

# 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 (2016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

### (2016年版)

Quanguo Jisuanji Dengji Kaoshi Siji Jiaocheng  
——Shujuku Yuanli

教育部考试中心

主编 杨冬青  
参编 邵佩英 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(2013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、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、数据库系统的体系结构和安全性、数据库技术的发展等。本书的编写目标是使通过四级数据库原理的考试合格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,并了解数据库技术的发展。

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:2016年版.数据库原理 / 教育部考试中心编. --北京:高等教育出版社, 2015.12

ISBN 978-7-04-027655-8

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(2015)第 273426 号

策划编辑 何新权  
责任印制 毛斯璐

责任编辑 柳秀丽

封面设计 杨立新

版式设计 马云

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 三河市华骏印务包装有限公司  
开 本 787mm×1092mm 1/16  
印 张 14  
字 数 320 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.landrao.com>  
<http://www.landrao.com.cn>  
版 次 2015 年 12 月第 1 版  
印 次 2015 年 12 月第 1 次印刷  
定 价 30.00 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 27655-00

# 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,2012 年报考人数已达 549 万人。截至 2012 年年底,全国计算机等级考试共开考 36 次,考生人数累计达 4 933 万人,有 1 876 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。

2013年3月

# 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(2013年版)编写的。

本教材以原全国计算机等级考试三级教程——数据库技术(2009版)为基础,根据四级数据库原理考试大纲(2013年版)和计算机专业基础课程数据库原理的教学要求,删掉了计算机基础知识、数据结构与算法和操作系统的相关章节,将教材内容集中于关于数据库系统基本概念、原理和技术的介绍上。

本教程主要内容包括:数据库原理概述、数据模型和数据库系统的模式结构、关系数据库模型和关系数据库系统、关系数据库标准语言 SQL、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、数据库系统的体系结构和安全性以及数据库技术的发展。

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,并了解数据库技术的发展。

本教程由北京大学杨冬青教授主编,参加编写的人员包括:邵佩英(第1~5章),杨冬青(第6、7章),王文杰(第8、9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	2.4.1 层次模型和网状模型 .....	17
1.1 数据库技术基本概念 .....	1	2.4.2 关系模型 .....	18
1.1.1 信息、数据与数据处理 .....	1	2.4.3 面向对象模型 .....	19
1.1.2 数据库、数据字典、数据库 管理系统、数据库系统 .....	2	2.4.4 对象-关系数据模型 .....	20
1.1.3 数据库系统中的人员 .....	4	2.5 数据库系统的模式结构 .....	20
1.2 数据管理技术的产生与发展 .....	4	2.5.1 数据库系统中模式、实例和数据库 状态 .....	21
1.3 数据库方法与文件系统方法 .....	5	2.5.2 数据库系统的三级模式结构 .....	23
1.3.1 数据库方法的优势 .....	5	2.5.3 数据库的两层映像与数据 独立性 .....	24
1.3.2 数据库与文件系统 .....	6	2.6 小结 .....	26
1.4 数据库应用简史 .....	7	习题 .....	26
1.4.1 使用层次和网状数据模型的早期数据 库应用 .....	7	参考答案 .....	27
1.4.2 关系模型数据库提供应用灵活性 .....	7	第 3 章 关系数据模型和关系数据库 系统 .....	28
1.4.3 面向对象数据库的应用 .....	8	3.1 关系数据库系统概述 .....	28
1.4.4 在 Web 上交换数据实现电子商务 .....	8	3.1.1 关系数据库系统的发展历史 .....	28
1.4.5 为新应用扩展数据库功能 .....	8	3.1.2 关系数据模型 .....	29
1.4.6 数据库技术与信息检索 .....	9	3.2 关系模型的数据结构 .....	30
1.5 数据库技术的研究领域 .....	9	3.2.1 关系模型的数据结构和基本 术语 .....	30
1.6 小结 .....	10	3.2.2 关系的形式定义和关系数据库对关 系的限定 .....	33
习题 .....	11	3.2.3 关系数据库中常用的表示法 .....	35
参考答案 .....	11	3.3 关系模型的完整性约束 .....	36
第 2 章 数据模型和数据库系统的模式 结构 .....	12	3.3.1 关系模型完整性约束的分类 .....	36
2.1 数据模型和数据模型组成的 要素 .....	12	3.3.2 实体完整性约束 .....	36
2.1.1 数据模型的概念 .....	12	3.3.3 参照完整性约束 .....	37
2.1.2 数据模型组成的要素 .....	13	3.3.4 用户定义完整性约束 .....	38
2.2 数据模型的分类 .....	13	3.3.5 关系模型完整性约束的检查 .....	39
2.3 概念数据模型——E-R 模型 .....	14	3.4 关系操作语言——关系代数 .....	40
2.3.1 概念数据模型的产生和基本 概念 .....	14	3.4.1 关系代数操作的分类 .....	40
2.3.2 概念模型的一种表示方法—— E-R 图 .....	16	3.4.2 基于传统集合论的关系操作 .....	42
2.4 常用的逻辑数据模型 .....	17	3.4.3 一元的专门关系操作 .....	43
		3.4.4 二元的专门关系操作 .....	46

