

吴洪潭 主编 王德林 陆惠娟 叶含笑 刘继来 编著

[illegible]

國防工業出版社

<http://www.ndip.com.cn>

数据库原理

吴洪潭 主编
王德林 陆惠娟 编著
叶含笑 刘继来

国防工业出版社

·北京·

图书在版编目(CIP)数据

数据库原理/吴洪潭主编. —北京:国防工业出版社,
2003.1

ISBN 7-118-03058-9

I. 数... II. 吴... III. 数据库系统 - 理论
IV. TP311.13

中国版本图书馆 CIP 数据核字(2002)第 097145 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

北京奥隆印刷厂印刷

新华书店经售

*

开本 787 × 1092 1/16 印张 14 1/4 322 千字

2003 年 1 月第 1 版 2003 年 1 月北京第 1 次印刷

印数:1—4000 册 定价:20.00 元

(本书如有印装错误,我社负责调换)

前 言

数据库技术是现代计算机应用的基础和核心,经过几十年的发展,数据库技术已经进入了成熟期。在今天,数据库技术的应用非常广泛。计算机的商业应用几乎都与数据库有关,从考生报名、网上检索资料,到企业的信息管理、银行的业务处理等,都是数据库在发挥着关键的作用。因此,数据库技术是信息社会的基石之一。

在高校中,数据库原理早已成为计算机专业的必修课,同时也是其他非计算机专业的选修课程。由于历史的原因,目前大部分数据库教材都偏重于理论,缺乏实用性,存在着计算机专业大学生学了数据库课程,而不会实际应用数据库的现象。本书的最大特色是,除了介绍数据库的一般原理、理论之外,较大篇幅介绍了 Oracle 数据库的入门知识,同时简要介绍了新一代数据库应用快速开发工具 PowerBuilder 和 Delphi。

本书由吴洪潭主编,王德林、陆惠娟、叶含笑、刘继来共同参与编著。吴洪潭编写了第 1、2、3、8 章,王德林编写了第 7 章,陆惠娟编写了第 5、6 章,叶含笑编写了第 4 章,刘继来编写了第 9 章。

本书能够完成,应该感谢浙江省招生办公室的宣成高工(教授级)和丁文硕士提供的指导和帮助,以及杭州泰兴数据工程公司的卫东亮经理和国防工业出版社的江洪湖先生大力的支持。在此谨表感谢。

由于水平有限,本书难免会存在缺点和错误,恳请广大读者批评指正。

编 者

2002 年 9 月

于中国计量学院

内 容 简 介

本书系统地介绍了关系数据库的基本概念、基本原理、基本方法和应用实践。内容包括:数据库概论、关系数据库、SQL 语言、关系数据库设计理论、数据库保护、数据库设计、数据库新技术、Oracle 数据库入门和新一代数据库应用快速开发工具。

本书的特点是把数据库理论与当前流行的数据库系统 Oracle 相结合。本书可以作为大学计算机专业、信息管理专业及其他专业有关数据库课程的教材,也可作为从事计算机开发与应用的科研人员的数据库自学和参考用书。

目 录

第 1 章 数据库概论	1
1.1 什么是数据库	1
1.1.1 信息	1
1.1.2 数据	2
1.1.3 数据库	2
1.2 数据库的发展历史	3
1.2.1 人工管理阶段(20 世纪 50 年代中期以前)	3
1.2.2 文件系统阶段(20 世纪 50 年代后期到 60 年代中期)	3
1.2.3 数据库系统阶段(20 世纪 60 年代后期至今)	3
1.3 数据库系统	5
1.3.1 数据库系统的组成	5
1.3.2 数据库系统的三级模式结构	6
1.3.3 数据库管理系统	7
1.4 数据模型	9
1.4.1 三个世界	9
1.4.2 数据模型的三要素	11
1.4.3 实体模型	11
1.4.4 几种主要的数据模型	14
习题	20
第 2 章 关系数据库	21
2.1 关系模型的基本概念	21
2.1.1 笛卡尔积(Cartesian Product)	21
2.1.2 关系(Relation)	22
2.1.3 码(Key)	23
2.1.4 关系模式(Relation Schema)	23
2.1.5 完整性约束(Integrity Constrains)	25
2.2 关系代数	26
2.2.1 传统的集合运算	26
2.2.2 专门的关系运算	27
2.3 关系演算	33
2.3.1 元组关系演算	33
2.3.2 域关系演算	35

