

高等专科学校教材

数据库概论

史九林

西安电子科技大学出版社

87
71

高等专科学校教材

数据库概论

史九林

西安电子科技大学出版社

1988

内 容 简 介

本书主要讲述数据库系统的基本概念、基本原理和基本方法。全书共八章。头三章介绍数据库的基本思想、基本概念和基本原理；第四、五、六章分析和讨论了IMS、DBTG、System R和dBASEⅡ几个典型系统；第七章讨论了数据库的逻辑设计方法；最后一章讨论了安全性、完整性、并发控制和恢复技术等一些数据库系统的共同性问题。各章均配有适量的习题。

本书主要用作大专计算机软件专业的教材，也可作一般计算机专业人员为扩大数据库系统知识的参考书或自学读本。

高等专科学校教材

数据库概论

史 九 林

责任编辑 谭玉瓦

西安电子科技大学出版社出版

西安电子科技大学印刷厂印刷

陕西省新华书店发行 各地新华书店经售

开本 787×1092 1/16 印张 11 字数 266 千字

1988年12月第1版 1988年12月第1次印刷 印数 1-7 000

ISBN7-5606-0052-2/TP·0016

定价 2.25 元

出版说明

根据国务院关于高等学校教材工作分工的规定,我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作。由于各有关院校及参与编审工作的广大教师共同努力,有关出版社的紧密配合,从1978年至1985年,已编审、出版了两轮教材,正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应“三个面向”的需要,贯彻“努力提高教材质量,逐步实现教材多样化,增加不同品种、不同层次、不同学术观点、不同风格、不同改革试验的教材”的精神,我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会,在总结前两轮教材工作的基础上,结合教育形势的发展和教学改革的需要,制订了1986~1990年的“七五”(第三轮)教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿,是从通过教学实践、师生反映较好的讲义中经院校推荐,由编审委员会(小组)评选择优产生出来的。广大编审者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量,作出了不懈的努力。

限于水平和经验,这批教材的编审、出版工作还会有缺点和不足之处,希望使用教材的单位,广大教师和同学积极提出批评建议,共同为不断提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前 言

本教材系按电子工业部的工科电子类专业教材 1986—1990 年编审出版规划，由大专计算机专业软件编审小组征稿，推荐出版。责任编委是阎墨云。

本教材由南京大学史九林主编，西北工业大学徐秋元教授主审。

本课程教学时数为 66 学时，其中 60 学时用于课堂教学，6 小时用于实习，另外再安排一周时间的课程设计。全书共八章。头三章着重介绍和讨论数据库的基本思想、基本概念和基本原理；第四、五、六章通过几个典型系统(IMS、DBTG、System R 和 dBASE II)的分析和介绍，力图使学生掌握数据库系统应用的具体方法；第七章讨论和介绍了数据库应用设计的一般方法，特别是数据库的逻辑设计方法；最后一章讨论了数据库系统的共同性问题，即安全性、完整性、并发性和恢复技术。各章配有适当数量的习题，供学生学习时练习。在师资力量和学生接受能力许可的情况下，还可在第六章中增加关系数据库理论(数据依赖的公理系统)的教学内容。

本书旨在使学生初步掌握数据库系统的基本概念、基本原理和基本方法。通过对本书的学习，配合上机实习和课程设计，培养学生具备运用数据库方法解决实际应用课题的初步能力。学生最好具备 COBOL 语言、操作系统、数据结构和离散数学的基本知识。

本书编写过程中，得到了教材编审组、南京大学计算机系、南京有线电厂职工大学等有关方面的积极鼓励，大力支持和热情帮助。在手稿处理过程中还得到了黄雪玉老师的不少帮助，史今同志协助腾写了大部分书稿。对他们，作者一并表示衷心的感谢。

由于时间仓促、经验不足、知识尚浅，加之数据库技术发展很快，书中谬识之处在所难免，敬请读者批评指正。

编 者

目 录

前 言

第一章 绪论

§ 1.1 信息、数据和数据处理	1
1.1.1 信息和数据	1
1.1.2 数据处理	2
§ 1.2 计算机数据管理技术的进展	2
1.2.1 程序管理方式	3
1.2.2 文件系统方式	3
1.2.3 数据库管理方式	5
§ 1.3 数据库系统方法	5
1.3.1 什么是数据库	5
1.3.2 数据库系统的主要特征	7
1.3.3 数据库系统的体系结构	9
1.3.4 数据库系统的活动周期	12
§ 1.4 数据库发展史的回顾	14
习题一	15

第二章 数据模型

§ 2.1 三个世界的假说	16
2.1.1 现实世界	16
2.1.2 信息世界	17
2.1.3 数据世界	18
§ 2.2 信息模型	19
2.2.1 信息结构	19
2.2.2 E-R方法	19
§ 2.3 数据模型	20
2.3.1 数据间的逻辑联系	20
2.3.2 数据模型	21
§ 2.4 数据与模型	24
§ 2.5 数据的存储结构	25
2.5.1 存储记录	25
2.5.2 实现联系的物理方法	26
习题二	28

第三章 数据库管理系统

§ 3.1 数据库管理系统的功能	29
§ 3.2 数据描述语言	30
3.2.1 模式描述语言	31
3.2.2 子模式描述语言	31
3.2.3 物理模式描述语言	32
§ 3.3 数据操纵语言	33

