

高等学校培养应用型人才教材——计算机系列



# 数据库系统 原理与设计

陈雁 主编

周如意 滕刚 王文 李武韬 编著



中国电力出版社

[www.infopower.com.cn](http://www.infopower.com.cn)

高等学校培养应用型人才教材——计算机系列

# 数据库系统 原理与设计

陈雁 主编

周如意 滕刚 王文 李武韬 编著

中国电力出版社

M595/08

## 内 容 提 要

本书为高等学校培养应用型人才教材之一，由有丰富教学经验的一线教师编写。本书从技术的先进性、实用性出发，介绍了数据库系统的产生与发展、数据库系统的特点、数据库系统的基本概念、关系数据库、关系数据库方法、数据库操作、关系数据库标准语言 SQL、关系规范化理论、SQL Server 2000 等内容，还给出一个综合实例。针对培养应用型人才教学的特点，本书还有大量实验和实训内容。

本书可作为高等院校计算机相关专业的数据库课程教材，也可供从事计算机软件工作的科技人员和工程技术人员如管理信息系统、电子商务网站开发人员等参阅。

### 图书在版编目 (CIP) 数据

数据库系统原理与设计 / 陈雁编著. —北京: 中国电力出版社, 2003

高等学校培养应用型人才教材——计算机系列

ISBN 7-5083-1523-5

I. 数... II. 陈... 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 (2003) 第 115736 号

中国电力出版社出版、发行  
(北京三里河路 6 号 100044 <http://www.infopower.com.cn>)

汇鑫印务有限公司印刷

各地新华书店经售

\*

2004 年 2 月第一版 2004 年 2 月北京第一次印刷

787 毫米×1092 毫米 16 开本 15.5 印张 346 千字

册数 0001—5000 册 定价 23.00 元

版 权 所 有 翻 印 必 究

(本书如有印装质量问题, 我社发行部负责退换)

# 高等学校培养应用型人才教材——计算机系列

## 编 委 会

主任委员：

宗 健 常明华

副主任委员：

顾元刚 陈 雁 杨翠南 林全新 华容茂 曹泰斌

魏国英 邵晓根 庄燕滨 邓 凯 吴国经 常晋义

许秀林 谢志荣 张家超 陶 洪 龚兰芳 刘广峰

丁 雁 方 岩 王一曙

委 员：（以姓氏笔画为序）

丁志云 及秀琴 石振国 李 翊 吕 勇 朱宇光

任中林 刘红玲 刘 江 刘胤杰 许卫林 杨劲松

杨家树 杨伟国 郑成增 张春龙 闵 敏 易顺明

周维武 周 巍 胡顺增 袁太生 高佳琴 唐学忠

徐煜明 曹中心 曾 海 颜友钧

# 序 言

进入 21 世纪,世界高等教育已从精英教育走向了大众教育。我国也适应这一潮流,将高等教育逐步推向大众化。培养应用型人才已成为国家培养国际人才的重要组成部分,且得到了社会各界的广泛支持。于是一大批有规模、有实力、规范化、以培养应用型人才为己任的高等学校得到了长足发展。这类高校办学的一个显著的特点是按照新时代需求和当地的需要来培养学生,他们重视产学研相结合,并紧密地结合当地经济状况,把为当地培养应用型人才作为学校办学的主攻方向。

这类学校的教学特点是:在教授“理论与技术”时,更注重技术方法的教学。在教授“理论与实践”时,更注重理论指导下的可操作性,更注意实际问题的解决。因此,这些学生善于解决生产中的实际问题,受到地方企事业单位的普遍欢迎。

为满足这类高校的教学要求,达到培养应用型人才的目的,根据教育部有关重点建设项目的要求和相关教学大纲,我们组织了多年在这类高校中从教,并具有丰富工程经验的资深教授、高级工程师、教师来编写这套教材。