习题	36
第3章 SQL语言	38
3.1 SQL概述	38
3.2 SQL数据定义	39
3.2.1 基本表	39
3.2.2 索引	42
3.2.3 视图	43
3.3 SQL数据查询	45
3.3.1 简单查询	47
3.3.2 多表查询	51
3.3.3 聚合和分组	56
3.3.4 子查询	60
3.3.5 空值	65
3.4 SQL数据操纵	66
3.4.1 插入数据	66
3.4.2 删除数据	67
3.4.3 修改数据	68
3.4.4 更新视图	69
3.5 SQL数据控制	70
3.5.1 安全控制	70
3.5.2 事务管理	71
3.6 嵌入式SQL	74
3.6.1 一般格式	74
3.6.2 SQLCA	74
3.6.3 主变量	75
3.6.4 游标	75
3.6.5 动态SQL	77
习题	79
第4章 关系数据库设计理论	81
4.1 问题的提出	81
4.2 函数依赖	84
4.2.1 基本概念	84
4.2.2 函数依赖的逻辑蕴涵	86
4.2.3 码	86
4.3 Armstrong 公理系统	87
4.3.1 Armstrong 公理	87
4.3.2 公理的推论	87
4.3.3 公理的完备性	88
4.3.4 属性闭包的计算	89

4.3.5 函数依赖集的等价、覆盖和最小依赖集	90
4.4 多值依赖与连接依赖	92
4.4.1 多值依赖	92
4.4.2 连接依赖	94
4.5 关系模式的分解	95
4.5.1 无损连接性	95
4.5.2 依赖保持性	97
4.6 关系模式的规范化	98
4.6.1 第1范式(1NF)	98
4.6.2 第2范式(2NF)	99
4.6.3 第3范式(3NF)	100
4.6.4 BCNF	102
4.6.5 第4范式(4NF)	104
4.6.6 第5范式(5NF)	104
习题	105
第5章 数据库保护	107
5.1 数据库的安全性	107
5.1.1 数据库安全性的含义	107
5.1.2 数据库安全控制的一般方法	107
5.2 完整性控制	112
5.2.1 数据库完整性的含义	112
5.2.2 完整性规则的组成	112
5.2.3 完整性约束条件	113
5.2.4 实现参照完整性	115
5.3 并发控制与加锁	117
5.3.1 数据库并发性的含义	117
5.3.2 并发控制概述	118
5.3.3 封锁及封锁协议	119
5.3.4 封锁出现的问题及解决方法	122
5.4 数据库的恢复	124
5.4.1 数据库恢复的含义	124
5.4.2 数据库恢复的原理及其实现技术	125
5.4.3 数据库的故障和恢复策略	127
习题	129
第6章 数据库设计	130
6.1 数据库设计概述	130
6.1.1 什么是数据库设计	130
6.1.2 数据库设计的特点	131
6.1.3 数据库设计的方法	131

