

数据库系统教程

宋国治 编
傅国光 主编
傅国光 编

SHUJUKU
XITONG
JIAOCHENG

中国铁道出版社

数据库系统教程

张健沛 戴宗荫 编著
李玉忱 杨 静

中国铁道出版社

1995年·北京

(京) 新登字 063 号

内 容 简 介

本书从原理、技术和应用的角度,对数据库系统的基本概念、基本结构和基本应用作了详细介绍,并以典型的 Oracle 数据库系统作为实例,讲述了它的组成原理、工作方式、SQL * PLUS 语言、应用程序设计及安全管理。

本书包括十二章内容,自然地分为两部分。第一部分以数据库理论为主,着重讲述数据库系统的基本概念、关系代数理论、查询语言、规范化理论、查询优化及其数据库系统设计的基本技术和方法。第二部分以目前国内外较为流行的 Oracle 数据库系统为主,对其实现方法和实际应用作了详细介绍。

本书的特点是:概念清楚,系统性强;突出重点,理论和实际并重;由易到难,面向应用。

本书可供高等学校计算机专业、信息工程专业的本科生学习使用,也可供从事计算机软件工作的科技人员及其他有关人员学习参考。

数据库系统教程

张健沛 戴宗荫 李玉忱 杨静 编著

中国铁道出版社出版发行

(北京市东单三条 14 号)

责任编辑 郭宇 封面设计 马利

各地新华书店经售

北京市燕山联营印刷厂印

开本: 787×1092 毫米 1/16 印张: 16.75 字数: 403 千

1995 年 8 月 第 1 版 第 1 次印刷

印数: 1—5000 册

ISBN7-113-02094-1/TP·218 定价: 18.00 元

前 言

数据库技术自 60 年代后期出现以后，一直活跃在数据处理领域。20 多年来，数据库技术得到了迅速的发展，已经成为计算机科学的一个重要分支。由于数据库技术已经成为开发各种信息系统的重要工具，因它的推广使用，使得计算机应用深入到工农业生产、商业管理、办公自动化、科学研究和工程技术的各个领域。随着社会各个领域的迫切需要和广泛应用，人们对数据库技术重要性的认识越来越明确，对它的应用和学习要求也越来越迫切。

本书是作者多年从事数据库系统教学和科研的基础上编著的。在编写过程中，本着由易到难，循序渐进，理论与实践并重，突出重点的原则，力求作到概念清楚，深入浅出，理论联系实际，面向应用，重点突出应用最广泛的关系数据库系统，对它的概念、原理作了较为深入的讨论。为了突出实际应用，本书还较为详细地介绍了目前国内外较为流行的 Oracle 关系数据库，以使初学者在学习过程中能更好地把理论与实际结合起来，从而达到更好的学习效果。另外，在内容选材上，我们还注意尽量反映这一领域的新方法、新技术，以使学生对数据库领域的前沿动态有初步的了解。

全书共分十二章，第一章介绍数据库系统的基本概念，基本原理。第二至四章介绍当前最流行、应用最广泛、理论研究最多的关系数据库系统，对其基本概念、关系代数理论、查询语言、规范化理论、查询优化等都作了充分的介绍。第五章介绍数据库设计的基本技术和方法，以指导应用。第六章介绍数据库保护技术。第七章介绍数据库系统的研究发展方向。第八至十二章介绍目前国内外较为流行的 Oracle 数据库系统，对 Oracle 数据库系统的体系结构、SQL * PLUS 语言、PRO * C 预编译接口、应用生成器 SQL * FORM，以及 Oracle 数据库管理的原理及实现方法都作了充分的介绍。

本书第一、二、四、六、七章由张健沛编写；第三、五章由李玉忱编写；第八、九、十章由杨静编写；第十一、十二章由戴宗荫、尹学义编写。全书由张健沛负责结构的设计和内容的组织。

由于水平有限，加之时间匆忙，书中难免有错误和不妥之处，敬请读者批评指正。

编 者

1995 年 3 月

目 录

第一章 数据库系统导论	1
1.1 引 言	1
1.1.1 数据与数据处理	1
1.1.2 数据管理技术的发展过程	3
1.1.3 什么是数据库	4
1.1.4 数据库系统的主要特征	6
1.2 现实世界的描述	9
1.2.1 现实世界	9
1.2.2 信息世界	10
1.2.3 计算机世界	10
1.2.4 常用术语的说明	11
1.3 数据模型	11
1.3.1 层次模型	12
1.3.2 网状模型	12
1.3.3 关系模型	13
1.3.4 各模型比较	14
1.4 数据库系统结构	15
1.4.1 数据库系统结构的基本概念	15
1.4.2 数据库系统的组成	17
1.4.3 数据库管理系统	19
1.4.4 数据库语言	21
1.4.5 数据库管理员	24
1.4.6 数据字典	24
1.4.7 数据库系统的工作流程	26
习 题	27
第二章 关系数据库	29
2.1 关系数据库的基本概念	29
2.1.1 笛卡儿乘积	29
2.1.2 关 系	30
2.1.3 关 键 字	30
2.1.4 关系模式	30
2.1.5 关系模型	31
2.1.6 关系数据库	31