3.3.1 数据操纵语言的结构	33
3.3.2 DML与程序设计语言的关系	34
§ 3.4 数据库管理系统的组成	35
习题三	36

第四章 层次数据库系统

§ 4.1 层次模型的一般讨论	37
4.1.1 层次模型的设计	37
4.1.2 层次数据库上的操作	40
§ 4.2 IMS数据库	41
4.2.1 物理数据库及其描述	41
4.2.2 IMS数据模型和数据库	44
§ 4.3 IMS外部数据库	45
4.3.1 逻辑数据库及其描述	45
4.3.2 IMS外部数据模型和外部数据库	46
§ 4.4 IMS数据库的操作	48
4.4.1 有关的基本概念	48
4.4.2 DL/1语言功能	50
4.4.3 应用程序的设计	53
§ 4.5 IMS存储结构	57
4.5.1 数段的存储表示	58
4.5.2 存取方法	58
4.5.3 小结	63
习题四	64

第五章 网络数据库系统

§ 5.1 网络模型的一般讨论	65
5.1.1 网络数据模型的设计	65
5.1.2 网络数据库上的操作	66
5.1.3 DBTG规范简介	67
§ 5.2 DBTG数据模型	68
5.2.1 DBTG系的概念	68
5.2.2 系结构的实现	70
§ 5.3 DBTG数据库存取策略	72
5.3.1 记录的定位	73
5.3.2 系值的排序	75
5.3.3 成员记录的属籍	76
5.3.4 系值的选择	77
§ 5.4 模式描述语言	78
5.4.1 模式的结构	79
5.4.2 数据描述语言DDL	79

5.4.3 一个模式的例子	81
§ 5.5 DBTG 子模式及其描述语言	83
5.5.1 子模式的设计	83
5.5.2 子模式的结构	84
5.5.3 子模式描述语言	85
§ 5.6 DBTG 数据库的操纵	87
5.6.1 运行单位和当前状态	87
5.6.2 DBTG 数据操纵语言	88
5.6.3 应用程序结构	93
习题五	94
第六章 关系数据库系统	
§ 6.1 关系数据库的一般意义	96
§ 6.2 关系模型的基本概念	96
6.2.1 关系的数学意义	97
6.2.2 关系的性质	98
6.2.3 关系模型	98
§ 6.3 关系数据库上的操作	101
6.3.1 基于关系代数的子语言	101
6.3.2 基于关系演算的子语言	107
6.3.3 基于图的子语言	109
§ 6.4 System R 简介	113
6.4.1 System R 的体系结构	114
6.4.2 System R 数据库管理系统	115
6.4.3 SQL 语言	117
§ 6.5 关系规范化问题	124
6.5.1 关系规范化的重要意义	124
6.5.2 函数依赖的概念	125
6.5.3 关键词	127
6.5.4 关系的规范化	127
§ 6.6 小型关系数据库 dBASE-Ⅱ 简介	129
6.6.1 dBASE-Ⅱ 的体系结构	129
6.6.2 dBASE-Ⅱ 数据库的建立和 装入	131
6.6.3 dBASE-Ⅱ 数据库上的操作	134
习题六	137

第七章 数据库设计初步

§ 7.1 数据库设计的一般讨论	139
7.1.1 设计准则	139
7.1.2 设计过程	140
7.1.3 设计方法	141
7.1.4 数据字典	143
§ 7.2 逻辑设计——分步法	144
7.2.1 分步法的设计大纲	144
7.2.2 分步法示例演示	144
§ 7.3 物理设计	149
7.3.1 物理设计的一般内容	150
7.3.2 物理设计原则	150
§ 7.4 数据库的实施	151
7.4.1 数据库系统的选择	151
7.4.2 数据库的建立和装入	151
习题七	152

第八章 数据库保护

§ 8.1 数据库保护的一般意义	153
§ 8.2 完整性控制	154
8.2.1 完整性约束条件	154
8.2.2 几个系统的完整性约束	155
§ 8.3 并发控制	156
8.3.1 并发操作的几种异常现象	156
8.3.2 并发控制方法	158
8.3.3 几个系统的并发控制	159
§ 8.4 安全性控制	161
8.4.1 安全性控制的一般讨论	161
8.4.2 几种常见的不安全的例子	161
8.4.3 安全性控制的一般方法	162
8.4.4 几个系统的安全性控制	164
§ 8.5 数据库恢复	166
8.5.1 恢复的一般原则	166
8.5.2 恢复的一般方法	167
习题八	169
参考文献	169

第一章 绪 论

人类社会是物质的，存在着一个强大的物质流；同时也是信息的，即还存在着一个强大的信息流。社会信息化本身就意味着人类对信息流认识的深化和对信息控制利用能力的增长。这种深化和增长愈快，信息化的程度就愈高，社会也就愈现代化。计算机和数据库技术的广泛应用正是社会信息化的集中体现。因此，本章首先讨论这个问题。

§ 1.1 信息、数据和数据处理

我们赖以生存的世界是一个物质的世界，也是一个信息的世界。这有两方面的特征：其一，物质的存在伴随着信息的存在。物质是信息存在的前提和基础；信息则是物质的抽象反映。其二，物质的变化(表现形式的变更，数量的增减，地域的变迁，……)能引起信息的变化和传播。因此，为了有效地利用信息，就十分有必要研究信息的性质，表示信息的手段以及处理信息的方法，以促进社会信息化的进程，使之具有更高的水平。

1.1.1 信息和数据

信息是一个抽象的概念。目前还很难给出确切的，具有广泛意义的一般性定义。从不同的角度出发，对信息也有不同的理解。这里，作者仅企图从计算机信息处理这一范畴加以讨论和研究。