6.1.4 数据库设计的步骤	132
6.2 需求分析	135
6.2.1 需求分析的任务	135
6.2.2 系统需求分析方法	137
6.3 概念结构设计	139
6.3.1 概念模型的特点	139
6.3.2 概念结构设计的方法与步骤	139
6.3.3 数据抽象与局部 E-R 模型设计	140
6.3.4 视图的集成	142
6.4 逻辑结构设计	145
6.4.1 逻辑结构设计的任务和步骤	145
6.4.2 E-R 图向数据模型的转换	146
6.4.3 数据模型的优化	147
6.4.4 设计用户子模式	148
6.5 数据库物理设计	149
6.5.1 确定数据库的物理结构	149
6.5.2 评价物理结构	151
6.6 数据库实施	151
6.6.1 建立实际数据库结构	151
6.6.2 装入数据	152
6.6.3 应用程序编码与调试	152
6.6.4 数据库试运行	153
6.6.5 整理文档	153
6.7 数据库运行和维护	153
6.7.1 维护数据库的安全性与完整性	154
6.7.2 监测并改善数据库性能	154
6.7.3 重新组织和构造数据库	154
习题	155
第 7 章 数据库新技术	156
7.1 分布式数据库	156
7.1.1 分布式数据库的定义	156
7.1.2 分布式数据库的特点	157
7.1.3 分布式数据库管理系统	158
7.1.4 分布式数据库的体系结构	160
7.1.5 分布式数据库设计	161
7.1.6 分布式查询	164
7.2 面向对象数据库	165
7.2.1 面向对象数据模型	165
7.2.2 面向对象数据库语言	167

7.3 数据仓库	168
7.3.1 什么是数据仓库	169
7.3.2 为什么需要数据仓库	170
7.3.3 数据仓库的体系结构	170
习题	171
第8章 Oracle 数据库入门	172
8.1 什么是 Oracle 数据库	172
8.1.1 Oracle 数据库的概念	172
8.1.2 Oracle 数据库的特点	173
8.2 Oracle 服务器	173
8.2.1 Oracle 数据库的基本结构	173
8.2.2 表空间	174
8.2.3 Oracle 系统的组成	175
8.2.4 Oracle 实例	177
8.3 Oracle 8 for Windows NT 的安装	177
8.4 Oracle 的常见操作	180
8.4.1 Oracle 服务器的配置	180
8.4.2 Oracle 的登录	181
8.4.3 注册新用户	184
8.4.4 启动与关闭数据库	185
8.4.5 存储管理器	188
8.4.6 模式管理器	190
8.5 创建 Oracle 数据库	191
8.5.1 数据库助手	192
8.5.2 选择创建方式	192
8.5.3 设置数据库信息	193
8.5.4 设置控制文件参数	193
8.5.5 建立表空间	193
8.5.6 设置重做日志文件参数	194
8.5.7 设置 SGA 参数	194
8.5.8 创建数据库	196
8.6 SQL * PLUS	197
8.6.1 登录到 SQL * PLUS	197
8.6.2 SQL * PLUS 编辑器	198
8.6.3 SQL * PLUS 的使用举例	199
习题	202
第9章 新一代数据库应用快速开发工具	204
9.1 PowerBuilder	204
9.1.1 PowerBuilder 的特点	205

9.1.2 PowerBuilder 8.0 的开发环境	206
9.1.3 PowerBuilder 8.0 的开发步骤	208
9.2 Delphi	208
9.2.1 Delphi 的特点	209
9.2.2 Delphi 6 的新特征	210
9.2.3 Delphi 6 的开发环境	211
9.2.4 Delphi 6 的开发步骤	213
习题.....	215
参考文献.....	216

第 1 章 数据库概论

数据库是计算机科学的一个重要分支。数据库技术的产生和发展虽然只有 40 来年的时间,然而数据库的理论和应用却取得了巨大的成功,极大地促进了计算机应用向各行各业的渗透。本章主要介绍数据库的基本概念和特点、数据库的发展历史,以及数据模型等基本知识。

1.1 什么是数据库

什么是数据库?数据库的本质是存储有用的信息和数据。因此我们从信息和数据的基本概念开始着手讨论数据库。

1.1.1 信息

人类的一切活动都离不开信息和数据。信息是一种重要的资源,信息与能源、材料构成了现代社会的三大支柱。信息是目前许多学科广泛使用的概念。在不同的领域中,其含义有所不同。一般认为,信息是数据和消息中所包含的意义,是关于现实世界事物的存在方式或运动状态的反映的综合。例如,电脑,它由显示器、主机、键盘和鼠标等组成,这都是关于电脑的信息,关于电脑的存在状态的反映,是从不同角度反映电脑这个事物。信息有以下 4 个重要特征。

