

21世纪大学计算机基础规划教材

数据库原理及应用

张文祥 杨爱民 主 编
薛立新 陈荣春 副主编

- 将数据库理论与当前流行的数据库系统Oracle相结合,理论联系实际剖析了数据库系统的理论及应用知识。
- 系统地介绍了关系数据库的基本概念、基本原理、基本方法和应用实例。讲解的内容由浅入深,易于理解,通俗易懂。
- 既可作为普通高校、成人院校本科计算机专业和信息管理专业的教材,也可作为高职高专相关专业的教材以及相关领域技术人员的参考书或培训教材。



中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

21 世纪大学计算机基础规划教材

数据库原理及应用

张文祥	杨爱民	主 编
薛立新	陈荣春	副主编
	汪 沁	参 编

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

内 容 简 介

本书是根据教育部制定的关于计算机科学与技术及相关专业学生的培养目标而编写的。本书把数据库理论与当前流行的数据库系统 Oracle 相结合,理论联系实际剖析了数据库系统的理论及应用知识。

本书系统地介绍了关系数据库的基本概念、基本原理、基本方法和应用实例。内容包括:数据库系统概述、关系数据库理论基础、数据库的设计、Oracle 系统概述及安装、Oracle 数据库服务器管理、SQL 语言基础、Oracle PL/SQL 语言、数据库的安全管理、并发控制和数据库技术的发展等内容。

本书既可作为普通高校、成人院校本科计算机专业和信息管理专业的教材,也可作为高职高专相关专业的教材以及相关领域技术人员的参考书或培训教材。

图书在版编目(CIP)数据

数据库原理及应用/张文祥,杨爱民主编. —北京:
中国铁道出版社, 2006. 6
(21 世纪大学计算机基础规划教材)
ISBN 7-113-07078-7

I. 数... II. ①张... ②杨... III. 数据库系统—高
等学校—教材 IV. TP311.13

中国版本图书馆 CIP 数据核字(2006)第 075961 号

书 名: 数据库原理及应用

作 者: 张文祥 杨爱民 薛立新 陈荣春 汪 沁

出版发行: 中国铁道出版社(100054,北京市宣武区右安门西街8号)

策划编辑: 严晓舟 秦绪好

责任编辑: 苏 茜 翟玉峰 郑 双

封面设计: 薛 为

封面制作: 白 雪

责任校对: 刘 洁

印 刷: 河北省遵化市胶印厂

开 本: 787×1092 1/16 印张: 14.25 字数: 340 千

版 本: 2006 年 8 月第 1 版 2006 年 8 月第 1 次印刷

印 数: 1~5 000 册

书 号: ISBN 7-113-07078-7/TP·1824

定 价: 20.00 元

版权所有 侵权必究

本书封面贴有中国铁道出版社激光防伪标签,无标签者不得销售

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社计算机图书批销部调换。

前 言

随着计算机技术与网络通信技术的发展,数据库技术已成为信息社会中对大量数据进行组织与管理的最有效的技术手段及软件技术,是网络化信息管理系统的基础。数据库技术产生于 20 世纪 60 年代末,是计算机软件的一个重要分支,发展至今已有 30 多年的历史,它不仅形成了完整的理论体系,而且随着硬件技术与软件技术的快速发展而不断更新和完善,并广泛应用于社会各个领域。目前绝大多数计算机应用系统都需要数据库技术的支持,数据库系统已经成为信息系统的基础和核心。为此各高校结合自己专业的特点纷纷开设了数据库原理与应用课程,以适应社会的需求。

目前,国内高校该类课程的教材内容多以数据库自身的体系为脉络,注重数据库原理公式的推导和介绍,一般比较偏重于数据库理论。但对于信息专业和应用型普通高校计算机专业,尤其是计算机方向高职高专的大学生来说,既有完整充足的数据库理论,又有把理论和编程思路与数据库相结合的实际应用,才是该层次学生更需要的教材。基于上述出发点,本书以关系数据库为基础,介绍了数据库的基础理论和基本技术,包括数据库体系结构、数据模型、数据库设计方法、数据库安全保护及并发控制技术等,同时本书还侧重于以大型数据库 Oracle 系统为例,全面介绍了 Oracle 数据库的创建、管理与应用,通过标准查询语言 SQL 操作 Oracle 数据库的方法。力求通过本书的讲解及上机实践,让学生既能理解关系型数据库的理论,又能掌握关系型数据库 Oracle 系统的实际应用方法。