非正式地说，信息是事物特征方面以及诸事物之间相互联系方面的一种抽象反映。这种反映能被人们(通过各种途径和方法)认识和理解，并将作为知识用来识别事物或进行推理，从而达到认识世界，改造世界和支配世界的目的。这恐怕也是社会信息化的实际意义所在。

我们在谈论“事物”的时候应当认识到：它比“物质”有更广泛的意义。即它不仅指可触及的具体物质，而且也指那些不可触及的抽象概念。故而，信息又可以看成是客观世界的真实反映。或者说，信息反映了客观世界。

因为处理、传播和使用的需要，常常要用文字、符号、图象或声音表示和记录信息。这就是数据。所谓数据乃是这种文字、符号、图象和声音的有意义的组合。这种组合具体生动地表示出信息的内容。就计算机处理而论，数据则是它能处理的一切对象和处理获得的结果。只是它们有外部表示形式(对人而言)和内部表示形式(对机器而言)的区别而已。

一提到数据，也许有人会产生某种误解，即仅仅把表示量值概念的数值数据看成是数据的全部内容。其实，在表示信息方面却不仅限于此。因为信息有更广泛的内容和更普遍的意义，所以，在我们谈论数据的时候，还应当包括表示“陈述”意义的数据。例如：“1986年，我国钢产量突破5000万吨大关。”这里有表示数量的数字，也有陈述概念的文字。后者称为非数值数据。因此，数值数据和非数值数据之总和才是我们所论之数据的全部内容。

信息和数据是两个相互联系，相互依存，但又相互区别的概念。信息以数据为载体而表现；数据则是表示信息的一种手段。一定形式的数据表示某一确定的信息。反之，某一信息都可以用不同的形式的数据表示之。例如，“教室里有10个人。”可以有另一种数据形式：

“There are ten persons in the classroom.”它们表示了同一信息内容。这是信息的稳定性特征。

数据确切地表示信息，信息也只有表示出来才有实际应用意义。所以我们把对信息的研究和处理转而对数据进行研究和处理。正因为如此，我们常常等效地使用这两个名词。

1.1.2 数据处理

在人类的社会(政治的、经济的和文化的)活动中，信息总居于非常重要的地位。可以毫不夸张地说，人类的全部活动都是在信息的支配和指导下去进行的。因而人类社会机体中有一类部门是专门从事信息处理的，如统计部门、政府部门、邮局、银行，……，它们要花费大量的人力、物力和时间去从事信息的收集、综合、加工，最后获得精细的信息，作为指导活动的依据，形成一个围绕信息的“数据—信息—决策—活动”的信息反馈周期(见图 1-1)。在社会信息化的今天，有意识地、准确地利用这种规律将有利于使自己立于不败之地。

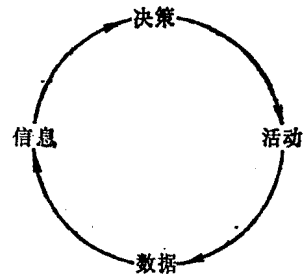


图 1-1 信息反馈周期

读者可以看出，在这个周期中信息是重要的，而数据和对数据的处理是它的基础。所谓数据处理则是指对“原料”数据的综合加工以导出“产品”数据的过程。这一过程包括对数据的收集、记载、分类、排序、存储、计算或加工、传输、制表和传播、递交等若干处理阶段。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，以及内在联系的有具体意义的数据；是信息的特定形态。

人类处理数据是社会活动的需要和社会信息化的必然。实际上，数据处理这一特殊的社会活动很早就已经开始，其历史可以追溯到远古时代。自有人类出现就开始了数据处理活动。原始人类的结绳记事，累石记数便是最早的数据处理形式和借助的“工具”。随着人类生产力的发展和文明的进步，处理工具的发明和利用，数据处理方法也得到了相应的发展。就其整个发展过程来看，经历了：手工数据处理，机械数据处理和电子数据处理等三个主要阶段。电子数据处理阶段的主要特征是电子计算机的发明和应用。它使数据处理实现了自动化。

§ 1.2 计算机数据管理技术的进展

在电子数据处理方式下，利用计算机系统作为处理工具，通常是把“原料”数据和对数据的处理方法或过程的描述——算法——输入系统；然后由系统自动地执行处理任务，最后输出处理结果(见图 1-2)。对某一特定任务而言，算法是相对稳定的。这通常是根据处理要求编制的程序。而原料数据和结果数据则随不同时间的处理请求和方式各异，而且数据量都很可观。如果这些数据不按某种确定的规则输入和输出就会使数据失去存在的意义，或者使算法表现得十分复杂，以至于不能实现既定的处理任务。我们所说的“规则”即是对数据的一种管理原则和方

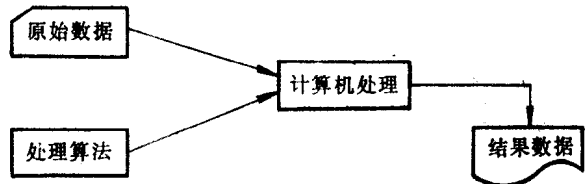


图 1-2 计算机数据处理示意图

法——数据管理技术。

如所周知，数据管理技术的优劣对数据处理方式和工效有着直接的影响。因此，自计算机进入数据处理领域之日起，许多计算机科学家，特别是软件工作者就展开了对这一技术的研究和实践，并且取得了重大的进展，提出了许多卓有成效的数据管理技术。数据库系统方法是这一技术的顶峰。回顾其发展史，大致可分 3 个时期 4 个阶段。

