

全国计算机等级 考试四级教程



教育部考试中心

——数据库原理（2018年版）

高等教育出版社



全国计算机等级考试四级教程

——数据库原理 (2018年版)

Quanguo Jisuanji Dengji Kaoshi Siji Jiaocheng
——Shujuku Yuanli

教育部考试中心

主编 杨冬青

参编 王文杰

高等教育出版社·北京

内容提要

本书根据教育部考试中心制订的《全国计算机等级考试四级数据库原理考试大纲(2018年版)》编写而成。主要内容包括数据库原理概述,数据模型和数据库系统的模式结构,关系数据模型和关系数据库系统,关系数据库标准语言 SQL,SQL与数据库程序设计,关系数据库的规范化理论与数据库设计,数据库系统实现技术,分布式、对象-关系、NOSQL数据库,数据库应用及安全性等。本书的编写目标是使通过四级数据库原理考试的考生能够掌握数据库系统基本概念和主要特征,了解各种主要数据模型,尤其是要深入理解关系数据模型,掌握关系数据语言,深入理解关系数据理论,掌握数据库设计方法,具有数据库设计能力,理解数据库管理的基本概念和数据库系统实现的核心技术,并了解随着数据库技术的发展而形成的新型数据库系统及其应用。

本书可供报考全国计算机等级考试四级数据库原理的考生使用,也可用作普通高等学校计算机专业基础课程教材或参考书。

图书在版编目(CIP)数据

全国计算机等级考试四级教程:2018年版.数据库原理 / 教育部考试中心编.--北京:高等教育出版社,2017.11

ISBN 978-7-04-048864-7

I. ①全… II. ①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. ①TP3

中国版本图书馆 CIP 数据核字(2017)第 272211 号

策划编辑 何新权 责任编辑 何新权 袁畅 封面设计 李小璐 版式设计 马云
责任校对 刁丽丽 责任印制 田甜

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 北京铭传印刷有限公司
开 本 787mm×1092mm 1/16
印 张 15
字 数 360 千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>
版 次 2017 年 11 月第 1 版
印 次 2017 年 11 月第 1 次印刷
定 价 32.00 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 48864-00

积极发展全国计算机等级考试 为培养计算机应用专门人才、促进信息 产业发展作出贡献

(序)

中国科协副主席 中国系统仿真学会理事长
第五届全国计算机等级考试委员会主任委员
赵沁平

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。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,迅速改变着人们的工作、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,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。软件产业是增长快速的朝阳产业,是具有高附加值、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方向发展。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产业。

为了适应知识经济发展的需要,大力促进信息产业的发展,需要在全民中普及计算机的基本知识,培养一批又一批能熟练运用计算机和软件技术的各行各业的应用型人才。

1994 年,国家教委(现教育部)推出了全国计算机等级考试,这是一种专门评价应试人员对计算机软硬件实际掌握能力的考试。它不限制报考人员的学历和年龄,从而为培养各行业计算机应用人才开辟了一条广阔的道路。

1994 年是推出全国计算机等级考试的第一年,当年参加考试的有 1 万余人,2017 年报考人数已达 620 万人。截至 2017 年年底,全国计算机等级考试共开考 50 次,考生人数累计达 7 665 万人,有 2 885 万人获得了各级计算机等级证书。

事实说明,鼓励社会各阶层人士通过各种途径掌握计算机应用技术,并通过等级考试对他们的能力予以科学、公正、权威性的认证,是一种比较好的、有效的计算机应用人才培养途径,符合我国的具体国情。等级考试同时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

形式都给予了充分肯定。

计算机技术日新月异。全国计算机等级考试大纲顺应技术发展和需求的变化,从2010年开始对新版考试大纲进行调研和修订,在考试体系、考试内容、考试形式等方面都做了较大调整,希望等级考试更能反映当前计算机技术的应用实际,使培养计算机应用人才的工作更健康地向前发展。

全国计算机等级考试取得了良好的效果,这有赖于各有关单位专家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付出的大量心血和辛勤劳动,他们为这项工作的开展作出了重要的贡献。我们在此向他们表示衷心的感谢!

