用,分为概念数据模型、数据结构模型和物理数据模型三类。

概念数据模型(Conceptual Data Models),简称概念模型,是按用户的观点对数据和信息进行建模,是现实世界到信息世界的第一层抽象,强调其语义表达功能,易于用户理解,是用户与数据