1.2.1 程序管理方式

数据管理技术发展的第一个时期（大约在 1960 年初之前）为程序管理方式时期。这里的程序是指用户应用程序。用户在编制程序的同时确定自己对数据的管理原则和方式。因此，不仅要在程序中表现处理算法，还要表现对数据的管理原则。不仅要考虑数据的逻辑定义，还要考虑数据的物理特征。因而程序流程和数据结合为一个不可分割的整体。

实质上，程序管理方式即是自由管理方式，用户有绝对的自由权。不同用户有不同的管理方法，就是同一用户的不同程序也有不同的办法。所以，在这一时期根本就不存在数据管理技术。有的只是程序员“艺术家”的匠心。他们总是在一个没有任何称为系统软件的裸机上工作。其结果是：数据和程序之间有极强的依赖性，同时，程序员往往要做多次重复工作，不利于提高程序设计的效率。

1.2.2 文件系统方式

把数据组织成文件，由一个专门的软件系统——文件管理系统——统一实施管理是数据管理技术的重大进步，并使数据处理与信息处理更加接近。文件组织方式就是按照统一的规则和方法来组织和存取数据的技术。这是数据管理技术发展史上的第二个时期。

一般地，文件是一组具有相同性质和结构的记录集合。记录又由某些相关数据项组成。它能完整地表示一个处理对象。文件系统方式规定：可以对文件组织中每一级数据命名，分别为文件名，记录名和数据项名。这样，程序员就可以在他的程序中按名存取数据。

文文件按其组织方式的不同分为：顺序文件、索引文件、直接文件、相对文件和倒排文件。它们各自适应不同的存储设备和用户处理要求。

文件系统还规定了对文件的存取方法：顺序存取方法、随机存取方法和动态存取方法。各自适应于不同的文件组织和设备。有关文件组织、设备和存取方法的适配关系见表 1-1。

由此可见，数据按文件系统的规则组织就可以离开使用它的程序自立存储，而由文件管理系统负责统一管理和维护。文件管理系统是一个独立的系统软件，是应用程序与数据文件之间的一个接口。应用程序通过文件管理系统建立和存取文件（见图 1-3）。用户程序只要给出数据的逻辑定义以及文件组织方式，存储用的设备类型和操作中用到的

表 1-1 文件组织、设备和存取方法的适配关系

文件组织	存储设备	存取方法
顺序文件	顺序存取设备 直接存取设备	顺序存取方法
索引文件	直接存取设备	顺序存取方法 随机存取方法 动态存取方法
相对文件	直接存取设备	顺序存取方法 随机存取方法 动态存取方法
直接文件	直接存取设备	随机存取方法
倒排文件	直接存取设备	随机存取方法

的存取方法等信息就足够了。而数据存取中的有关各种细节都由文件管理系统负责确定、执行和完成。如在图 1-3 中，程序 A 通过文件管理系统与磁盘上的文件 A₁ 和磁带上的文件 A₂ 接口；如此等等。程序 A、B、C 不再需要考虑更多关于文件存取操作上的任何细节。

文件系统有一个较长的发展时期，至今人们还乐于使用。这主要是因为它的简易性和灵活性。我们可以把其发展分成两个阶段：

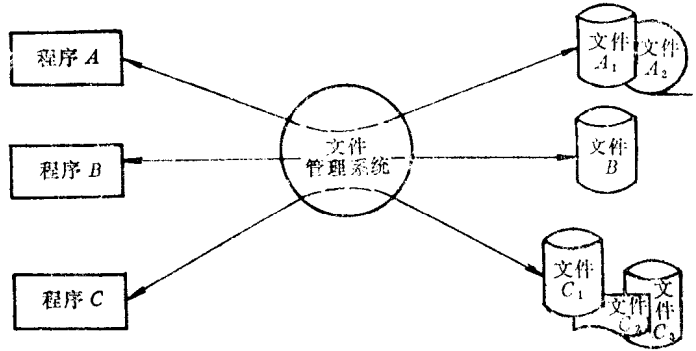


图 1-3 文件系统工作示意图

1. 独立文件阶段：这一阶段的时间并不长。其特征是，文件

管理系统是一个专用程序。它不能提供和支撑不同程序对同一组文件的共享能力。文件只能为用户私有，甚至只能为某一特定程序专用。其结果是使文件处于一种分散状态；数据利用率不高；还造成大量的数据冗余（主要是同一数据的多副本）以及由此带来的许多弊病。不同文件中同一性质之数据的一致性不易保证；不同文件之间的逻辑关系无法直接建立。特别是程序和数据之间存在着不可解决的相关性问题。因为文件仅仅是“纯”数据的存储，而设备的选择，文件组织和存取方法的确定，数据操作的逻辑意义仍要求由程序指明和提供信息，所以，文件一旦离开对应的程序便失去存在的价值，成为一堆不可知的“废物”。

2. 共享文件阶段：计算机操作系统的运行，特别是多道程序设计系统使文件进入共享阶段。由于文件系统成为操作系统的主要功能之一，文件管理系统也成为其软件组成，这样，操作系统对在其上运行的用户作业建立的文件实施统一管理，使文件共享成为可能。但是，这种共享是有条件的，即需要在多方面保持一致。