2.2	关系模型及其描述	31
2.2.1	关系模型	31
2.2.2	关系模式	32
2.2.3	关系子模式	33
2.2.4	关系模型的完整性规则	34
2.3	关系数据库语言	34
2.3.1	关系代数	35
2.3.2	关系演算语言	42
2.3.3	QBE 语言	46
2.3.4	关系数据库语言的评价	48
	习 题	48
第三章	关系数据库设计理论	50
3.1	引 言	50
3.2	关系模式的规范化	52
3.2.1	函数依赖	52
3.2.2	第一范式	54
3.2.3	第二范式	55
3.2.4	第三范式	56
3.2.5	BCNF	57
3.3	函数依赖的公理系统	58
3.3.1	引 言	58
3.3.2	Armstrong 公理	59
3.3.3	函数依赖集的等价与覆盖	64
3.4	关系模式的分解	67
3.4.1	模式分解的等价问题	67
3.4.2	无损联接分解	68
3.4.3	分解的函数依赖保持性	72
3.4.4	模式分解算法	74
3.5	多值依赖和关系的 4NF	78
3.5.1	多值依赖定义	78
3.5.2	多值依赖公理	80
3.5.3	第四范式 (4NF)	81
3.6	联接依赖和关系的 5NF	82
3.6.1	联接依赖	82
3.6.2	第五范式	84
	习 题	85
第四章	查询优化	87
4.1	查询优化概述	87
4.2	查询优化的一般策略	88

4.3	基于关系代数表达式的优化算法	89
4.3.1	关系代数表达式的等价变换规则	89
4.3.2	关系代数表达式的优化算法	90
4.4	分解查询的优化方法	92
4.4.1	分解处理	92
4.4.2	结局处理	95
4.5	连接运算的优化	97
4.5.1	实现二元连接运算的几种常用方法	97
4.5.2	n元连接的优化	99
	习 题	105
第五章	数据库设计	106
5.1	数据库设计概述	106
5.1.1	数据库设计的主要内容	106
5.1.2	数据库设计方法简介	107
5.1.3	数据库设计步骤	108
5.2	需求分析	110
5.3	概念模型设计	112
5.3.1	设计局部 E-R 模型	113
5.3.2	将各局部 E-R 模型综合成总体 E-R 模型	114
5.4	逻辑结构设计	118
5.4.1	概念模型向网状模型转换	118
5.4.2	概念模型向关系模型转换	120
5.4.3	规范化理论的应用	121
5.4.4	模型的评价与改进	122
5.5	数据库物理设计	123
5.5.1	数据簇集设计	123
5.5.2	索引的选择	126
5.6	数据库的实施与维护	128
5.6.1	数据库数据的装入	128
5.6.2	数据库的试运行	129
5.6.3	数据库的运行和维护	130
	习 题	130
第六章	数据库保护	133
6.1	引 言	133
6.2	数据库的安全性	133
6.2.1	用户标识和鉴定	134
6.2.2	授权规则	134
6.2.3	数据分级法	135
6.2.4	数据加密	135

6.3	数据库的完整性.....	136
6.3.1	完整性的破坏原因	136
6.3.2	完整性规则	137
6.4	数据库的并发控制.....	137
6.4.1	排它型封锁和 PX 协议	139
6.4.2	可串行化、死锁和活锁	139
6.4.3	时间标志法	140
6.5	数据库的恢复.....	141
6.5.1	数据库故障和数据库恢复	142
6.5.2	恢复方法	142
6.6	数据库的再组织.....	143
6.6.1	再组织的必要性	143
6.6.2	影响再组织点的因素	143
6.6.3	再组织点的选取	144
	习 题.....	144
第七章	数据库系统的研究与发展.....	145
7.1	分布式数据库系统.....	145
7.1.1	分布式数据库系统概述	145
7.1.2	分布式数据库系统的组成	149
7.2	面向对象数据库系统.....	154
7.2.1	传统的关系数据库系统的不足之处	155
7.2.2	对象概念的实质	155
7.2.3	面向对象数据库技术的产生	156
7.2.4	面向对象模型	156
7.2.5	面向对象数据库管理系统	157
7.3	一些新型的数据库.....	158
7.3.1	演绎数据库	158
7.3.2	模糊数据库	159
7.3.3	模糊演绎数据库	160
7.3.4	巨型数据库	161
7.3.5	主动数据库	161
7.3.6	多媒体数据库	162
7.3.7	集成的工程数据库	163
7.3.8	联邦数据库	163
7.3.9	时态数据库	163
	习 题.....	164
第八章	Oracle 关系数据库系统概述	165
8.1	Oracle 系统软件结构.....	165
8.1.1	Oracle RDBMS	166