3.4.5 扩充的关系操作 .....	48	5.2 数据库程序设计 .....	85
3.5 小结 .....	50	5.2.1 数据库程序设计概述 .....	85
习题 .....	51	5.2.2 数据库程序设计方法 .....	85
参考答案 .....	53	5.2.3 数据库程序设计中的交互序列 .....	86
<b>第4章 关系数据库标准语言</b>		5.3 SQL的存储过程与SQL/PSM .....	86
<b>SQL(一)</b> .....	54	5.3.1 SQL存储过程与函数概述 .....	86
4.1 SQL概述 .....	54	5.3.2 SQL/PSM .....	87
4.1.1 结构化查询语言SQL .....	54	5.3.3 创建、执行和删除存储过程和函数 .....	88
4.1.2 SQL的语言特点 .....	55	5.3.4 存储过程和函数示例 .....	89
4.1.3 SQL的数据类型 .....	56	5.4 SQL的触发器 .....	90
4.1.4 SQL对关系数据库模式的支持 .....	57	5.4.1 触发器概述 .....	90
4.1.5 SQL语言的组成和语句类型 .....	58	5.4.2 创建触发器 .....	90
4.2 SQL的数据定义 .....	59	5.4.3 激活触发器 .....	91
4.2.1 模式的定义和删除 .....	60	5.4.4 删除触发器 .....	92
4.2.2 基本表的定义、删除和修改 .....	60	5.4.5 触发器示例 .....	92
4.2.3 索引的建立和删除 .....	63	5.5 嵌入式SQL .....	93
4.2.4 域的建立和删除 .....	64	5.5.1 嵌入式SQL概述 .....	93
4.3 SQL的数据查询 .....	64	5.5.2 使用嵌入式SQL时应解决的三个问题 .....	94
4.3.1 简单查询 .....	65	5.5.3 嵌入式SQL程序的组成 .....	95
4.3.2 连接查询 .....	67	5.5.4 在嵌入式SQL中使用游标检索多个元组 .....	96
4.3.3 嵌套查询 .....	68	5.6 动态SQL .....	97
4.3.4 集合查询 .....	69	5.6.1 动态SQL的概念和作用 .....	97
4.3.5 SQL中的连接表和外连接 .....	70	5.6.2 动态SQL的语句类型和执行方式 .....	98
4.4 SQL的数据修改 .....	71	5.7 SQL的其他功能 .....	99
4.4.1 插入数据 .....	72	5.8 小结 .....	100
4.4.2 删除数据 .....	72	习题 .....	101
4.4.3 更新数据 .....	73	参考答案 .....	102
4.5 SQL的数据控制 .....	73	<b>第6章 关系数据库的规范化理论与数据库设计</b> .....	103
4.5.1 权限授予 .....	73	6.1 “不好”的关系模式中存在的	
4.5.2 权限收回 .....	75	问题 .....	104
4.6 小结 .....	76	6.2 函数依赖 .....	104
习题 .....	76	6.2.1 函数依赖的定义 .....	104
参考答案 .....	78	6.2.2 函数依赖的逻辑蕴涵 .....	105
<b>第5章 关系数据库标准语言</b>		6.2.3 码 .....	106
<b>SQL(二)</b> .....	79	6.2.4 函数依赖的公理系统 .....	106
5.1 SQL的视图 .....	79		
5.1.1 视图的概念和定义 .....	79		
5.1.2 视图的查询 .....	82		
5.1.3 视图的修改 .....	83		
5.1.4 视图的作用 .....	84		