另一种文件共享方法是建立一个公用文件处理软件——专用程序，也称应用软件包，负责维护一组文件。再提供一种专用命令语言，使用户可以通过命令存取任何文件（如图 1-4）。这种思想已经具有数据库的某些概念，但只是其雏型。在许多未装备数据库系统软件的计算机系统上，人们常常采用此法实现数据库的思想。

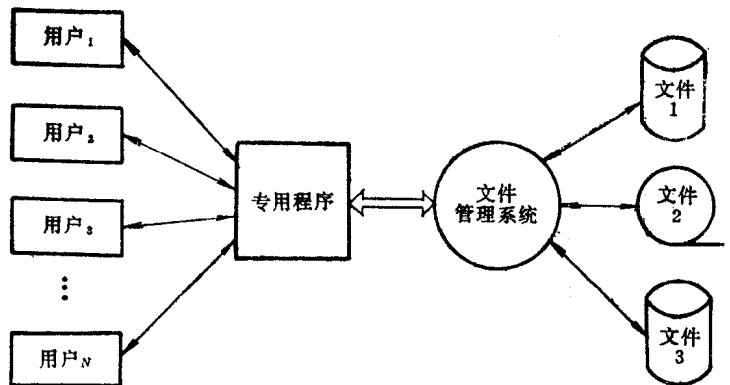


图 1-4 共享文件方式工作示意图

综上所述，文件系统已初步解决了数据管理问题，提供了一种技术。但是，还存在某些致命的弱点：①文件只能表示事物，而不能表示事物之间的联系；②文件的数据结构一旦定义便不可改变；③数据的

保护只能以文件为单位；④最需要的一点是，程序和数据之间的数据依赖性，因而它是一种以程序为中心的数据管理方式。数据库正是在这些问题上与之不同。

1.2.3 数据库管理方式

数据库方式是在文件系统的基础上发展起来的最新技术。它企图克服文件系统的弱点，提供一种完美的、高级的数据管理技术，以实现数据集中统一的、独立的管理，使数据的存储和维护不受任何用户的影响。

数据库方式的根本目的在于提高应用程序生产率，同时，也是计算机工业发展的必然结果。特别是硬件设计技术的不断提高；存储容量愈来愈大；价格愈来愈低，整机性能价格比愈来愈好，均为之提供了良好的物理环境。而软件技术的发展和不断创新是数据库技术得以成功的直接因素。

§ 1.3 数据库系统方法

为使读者尽快对数据库技术有一个概观性了解，本节首先简要地回答读者可能会提出的几个问题：

- ① 什么是数据库？
- ② 为什么要发展和应用数据库？
- ③ 数据库系统有怎样的体系结构？
- ④ 数据库系统是如何工作的？

1.3.1 什么是数据库

数据库这一术语流行甚广。那么，倒底什么是数据库呢？不妨先看一个简单的例子。假定有一张如图 1-5 所示的表 PERSONAL 和两个用户 U_1 和 U_2 。 U_1 处理有关人事数据； U_2 处理电话通信数据。它们各自根据需要从表中获得数据，但不是表中的全部。因此，作为数据库系统，其宗旨就是维护和管理 PERSONAL 表，并提供对两个用户的服务。

由上可见，数据库是为满足某一具体单位或组织中多种用户的多种应用需要，在计算机系统中按照一定数据模型组织、存储和应用的互相联系的数据集合。图 1-6 将给我们一个关于数据库的大致印象。