我们相信,在21世纪知识经济和加快发展信息产业的形势下,在教育部考试中心的精心组织领导下,在全国各有关专家的大力配合下,全国计算机等级考试一定会以“激励引导成才,科学评价用才,服务社会选材”为目标,服务考生和社会,为我国培养计算机应用专门人才的事业作出更大的贡献。

前 言

全国计算机等级考试四级教程,暨计算机专业基础课程教材——《数据库原理》,是根据教育部考试中心组织和实施的考试及制订的《全国计算机等级考试四级数据库原理考试大纲(2018年版)》编写的。

本教材以《全国计算机等级考试四级教程——数据库原理(2017年版)》为基础,根据新版考试大纲的要求,反映数据库技术的新进展,对教材内容进行了适当的扩充和修改。

本教程主要包括:数据库原理概述、数据模型和数据库系统的模式结构、关系数据模型和关系数据库系统、关系数据库标准语言 SQL 与数据库程序设计、关系数据库的规范化理论与数据库设计、数据库系统实现技术、分布式数据库与对象-关系数据库和 NOSQL 数据库、数据库应用及安全性。

通过学习,要求计算机等级考试四级的合格考生能掌握数据库系统基本概念和主要特征,了解各种主要数据模型,尤其是要深入理解关系数据模型,掌握关系数据语言,深入理解关系数据理论,掌握数据库设计方法,具有数据库设计能力,理解数据库管理的基本概念和数据库系统实现的核心技术,并了解随着数据库技术的发展而形成的新型数据库系统和数据库应用。

本教程由北京大学杨冬青教授主编,参加编写的人员包括:王文杰(第 1、2、3、8、9 章),杨冬青(第 4、5、6、7 章)。中国科学院研究生院罗晓沛教授对教程进行了审阅。

由于编写时间仓促,编者水平有限,疏漏之处在所难免,望读者提出宝贵意见,以便修订时改正。

编 者

目 录

第 1 章 数据库原理概述	1
1.1 数据库技术基本概念	1
1.1.1 信息与数据	1
1.1.2 数据库、数据字典、数据库 管理系统、数据库系统	2
1.1.3 数据库系统中的人员	4
1.2 数据库数据管理方法的特点	5
1.2.1 数据管理技术发展的三个 阶段	5
1.2.2 数据库方法的特点	5
1.3 数据库管理技术的发展历程	8
1.4 数据库的应用	10
1.4.1 多媒体数据库	10
1.4.2 移动数据库	11
1.4.3 空间数据库	11
1.4.4 电子商务	11
1.4.5 ERP 和 CRM	11
1.4.6 数据库技术与信息检索	11
1.5 数据库技术的研究领域	12
1.6 小结	13
习题	13
参考答案	14
第 2 章 数据模型和数据库系统的 模式结构	15
2.1 数据模型和数据模型组成的 要素	15
2.1.1 数据模型的概念	15
2.1.2 数据模型组成的要素	16
2.2 数据模型的分类	16
2.3 概念数据模型——E-R 模型	17
2.3.1 概念数据模型的基本概念	17
2.3.2 概念模型的一种表示方法 ——E-R 图	20
2.4 常用的逻辑数据模型	22
2.4.1 层次模型和网状模型	22
2.4.2 关系模型	23
2.4.3 面向对象模型	23
2.4.4 对象-关系数据模型	24
2.5 数据库系统的模式结构	25
2.5.1 数据库系统中模式、实例 和数据库状态	25
2.5.2 数据库系统的三级模式 结构	28
2.5.3 数据库的两层映像与数据 独立性	29
2.5.4 DBMS 语言	30
2.6 小结	31
习题	32
参考答案	32
第 3 章 关系数据模型和关系数据 库系统	33
3.1 关系数据库系统概述	33
3.1.1 关系数据库系统的发展 历史	33
3.1.2 关系数据模型	34
3.2 关系模型的数据结构	36
3.2.1 关系模型的数据结构和基本 术语	36
3.2.2 关系的形式定义和关系数据 库对关系的限定	38
3.2.3 关系数据库中常用的 表示法	41