6.3 1NF、2NF、3NF 和 BCNF .....	107	7.5 小结 .....	151
6.3.1 1NF 及进一步规范 .....	107	习题 .....	152
6.3.2 2NF .....	107	参考答案 .....	153
6.3.3 3NF .....	108	<b>第 8 章 数据库系统的体系结构和安</b>	
6.3.4 BCNF .....	108	<b>全性</b> .....	154
6.4 多值依赖和 4NF .....	109	8.1 概述 .....	154
6.4.1 多值依赖 .....	109	8.2 分布式数据库系统体系结构 .....	155
6.4.2 4NF .....	111	8.2.1 分布式数据库系统的主要特点 .....	155
6.5 关系模式的分解 .....	111	8.2.2 分布式数据库的设计技术 .....	156
6.5.1 模式分解的等价标准 .....	112	8.2.3 分布式数据库中的查询处理 .....	157
6.5.2 关于模式分解的几个事实 .....	114	8.2.4 分布式数据库系统的并发控制 .....	158
6.6 数据库设计 .....	115	8.2.5 分布式恢复 .....	159
6.6.1 设计过程概览 .....	115	8.3 客户机/服务器数据库体系	
6.6.2 概念结构设计 .....	115	结构 .....	160
6.6.3 逻辑结构设计 .....	117	8.3.1 客户机/服务器计算模式 .....	160
6.6.4 物理结构设计 .....	118	8.3.2 面向 Web 应用的数据库体系	
6.7 规范化理论在数据库设计中的		结构 .....	160
应用 .....	119	8.3.3 面向 Web 应用的软件开发 .....	161
6.8 小结 .....	120	8.4 安全数据库以及数据库的	
习题 .....	120	安全性 .....	164
参考答案 .....	122	8.4.1 数据库安全性问题概述 .....	165
<b>第 7 章 数据库系统实现技术</b> .....	123	8.4.2 基于授予和收回权限的自主访问	
7.1 数据库管理系统概述 .....	123	控制 .....	167
7.1.1 数据库管理系统的基本功能 .....	124	8.4.3 多级安全性的强制访问控制 .....	168
7.1.2 数据库管理系统的主要成分和		8.4.4 基于角色的访问控制 .....	169
工作流程 .....	124	8.4.5 统计数据库的安全性 .....	170
7.2 存储管理 .....	125	8.4.6 加密 .....	170
7.2.1 物理存储介质简介 .....	125	8.4.7 可信计算机系统评估标准 .....	172
7.2.2 数据存储组织 .....	128	8.5 小结 .....	173
7.2.3 缓冲区管理 .....	129	习题 .....	174
7.2.4 数据字典 .....	129	参考答案 .....	175
7.2.5 索引结构 .....	130	<b>第 9 章 数据库技术的发展</b> .....	176
7.3 查询处理 .....	135	9.1 对象及对象-关系数据库 .....	176
7.3.1 查询处理概述 .....	135	9.1.1 面向对象数据库基本概念 .....	177
7.3.2 查询执行 .....	136	9.1.2 对象数据库标准、语言和设计 .....	179
7.3.3 查询优化 .....	138	9.1.3 对象-关系数据库模型 .....	180
7.4 事务管理 .....	140	9.2 几种应用数据库系统 .....	182
7.4.1 事务的概念和特性 .....	140	9.2.1 主动数据库 .....	182
7.4.2 故障恢复 .....	143	9.2.2 移动数据库 .....	184
7.4.3 并发控制 .....	145	9.2.3 多媒体数据库 .....	185

9.3 数据仓库 .....	187	9.5 小结 .....	197
9.3.1 数据仓库基本概念 .....	187	习题 .....	198
9.3.2 数据仓库的数据模型 .....	190	参考答案 .....	199
9.3.3 数据库仓库体系结构 .....	191	附录 1 全国计算机等级考试四级数据库	
9.3.4 联机分析处理 .....	192	原理考试大纲(2013年版) .....	200
9.4 数据挖掘 .....	193	附录 2 全国计算机等级考试四级数据库	
9.4.1 数据挖掘主要内容 .....	194	原理样卷及参考答案 .....	202
9.4.2 关联分析 .....	195	参考文献 .....	211
9.4.3 分类和聚类 .....	196		