本书共分 10 章,第 1 章主要介绍了数据库系统的基本概念、数据库系统的体系结构及数据模型;第 2 章主要介绍了关系数据库理论,包括关系数据结构、关系演算理论、函数依赖及关系规范化理论和方法;第 3 章主要介绍了数据库的概念设计、逻辑设计和物理设计的理论与方法;第 4 章介绍了 Oracle 数据库系统的安装、设置和连接;第 5 章主要介绍了 Oracle 数据库服务器的配置、管理和使用;第 6 章主要介绍了 SQL 语言对数据进行定义、查询、更新和删除的方法;第 7 章主要介绍 Oracle PL/SQL 语言的使用;第 8 章主要介绍数据库的安全管理技术和实用工具;第 9 章结合 Oracle 数据库系统介绍了数据库的并发控制技术;第 10 章介绍了数据库技术新的发展。本书为帮助学生对数据库理论的理解和应用,每章后都配有适量的习题供学生练习。

本书主编为张文祥、杨爱民,副主编为薛立新、陈荣春,参编的老师还有汪沁。具体任务分工是张文祥编写第 1 章和第 2 章,杨爱民编写第 3 章、第 8 章、第 9 章,薛立新编写第 4 章、第 5 章,陈荣春编写第 6 章、第 7 章,汪沁编写第 10 章,另外本书的审稿工作由张文祥老师负责。

限于编者水平有限,书中难免有错误和不妥之处,敬请广大读者和专家批评指正。

编 者

2006 年 5 月

目 录

第 1 章 数据库系统概述	1
1.1 数据库技术的产生与发展	1
1.1.1 人工管理阶段	1
1.1.2 文件系统管理阶段	1
1.1.3 数据库管理阶段	2
1.2 数据库系统	4
1.2.1 数据库系统组成	4
1.2.2 数据库系统的效益	6
1.3 数据库管理系统	7
1.3.1 DBMS 的主要功能	7
1.3.2 DBMS 的组成	8
1.3.3 DBMS 的工作过程	9
1.4 数据模型	10
1.4.1 信息和数据	11
1.4.2 数据模型的三个层次	11
1.4.3 信息世界中的基本概念	12
1.4.4 概念模型的 E-R 模型表示方法	13
1.4.5 数据库层次的数据模型	16
1.4.6 物理数据描述	20
1.5 数据库系统的模式结构	20
1.5.1 数据库系统的三级模式结构	20
1.5.2 数据库的二级映像功能与数据独立性	22
1.5.3 小结	22
1.6 数据库系统的体系结构	22
1.6.1 单用户数据库系统	23
1.6.2 主从式结构的数据库系统	23
1.6.3 分布式结构的数据库系统	23
1.6.4 客户机/服务器结构的数据库系统	24
思考与练习	25
第 2 章 关系数据库理论基础	27
2.1 关系数据库概述	27
2.1.1 数据结构	27
2.1.2 关系操作	27

2.1.3 完整性.....	28
2.2 关系数据结构.....	28
2.2.1 关系的定义及性质.....	28
2.2.2 关系模式与关系数据库.....	29
2.2.3 关系的完整性规则.....	30
2.3 关系代数.....	31
2.3.1 传统的集合运算.....	31
2.3.2 专门的关系运算.....	33
2.3.3 关系代数表达式及其应用实例.....	37
2.3.4 扩充的关系代数操作.....	39
2.4 关系演算.....	40
2.4.1 元组关系演算.....	40
2.4.2 域关系演算.....	42
2.4.3 关系运算的安全性和等价性.....	42
2.5 查询优化.....	43
2.5.1 关系代数表达式的优化问题.....	43
2.5.2 关系代数表达式的等价变换规则.....	44
2.5.3 优化的一般策略.....	46
2.5.4 优化算法.....	47
2.6 函数依赖.....	49
2.6.1 问题的提出.....	49
2.6.2 函数依赖定义.....	51
2.6.3 码.....	52
2.7 关系的规范化.....	52
2.7.1 第一范式.....	52
2.7.2 第二范式.....	53
2.7.3 第三范式.....	54
2.7.4 BCNF.....	54
2.7.5 模式分解.....	55
思考与练习.....	56
第3章 数据库的设计.....	58
3.1 数据库设计的基本步骤.....	58
3.2 需求分析.....	60
3.2.1 需求描述与分析.....	60
3.2.2 需求分析阶段的输入和输出.....	60
3.2.3 需求分析的内容与方法.....	61
3.2.4 需求分析的步骤.....	61