1. 信息的来源

信息源于物质和能量。一切事物,包括自然界和人类都产生信息,信息是物质和能量形态的反应,不可能脱离物质而存在。

2. 信息的感知

人类对客观事物的感知,可以通过感觉器官,也可以通过各种仪器和传感器,不同的信息源有不同的感知方法。

3. 信息的存储

动物用大脑存储信息即记忆,人类用文字提供长期存储信息的手段。计算机的发展进一步扩大了信息存储的范围。

4. 信息的加工、传递和再生

人类对收集到的信息可以进行取舍整理。从客观世界收集的信息或加工后的信息,可以通过各种形式传递,如语言、文字、电话、通信和网络等。这 4 大特征构成了信息的最重要的自然属性。作为信息的社会属性,信息已经成为社会上各行各业不可缺少的重要资源之一。

1.1.2 数据

几乎和信息同样广泛使用的另一个概念是数据,所谓数据,通常指用符号记录下来的可加以鉴别的信息。数据的概念包括两个方面:一是数据内容为事物特性的反映或描述;二是数据为符号的集合。

我们从数据的内容可以看到信息和数据之间的固有的联系:数据是信息的物理符号表示或载体,信息是数据的内涵,是对数据的语义解释。

但在另一方面,数据和信息又有区别,某一具体的信息和表示它的数据的这种对应关系又因环境而变化,数据能表示信息,但并非任何数据都能表示信息,有的数据可能完全没用处,称为数据垃圾。同一信息可能有不同的符号表示,同一数据也可能有不同的解释。因此,信息是人们消化理解了的数据,信息是抽象的,不随数据设备所决定的数据形式而改变,而数据的表示方式及存在方式却是客观现实的。

数据概念中的符号不仅包含数据符号,而且包含文字、图像和其他特征符号。其存储方式不仅可以在纸上,大量的是在磁记录、光刻等记录形式上,现在所谓的多媒体数据库技术的研究,正是这种数据概念拓宽的自然结果。

1.1.3 数据库

数据库,英文为 Database,这个名词起源于 20 世纪 50 年代,顾名思义,就是存放数据的仓库,这样理解是不确切的,实际上数据仓库已经成为数据库技术中的另一个专用名词,是数据库技术的一个新的应用领域。数据库的一般定义为:存储在计算机内的、有组织的、可共享的数据集合。其作用主要是共享数据库中的资源信息。数据库有以下几个特点。

1. 数据结构化

在数据库系统中,数据不再像文件系统中的数据那样从属于特定的应用,而是面向全组织的复杂的数据结构,数据的结构化是数据库区别于文件系统的根本特征。

2. 数据共享

数据库系统中的数据可供多个用户、多种语言和多个应用程序共享,这是数据库技术的基本特性,数据共享大大减少了数据的冗余度和不一致性,大大提高了数据的利用率和工作效率。

3. 数据独立性

数据独立性包括数据的物理独立性和逻辑独立性。用户的应用程序与存储在磁盘上的数据库的数据是相互独立的,这就是数据的物理独立性;同时用户的应用程序与数据库的逻辑结构是相互独立的,这就是数据的逻辑独立性;它不会因一方的改变而改变,这大大地减少了应用程序设计和数据库维护的工作量。

数据库技术所研究的问题是如何科学地组织和储存数据,如何高效地获取和处理数据。在计算机应用中,数据处理占的比重最大,而数据库系统是数据处理的核心,它的效能往往决定了整个计算机应用的经济效益。