在这套教材的编写中,我们提倡“实用、适用、先进”的编写原则和“通俗、精练、可操作”的编写风格,以解决多年来在教材中存在的过深、过高且偏离实际的问题。

**实用**——本套教材重点讲述本行业中最广泛应用的知识、方法和技能。使学生学习后能胜任岗位工作,切实符合当地经济建设的需要和社会需要。

**适用**——本套教材是以工程技术为主的教材,所以它适用于培养应用型人才的所有高校(包括本科、专科、技术学院、高职等),既符合此类学生的培养目标,又便于教师因材施教。

**先进**——本套教材所选的内容是当今的新技术、新方法。使学生在掌握经典的技术和方法之后,可用教材中的新技术、新方法去解决工程中的技术难题,为学生毕业后直接进入生产第一线打下坚实的基础。

**通俗**——本套教材语言流畅、深入浅出、容易读懂。尽量避开艰深的理论和长篇的数学推导,尽量以实例来说明问题,在应用实例中掌握理论,使学生轻松掌握所学知识技能,达到事半功倍的效果。

**精练**——本套教材选材精练、详细而不冗长、简略得当,对泛泛而谈的内容将一带而过,对学生必须掌握的新技术、新方法详细讲,讲透、讲到位,为教师创造良好的教学空间和结合当地情况调整教学内容的余地。

**可操作**——本套教材所有的实例均是容易操作的,且是有实际意义的案例。把这些案例连接起来,就是一个应用工程的实例。通过举一反三的应用,使学生能够在更高层次上创造性地应用教材中的新思想、新技术、新方法去解决问题。

本套教材面向培养应用型人才的高等学校,同时也可作为社会培训高级技术人才的教材和需要加深某些方面知识技能的人员的自学教材。

编委会

# 前 言

数据库技术自从 20 世纪 60 年代产生至今，只有短短的几十年，却得到了十分广泛的应用，在社会经济活动中担任着十分重要的角色，发挥着越来越大的作用。随着网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等的发展及其相互渗透、相互结合，数据库技术也得到了迅速的发展，已经成为计算机科学的一个重要组成部分，并成为信息领域的支撑技术，也必然地成了高校计算机专业的必修课程。

本教材针对应用型高校学生的特点，结合大量的实例和作者的教学体会，对数据库基本原理、基本概念、应用技术和当前的研究发展状况进行了全面的阐述。在编撰时，我们力求由浅入深、循序渐进，既注重体系的完整性，又特别注意理论与实践的结合，注重学生实践能力的培养。本教材以 Microsoft SQL Server 为平台，介绍了数据库的建立、修改、查询等操作，介绍了 SQL Server 数据库的保护技术，还以一个图书管理实例为主线，贯穿全书始末，把各种数据库技术的知识要点串成一个逻辑严密的整体，并以第 7 章数据库系统设计实例为总结，对该实例进行规范、完善，给学生一个完整概念和示例。本书在每章后还配有习题，并有专门章节对课程的实验、实训给出要求和指导。这些，也许可以算作本教材的特色。

本教材可作为应用型本科院校、高等专科学校、高等职业学校及成人高等学校计算机类专业或信息类相关专业的教材，可作为非计算机专业学生的选修课或辅修课的教材，也可作为计算机应用人员和工程技术人员的自学参考书。

本教材由陈雁主编，第 1 章由李武韬撰写；第 2 章、第 9 章由陈雁撰写；第 3 章、第 8 章由滕刚撰写；第 4 章、第 6 章由周如意撰写；第 5 章、第 7 章由王文撰写。全书由陈雁统稿。

苏州职业大学冯矢勇高级工程师在百忙中抽出时间担任主审，认真细致地审阅了全书，提出了许多宝贵的修改意见和建议。编写过程中，还得到了中国电力出版社的帮助与指导，得到了作者所在单位的领导和同行的支持，特别是得到陈媛、范广慧两位青年教师的帮助，在此一并表示衷心的感谢。

囿于我们的知识水平，本教材难免存在许多不足，敬盼使用者不吝赐教，以便在再版时加以完善。

编 者

# 目 录

## 序 言 前 言

第 1 章 数据库的基本概念 .....	1
1.1 数据、信息与信息处理 .....	1
1.2 数据管理技术的发展 .....	3
1.3 数据库及其特点 .....	6
1.4 数据模型 .....	9
1.5 数据库系统的结构 .....	17
1.6 数据库系统的组成 .....	19
1.7 数据库管理系统 .....	22
习题 .....	25
第 2 章 关系数据库 .....	26
2.1 关系数据结构 .....	26
2.2 关系操作 .....	30
2.3 关系的完整性 .....	31
2.4 关系代数 .....	33
2.5 关系演算 .....	45
2.6 关系代数的优化 .....	45
习题 .....	47
第 3 章 关系数据库标准语言——SQL .....	49
3.1 SQL 语言概述 .....	49
3.2 SQL 的数据定义 .....	52
3.3 SQL 的数据查询 .....	60
3.4 SQL 的数据更新 .....	90
3.5 视图 .....	96
3.6 SQL 的数据控制 .....	100
习题 .....	103
第 4 章 关系数据库的规范化理论 .....	106
4.1 关系规范化问题的提出 .....	106
4.2 函数依赖 .....	108
4.3 关系模式的范式 .....	113
4.4 关系模式的规范化过程 .....	117