首先，图 1-6 中的圆柱体形表示硬件支撑，即计算机。因为图中只强调数据的存储，故只出现表示磁盘的符号。硬件仅作为一

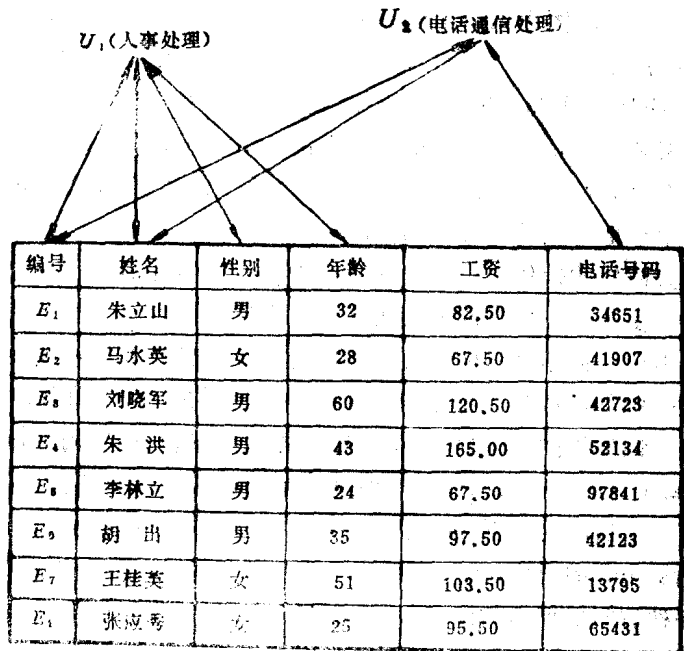


图 1-5 一个数据库的简单例子

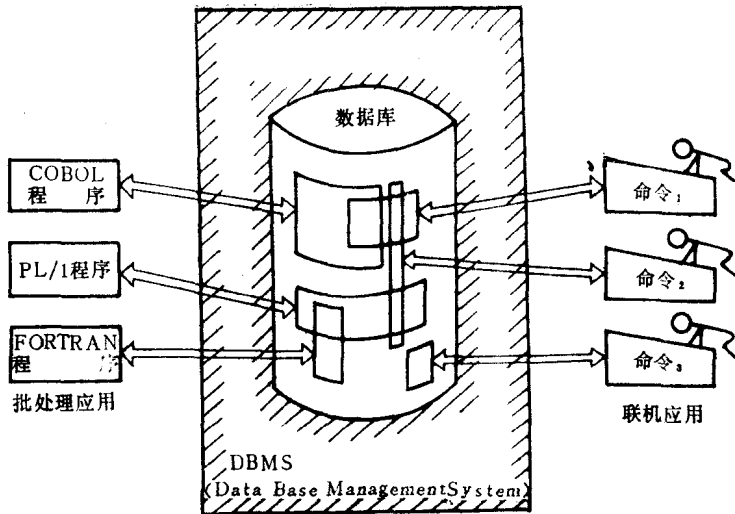


图 1-6 数据库系统的印象

种物质基础在系统中存在。

其次,存储在计算机中的是数据,如工厂中的产品数据;学校中的学生档案数据;政府部门中的计划统计数据等等。任何一个组织中的数据都将按照实际应用处理要求,有机地组织成一个整体,构成数据库。这种整体性能反映出数据的自然属性和有机联系,从数据的角度反映出那个组织的业务处理职能和相互关系。数据库中的数据是一种处理用的中间数据,又称业务数据(Operational data)。虽然它被存储在输入/输出设备上,又能被用户存取,但它却与输入/输出数据有本质的不同。

第三,图中示出了两类用户:批处理用户和联机用户。这意味着存在对数据库的操作请求,意味着与用户之间的关系。批处理用户是指专业程序员。他们能使用程序设计语言(如 COBOL PL/I FORTRAN, 汇编语言)编制程序存取数据库,并作某种应用处理。因为他们需要大面积地存取数据库,每次处理的数据量比较大,故称批处理。联机用户多为非计算机专业人员,如组织中的领导,工程技术人员或各级业务管理人员。他们通常只需要从数据库中获取有关的综合性信息,如统计资料,查询某些特定意义的数据库等等。为了向他们提供十分简便的操作,一般设立命令语言,或称查询语言,参数语言。

多个用户可以同时使用同一数据库,即共享。但是,每个用户几乎总是只存取其中的一个部分,而不是全体。各用户面对的数据又可以任意地交叉或重叠。

此外,还有一种用户,那就是系统用户,但未在图中示出。他的应用在于对数据库实施整体性的维护,其操作对象是整个数据库。这一特殊用户便是数据库管理员,简称 DBA(Data Base Administrator),他是数据库的责任维护者。

最后,讨论图 1-6 中的 DBMS。DBMS 是数据库管理系统(Data Base Management System)的简称,是一个系统软件。其职能是对数据库实施正常维护,接受并完成用户提出的各种访问数据库的请求。也即是说,用户不能直接对数据库进行无约束的存取操作,而必须通过 DBMS。因此 DBMS 为用户提供了数据库操作的公共的、统一的方法和途径,是数据库与用户之间的一个标准接口。

可见，计算机硬件、DBMS、数据和用户等四者构成了一个完整的系统——数据库系统。图 1-6 并非十分精确，但数据库系统的意义已尽在其中了。

1.3.2 数据库系统的主要特征

人们之所以热衷于数据库技术的应用和研究，主要是因为它具有许多独特的优越性。也正是从这些优越性上区别于文件系统，提供了强有力的数据管理能力。归结起来主要是：

1. 数据集中控制特征：这是相对于文件技术而言的。因为文件完全处于一种分散状态，造成管理上的困难和处理上的不方便，所以数据库系统首先要克服这些不足。集中控制特征主要体现在两个方面：

(1) 由 DBA 统一管理和维护数据。DBA 由一个人或一组人充任，他们对所在组织的数据库负责。DBA 不仅应具有精湛的计算机技术，而且还应全面了解组织内的业务处理的职能和过程，以及其间的各种关系。他们负责统一设计、规划和建立数据库；经常监督和维修数据库；合理协调和平衡用户对数据库的应用要求；为用户提供良好的、统一的服务。这样就使数据库有“最好”的性能。

(2) 数据库中的数据是集成化的数据。在文件系统中，为了处理的需要，常常在多个文件中存储同一数据的不同副本。这不仅多占用存储空间，而且还会引起许多其它弊端。集成化的意义在于统一存储文件中的公共数据，而不是诸文件的简单组合。同时，系统能按照一定的规则为用户动态地组织和提供数据，以满足用户的各种数据要求。

集中控制特征产生的效果是令人鼓舞的。首先，压缩了数据的冗余度。同一数据的多副本存储是数据冗余的典型现象。不仅如此，可导数据的存储也是数据冗余的例子。压缩冗余可以减低数据维护的难度。

其次，可以一定程度地避免数据不一致性的发生。所谓不一致性是指数据的不相容性和矛盾性。数据冗余是引起不一致性的根源之一。对于非冗余的数据也可能在操作中产生不一致性，这主要由对数据库的并发操作所致。

第三，可以实施标准化。随着信息处理自动化和信息的广泛利用，标准化问题显得愈来愈重要。简单地讲，就是建立“共同语言”。这不仅能使信息管理得到简化，提高性能，而且还能使信息更具通用性的品格。