数据库这门学科与其他基础软件、系统软件有着密切的联系。例如操作系统,数据库技术就是在操作系统的文件系统的基础上发展起来的,而且数据库系统本身就是在操作

系统的支持下才能工作。数据库和数据结构的联系也很密切,数据库技术不仅要用到数据结构的知识,而且丰富了数据结构的内容。编译方法在数据库技术中的应用也很广泛,在数据库系统中要定义许多数据语言。程序设计是使用数据库系统的最基本方式,因为数据库中大量的应用程序都是用高级语言加上数据库的操纵语言编写的。集合论、数理逻辑是关系数据库的理论基础,很多概念、术语、思想都直接用到关系数据库中。

1.2 数据库的发展历史

在讨论数据库的发展历史之前,我们先来回顾一下数据管理技术的发展历程。数据管理的发展经历了人工管理、文件系统和数据库 3 个阶段。

1.2.1 人工管理阶段(20 世纪 50 年代中期以前)

在人工管理阶段,计算机主要应用于科学计算,对于数据保存的需求尚不迫切,数据的管理是靠人工进行的,计算机不保存数据,也没有专用的软件对数据进行管理,只有程序(Program)的概念,没有文件(File)的概念,一组数据对应一个应用程序,数据存在大量重复存储现象,如图 1.1 所示。

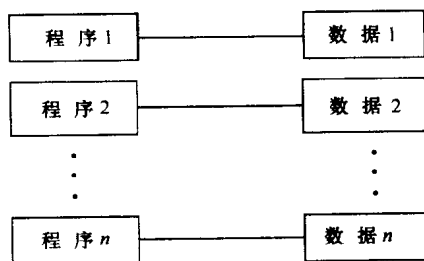


图 1.1 人工管理阶段

1.2.2 文件系统阶段(20 世纪 50 年代后期到 60 年代中期)

由于计算机技术的发展,硬件方面有了可以直接存取的外部存储设备,软件方面有了操作系统中专门管理数据的文件系统。数据的管理是以独立的数据文件形式存放,并可按记录存取。在文件系统阶段,一个应用程序可以处理多个数据文件,文件系统在程序与数据之间起到接口的作用,使程序和数据有了一定的独立性,如图 1.2 所示。这使得程序

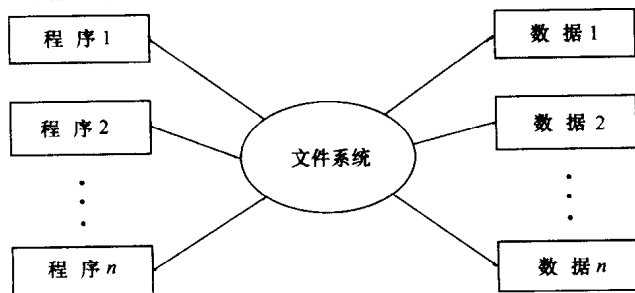


图 1.2 文件系统阶段

员可以集中精力于算法,不必过多地考虑物理细节,因此在这一时期各种数据结构和算法得到了充分的发展,大大丰富了计算机科学,今天的数据库也正是在文件系统的基础上发展起来的。但是,文件系统的致命缺陷是数据文件之间缺乏有机的联系,数据与程序之间缺乏独立性,不能有效地共享相同的数据,从而造成数据的冗余度大和不一致性,给数据的修改和维护带来了困难。

1.2.3 数据库系统阶段(20 世纪 60 年代后期至今)

随着计算机技术的迅速发展和广泛应用,磁盘技术取得重要进展,数据管理中数据量急剧增长,对数据共享和数据管理提出了更高的要求,文件系统已经不能满足应用的需求,导致了数据库技术的产生。数据库与应用程序的关系如图 1.3 所示。

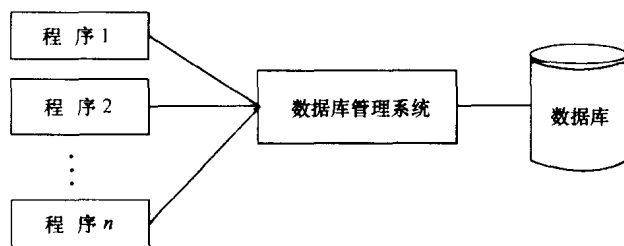


图 1.3 数据库系统阶段

1968 年,IBM 公司推出的商品化软件数据库信息管理系统 IMS(Information Management System),这是数据库技术的第 1 次飞跃,揭示了数据库技术的序幕;这是一种树型结构的层次模型,曾在 20 世纪 70 年代商业上广泛应用。