4.5	关系模式分解	119
	习题	123
<b>第5章</b>	<b>数据库设计</b>	<b>125</b>
5.1	概述	125
5.2	需求分析	128
5.3	概念结构设计	134
5.4	逻辑结构设计	142
5.5	物理结构设计	145
5.6	数据库的实施	149
5.7	数据库的运行和维护	150
	习题	152
<b>第6章</b>	<b>数据库保护</b>	<b>153</b>
6.1	数据库的安全性保护	153
6.2	数据库的完整性保护	160
6.3	数据库的并发控制	170
6.4	数据库的恢复	183
	习题	190
<b>第7章</b>	<b>数据库系统设计实例</b>	<b>193</b>
7.1	概述	193
7.2	实例分析	200
<b>第8章</b>	<b>数据库系统的研究与发展</b>	<b>219</b>
8.1	传统数据库系统的局限性	219
8.2	新一代数据库技术的特点	220
8.3	新一代数据库系统	221
<b>第9章</b>	<b>实验、实训大纲</b>	<b>230</b>
9.1	实验计划和实施要求	230
9.2	前期准备阶段的实验	231
9.3	基本操作阶段的实验	232
9.4	技术提高阶段的实验安排	236
9.5	数据库应用实训大纲	237



# 第 1 章 数据库的基本概念

信息在现代社会和国民经济发展中所起的作用越来越大，信息资源的开发和利用水平已成为衡量一个国家综合国力的重要标志之一。在计算机的三大主要应用领域：科学计算、数据处理和过程控制中，数据处理是一个重要方面。数据库技术就是作为数据处理中的一门技术而发展起来的，是计算机软件学科的一个重要的分支。数据库技术自 20 世纪 60 年代诞生以来，为计算机收集、存储、加工和利用数据提供了全面的支持，并起到了十分关键的作用，它是当今计算机信息系统的核心技术。

随着 Internet 的发展和经济全球化、市场一体化进程的不断加快，对信息资源的利用已成为各企业或组织在激烈的竞争环境下生存和发展的关键，因此，如何建立一个行之有效的信息系统，保存和充分利用企业或组织内部信息（如企业在生产、销售、财务、库存、人事管理等各种业务活动中所发生的各种数据）和外部信息，以帮助企业做好经营管理工作、增强竞争力，是每个企业或组织都在考虑的问题，而信息系统的核心和基础就是数据库。世界上许多著名企业和组织的实践，都说明了利用数据库技术进行信息管理和处理对企业有着十分重要的意义。

## 1.1 数据、信息与信息处理

信息处理是计算机应用中用得最广泛的领域，而数据是计算机实际处理的对象，数据经处理后会转换成更能反映事物本质的信息。在介绍数据库之前，我们先对数据和信息两个基本概念做一说明。

### 1.1.1 数据

对数据这个概念，人们往往理解为数值。其实，数值数据只是数据的一个子集。凡是需要数量表示的事物，都要用到数据，这是我们熟悉的。但还有大量的事物，不仅需要数量描述，还要有“陈述”表达。比如关于天气的描述，除温度、风力等用数值描述外，阴、晴、雨等则要用文字陈述。简而言之，数据是事实的反映和记录。这里的数据是广义的数据，包括数字、字符串、报表和图形等。数据可称为客观事物（客体）属性的记录。

### 1.1.2 信息

现在人类已进入了信息时代，信息概念变得越来越复杂，对信息这个词很难给出精确、全面的定义。有人将信息解释为人得到的知识，有人称信息是人与外界相交换的内容，有人则把通过口头、通信装置或书面传达的消息、情报都称作信息。于是人们常常把数据、资料、知识、消息等统称为信息。信息在自然界、社会中以及人体自身都广泛存在着。人类进行的每一次社会实践、生产实践和科学实验都在接触信息、获得信息、处理信息和利

用信息。根据 ISO (International Standards Organization) 和 ANSI, 可将信息定义为“人借助于在数据的表示中所用的已知约定来赋予数据的含义”。可以从不同角度来说明信息, 我们主要是从信息处理角度来谈论信息。信息有以下特征:

(1) 信息具有表征性。信息是表征事物状态、属性和运动特性的一种普遍形式。比如对一个企业生产运营活动的各环节及整个过程的表征和描述。

(2) 信息具有可用性、可处理性和可替代性。人的感觉是信息的获取, 而人的思维则是一种信息处理。信息有描述型、加工型、预测型和控制决策型。

(3) 信息有可传递性、可存储性和共享性。

(4) 信息与其表现符号的不可分离性和信息的可转换性。

(5) 信息是由人主观重新定义的数据, 所以是与人的主观愿望相联系的。但信息的事实性是主要的。因为信息源于物质和能量, 它不可能脱离物质而存在, 信息的传递需要物质载体, 信息的获取和传递要消耗能量。如信息可以通过报纸、书刊、电台、电视和计算机网络等载体进行传递。

### 1.1.3 信息与管理、决策

在商业、企业管理活动中, 有大量的数据与信息要表示、传递和处理。对一个企业来说, 没有信息就无法管理, 信息不够、不准、不及时就难以决策。这里简单说明一下企业中信息与管理、决策的关系。