第四，可以平衡不同用户对数据库的性能要求。不同用户对数据库的实际性能要求往往不同；但是，从用户心理来说总希望自己得到最快最好的服务。然而，对整个系统来说应当考虑的问题是要保证系统有“最好”的性能。当数据置于集中控制之下时将使之成为可能。

2. 数据结构化特征：这一特征是数据能被集中控制的保证和前提。数据模型在数据库系统中的应用是数据结构化特征的体现。它从两个方面表现数据：其一是数据的逻辑表示，即表示事物；其二是数据间联系的逻辑表示，即表示事物之间的某些关系。这样就比较全面地表示了组织。相对而言，文件是无结构的。

关于结构化的意义，可以一例说明之。假定有一组学生数据和一组课程数据（见图 1-7(a)）。现在的处理问题是：“哪些学生选修了哪些课程？”“成绩是多少分？”若采用文件方式，则必须建立一个如图 1-7(b) 的文件。这时的一个严重问题是图 1-7(a) 中的数据在 (b) 中大量重复。若建立另一组数据“得分”（见图 1-8(a)），则可以使冗余大大压缩。其故是因为我们用“联系”的方法把三组数据联结为一个整体。这就成为一个数据模型（见图 1-8(b)），即有

了一种结构。结构化是数据库方法的中心和生命，可见其重要性了。由此，许多数据库的特征便应运而生了。

S(学生)

学号	姓名	...
E ₁	王 英	...
E ₂	李 明	...
E ₃	钱 刚	...
E ₄	朱 朋	...
E ₅	刘 晓	...

C(课程)

课程号	课程名称	...
C ₁	操作系统	...
C ₂	数据结构	...
C ₃	数据库概论	...

(a)

SC(学生选课)

学号	姓名	...	课程号	课程名称	...	得分
E ₁	王 英	...	C ₁	操作系统	...	98
E ₁	王 英	...	C ₂	数据结构	...	75
E ₂	李 明	...	C ₁	操作系统	...	100
E ₂	李 明	...	C ₃	数据库概论	...	82
E ₃	钱 刚	...	C ₂	数据结构	...	94
E ₃	钱 刚	...	C ₁	操作系统	...	67
E ₃	钱 刚	...	C ₃	数据库概论	...	93
E ₄	朱 朋	...	C ₁	操作系统	...	99
E ₅	刘 晓	...	C ₃	数据库概论	...	78

(b)

图 1-7 文件方式的无结构性

3. 数据共享特征：数据库如果没有共享性质就将失去存在的意义和先进性。一个主要的理由是：数据集中存储和管理的目的本身就是为了方便地提供多种用户的多个应用使用数据。所谓共享有四方面的内容：

- (1) 当前所有用户可以同时存取同一数据库。
- (2) 未来出现的用户与当前用户可以同时存取同一数据库。
- (3) 用户可以通过多种不同程序设计语言编写的程序或命令语言与同一数据库接口。
- (4) 每一个用户存取数据库的部分可以以任意方式交叉或重叠，乃至于是同一块数据。

4. 数据保护特征：一个数据库是否实用，其安全可靠是首要问题。因为数据库的共享特征，其保护显得更为重要。数据库的保护涉及许多问题。这里仅限于数据库系统的自我保护能力，主要是：安全性保护，完整性保护，并发控制，以及故障发现和恢复等四个方面。详细讨论留待第八章。

5. 数据独立性特征：数据独立性和前面提到的数据依赖性都是程序和数据之间的一种性质，并统称为数据相关性。所谓数据独立性即是数据与程序之间的无关性。与文件系统相反，数据库系统保证数据的独立存储和维护，使得数据和程序中的一方发生变动时而不影响另一方。故又称这种相关性为以数据为中心的。数据独立性包括两方面的内容：逻辑

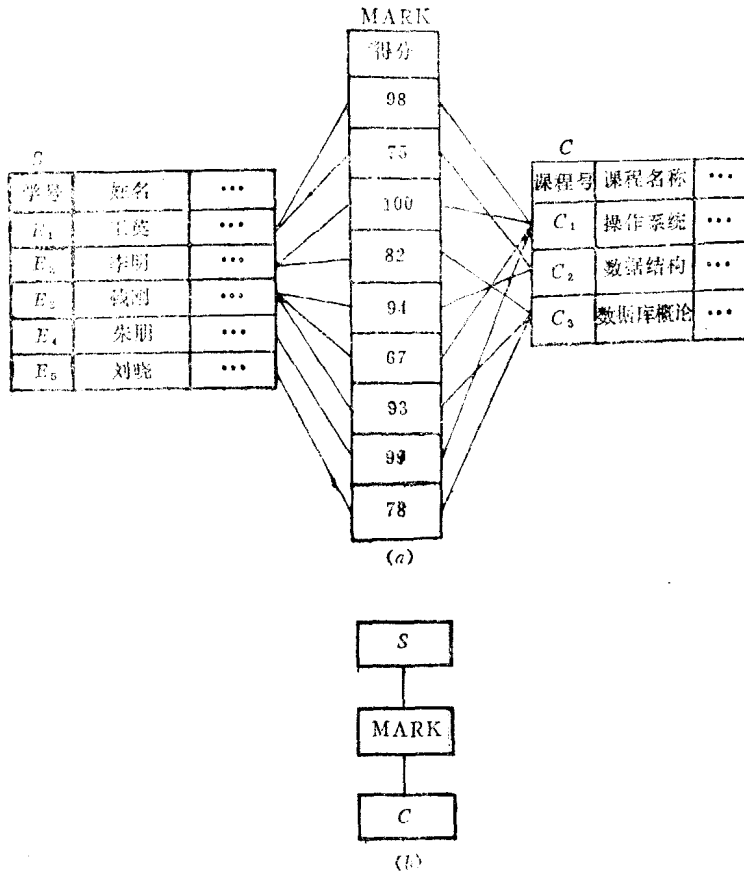


图 1-8 数据库方式的结构性

独立性和物理独立性。前者保证在数据库的逻辑结构发生变动(定义的修改, 新数据类型的加入, 逻辑联系的改变等)时, 不影响原有应用程序对数据库的存取。后者保证在数据库的物理结构发生变化(设备更换, 空间易位, 物理组织调整, 存取方法的改变等)时, 对应用程序同样不产生任何影响。