3.2.5 数据字典.....	63
3.3 概念设计.....	65
3.3.1 概念设计的必要性及要求.....	65
3.3.2 概念设计的方法与步骤.....	66
3.3.3 数据抽象.....	67
3.3.4 E-R 模型的操作.....	69
3.3.5 采用 E-R 方法的数据库概念设计.....	71
3.4 逻辑设计.....	77
3.4.1 逻辑设计环境.....	77
3.4.2 逻辑设计的步骤.....	77
3.4.3 从 E-R 图向关系模型转换.....	78
3.4.4 设计用户子模式.....	80
3.4.5 对数据模型进行优化.....	80
3.5 数据库的物理设计.....	82
3.5.1 数据库设计人员需掌握的物理设计知识.....	82
3.5.2 数据库物理设计的主要内容.....	82
3.5.3 物理设计的性能评价.....	85
思考与练习.....	85
第 4 章 Oracle 系统概述及安装.....	87
4.1 Oracle 系统概况.....	87
4.1.1 Oracle 系统的特点.....	87
4.1.2 Oracle 产品结构.....	88
4.2 Oracle 数据库系统的安装.....	90
4.2.1 Oracle 数据库系统服务器端的安装.....	90
4.2.2 Oracle 数据库系统客户端的安装.....	95
4.3 客户端和服务端端的连接.....	99
4.3.1 服务器端的设置.....	99
4.3.2 客户端的设置.....	100
思考与练习.....	104
第 5 章 Oracle 数据库服务器管理.....	105
5.1 Oracle 数据库服务器的体系结构.....	105
5.1.1 Oracle RDBMS 文件结构.....	105
5.1.2 Oracle 数据库的逻辑结构.....	107
5.1.3 Oracle 数据库的存储结构.....	110
5.2 管理服务器的配置.....	111
5.2.1 数据库服务器的管理模式.....	111
5.2.2 管理服务器的配置.....	112

5.2.3	管理服务器的启动与关闭	114
5.3	数据库服务器的管理用户	115
5.3.1	system 用户	115
5.3.2	sys 用户	115
5.3.3	scott 用户	115
5.3.4	sys 和 system 用户的比较	116
5.4	管理工具的使用	116
5.4.1	“SQL Plus” 的使用	116
5.4.2	“SQL Plus Worksheet” 的使用	117
5.4.3	“Enterprise Manager Console” 的使用	119
5.5	数据库服务器的启动与关闭	121
5.5.1	服务器的关闭	121
5.5.2	服务器的启动	123
	思考与练习	124
第 6 章	SQL 语言基础	126
6.1	SQL 概述	126
6.2	SQL 数据定义	127
6.2.1	基本表的创建、修改、删除及重命名	127
6.2.2	索引的定义和删除	130
6.2.3	视图的定义和删除	131
6.3	SQL 数据查询	132
6.3.1	查询命令 (SELECT)	132
6.3.2	简单查询	133
6.3.3	表连接操作	135
6.3.4	集合运算	137
6.3.5	聚合和分组查询	138
6.3.6	子查询	138
6.4	SQL 数据操纵	140
6.4.1	向表中插入新行 (记录)	140
6.4.2	表中记录更新 (UPDATE)	141
6.4.3	删除表记录 (DELETE)	141
6.5	SQL 数据控制	142
6.6	嵌入式 SQL	143
6.6.1	嵌入式 SQL 语言简介	143
6.6.2	嵌入式 SQL 语法格式	143
6.6.3	使用游标处理多条记录	144

6.6.4 动态 SQL 语句	146
思考与练习	147
第 7 章 Oracle PL/SQL 语言	150
7.1 PL/SQL 语法基础	150
7.1.1 PL/SQL 程序结构	150
7.1.2 PL/SQL 语法要素	150
7.2 PL/SQL 流程控制	154
7.2.1 条件控制语句	154
7.2.2 循环控制语句	155
7.3 PL/SQL 程序应用	158
7.3.1 PL/SQL 函数	158
7.3.2 PL/SQL 过程	159
7.3.3 PL/SQL 软件包	161
7.3.4 PL/SQL 触发器	163
7.3.5 PL/SQL 中的事务处理	164
7.4 游标	164
7.4.1 定义游标	165
7.4.2 打开与关闭游标	165
7.4.3 从游标提取数据	165
思考与练习	166
第 8 章 数据库的安全管理	167
8.1 数据库的安全性（用户鉴别、特权、角色、审计）	167
8.1.1 数据库的存取控制和用户的建立	168
8.1.2 特权和角色	171
8.1.3 审计	176
8.2 数据完整性（数据库触发器）	176
8.2.1 完整性约束	177
8.2.2 数据库触发器	177
8.3 Oracle 数据库的安全管理	180
8.3.1 系统安全性与授权	180
8.3.2 审计工具	184
8.3.3 利用视图实施安全性控制	186
8.4 Oracle 数据库的备份和恢复	187
8.4.1 数据库的卸载实用程序（EXPROT）	187
8.4.2 数据库的装载实用程序（IMPORT）	190
思考与练习	192