3.3 关系模型的完整性约束	42
3.3.1 关系模型完整性约束的分类	42
3.3.2 实体完整性约束	42
3.3.3 参照完整性约束	42
3.3.4 用户定义完整性约束	44
3.3.5 关系模型完整性约束的检查	45
3.4 关系操作语言——关系代数	46
3.4.1 关系代数操作的分类	46
3.4.2 基于传统集合论的关系操作	48
3.4.3 一元的专门关系操作	49
3.4.4 二元的专门关系操作	52
3.4.5 扩展的关系操作	54
3.5 小结	56
习题	57
参考答案	59

第4章 关系数据库标准语言

SQL	60
4.1 SQL 概述	60
4.1.1 结构化查询语言 SQL	60
4.1.2 SQL 的语言特点	61
4.1.3 SQL 的数据类型	62
4.1.4 SQL 对关系数据库模式的支持	63
4.1.5 SQL 语言的组成和语句类型	64
4.2 SQL 的数据定义	65
4.2.1 模式的定义和删除	65
4.2.2 基本表的定义、删除和修改	66
4.2.3 索引的建立和删除	68
4.2.4 域的建立和删除	69
4.3 SQL 的数据查询	70
4.3.1 简单查询	70
4.3.2 连接查询	73

4.3.3 嵌套查询	74
4.3.4 集合查询	76
4.3.5 SQL 中的连接表和外连接	76
4.4 SQL 的数据修改	78
4.4.1 插入数据	78
4.4.2 删除数据	78
4.4.3 更新数据	79
4.5 SQL 的视图	80
4.5.1 视图的概念和定义	80
4.5.2 视图的查询	82
4.5.3 视图的修改	83
4.5.4 视图的作用	84
4.6 SQL 的数据控制	85
4.6.1 权限授予	85
4.6.2 权限收回	87
4.7 小结	88
习题	88
参考答案	90

第5章 SQL 与数据库程序设计

91	
5.1 数据库程序设计概述	91
5.2 ODBC 和 IDBC	92
5.2.1 开放数据库互连 ODBC	92
5.2.2 Java 数据库连接 JDBC	94
5.3 SQL 的存储过程与函数	97
5.3.1 SQL 存储过程与函数概述	97
5.3.2 SQL/PSM	97
5.3.3 创建、执行和删除存储过程和函数	98
5.3.4 存储过程和函数示例	99
5.4 SQL 的触发器	100
5.4.1 触发器概述	100
5.4.2 创建触发器	101
5.4.3 激活触发器	101
5.4.4 删除触发器	102
5.4.5 触发器示例	102
5.5 嵌入式 SQL	103

5.5.1 嵌入式 SQL 概述	103	6.7.2 概念结构设计	125
5.5.2 使用嵌入式 SQL 时应解决的三个问题	103	6.7.3 逻辑结构设计	126
5.5.3 嵌入式 SQL 程序的组成	104	6.7.4 物理结构设计	128
5.5.4 在嵌入式 SQL 中使用游标检索多个元组	105	6.8 规范化理论在数据库设计中的应用	128
5.6 动态 SQL	107	6.9 小结	129
5.6.1 动态 SQL 的概念和作用	107	习题	130
5.6.2 动态 SQL 的语句类型和执行方式	107	参考答案	131
5.7 小结	109	第 7 章 数据库系统实现技术	132
习题	109	7.1 数据库管理系统概述	132
参考答案	111	7.1.1 数据库管理系统的基本功能	133
第 6 章 关系数据库的规范化理论与数据库设计	112	7.1.2 数据库管理系统的主要成分和工作流程	133
6.1 “不好”的关系模式中存在的问题	113	7.2 存储管理	134
6.2 函数依赖	113	7.2.1 物理存储介质简介	134
6.2.1 函数依赖的定义	113	7.2.2 磁盘存储器	136
6.2.2 函数依赖的逻辑蕴涵	114	7.2.3 加速对辅助存储器的访问	137
6.2.3 码	115	7.2.4 数据存储组织	138
6.2.4 函数依赖的公理系统	115	7.2.5 缓冲区管理	139
6.3 1NF、2NF、3NF 和 BCNF	116	7.2.6 数据字典	139
6.3.1 原子域和第一范式 (1NF) 及进一步规范	116	7.2.7 索引结构	140
6.3.2 第二范式 (2NF)	117	7.3 查询处理	145
6.3.3 第三范式 (3NF)	117	7.3.1 查询处理概述	145
6.3.4 Boyce-Codd 范式 (BCNF)	118	7.3.2 查询执行	147
6.4 多值依赖和 4NF	119	7.3.3 查询优化	149
6.4.1 多值依赖	119	7.4 事务管理	152
6.4.2 第四范式 (4NF)	120	7.4.1 事务的概念和特性	152
6.5 关系模式的分解	121	7.4.2 故障恢复	154
6.5.1 模式分解的等价标准	121	7.4.3 并发控制	156
6.5.2 关于模式分解的几个事实	124	7.5 小结	163
6.6 更多的范式	124	习题	163
6.7 数据库设计	125	参考答案	165
6.7.1 设计过程概览	125	第 8 章 分布式、对象-关系、NOSQL 数据库	166
		8.1 分布式数据库系统	166