企业的活动分为生产活动和管理活动。在生产活动中, 流动的是物, 从输入、转换到输出, 是一股物流, 这是企业生产经营活动的主体流程。当然参与这个流程的还有人、设备、资金、能源、交通工具等因素。企业的管理活动是指用组织、计划、领导、控制和协调等各种基本行动, 来有效地利用人力、材料、资金、设备和方法等各种资源, 发挥最高效率, 以实现一个组织所预定的目标和任务。有的学者把管理过程划分为: 计划—组织—控制三个阶段, 通过这样的管理过程达到管理目标。

在管理活动中流动的是信息, 从输入、转换到输出是一股信息流。信息流是伴随物流产生的, 对生产活动和物流起着主导作用, 不充分发挥信息流的主导作用, 会导致物流的混乱。从控制论观点看, 管理过程是信息的收集、传递、加工、判断和决策的过程。在管理过程的三个阶段都离不开信息, 所以要搞好管理工作, 必须重视研究如何搞好信息管理。信息流能根据生产的规律和环境条件, 按照管理者的意志来规划和调节物流的数量、方向、速度和目标, 使其按照一定的目的和规则运动。信息被中断或传递不及时、不准确, 都会给生产带来损失。所以对企业来说, 信息最重要的资源是企业经营决策的前提, 也是提高企业管理水平的基础。举例来说, 有一家网上书店, 希望给其会员提供优质的客户服务, 该书店保留了每个会员及其网上买书的信息, 由此可以推断不同会员的喜好, 并有针对性地给会员提供在线新书信息, 以提高网上图书的销售量。

### 1.1.4 数据处理和数据管理

数据处理包括对各种数据进行的收集、存储、加工与传播等一系列活动。其中数据的收集是指在数据的发生处将它们读入到计算机中。例如, 在各种预售火车票或飞机票等的

订票点都配备了计算机终端，并由售票员将售票数据输入到计算机中；而在各类百货公司的收银台上，收银员也往往会将商品的销售数据通过计算机终端输入到公司的计算机中；又如大部分的超市一般都配备了条形码阅读器，以便于快速地将商品的销售数据输入到计算机中，等等。数据的存储是指将收集到的数据经过整理后用计算机的存储介质保存起来，以备今后使用；数据的加工是指从某些已知的数据出发，推导加工出一些新的数据的过程；数据的传播就是利用计算机通信设备将数据传送给需要的人员。

数据处理也称为信息处理，因为当把客观事物表示成数据后，这些数据便被赋予了特定的含义，而对这些数据进行加工处理后又可以形成新的数据，这些新的数据又表示了新的信息，从而为人们提供了不必直接观察和度量事物就可以获得有关信息的手段。在数据处理（或信息处理）的各种活动中，数据加工是最重要的数据处理活动。这些对数据进行加工处理的活动，有的只涉及到简单的数学运算或逻辑运算，如根据每一产品的价格计算所有产品的平均价格，又如将商品单价与销售数量相乘得到销售金额，等等。这些活动通常在企业经营管理的事务处理活动过程中就能完成，并完全可以在预先设计好的计算机程序的指挥下自动实现，这种比较简单的数据加工处理活动称为联机事务处理（Online Transaction Processing，简称 OLTP）。另外一些数据加工活动涉及的计算比较复杂，因而必须在人的参与和支配下才能完成，这种加工处理活动称为联机分析处理（Online Analytical Processing，简称 OLAP）。

随着数据处理量的日渐增大，需要对数据进行分类、组织、编码、存储、检索和维护，这样的过程称为数据管理，它是数据处理业务的基本环节。数据管理技术的优劣，将直接影响到数据处理的效率。

## 1.2 数据管理技术的发展

数据管理技术与数据处理方式有着密切联系，且直接影响着数据处理的效率。在数据处理的手工阶段和机械阶段其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后，原来的那套手工管理方式已不能适应计算机自动处理数据的需要了。而且，随着需要管理的数据量的急速增长，如果仍然沿用原来那套手工管理方式，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。计算机数据管理技术的发展经历了以下三个阶段。

### 1. 人工管理阶段（20世纪50年代中期以前）

这一阶段，计算机主要应用于科学计算。在硬件方面，外存没有直接存取存储设备，只有纸带、卡片和磁带等；在软件方面，只有汇编语言，没有操作系统和数据管理方面的软件，数据处理方式是批处理。人工管理阶段数据管理的主要特点是：

（1）数据不保存。由于计算机主要用于科学计算，所以数据是不需要保存的。当用户需要计算某一课题时，就临时将有关数据输入内存，计算完毕后输出运算结果，并释放相应内存空间。