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. 数据处理的基本环节

人们用数据表示原始信息,称为源数据。然后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”可理解为是在计算机存储设备中按一定格式存放数据的仓库。这个名词起源于20世纪中期,当时美军为了指挥作战需要建立了一个高级军事情报基地,把收集到的各种情报存储在计算机中,并称之为“数据库”。后来随着数据库技术的产生,人们引申并沿用了“数据库”这个名词,赋予它更加深层的含义。

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。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。

#### 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。

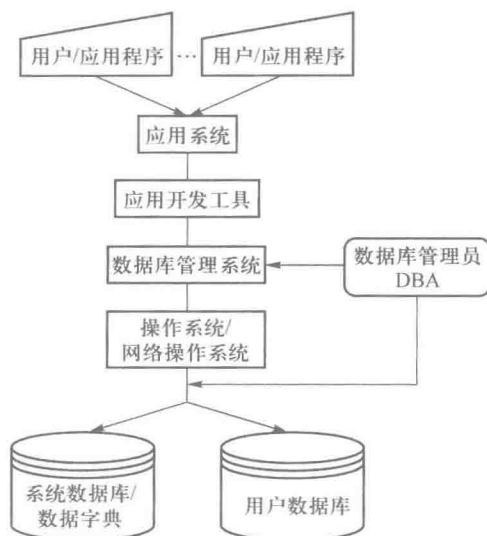


图 1.1 一个简化的数据库系统示意图

### (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)。
- 具有与数据库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定、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。

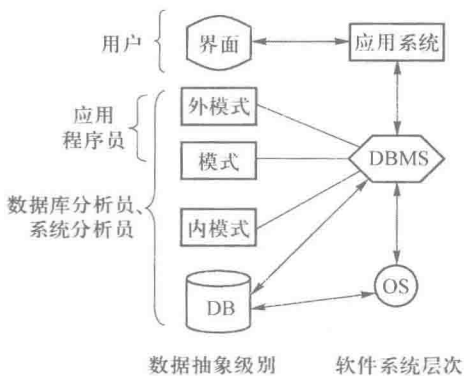


图 1.2 数据库系统中各类人员的数据视图

## 1.2 数据管理技术的产生与发展

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。数据管理技术是应数据管理任务的需求而产生的。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

为了解决多用户、多应用共享数据,使数据为尽可能多的应用服务,一种新的数据管理技术——数据库技术应运而生,并发生了对数据库技术有着奠基作用的三件大事:

① 1968年美国的IBM公司推出了第一个数据库管理系统IMS,它是基于层次模型的数据库管理系统。

② 1969年美国数据库系统语言协会(CODASYL)的数据库任务组(DBTG)发表了网状数据模型的DBTG报告。

③ 1970年美国IBM公司的高级研究员E. F. Codd提出了关系数据模型及其相关概念,奠定了关系数据库的理论基础。

## 1.3 数据库方法与文件系统方法

### 1.3.1 数据库方法的优势

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

数据库系统不仅包括数据库数据本身,还包括定义和说明数据库数据的结构和约束条件。

这些定义和说明称为数据字典或元数据 (Meta Data) 被保存在系统目录中。

## 2. 数据结构化

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。

## 3. 数据共享性高、冗余度小、易扩充

数据库数据不再是面向某一应用,而是面向整个系统,为用户提供更广泛的数据共享,并大大减小数据的冗余度,避免数据之间的不相容性和不一致性,而且使得数据库系统具有弹性大、易扩充的特点。

## 4. 数据独立性高

在数据库方法中,存储在数据库的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,简称数据独立性。

## 5. 由 DBMS 进行统一的数据管理和控制

数据库对系统中的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 提供数据库数据的安全性 (Security) 保护,限制非授权的访问。DBMS 提供安全和权限管理子系统,由 DBA 给不同用户授予不同访问数据库的权限,维护数据库的安全性。

② 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,具有从硬件和软件的故障中进行恢复的能力。

③ 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为自己持有的数据,根据自己业务规则的需要以及根据数据模型的内在规则,建立某些完整性约束 (Integrity Constraints)。DBMS 应该具有提供定义和确保这些约束检查的能力。