第 9 章 并发控制	193
9.1 事务的基本概念	193
9.1.1 事务的定义	193
9.1.2 事务的性质	194
9.2 并发控制	194
9.2.1 丢失修改 (Lost Update)	195
9.2.2 不可重复读 (Non-Repeatable Read)	195
9.2.3 读“脏”数据 (Dirty Read)	196
9.3 封锁及封锁协议	196
9.3.1 封锁	196
9.3.2 封锁协议	197
9.3.3 活锁和死锁	199
9.4 并发调度的可串行性及两段锁协议	201
9.4.1 并发调度的可串行性	201
9.4.2 两段锁协议	203
9.5 封锁的粒度	204
9.5.1 多粒度锁协议	204
9.5.2 意向锁	205
9.6 Oracle 的并发控制机制	206
思考与练习	207
第 10 章 数据库技术的发展	208
10.1 概述	208
10.2 新一代数据库系统	208
10.2.1 面向对象数据库系统宣言	208
10.2.2 第三代数据库系统宣言	208
10.2.3 数据库技术与相关技术结合形成新型的数据库系统	210
思考与练习	211
附录 上机实验指导	212
参考文献	218

第 1 章 数据库系统概述

1.1 数据库技术的产生与发展

数据库技术是一门研究数据管理的技术。它是随着计算机应用领域由科学计算发展到数据处理而产生的一种技术，它的发展是与计算机应用领域对数据处理的速度和规模要求及计算机硬件和软件技术的发展支持分不开的。用计算机实现数据管理经历了三个发展阶段：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 年代后期至 60 年代中后期）和数据库系统阶段（20 世纪 60 年代末开始）。

1.1.1 人工管理阶段

早期的计算机外存没有磁盘等直接的存储设备，也缺少相应的软件支持，使用计算机进行数据处理时，要将原始数据和程序一起输入主存，运算处理后将结果数据输出，数据处理的方式基本上是批处理。

这个时期数据管理的特点是：

（1）数据不保存。这个时期处理的数据量不大，不需要保存。

（2）数据的独立性差。程序员设计应用程序时面对的是裸机，不仅要设计处理数据的操作步骤，数据的组织方式也必须由程序员自行设计与安排，数据与程序不具有独立性，一旦数据发生改变，就必须由程序员修改程序。由于各应用程序处理的数据之间毫无联系，不同程序处理的数据之间会有相同数据的重复，编程效率比较低，处理过程人工干预比较多。

（3）只有程序（Program）的概念，没有文件（File）的概念。即使有文件，也大多是顺序文件。由于没有数据管理软件，基本没有文件的概念。

（4）数据面向应用。一组数据对应于一个程序。

1.1.2 文件系统管理阶段

随着计算机软硬件的发展，外存已有磁盘、磁鼓等直接存储设备。软件方面有了高级语言和操作系统。操作系统中的文件管理系统提供了管理外存数据的功能。文件管理系统的方式就是把相关的数据组织成数据文件，以记录为单位，以文件名的方式存储在磁盘上。在程序中以文件名和数据记录方式存取数据，不必考虑数据的具体存储位置。

这一阶段数据管理的特点是：

（1）数据可长期保存在磁盘上。用户可随时通过程序对文件进行查询、修改和删除等处理。

（2）数据的物理结构与逻辑结构有了区别，功能较简单。程序与数据之间有物理的独立性，程序只需通过文件名存取数据，不必关心数据的物理位置，数据的物理位置变动不一定影响程序，数据的物理结构与逻辑结构的转换由操作系统的文件管理系统完成，程序员不必过多地考虑数据的存放地址，而把精力放在算法的设计上即可。

(3) 文件的形式已多样化, 有索引文件、链接文件和直接存取文件等, 因而对文件的记录可顺序访问。但文件之间是独立的, 联系要通过程序去构造, 文件的共享性差。

(4) 数据独立于程序。有了存储文件以后, 数据不再属于某个特定的程序, 一定程度上可以共享。但文件结构的设计仍然是基于特定的用途, 程序仍然是基于特定的物理结构和存取方法编制的, 因此, 当数据的物理结构改变时, 仍需要修改程序。

(5) 对数据的存取以记录为单位。

文件系统管理阶段是数据管理技术发展的重要阶段, 但由于数据管理规模的扩大, 数据量的急剧增加, 逐渐显露出很多缺陷, 主要表现在:

① 数据冗余性 (Redundancy): 由于文件之间缺乏联系, 造成每个应用程序都有对应的文件, 就可能出现同样的数据在多个文件中重复存储。

② 不一致性 (Inconsistency): 这往往是由数据冗余造成的, 在进行更新操作时, 稍不谨慎, 就可能使同样的数据在不同的文件中不一样。

③ 数据联系弱 (Poor Data Relationship): 这是文件之间独立、缺乏联系造成的。

由于这些原因, 促使人们研究一种新的数据管理技术, 以克服文件系统的缺陷, 这就是 20 世纪 60 年代末产生的数据库技术。

1.1.3 数据库管理阶段

从 20 世纪 60 年代后期开始, 计算机用于信息处理的规模越来越大, 对数据管理的技术提出了更高的要求, 此时开始提出计算机网络系统和分布式系统, 出现了大容量的磁盘, 文件系统已不再能胜任多用户环境下的数据共享和处理。

这个时期磁盘技术取得了重大进展, 大容量 (数百兆字节以上) 和快速存取的磁盘陆续进入市场, 成本下降了很多, 为数据库技术的实现提供了物质条件。各种数据库系统也相继问世, 首先是 1968 年美国 IBM 公司推出的层次模型的 IMS 数据库系统, 然后是 1969 年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, 它们为统一管理与共享数据提供了有力的支持。这两种数据库系统由于都是由文件系统发展而来的, 数据结构比较简单, 程序受数据库文件中物理结构的影响较大, 用户在使用数据库时需要对数据的物理结构有详细的了解, 这对数据库的使用造成很多困难。同时, 由于数据结构过于烦琐, 影响了复杂数据结构的实现。1970 年起, 美国 IBM 公司 E.F.Codd 博士连续发表一系列论文, 奠定了关系数据库的理论基础, 关系数据库在 20 世纪 70 年代中期至 20 世纪 80 年代得到了充分发展, 它具有简单的结构方式与较少的物理表示, 使用与操作符合人们日常的处理方式并且非常方便, 因此在 20 世纪 80 年代初逐步取代层次与网状数据库系统成为数据库系统的主导, 如 20 世纪 80 年代初出现的一批商品化的关系数据库系统, 如 Oracle、SQL/DS、DS、DB2、INGRES、INFORMIX、UNIFY 和 DBASE 等, 被广泛用于数据库查询的 SQL 语言在 1986 年被美国 ANSI 和国际标准化组织 (ISO) 采纳为关系数据库语言的国际标准。

与文件系统相比, 数据库系统克服了文件系统的缺陷, 提供了对数据更高级更有效的管理。概括起来, 数据库技术的管理方式具有以下特点。

(1) 数据的结构化: 数据库是存放在磁盘等直接存储的外存上的数据集合, 是按一定

的数据结构（数据模型）组织起来的。数据库系统与文件系统相比，文件系统文件内部数据间有联系，但文件之间不存在联系，从总体上看数据是没有结构的。而数据库中的文件数据是相互联系的，从总体上遵循一定的结构形式。数据库正是通过文件之间的联系反映现实世界事物间的自然联系的。

（2）数据共享：数据库的数据是面向整个应用系统中全体用户的，要考虑所有用户的数据要求，数据库中包含了所有用户的数据成分，不同用户可以使用其中部分数据，也可以使用同一部分数据。这样数据不再面向特定的某个或多个应用，而是面向整个应用系统。数据冗余明显减少，实现了数据共享。

（3）减少了数据的冗余和不一致性：由于数据库包含了整个系统所有用户的数据，用户操作的数据是通过数据库管理系统从数据库中映射出来的某个子集，不是独立的文件，实际上所有用户使用的是物理存储的一个文件，这就减少了数据冗余及不一致性。

（4）有较高的数据独立性：在数据库系统中，为保证数据独立性，系统提供映像功能，确保数据存储方式的改变不会影响到应用程序。数据库结构分成用户逻辑结构、整体逻辑结构和物理结构（见图 1-1）。数据独立性包括物理独立性和逻辑独立性。物理独立性是指数据库的物理结构（即数据的组织、存储、存取方式和外部存储设备等）发生变化时，不会影响到整体逻辑结构和用户的逻辑结构，由于应用程序是根据用户的逻辑结构编写的，所以应用程序不必改动，这样数据库就达到了物理独立性。逻辑独立性是指数据库的整体逻辑结构改变时，由数据库管理系统改变整体逻辑结构与用户逻辑结构之间的映像，使用户逻辑结构不变，从而应用程序不变，这样就实现了数据库的逻辑独立性。