8.1.2 Oracle 数据库的结构	167
8.2 Oracle 产品结构	167
8.2.1 Oracle RDBMS 产品构成	168
8.2.2 Oracle 外围产品	168
习 题	169
第九章 SQL * PLUS 语言	170
9.1 概 述	170
9.1.1 SQL 语言	170
9.1.2 SQL * PLUS 命令	170
9.1.3 运行 SQL * PLUS	170
9.2 数据定义	171
9.2.1 定义表	171
9.2.2 修改表结构	173
9.2.3 删除表定义	173
9.2.4 表的重新命名	174
9.3 数据操纵	174
9.3.1 插入数据	174
9.3.2 修改数据	175
9.3.3 删除数据	176
9.4 数据查询	176
9.4.1 简单查询	177
9.4.2 空值的处理	179
9.4.3 表连接操作	181
9.4.4 集合运算	182
9.4.5 子查询	184
9.4.6 树型信息的查询	187
9.5 视图	192
9.5.1 视图的定义	192
9.5.2 视图查询	193
9.5.3 视图更新	193
习 题	194
第十章 Oracle 应用程序设计	195
10.1 概 述	195
10.2 预编译接口	195
10.3 PRO * C	197
10.3.1 程序首部	197
10.3.2 程序体	201
10.3.3 动态 SQL 语句	208
习 题	212

第十一章 应用生成器 SQL * FORM	213
11.1 SQL * FORM 的基本概念	213
11.1.1 FORM 的组成	213
11.1.2 当前的概念和事务	214
11.1.3 SQL * FORM 中的功能键和窗口	214
11.2 FORM 的定义和屏幕设计	215
11.2.1 FORM 的生成和定义	215
11.2.2 块级设计	216
11.2.3 域级设计	219
11.2.4 屏幕设计	223
11.3 触发器设计	225
11.3.1 触发器的基本概念	225
11.3.2 触发器的类型和用途	227
11.3.3 编写触发器的命令	230
11.3.4 触发器的定义过程	233
11.4 触发器的运行与结果	235
11.4.1 触发器的执行	235
11.4.2 一个触发器的执行过程	235
11.4.3 触发器执行的结果与后果	235
11.4.4 一个触发器的执行结果及原因	236
习 题	237
第十二章 Oracle 数据库管理	238
12.1 Oracle 数据库存储组织	238
12.1.1 存储结构	238
12.1.2 存储管理	239
12.2 Oracle 安全管理	243
12.2.1 系统安全性授权	243
12.2.2 审 计	247
12.2.3 利用视图实施安全性控制	250
12.3 Oracle 数据库并发控制	250
12.3.1 事 务	250
12.3.2 并发控制	251
12.4 Oracle 数据库恢复	253
12.4.1 BI 文件的恢复	253
12.4.2 AI 文件的恢复	255
习 题	255
参考文献	256

第一章 数据库系统导论

1.1 引言

1.1.1 数据与数据处理

人们在日常生活中接触到大量数据，如银行帐目数据、学校的教学管理数据、企业的生产管理和产品销售数据等等。这些数据都是人们生活中所离不开的，它们可以被视为人类社会中一种极其重要的资源。

然而这些数据之所以有价值，就是因为这些数据是表现信息，是载荷信息的物理符号。简单地讲，信息是向人们提供关于现实世界新的事实的知识，它反映了客观事物的物理状态。例如，某一国家的汽车产量是 30 万辆，这个 30 万辆是一个数据，而它又表示了个国家生产汽车的能力这一信息。由此可见，信息是用数据来表示的，但二者是即不可分离，又有一定区别的概念。一方面并非任何数据都能表示信息，信息只是消化了的数据，信息是依赖于数据而存在的。另一方面，信息是更基本的、直接反映现实的概念，数据是信息的具体表现。

对于数据进行综合推导，得出新的数据，这些新的数据则表示了新的信息。例如，在商品生产中，生产管理者根据某种商品历年销售数量及最近的市场需求调查，获得了许多数据。再对这些数据进行加工，就得出了有关商品的市场预测信息。生产管理者就可以根据这些信息进行分析和评价，作出对该产品是增产、减产还是停产的决策，并指挥生产部门的生产活动。这种从收集数据，到加工成信息进行评价和决策，再指挥实践活动，从而产生新的数据的循环过程称为信息循环。