④ 提供并发控制功能。当多个用户同时对数据库中数据进行存取或修改时,可能会发生相互冲突和干扰而产生错误结果,并有可能破坏数据库的完整性。因此 DBMS 必须对多个用户的并发操作加以控制和协调。

⑤ 提供高效的数据查询处理与优化 (Query Processing and Optimization)。DBMS 的查询处理与优化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。

⑥ 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 就应该提供各种不同类型的用户界面。这些界面通常包括为偶尔访问的用户提供的查询语言、为应用程序开发人员提供的编程语言界面、为简单参与用户提供窗体和命令代码、为独立用户提供菜单驱动界面和自然语言界面。窗体风格界面和菜单驱动界面被称为图形用户界面 (Graphical User Interface, GUI)。有许多专用的语言和应用环境可以指定 GUI。现在向数据库提供 Web GUI 界面的功能,即由数据库提供 Web 支持,正变得越来越普遍。

### 1.3.2 数据库与文件系统

尽管数据库方法有很多优点,但是在某些情况下,使用数据库系统反而会导致不必要的开销,而使用传统的文件处理方式却不会蒙受这种损失。导致使用 DBMS 而增加额外开销的主要原因有:



- ① 初期对硬件、软件和培训的高额投资；
- ② DBMS 为了定义和处理数据而提供的普遍性导致的额外开销；
- ③ 为了提供安全性、并发控制、恢复和完整性功能而导致的额外开销。

如果数据库设计者和 DBA 没能正确地设计数据库,或是没能正确地实现数据库系统应用,还可能会出现另外的一些问题。因此,在下述情况下使用传统的文件方式处理数据可能会更加合适:

- ① 数据库应用比较简单,易于定义并且一般不会发生变化;
- ② 某些应用程序存在严格的实时性和专用性要求,而通用的 DBMS 对处理这样的应用开销过高,不合适;
- ③ 不需要多个用户访问数据。

## 1.4 数据库应用简史

本节将简单回顾使用数据库的应用的历史,并理解这些应用是促进新型数据库发展的主要动力。关于数据库技术的发展详见第 8、9 章。

### 1.4.1 使用层次和网状数据模型的早期数据库应用

许多早期数据库应用主要维护大型组织中的记录,如公司、大学、医院和银行等。在许多这样的应用中,存在着大量结构相似的记录。例如,在一个大学应用中,将会为每个学生、每门课程和每个成绩记录等保存相似的信息。

早期层次和网状数据模型(见 2.4.1 小节)的数据库系统的一个主要问题在于:数据记录的概念表示与其在磁盘上的物理存储位置混淆不清。例如,某个特定学生的成绩记录可能与该学生的基本信息记录在物理存储位置上相邻。尽管这样可以为原始查询提供非常有效的检索,因为数据库就是为“查询某学生的成绩信息”这样的查询而设计的,但是无法提供足够的灵活性。当出现新的查询时,例如要“查询课程与成绩的信息”就不能有效地检索相关的记录。尤其是当新的查询需要不同的存储组织以便有效地处理时,这种新的查询将难以有效实现。如果应用需求有所改变,要重新组织数据库也是很困难的。

早期层次和网状数据模型数据库系统的另一个缺点是其仅提供编程语言界面。这使得实现新查询和处理不仅费时而且代价不菲,这是因为必须编写、测试和调试新的程序。从 20 世纪 60 年代中期开始,一直持续到 70 年代和 80 年代,大多数此类数据库系统都是基于昂贵的大型机实现的。早期系统的主要类型基于三种主要范型:基于层次模型和网状模型的数据库系统以及倒排(inverted)文件系统。

### 1.4.2 关系模型数据库提供应用灵活性

关系模型(见 2.4.2 小节)是数学化的模型,它将数据的概念表示与其物理存储分离开来。关系数据模型还引入了高级查询语言,可以作为编程语言界面的一种替代,这样就可以快速编写新的查询。数据的关系表示类似于一张二维表。关系系统最初与早期系统面向同样的应用,但是关系系统试图提供一些灵活性,以便快速地进行新查询,并可在需求改变时重新组织数据库。