（2）数据由应用程序进行管理，数据与程序不具独立性。由于没有专门的软件进行数据的管理，数据是由应用程序自己管理的。程序员不仅要在程序中设计数据的逻辑结构，

还要规定包括数据的存储结构、存取方法和输入输出方式在内的物理结构。而当数据的物理结构或逻辑结构改变时，就必须修改相应的程序。

(3) 数据面向应用，不能共享。每个应用程序都只使用自己定义的一组数据，即使有些应用程序使用了相同的一部分数据，也都必须在应用程序中各自定义，如图 1-1 所示。

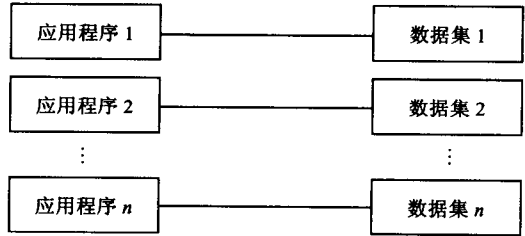


图 1-1 数据的人工管理

## 2. 文件系统阶段（20 世纪 50 年代后期~20 世纪 60 年代中期）

这一阶段，计算机不仅用于科学计算，也用于数据管理。在硬件方面，已经有了直接存取存储设备，如磁盘、磁鼓等；软件方面，出现了高级语言和操作系统，并在操作系统中包含了专门的数据管理软件——文件系统，数据处理方式有批处理和联机处理。文件系统阶段数据管理的主要优点是：

(1) 数据长期保存在外存的数据文件中，可以进行查询、插入、删除和修改操作。

(2) 数据文件由文件系统进行管理，程序员在程序中只需用文件名就可与数据打交道，因此不必关心数据的物理存放位置，负担大大减轻了。

(3) 程序和数据间有一定的独立性。数据文件由文件系统进行管理，程序和数据间由软件提供的存取方法进行转换，数据存储发生变化对程序的运行影响不大。

与人工管理阶段相比，文件系统阶段对数据的管理有了很大的进步，但一些根本性问题仍没有彻底解决，其主要缺点是：

(1) 数据冗余度大。由于一个数据文件只为某个特定的应用程序服务，不同的应用程序使用相互独立的数据文件，因此，相同的数据同时出现在几个数据文件中几乎是不可避免的。例如，职工的姓名、所在部门、工资等数据会同时出现在人事档案和工资文件中，而在一所大学中，教师的姓名，所在系等数据则可能同时出现在人事档案、教务档案、科研档案和工资等文件中。这种数据大量重复的现象，称为冗余，它降低了存储空间的有效利用率。

(2) 程序与数据的独立性较差。虽然有了文件系统对文件进行管理，但文件的逻辑结构仍然是在程序中定义的，所以，当文件的逻辑结构改变时，仍要修改相应的程序，程序与数据的独立性仍较差，如图 1-2 所示。

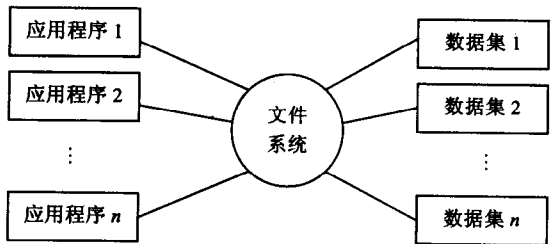


图 1-2 数据的文件系统

(3) 易造成数据的不一致性。由于同一数据重复存储在由不同的应用程序（用户）使用和维护的文件中，不能保证其更新的同

时性和正确性，以致造成同一数据在不同的文件中有不同的值。比如，某个教师因为某项科研工作有重大贡献而被晋升工资一级，但人事部门忘了通知财务部门，从而造成该教师的工资在人事档案和财务科的工资文件中具有不同的值。

这种缺点在规模较大的应用系统中尤其明显。美国在 20 世纪 60 年代执行阿波罗登月计划时委托 Rockwell 公司研制了一个基于磁带文件的零部件生产计划管理系统，共用了

18 盘磁带，其中 60% 的数据是冗余数据，只能以批处理方式工作，维护十分困难。该系统的状况曾一度成为实现阿波罗计划的严重障碍。

### 3. 数据库系统阶段（20 世纪 60 年代末至今）

数据库（DataBase）是计算机软件的一个重要分支，是在 20 世纪 60 年代后期发展起来的数据管理新技术。从 1968 年 9 月美国 IBM 公司发表其研制成功的信息管理系统（Information Management System, IMS）的第一个版本，1969 年 10 月美国的 CODASYL（Conference Data System Language 数据库系统语言协会）组织发表第一个 DBTG（DataBase Task Group）报告和美国 IBM 公司的研究员 E. F. Codd 于 1970 年开始发表一系列关系数据库（Relational DataBase）的论文以来，在提出关系数据库的方法直到今天的几十年的时间里，数据库技术有了惊人的发展，几乎成为各种计算机应用系统的核心部分。