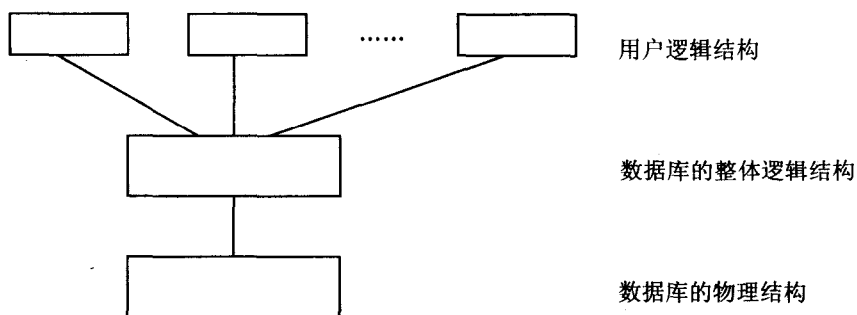


图 1-1 数据库结构

（5）方便的用户接口：数据库管理系统作为用户和数据库之间的接口，提供了数据库定义、数据库运行、数据库维护和数据控制等方面的功能；允许用户使用查询语言操作数据库，同时支持程序方式（用高级语言如 C、FORTRAN 等和数据库操纵语言编制的程序）操作数据库。

（6）数据控制功能：数据库管理系统提供了 4 个方面的数据控制功能。

① 数据完整性：是指保证数据库始终存储正确的数据。用户可设计一些完整性规则以确保数据值的正确性。例如可把数据值限制在某个范围内，并对数据值之间的联系进行各种检验。

② 数据安全性：保证数据的安全和机密，防止数据丢失或被窃取。

③ 数据库的并发控制：避免并发程序之间的相互干扰，防止数据库数据被破坏，杜绝提供给用户不正确的数据。

④ 数据的恢复：在数据库被破坏时或数据不可靠时，系统有能力把数据库恢复到最近某个时刻的正确状态。

对数据库的操作除了以记录为单位外，还可以数据项为单位，增加了系统的灵活性。

数据库阶段的程序和数据的联系可用图 1-2 表示。

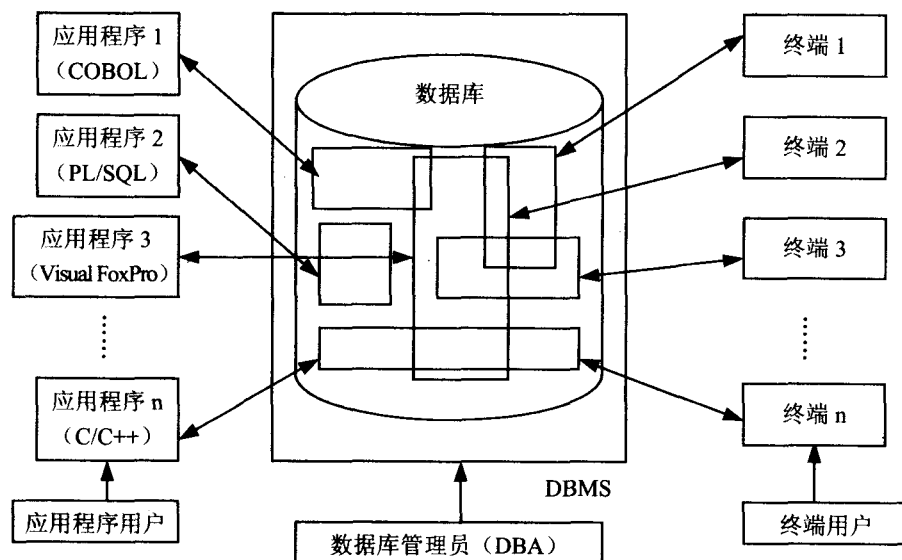


图 1-2 数据库阶段的程序和数据的联系

综上所述，数据库可以定义为：长期存储在计算机内有组织的、可共享的数据集合。数据库中的数据按一定的数据模型组织、描述和存储，具有较小的冗余度，较高的数据独立性和易扩展性，并可为各种用户共享。

从文件系统发展到数据库技术是信息处理领域的一个重大变化。在文件系统阶段，程序设计处于主导地位，数据只起着服从程序设计需要的作用；而在数据库方式下，数据开始占据了中心位置，数据的结构设计成为信息系统首先关心的问题，而利用这些数据的应用程序设计则退居到以既定的数据结构为基础的外围地域。

目前在国内外数据库应用已相当普遍，各行业都建立了以数据库技术为基础的大型计算机网络系统，并在国际互联网（Internet）的基础上建立了国际性联机检索系统，其应用深入到人类社会生活的各个领域，甚至家庭。为满足工程设计统计、人工智能、多媒体和分布式等不同领域的需要，在 20 世纪 80 年代中涌现出了如工程数据库、多媒体数据库、CAD 数据库、图形数据库、图像数据库、智能数据库、分布式数据库和面向对象的数据库等，特别是其中的面向对象的数据库由于其通用性强、适应面广而受到青睐，将会起主导作用。