8.1.1 分布式数据库系统的主要特点	167	9.1.4 联机分析处理	199
8.1.2 分布式数据库的设计技术	168	9.1.5 数据仓库与视图	202
8.1.3 分布式数据库中的查询处理	170	9.2 数据挖掘	202
8.1.4 分布式数据库系统的并发控制	170	9.2.1 数据挖掘主要内容	202
8.1.5 分布式恢复	172	9.2.2 关联分析	204
8.1.6 客户机/服务器数据库体系结构简介	172	9.2.3 分类和聚类	205
8.2 对象及对象-关系数据库	176	9.3 数据库的安全性	207
8.2.1 面向对象数据库基本概念	177	9.3.1 数据库安全性问题概述	207
8.2.2 对象数据库标准、语言和设计	180	9.3.2 基于授予和收回权限的自主访问控制	209
8.2.3 对象-关系数据库对 SQL 的扩展	182	9.3.3 多级安全性的强制访问控制	211
8.3 NOSQL 数据库简介	183	9.3.4 基于角色的访问控制	211
8.3.1 NOSQL 系统介绍	183	9.3.5 SQL 注入 (Injection)	212
8.3.2 CAP 原理	186	9.3.6 统计数据库的安全性	213
8.3.3 NOSQL 系统分类	188	9.3.7 加密	214
8.4 小结	190	9.3.8 可信计算机系统评估标准	215
习题	191	9.4 小结	216
参考答案	192	习题	217
第 9 章 数据库应用及安全性	193	参考答案	218
9.1 数据仓库	193	附录 1 全国计算机等级考试四级数据库原理考试大纲 (2018 年版)	219
9.1.1 数据仓库基本概念	193	附录 2 全国计算机等级考试四级数据库原理样题及参考答案	221
9.1.2 数据仓库的数据模型	196	参考文献	229
9.1.3 数据库仓库体系结构	198		

因此,构成一定含义的一组数据称为信息,信息是数据的内涵,信息通过数据进行描述,又是数据的语义解释。

尽管信息与数据两者在概念上不尽相同,但在某些不需要严格分辨的场合,也可以把两者不加区分地使用,例如信息处理也可以说成数据处理。

4. 数据处理

表示原始信息的数据,称为源数据。对这些源数据进行汇集、存储、综合、推导,从这些原始、杂乱、难以理解的数据中抽取或推导出新的数据,这些新的数据称为结果数据,它们表示了新的信息,是有价值、有意义的。提供给某些特定的人使用,可以作为某种决策的依据或用于新的推导。这一过程通常称为数据处理或信息处理。

众所周知,信息是有价值的,信息的价值与它的准确性、及时性、完整性和可靠性有关。因为信息的价值必须通过使用信息的决策者的行为结果来体现,所以,为了提高信息的价值,就要用科学的方法来管理用于表示信息的数据,这种科学的方法就是数据库技术。