这一阶段，计算机的应用更加广泛，管理规模日益扩大，数据量不断增加，而且多个应用要求能共享数据集合，这就给数据管理技术提出了更高的要求。与此同时，磁盘技术突飞猛进，出现了数百兆字节容量和快速存取的磁盘，而且价格也不高。但是由于文件系统的缺点，软件的编制与维护却越来越困难，成本也不断上涨，价格不断攀升。因而需要有新的数据管理技术来减轻程序员进行软件编制和维护的负担，降低软件的成本，这就导致了数据库系统的产生。数据库系统阶段数据管理的主要特点是：

（1）数据存放在数据库中，并用数据模型来描述数据本身及其数据间的联系。主要的数据库模型有层次模型、网状模型、关系模型和面向对象模型，当然这些数据模型是逐步发展起来的，并不是一开始就全部都有的。

（2）数据库中的数据由数据库管理系统进行管理，数据与程序的独立性很高。在数据库系统阶段，应用程序与数据的对应关系如图 1-3 所示。应用程序不是直接与数据库打交道的，而是通过数据库管理系统（DataBase Management System，简称 DBMS）对数据库进行操作的。数据的定义、

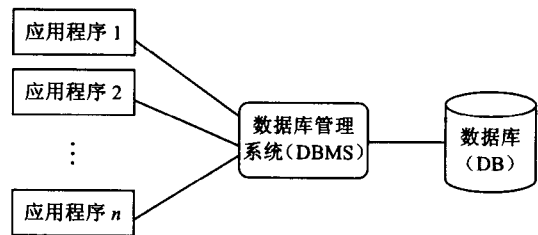


图 1-3 数据的数据库系统

查询、删除、插入、修改都由数据库管理系统负责，应用程序员不需要在程序中定义数据，当然也就不需要定义数据的逻辑结构和物理结构了。因此，当数据的逻辑结构或物理结构发生改变时，一般都不需要修改程序，数据和程序的独立性很强。

（3）数据是面向系统的，冗余度低，可以共享。这是数据库系统阶段的重大改进，数据不再面向某个应用程序而是面向整个系统，当前所有用户可以同时存取数据库中的数据。这样就减少了不必要的数据冗余，提高了数据的共享性，同时也节省了存储空间，避免了数据之间的不相容性与不一致性。

（4）数据库管理系统提供了以下几个方面的数据控制功能。

**数据安全性控制：**保护数据，防止对数据库的非法操作所引起的数据的丢失、泄露和破坏。

**数据完整性控制：**保证数据库中的数据永远是正确的、有效的和相容的。

并发控制：避免因多个用户并发进程同时存取、修改数据库时所引起的相互干扰，保证数据的正确性。

数据库的恢复：当数据库中的数据由于种种原因（如系统故障、介质故障和计算机病毒等）而变得不正确，或部分甚至全部丢失时，DBMS 有能力将数据库恢复到最近某时刻的一个正确状态。

日益增多的数据管理需求是促使数据库技术产生的客观需要，而计算机工业的飞速发展，计算机内外存容量的日益扩大，价格大幅度下降，可靠性越来越高，则是实现数据库的物质基础。

## 1.3 数据库及其特点

### 1.3.1 什么是数据库

所谓数据库就是存放在计算机中的、以一种合理的方法组织起来的、与公司或组织的业务活动和组织结构相对应的各种相关数据的集合，该集合中的数据可以为公司的各级经过授权的人员所共享。

该定义具有以下三种含义：

第一，数据库是存在于计算机中的、与公司或组织的业务活动和组织结构相对应的各种相关数据的一个仓库。

第二，存放在数据库中的数据是按一定的方式组织起来的，而不是杂乱无章地存放的。

第三，数据库是一个共享的信息资源，它可以被企业或组织中的多个经过授权的人员使用。

为便于读者进一步理解数据库，我们可以把数据库比喻成一个图书馆：

- 图书馆是用来存放图书的，而数据库是用来存放数据的；
- 图书馆中的书是存放在许多房间中的，而数据库中的数据是存放在计算机文件中的；
- 图书馆的图书是按照一定的规则存放在书架上的（例如按图书类别存放），数据库中的数据也是按照一定的方式组织起来的；
- 图书馆中的书可以被不同的读者借阅，数据库中的数据也可以被不同的用户使用。