1.2 数据库系统

1.2.1 数据库系统组成

数据库系统（DataBase System，简称 DBS）指在计算机系统中引入数据库后构成的系统，含义已经不仅仅是一组对数据进行管理的软件（即通常所说的数据库管理系统），也不仅仅是

一个数据库。一个数据库系统是一个实际可运行的，按照数据库方式存储、维护和向应用系统提供数据或信息支持的系统。它是存储介质、处理对象和管理系统的集合体，一般由数据库、硬件、软件、人员4部分构成。

1. 数据库 (DataBase, 简称 DB)

数据库是与一个特定组织的各项应用相关的全部数据的汇集。通常由两大部分组成：一部分是有关应用所需要的工作数据的集合，称为物理数据库，它是数据库的主体；另一部分是关于各级数据结构的描述，称为描述数据库，通常是由一个数据字典系统管理。

数据库构建主要是通过综合各个用户的文件，除去不必要的冗余，使之相互联系形成的数据结构，数据结构的实现取决于数据库的类型。

2. 硬件支持系统

硬件是数据库赖以存在的物理设备，包括 CPU、内存、外存和数据通道等各种存储、处理和传输数据的设备。对数据库系统来说，要求配备较大的内存，用来存放系统程序、应用程序，以及开辟系统和用户工作区缓冲区；外部存储一般要配备高速的、大容量的直接存取设备，如磁盘或光盘等；关注 I/O 存取速度、可支持终端数和性能稳定性等指标，在许多应用中还要考虑系统支持联网的能力和配备必要的后备存储设备等因素，此外还要求系统有较高的通道能力，以提高数据的传输速度。

3. 软件支持系统

主要包括操作系统、数据库管理系统、各种宿主语言和支持开发的实用程序等。数据库管理系统 (DataBase Management Systems, 简称 DBMS) 是管理数据库的软件系统，DBMS 是在操作系统 (OS) 支持下工作的，选用 DBMS 时还要考虑选择提供支持的操作系统；为开发应用系统，需要各种宿主语言 (如 COBOL、PL/I、FORTRAN、C 等) 及其编译系统，这些语言应与数据库有良好的接口。需要支持开发的实用程序，如报表生成器、表格系统、图形系统、具有数据库存取和表格 I/O 功能的软件以及数据字典等，它们是系统为应用开发人员和最终用户提供的高效率、多功能的交互式程序设计系统，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 20~100 倍。

4. 人员

管理、开发和使用数据库系统的人员主要有系统分析员、数据库管理员 (DBA)、应用程序员和用户。他们分别有不同的职责。

(1) 系统分析员：负责应用系统的需求分析和规范说明。他们要与用户及 DBA 配合，确定系统的软硬件配置并参与数据库各级模式的概要设计。

(2) 数据库管理员 (Data Base Administrator)：是对数据库系统监督和管理的人员，又称 DBA。

(3) 应用程序员：负责设计应用系统的程序模块，根据外模式编写应用程序和编写对数据库的操作过程程序。

(4) 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、表格操作和图形显示报表生成等，这些接口给用户提供了简明直观的数据表示。

大型的系统中由于数据库具有共享性,要想成功地运转数据库,需要配上 DBA 来维护和管理数据库,使之处于最佳的状态。DBA 可以是一个人或几个人组成的小组,其主要职责是:

(1) 决定数据库的信息内容和结构,确定某现实问题的实体联系模型,建立与 DBMS 有关的数据模型和概念模式。

(2) 决定存储结构和存取策略,建立内模式和模式/内模式映像。使数据的存储空间利用率和存取效率两方面都较优。

(3) 充当用户和 DBS 的联络员,建立外模式和外模式/模式映像。

(4) 定义数据的安全性要求和完整性约束条件,以保证数据库的安全性和完整性。安全性要求是用户对数据库的存取权限,完整性约束条件是对数据进行有效性检验的一系列规则和措施。

(5) 确定数据库的后援支持手段及制定系统出现故障时数据库的恢复策略。

(6) 监视并改善系统的“时空”性能,提高系统的效率。

(7) 当系统需要扩充和改造时,负责修改和调整外模式、模式和内模式。

总之,DBA 承担创建、监控和维护整个数据库结构的责任。由于职责重要和任务复杂,要求 DBA 兼有系统程序员和运筹学专家的素质和知识,一般由业务水平较高、资历较深的人员担任。