1.1.2 数据库、数据字典、数据库管理系统、数据库系统

数据库、数据字典、数据库管理系统和数据库系统是数据库技术密切相关的基本概念。

1. 数据库(Database, DB)

数据库是按一定结构组织并可以长期存储在计算机内的、在逻辑上保持一致的、可共享的大量相关联数据的集合,是存放数据的仓库。也就是说:数据库中的数据按一定的数据模型组织、描述和储存,具有较小的冗余度、较高的数据独立性和易扩展性,并可为在一定组织范围内的各种用户所共享。

数据库的规模可以是任意的,而且它的复杂程度也是有高有低的。例如,只包含名字和地址列表的数据库,可能只包含几百条记录,每条记录的结构也很简单。然而,一个大型图书馆的计算机管理可能会有几十万条按不同类别组织的记录。再有像 Facebook 网络和亚马逊网站的大型数据库,更加复杂,规模更大,可以有几亿甚至更多用户,再加上用户个人信息、商品信息等,占用的存储空间达到几十 TB(Terabyte, $1 \text{ TB} = 10^{12} \text{ Byte}$),需要上百人维护数据库,并保证数据库的最新状态。

2. 数据字典(Data Dictionary, DD)

数据库中的数据通常可以分为用户数据和系统数据两个部分。用户数据是用户使用的数据;系统数据也称数据字典,包括对数据库的描述信息、数据库的存储管理信息、数据库的控制信息、用户管理信息和系统事务管理信息等。所以,数据字典也称系统目录或元数据。

3. 数据库管理系统(Database Management System, DBMS)

数据库管理系统是位于用户与操作系统之间的一个定义(Defining)、操纵(Manipulating)、管理(Management)、构建(Constructing)和维护(Maintaining)数据库的系统软件,是数据库和用户之间的一个接口,并为不同用户和应用程序共享(Sharing)数据库提供便利。在 DBMS 的支持下,用户可以方便地定义数据库中的数据;可以对数据库的数据执行查询、插入、删除、更新等基本操作;统一管理和控制数据库的建立、运行和维护,以保证数据库中数据的安全性、完整性、多用户对数据的并发使用、发生故障后的系统恢复、事务支持;数据库的重组织和性能监视、分析;等等。

有关数据库管理系统的进一步阐述见第7章。

4. 数据库系统(Database System, DBS)

数据库系统是指在计算机系统中引入数据库后的系统,如图 1.1 所示,一般由用户数据库、操作系统、数据库管理系统、应用开发工具、应用系统、数据库管理员和数据库用户构成。应当指出的是,数据库的建立、使用和维护等工作只靠一个 DBMS 远远不够,还要有专门的人员来完成,这些人被称为数据库管理员(Database Administrator, DBA),详见 1.1.3 小节。

在一般不引起混淆的情况下常常把数据库系统简称为数据库。实际上,可以将数据库描述为更大范畴的、大型企业内部的信息系统的一部分。一个公司内部的信息技术(Information Technology, IT)部门设计和维护一个由各种计算机、存储系统、应用软件以及数据库组成的信息系统。

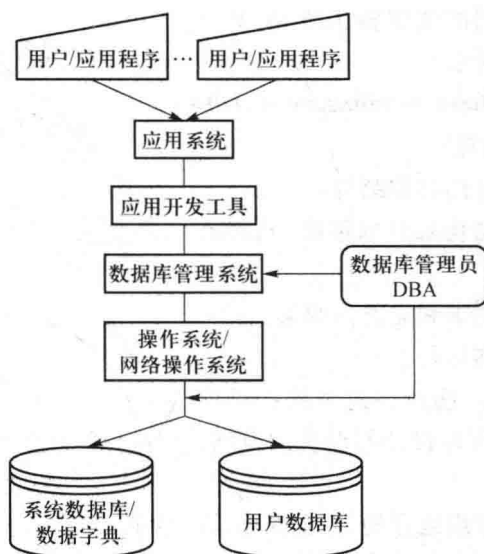


图 1.1 一个简化的数据库系统示意图

(1) 数据库系统中的硬件平台

包括计算机和网络设备。计算机是硬件的基础平台,常用的有微型机、小型机、中型机、大型机和巨型机;网络设备也已成为目前数据库系统的硬件平台,其结构形式主要有单机集中方式、客户机/服务器方式、浏览器/服务器方式及分布方式。