诸如此类围绕信息所做的系统化的操作，称为信息处理。一般地说，信息处理就是指信息的收集整理、加工、归并、分类、计算、汇编、编辑、存储和传播等一系列活动的总称。因为信息是用数据表示的，所以对信息的处理又具体地体现在对数据的处理上。通过对原始数据（即尚未作评价的各种信息）的处理，产生新的数据（新的信息），作为结果。这一处理，包括对数据的收集、记录、分类、排序、存储，计算、加工、传输、制表和递交等处理。经过处理的数据是精炼的数据，是能够反映事物或现象的本质和特征及其内在联系的数据。

数据处理的历史可以追溯到远古时代。原始人类的结绳记数，累石记数便是数据处理的雏形。随着社会生产和文明的日益发展，信息的概念就趋于越来越复杂和深化，信息已经支配着人类的整个社会活动，所以有人把现代社会称为信息社会。研究信息的形态、传输、处理和存储理论的信息科学也就应运而生。随着信息概念的深化和发展，数据处理也就得到了相应的发展，就整个数据处理的发展过程来看，大致可以分成三个阶段。

第一阶段是数据的手工处理阶段。这是自远古时代到 19 世纪末一直使用的数据处理方法。由于当时的社会生产力低下，科学技术不发达，因此，数据处理的方法也就保持着简单、低级的手工操作。人们只能借助于如算盘、手摇计算机、计算尺等低级的计算工具处理数据。

这种方法的效率低，能处理的数量少，而且可靠性差。随着生产力和科学技术的发展，依靠手工操作来处理数据的方法，已经不能适应社会的需要，因而人们着手进行计算工具的改革和发明创造。

第二阶段是数据的机械处理阶段。这一阶段的主要特征是使用了比第一阶段先进得多且比较有效的工具。例如，H·霍勒内斯在19世纪80年代发明了卡片制表机。这种机器能以自动方式进行卡片的穿孔、校验、分类、整理和制表等操作。这是跨入机械数据处理阶段的一个重要标志。机械设备的使用，使数据处理的能力有了很大的改进和提高。但是这种设备的性能和使用，仍然受着很大的限制。这一阶段仍然有着较多的手工操作。直到20世纪40年代中期，第一台电子计算机ENIAC问世以后，数据处理才进入了自动化的电子数据处理时代。

第三阶段是电子数据处理阶段。电子计算机（简称计算机）的出现，为数据处理展现了广阔的前景。它不仅处理速度快、存储容量大、输入输出灵活，而且把人的手工操作降低到最小程度。电子数据处理方法不仅适应了不断提高的社会生产力的需要，而且给社会生产力的发展以推动和促进。因此，计算机的应用是数据处理领域中的一场革命。

尽管这3种数据处理的方法一个比一个先进，但是目前世界上仍然存在着手工和机械处理方法，特别是手工处理方法还是屡见不鲜的。而机械方法相对来说目前很少使用了。这一方面是因为某些领域的某些部门还不适宜使用计算机，另一方面是因为世界上各个国家和地区的发展还不平衡。这种局面将随着超小型和微型计算机的发展而得到改善。

人类活动的整个历史，离不开对信息和数据的收集、保存和处理。像银行、统计、仓库等部门的工作人员，无时无刻不在和大量的数据打交道。然而，随着社会的不断进步，人类社会积累的信息，正以“几何级数”的速度增长。因此人们过去传统的、落后的数据处理方法，已经远远适应不了形势发展的需要。所以长期以来，如何妥善地保存和管理数据，一直是人们十分关注的课题。人们对数据处理现代化的要求日益迫切。

计算机的出现，为人类处理数据自动化提供了可靠的物质基础。初期，计算机只是单纯为科学计算服务，所以人们又把它称为数字计算机。随着计算机硬件和软件的发展，计算机硬件成本的不断降低和各种新型外部设备的问世，特别是大容量磁盘存储器的生产和应用，人们逐渐地认识到，除了科学计算之外，计算机还具有强大的处理数据的能力，因为计算机的外存储器是保存数据资料的理想场所。于是，计算机的应用走出了科学计算这一传统领域，在数据处理这个广阔的领域大显身手。60年代后期，数据库技术出现了。数据库技术研究对象，就是高效地、巧妙地进行数据处理，而又开销最小。