1.2.2 数据库系统的效益

数据库系统的应用,使计算机应用深入到社会的各个领域。这是因为从数据库系统可获得很大的效益,具体体现在以下几个方面:

(1) 灵活性。数据库容易扩充以适应新用户的要求,同时也容易移植以适应新的硬件环境和更大的数据容量。

(2) 简易性。由于精心设计的数据库能模拟企业的运转情况,并提供该企业数据逼真的描述,使管理部门和使用部门能很方便地使用和理解数据库。

(3) 面向用户。由于数据库反映企业的实际运转情况,因此基本上能满足用户的要求,同时数据库又为企业的信息系统奠定了基础。

(4) 数据控制。对数据进行集中控制,就能保证所有用户在同样的数据上操作,而且数据对所有部门具有相同的含意。数据的冗余减到最少,消除了数据的不一致性。

(5) 加快应用系统开发速度。程序员和系统分析员可以集中全部精力在应用处理设计上,而不必关心数据操纵和文件设计的细节,后援和恢复问题均由系统保证。

(6) 程序设计方便。数据库方法使系统中的程序数目减少而又不过分增加程序的复杂性,由于数据管理语言命令功能强,应用程序编写起来较快,进一步提高了程序员的生产效率。

(7) 修改方便。数据独立性使得修改数据库结构时尽量不损害已有的应用程序,使程序维护工作量大为减少。

(8) 标准化。数据库方法能促进建立整个企业的数据一致性和用法的标准化工作。

1.3 数据库管理系统

数据库管理系统 (DBMS) 是数据库系统中对数据进行管理的软件, 是数据库系统的核心组成部分。对数据库的一切操作, 包括定义、查询、更新及各种控制, 都是通过 DBMS 进行的。DBMS 是用户与数据库的接口, 用户要对数据库进行操作, 是由 DBMS 把操作从应用程序带到外部级、概念级, 再导向内部级, 进而操纵存储器中的数据。

DBMS 是针对某种数据模型设计的, 可以看成是某种数据模型在计算机系统上的具体实现。根据所采用数据模型的不同, DBMS 可以分成网状型、层次型、关系型、面向对象型等。但在不同的计算机系统中, 由于缺乏统一的标准, 即使同种数据模型的 DBMS, 它们在用户接口、系统功能等方面也常常是不同的。

1.3.1 DBMS 的主要功能

1. 数据库定义功能

DBMS 提供数据定义语言 (Data Definition Language, 简称 DDL) 定义数据库的结构, 包括外模式、内模式及其相互之间的映像, 定义数据的完整性约束、保密限制等约束条件。定义工作是由 DBA 完成的。在 DBMS 中有 DDL 的编译程序, 负责将 DDL 编写的各种模式编译成相应的目标模式。这些目标模式是对数据库的描述, 不是数据本身, 是数据库的框架 (即结构), 并被保存在数据字典中, 供以后进行数据操纵或数据控制时查阅使用。

2. 数据库操纵功能

DBMS 提供数据操纵语言 (Data Manipulation Language, 简称 DML) 实现对数据库的操作。基本的数据操作有 4 种: 检索、插入、删除和修改。DML 有两类, 一类是嵌入在宿主语言中使用, 例如嵌入在 COBOL、FORTRAN、C 等高级语言中, 这类 DML 称为宿主型 DML; 另一类是可以独立地交互使用的 DML, 称为自主型或自含型 DML。因而 DBMS 中必须包括 DML 的编译程序或解释程序。

3. 数据库运行控制功能

DBMS 对数据库的控制主要通过 4 个方面实现: 数据安全性控制、数据完整性控制、多用户环境下的并发控制和数据库的恢复机制。

(1) 数据库安全性控制是对数据库的一种保护。使用数据的用户, 必须向 DBMS 标识自己, 由系统确定是否可以对指定的数据进行存取。它的作用是防止未经授权的用户蓄意或无意地修改数据库中的数据, 以免数据泄露、被更改或破坏, 使企业蒙受巨大的损失。

(2) 数据完整性控制是 DBMS 对数据库提供保护的另一个重要方面。其目的是保持进入数据库中的存储数据的语义的正确性和有效性, 防止任何对数据造成违反其语义的操作。因此, DBMS 都允许对数据库中各类数据定义若干语义完整性约束, 并由 DBMS 强制实行。

(3) 并发控制是 DBMS 的第三类控制机制。数据库技术的一个优点是数据的共享性。但多个应用程序同时对数据库进行操作可能会破坏数据的正确性, 或者在数据库内存储了错误的数据, 或者用户读取了不正确的数据 (称为“脏”数据)。并发控制机制能防止上述情况发生, 正确处理好多用户、多任务环境下的并发操作。