1969 年,美国数据系统语言协会 CODASYL(Conference On Data System Language)下属的数据库任务组 DBTG(Data Base Task Group)发表了网状数据库系统的标准文本:DBTG 文本,它是数据库网状模型的基础。

20 世纪 70 年代是数据库技术蓬勃发展的年代。网状系统和层次系统占据了整个商用市场,而关系系统仅处于实验阶段。许多计算机公司研制了各种数据库管理系统,许多教授、专家发表了大量的理论论文,使数据库技术在实践和理论上都得到了飞速的发展。数据库技术也日趋成熟。1970 年,IBM 公司 San Jose 研究室的高级研究员 E.F.Code 发表了一篇奠基性的论文,题目是《大型共享数据库的关系模型》,他引进了关系代数,首次提出了数据库的关系模型的概念,为数据库技术的发展奠定了基础,这是数据库技术的第 2 次飞跃。他在 1981 年获得了计算机科学的最高荣誉——图灵奖。1974 年,IBM 公司在 IBM370 大型机上研制成功第 1 个关系数据库系统 System R。1974 年 IBM 公司的 Boyce 和 Chamberlin 提出了结构化查询语言(Structured Query Language)SQL,在 IBM 公司的关系数据库系统 System R 上实现。20 世纪 80 年代是关系数据库技术得到了进一步的发展,关系数据库(RDB)由于使用方便,逐步取代网状系统和层次系统,占领了市场。同时关系数据库的理论也日趋完善。目前流行的关系数据库系统有 IBM 公司的 DB2、ORACLE 公司的 Oracle、SYBASE 公司的 Sybase、Microsoft 公司的 Microsoft SQL Server,以及微机上关系数据库系统 Debase、FoxPro 等。

现在,数据库技术已成为计算机领域中最重要技术之一,它是软件科学中一个独立

的分支。它的出现使得计算机应用渗透到工农业生产、商业、行政、科学研究、工程技术和国防军事的各个部门。管理信息系统、办公自动化系统、决策支持系统等都是使用了数据库管理系统或数据库技术的计算机应用系统。

1.3 数据库系统

1.3.1 数据库系统的组成

数据库系统(Database System, 简称 DBS)是采用了数据库技术的计算机系统。通常由数据库及其硬件系统、软件、人员 3 部分组成。数据库系统的层次结构如图 1.4 所示。

1. 数据库及其硬件系统

数据库系统是建立在计算机系统之上的,数据库的硬件包括 CPU、内存、外部存储器、I/O 通道等硬件设备。但对数据库系统来说,有以下特殊的要求。

1) 足够大的内存

要有足够大的内存用于存放操作系统(OS)、数据库管理系统(DBMS)及数据库缓冲区。例如 Windows NT 上的 Oracle 8 数据库系统至少需要 128MB 内存才能正常工作,一般内存越大数据库性能越好,因为这样可以增大数据库的缓冲区,减少外存的访问次数。

2) 足够大容量的磁盘

数据库要存储大量的数据,磁盘是主要的计算机存储设备,数据库文件就存在磁盘上,因此也要有尽可能高的磁盘存取速度,另外还需要磁带或光盘作数据库的备份。其他要考虑的有可支持终端数、性能稳定性、安全性、联网能力等因素,CPU 速度虽然重要,但在数据库系统不再是数据库运算速度的决定因素。因为在数据处理中数据的存取速度主要是 I/O 所占的时间。