目前，在整个计算机应用领域中，数据处理还是最庞大、最活跃的应用领域。在发达国家中，数据处理约占80%~90%。计算机工业历史说明，计算机是在科学计算领域中诞生，却在数据处理领域中站稳了脚跟，找到了真正的立足点。而数据库系统是数据处理的核心机构，数据库管理系统的最新成果，反映了数据处理技术的最新水平。有人用这样一段话概括地说明了数据库系统的地位：“数据库管理是现代计算机系统提供的最重要的功能。事实上，其重要性已经达到了这样的程度，即它已普遍成为购买计算机的主要出发点”。

一个计算机化的电子数据处理系统（Electronic Data Processing System，简称EDP系统），包括了人和计算机系统。它执行输入、处理、存储、输出和控制功能，用于完成对数据的一系列操作过程。一个计算机化的电子数据处理系统，一般由三个基本系统组成，它们是电传

处理系统、数据管理系统及科学计算系统。如图 1.1 所示。

电传处理系统的基本功能是，在适当的地点和时间提供必要的数据库，即由它完成数据库提供、通信（传输）和传递。

数据管理系统完成对数据的组织、编目、定位、存储、检索和维护等功能。

科学计算系统作为数据处理系统的一部分，会大大地加强系统的处理能力。它强调使用推理过程产生信息，好象纸和笔一样，成为人们在分析问题时的有力工具。但是有些复杂的推理问题，单靠纸和笔是不能解决的，而要靠科学计算系统来完成，因为它有助于提高人们的逻辑推理和通信的功能。例如在一个企业中，它能够用来解决某些管理决策、计划编制和资源最优分配等问题。

由于数据处理系统的核心是数据，因此数据管理系统是计算机数据处理系统中最重要的组成部分，而数据库管理系统则是一种新的数据管理软件，它提供了一种新的数据管理方法。

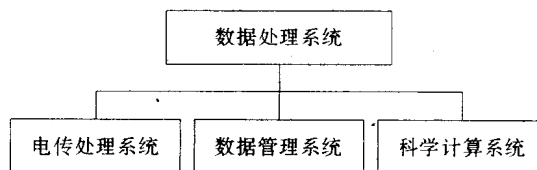


图1.1 数据处理系统

1.1.2 数据管理技术的发展过程

数据以及数据之间的关系可以从逻辑的物理两方面进行描述和组织。数据的逻辑描述是指按用户要求的数据间逻辑关系来组织和表达数据的，而物理描述则将涉及数据在计算机内的存储方式，是以存储数据的时空概念来描述数据间的关系的，前者可称为数据的逻辑结构，后者则称为数据的物理结构。

如何用计算机对数据进行组织、编目、定位、存储、检索和维护是数据管理的任务。数据管理是数据处理的中心问题。数据管理技术与数据处理方式有着密切的关系，并且直接影响着数据处理的效率。在数据处理的手工处理阶段和机械处理阶段，数据管理技术也是手工的。但是，在计算机进入数据处理领域之后，原来的那套手工管理方式就不能适应计算机自动处理数据的需要了。而且，在跨入信息社会的今天，需要管理的数据越来越多，如果仍然沿用原来那套方式，不仅需要耗费大量的人力，而且很难使这些数据发挥应有的作用。为此，许多计算机专家，特别是软件工作者，对数据管理技术做了大量的研究工作，使数据管理技术经历了一个由简单到复杂的不断完善的发展过程。

1.1.2.1 人工管理阶段（50年代中期以前）

在这一阶段中，计算机除了硬设备之外没有任何软件可供使用，放在用户面前的只有硬件的裸机。用户在这种裸机上进行数据处理，除了编制自己的课题程序外，还必须考虑数据的逻辑定义和组织，以及数据在计算机存储设备内的物理存储方式和地址。数据的引用是按物理地址进行的。这时，表示处理流程的程序和作为处理对象

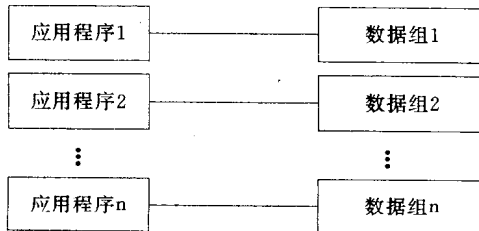


图1.2 人工管理阶段数据和程序之间的关系

的数据，相互结合成一个整体。数据的管理基本上是手工的、分散的，计算机无法在数据管理中发挥应有的作用。因此，严重地影响了计算机的使用效率。这一时期数据和程序之间的关系可以由图 1.2 表示。

1.1.2.2 文件系统阶段 (50年代后期到60年代中期)