由于数据库系统数据量都很大,加之 DBMS 丰富的功能使得自身的规模也很大,因此整个数据库系统对硬件资源提出了较高的要求,这些要求是:

- 有足够大的内存存放操作系统、DBMS 的核心模块、数据缓冲区和应用程序。
- 有足够大的磁盘等直接存取设备存放数据库,有足够的进行数据备份的设备。
- 系统有较高的通信能力,以提高数据传送率。

(2) 数据库系统中的软件平台

数据库系统的软件平台主要包括:

- DBMS:为数据库的建立、使用和维护而配置的软件。
- 支持 DBMS 运行的操作系统(OS)或网络操作系统(NOS)。
- 与数据库有接口的高级语言及其编译系统,便于开发应用程序。
- 以 DBMS 为核心的应用开发工具。
- 为特定应用环境开发的数据库应用系统。

(3) 人员

数据库系统中的人员包括:数据库管理员、系统分析员和数据库设计人员、应用程序员和最终用户。不同的人员涉及不同的数据抽象级别,具有不同的数据视图。详见 1.1.3 小节。

1.1.3 数据库系统中的人员

数据库系统中的人员包括数据库管理员、系统分析员和数据库设计人员、应用程序员和最终用户。不同的人员涉及不同的数据抽象级别,具有不同的数据视图,如图 1.2 所示。

1. 数据库管理员 (Database Administrator, DBA)

数据库管理员的职责包括:

- ① 确定数据库中的信息内容和结构。
- ② 确定数据库的存储结构和存取策略(选择索引)。
- ③ 定义数据的安全性要求和完整性约束。
- ④ 监控数据库的使用和运行。
- ⑤ 数据库的性能改进。选择创建及维护哪些索引,这个工作属于物理数据库设计与调优的范畴,而这也是 DBA 的职责之一。

- ⑥ 定期对数据库进行重组或重构,以提高系统的性能。

2. 系统分析员和数据库设计人员

系统分析员的职责包括:

- ① 负责应用系统的需求分析和规范说明。
- ② 确定系统的软硬件配置并参与数据库系统的概要设计。

数据库设计人员具体的职责包括:

- ① 参加用户需求调查和系统分析。
- ② 负责数据库中数据的确、数据库各级模式的设计。

在一般情况下,这两种工作可都由数据库管理员担任。

3. 应用程序员

应用程序员负责设计和编写应用系统的程序模块,并进行调试和安装。

4. 用户

这里用户是指最终用户(End User),他们通过应用系统的用户接口使用数据库。常用的接口方式有浏览器、菜单驱动、表格操作、图形显示、报表书写等。

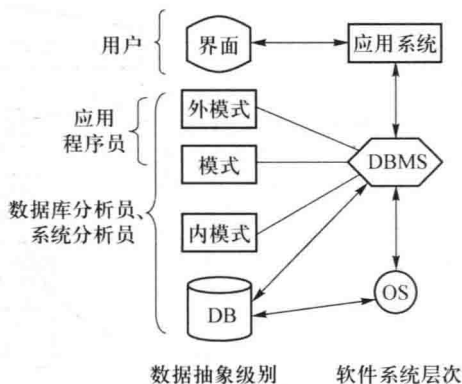


图 1.2 数据库系统中各类人员的数据视图

1.2 数据库数据管理方法的特点

1.2.1 数据管理技术发展的三个阶段

数据管理技术是指对数据的分类、组织、编码、存储、检索和维护的技术。数据管理技术的发展和计算机技术及其应用的发展联系在一起,经历了人工管理、文件系统和数据库系统三个阶段。表 1.1 对数据管理技术发展的三个阶段进行了比较。