2. 软件

数据库系统的软件主要包括操作系统 OS、数据库管理系统 DBMS、带有数据库接口的各种高级语言、应用开发系统和实用程序等。DBMS 是在操作系统的文件系统基础上发展起来的,是操作系统支持下管理数据库的软件系统,是数据库系统的核心软件。当前比较流行的应用开发工具主要有 PowerBuilder、Delphi、C++、Java、Visual FoxPro 等,这些高级语言应与数据库系统有良好的接口,称为第 4 代语言,它们为数据库应用系统的开发和应用提供了良好的环境,使用户提高生产率 20 倍~100 倍。使用第 4 代语言的人主要是终端用户,从初学者到专家都可以快速简便地用它编写程序解决用户的问题。

3. 人员

管理、开发和使用数据库系统的人员主要有数据库管理员、应用程序员和普通用户。数据库系统中不同人员涉及到不同的数据抽象级别,具有不同的数据视图。

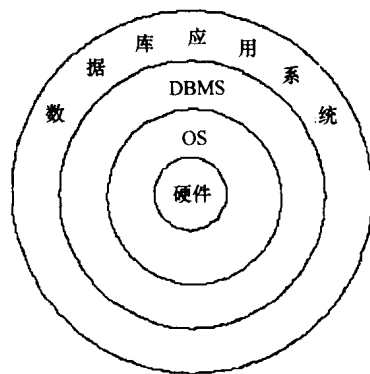


图 1.4 数据库系统的层次结构图

1) 普通用户

普通用户有应用程序和终端用户两类。它们通过应用系统的用户接口使用数据库,目前常用的接口方式有菜单驱动、表格操作、图形显示、报表生成等,这些接口使得用户的操作简单、易学易用,适合非计算机专业人员使用。

2) 应用程序员

应用程序员是负责设计和调试数据系统的应用程序的人员。他们通常使用 4GL 开发工具编写数据库应用程序,供普通用户使用。

3) 数据库管理员(Database Administrator, 简称 DBA)

DBA 在数据库管理中是极其重要的,即所谓的超级用户。DBA 全面负责管理、控制和维护数据库,使数据能被任何有权使用的人有效使用,DBA 可以是一个人,或几个人组成的一个小组。DBA 是面向应用的,对用户的需求和系统性能都要有充分的认识,要求他具有系统分析和企业业务等多方面的知识。DBA 必须要参与数据库系统的设计过程,决定数据库的内容、结构及用户的安全级别。在数据库的日常运行中,要监视数据库系统的运行情况,及时发现系统的潜在问题作相应的调整,在系统发生硬件或软件故障时,必须与系统服务商一起迅速排除故障,保证数据库系统及时正常运行。DBA 主要有以下 5 大职责:

- (1) 数据库的设计。
- (2) 数据库的存储和存取。
- (3) 数据库的运行。
- (4) 数据库的重组。
- (5) 数据库的维护。

1.3.2 数据库系统的三级模式结构

数据库系统通常采用三级模式结构:外模式、模式和内模式,如图 1.5 所示。

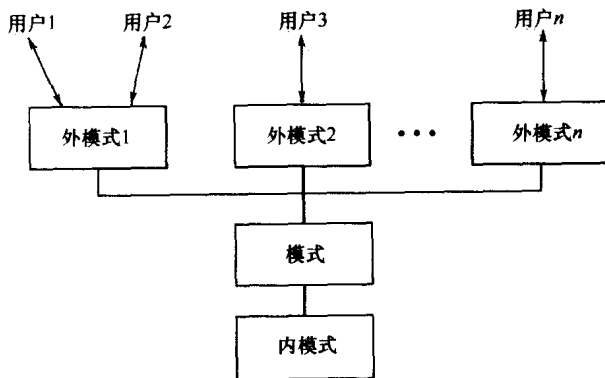


图 1.5 数据库系统的三级模式结构

1. 外模式(External Schema)

外模式是指用户所看到和使用的数据库,即局部逻辑结构,又称子模式或用户视图。一个数据库可以有多个外模式,由于用户的需求、数据的安全等方面的不同,可以用不同的外模式。每个用户都需要通过一个外模式来使用数据库,但不同的用户可以使用同一