数据独立性特征实质上是数据库系统开发和研究的目, 是提高应用程序生产率的根本, 至今还是研究的主要内容之一。

读者应从对这些主要特征的讨论弄清数据库系统与文件系统的区别; 从而也就不难理解数据库系统被广泛使用的道理了。

1.3.3 数据库系统的体系结构

从数据管理的角度来看, 数据库系统有一个严谨的体系结构, 以保证其功能的实现和特性的发挥。迄今, 世界上已有数以千计的系统在运行, 种类也不尽相同, 但是, 它们的基本结构大体一致, 都是三层结构。

1. 抽象级别和三层结构: 由前可知, 有两类人员与数据库有关: 用户和系统人员。用户主要是指程序员和终端操作员; 系统人员包括 DBA、系统设计员和系统分析员。他们对数

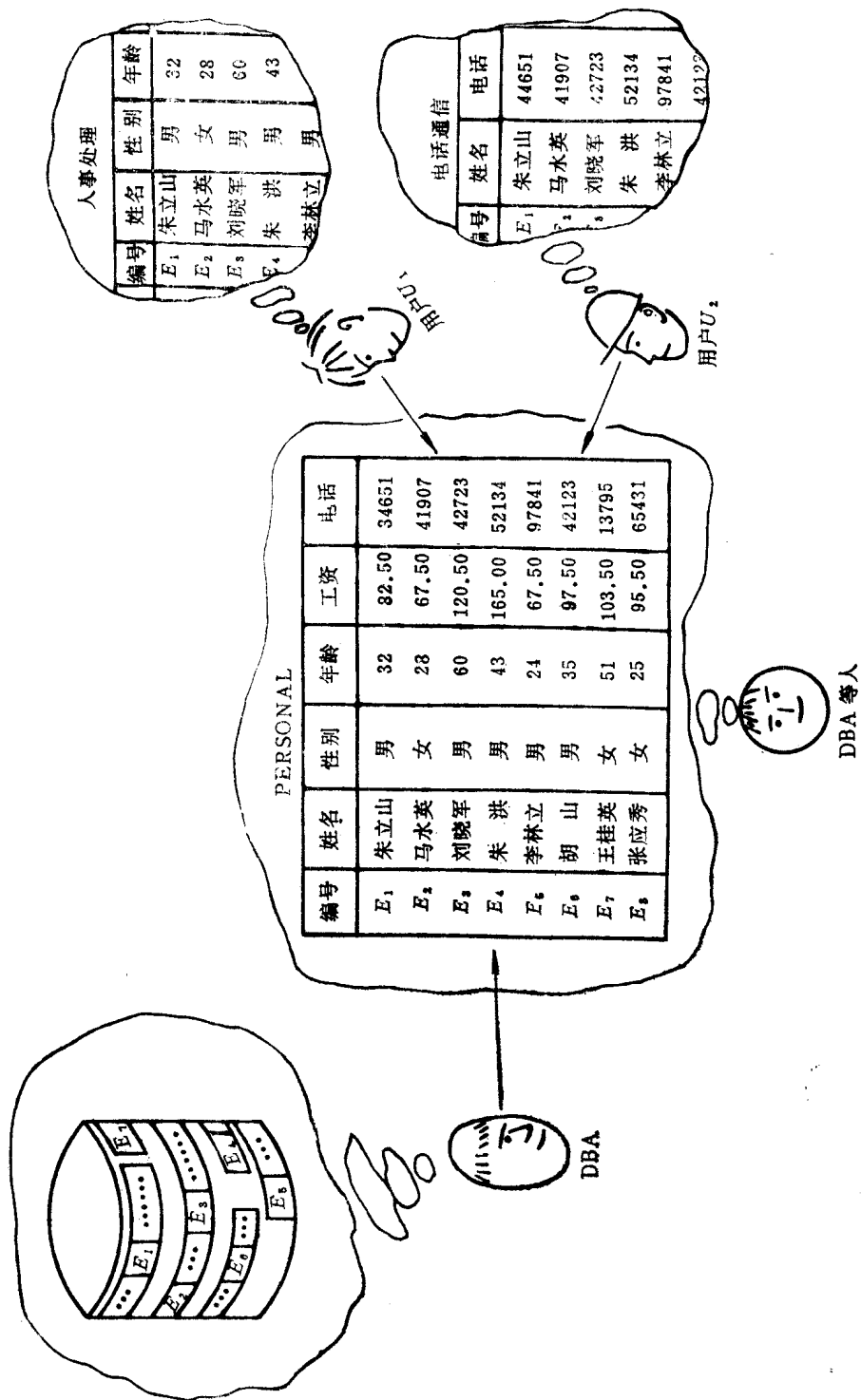


图 1-9 三种数据观下的数据库