表 1.1 数据管理技术发展的三个阶段的比较

		人工管理阶段(20世纪50年代中期以前)	文件系统阶段(20世纪50年代后期到60年代中期)	数据库系统阶段(20世纪60年代后期以来)
背景	应用背景	科学计算	科学计算、数据管理	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘、磁盘阵列
	软件背景	没有操作系统	有文件系统	有数据库管理系统
特点	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理
	数据的管理者	用户(程序员)	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用	现实世界中的某个部门、企业、组织等
	数据的共享程度	无共享、冗余度极大	共享性差、冗余度大	共享性高、冗余度小
	数据的独立性	不独立,完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构、整体无结构	整体结构化,用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

1.2.2 数据库方法的特点

与人工管理方法和文件系统方法相比较,数据库方法具有如下的优势。

1. 数据库系统自描述特点

数据库方法的一个基本特征是数据库系统不仅包括数据库本身,还包括对数据库的结构和约束条件的完整定义或描述。这些定义被存储在 DBMS 系统目录中,该目录中的信息包括:每个文件的结构信息、每个数据项的类型和存储格式,以及加在数据上的各种约束条件。存储在系统目录中的这些信息描述了主数据库的结构。

DBMS 软件和那些需要了解数据库结构信息的数据库用户,都需要使用系统目录。一个通用的 DBMS 软件包并不是为某个特定的数据库应用程序编写的。因此,应用程序或用户必须参考系统目录以了解某个指定数据库的文件结构,例如将要访问的数据的类型和格式等。

在传统的文件处理中,数据的定义一般作为应用程序自身的一部分。因此,这些程序也就局限于只能在某个专门的数据库中工作,而该数据库的结构也就在应用程序中进行声明。因此,文件处理软件只能访问一个指定的数据库,而 DBMS 软件则可以通过从系统目录中提取数据库的定义,然后使用这些定义来达到访问多个数据库的目的。

2. 程序与数据是分离的

在传统的文件处理中,数据文件的结构总是嵌入在访问它的程序之中。因此,文件结构的任何改变都需要同时改变访问这一数据文件的所有程序。与之相比,在大多数情况下,DBMS 访问程序则无须做这样的改变。这是因为存储在 DBMS 数据字典中的数据文件结构与访问它的程序之间是相互分离的。当程序访问数据库数据时,由 DBMS 通过查询数据字典确定被访问的数据。通常把这种性质称为程序-数据独立性(Program-Data Independence)。

在某些类型的数据库系统,如面向对象数据库和对象-关系系统中,用户可以把施加于数据上的操作定义作为数据库定义的一部分。这样,操作(Operation,也称为函数或方法)就可以分为两部分指定:操作的接口(或签名)部分包括该操作的名称和它所需要的变量(Argument)(或参数)的数据类型;操作的实现(或方法)部分可独立地指定,从而也可独立地进行改变,不会影响接口部分。用户的应用程序可以操作数据,通过操作的名称和参数调用这些操作即可,无须理会这些操作是如何实现的。这种特征称为程序-操作独立性(Program-Operation Independence)。

3. 数据结构化

DBMS 提供给用户的是数据的概念表示(Conceptual Representation),不包含数据存储和操作实现的细节。非形式化地说,数据模型(Data Model)是一种用于提供这种概念表示的数据抽象类型。数据库使用数据模型来描述整个组织的数据结构,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,但却隐藏了大多数数据库用户并不关心的数据存储和实现的细节。

4. 由 DBMS 进行统一的数据管理和控制

数据库对用户来说是可以同时存取(并发)、共享的资源,它由 DBMS 进行统一的数据管理和控制。为此,DBMS 还必须提供以下各方面数据控制功能:

(1) 支持数据的多视图

一个 DBMS 通常会有许多用户,而每个用户可能需要不同的数据库视角或视图(View)。一个视图可能是一个数据库的子集,也可能包含从数据库文件导出但未显式存储的虚数据(Virtual Data)。某些用户可能没有必要知道他们引用的数据是存储的还是导出的。因此,一个允许用户执行各种不同应用程序的多用户 DBMS 必须提供定义多视图的能力。