把数据组织成文件的形式,是计算机数据管理的重大进展,也是计算机在数据管理中直接发挥作用的开始。众所周知,无组织的一大堆数据是毫无意义的,也是没有实用价值的。该软件将数据按照一定的规则组织起来,成为一个有效的数据组合体,赋予它一个名字,称为文件名或文件标识,用户通过文件名来访问该文件。这时用户就不必过多地考虑物理细节了,而是由文件系统充当应用程序和数据文件之间的接口。这一时期应用程序和数据的关系如图1.3所示。

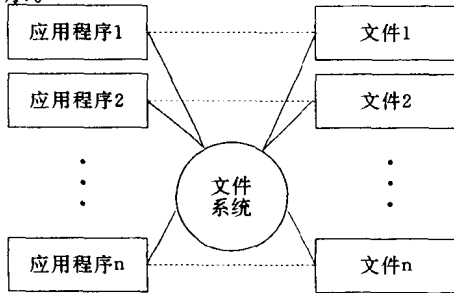


图1.3 文件系统阶段数据和程序之间的关系

虽然这一阶段较人工管理阶段有了很大的改进,但仍存在很大的弱点,即文件本身还是基本上对应于一个或几个应用程序,或者说数据还是面向应用的。从逻辑结构上讲,还与前一阶段没有什么变化,尽管程序不必要直接与文件打交道而有软件做为接口,但它仍然是一个不具有弹性的无结构的信息集合,存在着冗余度大、空间浪费及文件不易扩充等缺点,因此反映不了现实

世界事物之间的广泛的内在联系。

1.1.2.3 数据库系统阶段 (60年代后期开始)

这一时期,磁盘技术已经成熟,并作为主要外存而广泛使用。计算机硬件的价格大幅度下降,可靠性增强,为数据管理技术的发展奠定了物质基础。另外,计算机用于管理,其规模更加庞大,从而使数据量急剧增加。对数据进行集中控制,充分提供数据共享的要求日益迫切。

在这样的背景下产生了一种新的数据管理技术即数据库技术。数据库技术克服了以前所有管理方式的缺点,试图提供一种完善的、更高级的数据管理方式。它的基本思想是解决多用户数据共享的问题,实现对数据的集中统一管理,具有较高的数据独立性,并为数据提供各种保护措施。

这一时期,数据库管理软件作为用户与数据的接口、程序和数据库的关系如图1.4所示。

1.1.3 什么是数据库

数据库系统是计算机软件的一个重要分支,是近20年来发展起来的一门新兴学科。它和计算机网络、人工智能被称为当今计算机技术界的三大热门技术。目前,虽有人对数据库的设计原则和方法进行总结和探讨,使之通用化、标准化和理论化。但总的说来,它的概念、原理和方法还处于从工程实践向理论过渡的阶段,它的概念、原理和方法还在继续变化和发展。另外,数据库是一个很复杂的系统,涉及面很广,难以用简练的语言准确地概括其全部特征。因此,给它下一个确切的定义是困难的,所以我们先从简单分析入手,逐步认识什么是数据库。

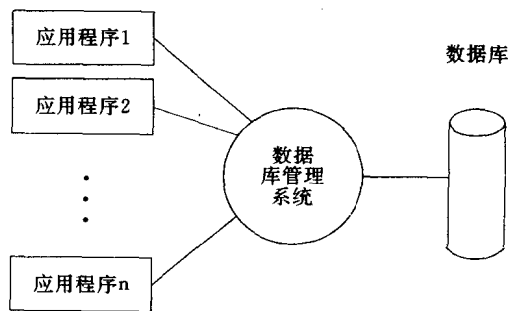


图1.4 数据库阶段数据和程序之间的关系

库,这个名词在日常生活中经常接触到。我们知道各种各样的库。例如:仓库、书库、金

库、血库等。稍微留心一下，我们就会发现，这些库都具有这样几个特点：

1. 可以存放现实世界中一定数量的东西。
2. 这些东西是有条理、有组织地存放在库中的，并由专人来进行管理。

例如，书库中的书就是有次序地摆在书架上的，并由图书管理员对它们进行管理。显然，库中的东西如果不是按照一定的次序存放，并由专人管理，那么就不能有效地利用库中的空间，也不可能方便地找出所需要的东西。

顾名思义，数据库就是存储数据的“仓库”。但它和我们前面讲的库是有所不同的。首先，数据不是存放在容器或空间中，而是存放在计算机的外存储器上（如磁盘），并且是有组织地存放的。数据的管理和利用是通过计算机的数据管理软件——数据库管理系统来完成的。因此，我们讲的数据库，不单是指存有数据的计算机外存，而是指存放在外存上的数据集合，以及管理它们的计算机软件的总和。这通常称为数据库管理系统。