可见，数据库与图书馆有很多相似之处，只是图书馆保管的对象是书，而数据库的对象是数据。当然，图书馆与数据库毕竟是不同的。例如，图书馆中的某一本书在某一时刻只能借给一个读者，其他读者想借阅的话必须等到前面的读者归还了该书以后才可以，而数据库中的某一数据却可以同时被多个用户查看。又如，图书馆中可以增加新书，也可以报废一些旧书，但却不可以修改某一本书的内容，而数据库中不仅可以增、删数据，还可以修改数据。总而言之，数据库是机械化的、可共享的、形式化定义的和集中控制的数据集合。对数据库的定义有不同说法，本书给出的是通俗说明。数据库本身不是孤立存在的，而是与其他部分构成数据库系统。在实际应用中，人们面对的是数据库系统。

### 1.3.2 数据库的主要特点

现在,从很小的微型机到很大的大型计算机都有数据库系统可以使用。当然,一个系统所提供的设施要由作为其基础的计算机的规模和能力来决定。一般而言,大型机上的系统是多用户系统,系统提供的设施非常完整,功能也极强;微型机上的系统大多是单用户系统,这种系统提供的设施往往是大型多用户系统的一个子集,即使设施相同,也没有大型多用户系统那么强的功能。

数据库系统已经成为各种计算机应用系统的核心部分。之所以如此,是因为数据库有着许多独特的特点,主要有以下几点。

#### 1. 数据被集成化

所谓“集成”是指若干不同性质数据文件的统一化,以完全地或部分地消除这些文件中的冗余。例如,一个数据库可能同时包含学生成绩文件(学号、姓名、英语成绩、数学成绩等)和档案文件(学号、姓名、出生年月、籍贯等)。在这两个文件中,学号、姓名这两项数据是重复的,因而存在冗余数据,在构造数据库时,应该尽可能减少这种冗余。今后我们将会看到,所谓集成化就是按照一定的数据模型来组织和存放数据,因此又称数据的结构化。这种结构化了的数据能反映数据之间的自然联系。而传统的文件是无模型的,结构非常简单,因此,不能反映数据之间的自然联系。所以,数据的集成化或结构化是数据库和文件之间的一个重要的本质差别,是实现数据的集中控制和减少数据冗余的前提和保证。

由于数据库是从一个组织的全部应用来全盘地考虑并集成数据的结构,所以数据库中的数据不再是面向个别应用,而是面向系统的。各个不同的应用系统所需的数据只是整体模型的一个子集。

#### 2. 具有数据独立性

所谓数据独立性就是数据与应用程序之间不存在相互依赖关系,也就是数据的逻辑结构、存储结构和存取方法等不因应用程序的修改而修改,反之亦然。这是数据库系统与文件系统之间的另一个重要区别,也是使得数据库系统结构复杂的一个重要原因。数据独立性分为物理独立性和逻辑独立性两级。

(1) 物理独立性:即数据的物理结构(或存储结构)的改变,如物理存储设备的更换、物理存储位置的变更、存取方法的改变等等,不影响数据库的逻辑结构,从而不致引起应用程序的修改。

(2) 逻辑独立性:数据库总体逻辑结构的改变,如修改数据的定义、增加新的数据类型、改变数据间的联系等等,无需修改原来的应用程序,这就是数据的逻辑独立性。遗憾的是,到目前为止,数据的逻辑独立性还没有完全彻底地实现。例如删除某个数据类型就会引起用到这些数据的应用程序的修改。

总之,数据独立性就是数据与程序之间的互不依存性。一个具有数据独立性的系统,称之为以数据为中心的系统或面向数据的系统。显然,数据库系统是以数据为中心的系统。

#### 3. 实现数据共享

数据共享是促成发展数据库技术的重要原因之一,也是数据库技术先进性的一个重要体现。数据库中数据的共享性主要体现在:



(1) 数据库中的数据可供多个应用程序用于不同的目的, 每个应用程序各有其自己的局部数据逻辑结构。即使同一数据, 在不同的应用程序中其用途也不一定相同。数据库中的数据不但可供现有的各个应用程序共享, 还可开发新的应用程序而无需附加新的数据, 实现新、老应用程序共享数据库中的数据。即使开发新的应用程序时, 需要往数据库中增加新的数据类型, 由于数据库具有数据独立性, 也无需修改原有的应用程序, 新、老应用程序仍可共享数据库中的数据, 这些应用程序甚至可以用不同的程序设计语言编写。

(2) 数据库可以为批处理用户和终端的即时用户共享。

#### 4. 数据的冗余度小

在非数据库系统中, 每个应用有它自己的文件, 从而造成了存储数据的大量重复, 称为数据冗余。在数据库系统中, 数据结构设计的不好, 同样会存在大量的重复数据。简单地说: 同一数据重复存储就是数据冗余。大量的数据冗余带来的弊病是:

(1) 浪费了存储空间;

(2) 为了更新某些冗余副本, 保持数据的一致性, 必须执行多次更新操作, 从而增加了不必要的机器时间;