(2) 支持数据共享和多用户事务处理

顾名思义,一个多用户 DBMS 必须能够允许多个用户同时访问数据库。当多个应用程序的数据在一个数据库中集成和维护时,这一特性就显得尤为重要。因此,DBMS 必须包括并发控制(Concurrency Control)软件,以确保若干个用户能以一种受控的方式更新相同的数据,从而保证更新的结果是正确的。这种类型的应用一般称为联机事务处理(On-Line Transaction Processing,

OLTP)应用。事务(Transaction)这个概念已经成为许多数据库应用的核心,多用户 DBMS 的一个基本任务即是保证并发事务可以正确而有效地执行。

(3) 具有可控的冗余

在文件系统和传统方法中多次存储相同数据的冗余(Redundancy)可能会引起一系列问题。比如,对于更新操作,可能会导致重复操作;相同数据被重复存储必将浪费存储空间,这对于大型数据库来说,将会造成非常严重的问题;表示相同数据的文件可能会出现不一致。

在数据库方法中,理想情况下,数据库设计应该将每个逻辑数据项只存储在数据库中的一个地方。这种方式避免了不一致性,并节省了存储空间。但在某些情况下,具有一定的可控冗余对于提高查询的性能也是十分有益的。在这种情况下,要求 DBMS 应该有能力控制这类冗余,以避免文件之间出现不一致。

(4) 提供数据库数据的安全性(Security)保护,限制非授权的访问

当多个用户共享一个大型数据库时,大多数用户都不可能授权访问数据库的所有信息。另外,有些用户可能只允许检索数据,而另外一些用户则既可以检索数据,又可以更新数据。因此,访问操作的类型(检索或更新)也必须加以控制。通常,赋予每个用户或用户组一个通过口令保护的账号,以此可获得合法的数据库访问。DBMS 应该提供一个安全和授权的子系统(Security and Authorization Subsystem),DBA 可以使用这个子系统创建账户,并指定对该账户的限制。然后,DBMS 将强制这些限制自动地执行。

(5) 为程序对象提供持久性存储

保持程序对象和数据结构的持久性存储是数据库系统的一个重要功能,这也是面向对象数据库系统(Object-Oriented Database System)出现的一个主要原因。编程语言一般都有复杂的数据结构,例如 C++中类的定义。一旦程序终止,这些程序变量的值一般会被丢弃。而对于 C++中的一个复杂对象,可以永久地存储在一个面向对象 DBMS 中。这样的一个对象就可以被认为具有持久性,因为它们可以在程序终止后继续存在,并可以被以后的另一个 C++程序直接检索。

(6) 提供高效查询处理的存储结构,支持高效的数据查询处理与优化

数据库系统必须提供高效的执行查询和更新的能力。因为数据库一般是存储在磁盘上的,DBMS 必须提供专门的数据结构来加快磁盘搜索以找到想要的记录。为此要使用称作索引(Index)的辅助文件。为了处理一个特定查询所需的数据库记录,必须将这些记录从磁盘复制到内存中。因此,DBMS 一般都有一个缓冲(Buffering)模块,用来维护主存缓冲区中的数据库部分。DBMS 也可以利用操作系统来完成磁盘数据的缓冲。

DBMS 的查询处理与优化(Query Processing and Optimization)模块基于现有的存储结构,负责为每个查询选择一种有效的查询执行方案。

(7) 提供数据库的备份(Backup)和恢复(Recovery)功能。DBMS 提供备份和恢复子系统,负责从硬件和软件的故障中进行恢复。

(8) 提供定义和确保数据库完整性(Integrity)检查。大多数数据库应用程序都需要根据自己业务规则的需要和数据模型的内在规则,为自己持有的数据建立某些完整性约束。DBMS 应该具有提供定义和确保这些约束检查的能力。

(9) 提供多用户界面。

因为数据库用户有各种类型,而且他们所具有的技术知识层次也各不相同。因此,DBMS 应