目前，对于什么是数据库还没有一个统一的、公认的定义。比较流行的有关数据库的文献中常见的数据库定义有如下 3 种：

1. DBTG 的定义。数据库是由一个特定的模式控制的所有记录、系和域组成的。如果有多个数据库，则每一个数据库必须有自己的模式。并假定不同数据库的内容是彼此无关的。

2. C. J. Date 的定义。它从三个方面描述数据库：①存储在磁鼓、磁盘或其它外存介质上的数据集合——这是指数据库自身；②存在以这种数据为背景而运行的若干个应用程序，可进行检索、修改、插入和删除等操作，另外可能有一些联机用户，利用远程终端与数据库相互作用；③数据库是集成的，即包括许多用户的数据，每个用户只享用其中的一小部分，且不同用户使用的部分以多种方式重叠，也就是单独的数据片能够被许多不同用户所共享。

3. J. Martin 的定义。数据库是存储在一起的相关数据的集合，这些数据去掉了有害的或不必要的冗余，为多种应用服务，数据的存储独立于使用它的程序；对数据库插入新数据，修改和检索原有数据均能按一种公用的和可控制方式进行；数据被结构化，为今后的应用研究提供基础。当某个系统中存在结构上完全分开的若干数据库时，则说该系统包含一个数据库集合。

综合上述 3 种定义，通常一个数据库系统包含下列内容：

1. 有一个结构化的相关数据的集合。在这个数据集合中没有有害的或不必要的冗余，能够为多种应用服务，它独立于应用程序而存在。这种结构化的数据集合就是数据库本身，是数据库系统的核心和管理对象。

2. 有一个负责数据库管理和维护的软件系统，称为数据库管理系统（DataBase Management System—DBMS）。它对数据库中数据的各种操作，提供一种公用的方法。它接受并完成用户程序或终端命令提出的访问数据库的各种请求，负责数据库的维护，保护数据库中的数据不受破坏。

3. 有一个（或一组）负责整个数据系统的建立、维护和协调工作的专门人员，这就是数据库管理员（Data Base Administrator—DBA）。他们负责对数据库系统的全面管理与监督，并对运行状况进行统计分析，不断改进数据库设计。

图 1.5 表示了数据库系统的用户、数据库管理系统、数据库管理员和数据库之间的相互关系。

另外，还应该指出，数据库保存的是属于企业和事业部门，团体和个人的有关数据。例

如，银行帐目数据、学校的教学管理数据、政府部门的管理、统计和计划数据等。特别应该

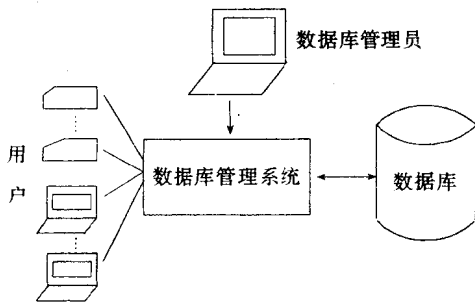


图1.5 数据库系统

强调的是，设计数据库保存这些数据的目的，不仅仅是为了扩展人们的记忆，而主要是帮助人们去控制与之相关的事物。因此，数据库往往不是孤立存在的，通常是一个更大的信息系统的一部分，两者的关系如图 1.6 所示。该图表示，人们从观测客观事物中得到大量信息，对这些信息进行记录、整理和归类（总称规范），然后将规范信息数据化，送入数据库中保存起来，其中一部分信息可能直接送入控制决策机构。另一方面，控制决策机构（它既可由一些人组成，又可是

一个自动控制系统）向数据库发出询问，并利用数据库响应后提供的信息（以及其它有关信息）作出决策，再行控制客观事物。例如，就一个企业的产品销售信息系统而言，客观事物可以是产品的销售情况和用户对该产品的反映情况等，控制决策机构指企业领导机关，策略是指提高产品质量扩大产品销售的措施。当数据库的管理人员把观测客观事物得到的信息规范化、数据化送入数据库后，企业领导机关可以通过询问数据库得到产品的销售情况和用户对产品的反映。根据这些情况，企业领导就可以提出改进产品销售，提高产品质量的措施，并在今后的生产经营中付诸实行（控制客观事物）。

数据库系统作为软件的一个分支，与其它基础软件和系统软件有密切的关系。它几乎涉及到软件的所有知识，是许多重要软件技术的综合应用。例如，数据结构、操作系统、编译技术、程序设计等知识在数据库都会用到。所以数据库系统是一门综合性的软件技术，是一门很有意义很有趣味的学科。要更好的研究掌握它，必须了解软件的各个方面，以便更加理解和认识到这些知识的内在联系，并在一种观念上将它们统一起来。