(3) 由于数据的不同副本可能处于不同的更新阶段, 从而可能给出不一致的信息。数据库则是从整体观点来组织和存储数据的, 数据是集成化、结构化的, 它是为多种应用所共享的, 从而大大减少了数据冗余。

#### 5. 避免了数据的不一致性

当本应相同的数据项在不同的应用中出现不同的值时, 我们就说出现了数据的不一致性。例如, 同一职工的工资, 如果在工资单上和人事档案中具有不同的值, 就出现了不一致性。数据的不一致性主要是由数据冗余引起的。数据库在理论上消除了数据冗余, 因而可以避免数据的不一致性。即使存在某些冗余, 数据库系统也提供了对数据的各种控制和检查, 保证在更新数据时, 同时更新所有的副本, 从而保证了数据的一致性。

#### 6. 可以实施安全性保护

数据的安全性主要指数据保密及防止数据的不合法使用。只有数据库管理员(DataBase Administrator, 简称 DBA)对数据库中的数据拥有完全的操作权限, DBA 可以规定各用户的权限。数据库系统保证对数据库的存取方法是惟一的, 每当用户企图存取敏感数据时, 数据库系统就进行安全性检查。对于数据库中每片数据的各种类型的操作(检索、修改、删除等等)数据库系统都可以实施不同的安全检查。

#### 7. 有利于保证数据的完整性

数据的完整性也就是数据的正确性。数据的不一致性是失去完整性的例子。当然, 数据冗余可能引起数据的不完整性, 但是, 没有数据冗余同样可能出现不正确的数据而使数据库失去数据完整性。例如, 在一个数值型数据中出现了字母、特殊符号等, 或一个工人一周的工作时间超过 200 小时等都是失去完整性的例子, 它们与是否存在数据冗余无关。数据库系统的集中控制可以避免这些情况的出现。它通过由 DBA 定义的完整性检查, 对每一次更新操作实施完整性检查, 保证数据的完整性。

数据的完整性检查对于多用户系统尤为重要。因为, 多个用户同时使用数据可能引起对数据的干扰。例如, 当用户 A 正在修改某一数据时, 别的用户去读这个正在被修改的数

据,就会引起错误。数据库系统在这种情况下也能维护数据的完整性。

#### 8. 可以发现故障和恢复正常状态

很难保证数据库在运行过程中不受到破坏。硬/软件的故障及用户操作的失误,随时有可能使数据库遭到局部性的或全局性的破坏。数据库系统有一套及时发现故障,并迅速地把数据库恢复到故障以前的正确状态的措施,如转储、日志、检查点等方法。具体内容将在以后介绍。

### 1.3.3 数据库与文件系统的联系与区别

数据库系统与文件系统既有联系又有显著的差别。首先,两者都有一定的数据管理功能和数据检索与更新等数据操纵功能,都是数据管理和数据处理的工具。但由于两者之间存在着下面所要介绍的显著的差别,文件系统是数据管理和数据处理的低级工具,数据库系统则是高级工具。其次,数据库最终还是通过文件系统以文件形式存放在物理存储设备上的,因此,文件系统是数据库系统的基础,数据库系统是文件系统的提高和发展。

两者的主要差别表现在:

#### 1. 控制方式

文件系统中的数据文件由各个应用程序员根据需要组织起来,并由各应用程序负责建立、使用和维护,因此其控制方式是分散的。而数据库中的数据文件是由数据库系统统一规划,按照一定的数据模型组织和建立的,由系统统一管理 and 集中控制。

#### 2. 数据结构

文件系统中的文件结构极其简单,不能反映它所代表的现实世界中各事物之间的内在联系。在文件系统中,各个数据文件的记录或数据项之间内在的逻辑联系必须由应用程序来管理和维护。数据库中的数据是一个组织(企业)的所有应用所需数据的集成,是按照一定的数据模型来组织和建立的,是结构化了的数据集合,它反映了各数据记录所代表的现实世界各事物之间存在的内在逻辑联系,并通过各种存取路径在物理上加以实现。

#### 3. 数据独立性

文件系统中的数据文件是面向应用的,它依赖于应用程序的存在而存在,应用程序的改变伴随着数据的逻辑结构、物理结构和存取方法的改变,从而必须重新组织和建立所需的数据文件,反之亦然。换句话说,文件系统是数据相关的,相反,数据库系统是面向数据的系统,它具有高度的数据独立性。

#### 4. 数据的维护

数据库系统由于数据冗余小,具有数据完整性约束的定义和检查设施来保证数据的完整性,且数据库系统还提供了转储、日志和检查点等多种设施,一旦系统发生故障很易发现、排除并恢复到故障前的状态。而文件系统本身无完整性约束的定义和检查功能,几乎无故障恢复功能。

## 1.4 数据模型

模型是对现实世界的模拟和抽象。在现实世界中我们会经常接触到模型,例如,在购