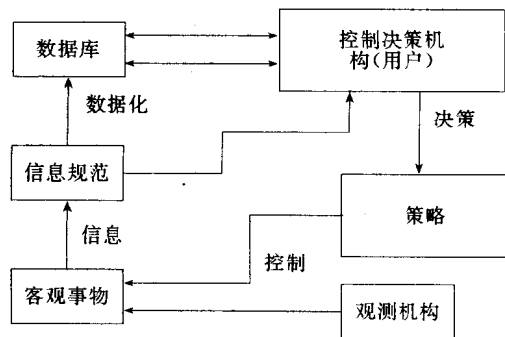


图1.6 信息控制系统

1.1.4 数据库系统的主要特征

数据库技术之所以能够在近 20 年内如此快速的发展，受到计算机科学界普遍的重视，成为引人注目的一门新兴学科，是因为它具有独特的特征。

1.1.4.1 实现数据共享

实现数据共享是促成数据库技术发展的一个重要原因，也是数据库技术先进性的一个重要体现。数据的共享性包括 3 个方面：

1. 当前的所有用户可以同时存取数据库的数据，这是十分清楚的。因为数据库的研制正是以当前用户为现实基础的。

2. 未来用户与当前用户可以同时存取数据库中的数据。也就是说，数据库不仅可以为当前已知的多个用户服务，也可为将要使用该数据库的任何新用户服务，而且可以同时为多个新用户服务。

3. 用户可以通过多种程序设计语言或命令语言与数据库接口。如 COBOL、PL/1、FORTRAN, 以及终端命令和数据库管理系统提供的数据库操纵语言等等, 从而可以用多种语言使用数据库。

1.1.4.2 减少数据的冗余度

在数据库技术产生以前的文件系统中, 用户根据各自的应用建立文件, 即使不同的应用程序需要使用的数据中也有许多是相同的, 也必须建立各自的文件, 其原因就在于文件系统中数据不能充分共享(文件系统中只提供文件一级的共享, 不提供记录级的共享)。因此可能使存储的数据存在着大量的重复, 这种现象称为数据的冗余。

下面我们通过一个例子来说明这个问题。

假设现在要管理的对象是一个学校, 它由人事子系统、业务子系统、总务子系统组成, 各子系统根据自己的特点及需要选择相应的数据格式, 存放自己感兴趣的数据。它由这样 3 种记录组成, 如图 1.7 所示。

人事记录	部 门	职 工 号	姓 名	性 别	年 龄	职 务	工 资	家 庭 出 身	本 人 成 份	政 治 历 史	社 会 关 系	
业务记录	部 门	职 工 号	姓 名	性 别	年 龄	职 务	工 资	学 历	专 业	外 语 程 度	教 学 历 史	科 研 历 史
总务记录	部 门	职 工 号	姓 名	性 别	年 龄	职 务	工 资	参 加 工 作 时 间	高 教 教 龄	家 庭 人 口	住 房 面 积	

图 1.7 学校的 3 个记录

在文件系统阶段, 由这 3 个记录的集合组成 3 个文件: 人事文件、业务文件、总务文件。这 3 个文件对应于 3 个不同的应用系统, 它们之间是互不联系的。在这 3 个文件中, 有许多项目如部门、职工号、姓名、性别、年龄、职务、工资是相同的。但是在文件系统阶段中, 数据与应用系统是一一对应的, 所以必须分别存储, 这样势必造成很多的数据冗余, 浪费存储空间。而这种冗余还会带来其它问题, 如在更新操作时, 必须要进行多次更新操作, 以更新某些冗余的副本, 增加不必要的机器时间。另外, 由于多个副本的存在, 系统容易出现不一致的信息。而在数据库中, 是从整体观点来组织数据的, 其数据不是面向个别的应用, 而是为多种应用所共享的, 从而避免了不必要的冗余。例如上面的例子在数据库中 3 个文件就不再是互不相干的了, 而是将这 3 个文件视为统一的整体。也就是说, 把这 3 个文件进行分解, 重新组合, 去掉重复部分, 构成一个整体。图 1.8 给出经过重新组合形成的数据结构。

从图中可以看出, 我们设计了各部门都要用的职工记录, 并带上了人事记录、业务记录和总务记录等, 这样就把整个组织的数据结构化了。这就要求在描述数据时, 不仅要描述数据本身, 还要描述数据之间的联系, 这种联系往往是通过存取路径来表现的。在这种情况下就大大地减少了数据的冗余度, 节省空间, 提高存取速度, 避免不相容。此外, 由于从整体