



中小學生信息科學知識

数据库系统

主 编：李广建

副主编：邓俊强 曹正斌



北京科学技术出版社

中华万有文库

总顾问 费孝通

总主编 季羨林

副总主编 柳 斌

科普卷·中小学生信息科学知识

数 据 库 系 统

编著 孙博阳

《中小学生信息科学知识》丛书编委会

主 编 李广建

副主编 邓俊强 曹正斌

编 委 高 聪 吴钢华 顾俊强

李广建 曹正斌 徐仁霖

耿 寿

北京科学技术出版社

中华万有文库

总 顾 问 费孝通
总 主 编 季羨林
副总主编 柳 斌

《中华万有文库》编辑委员会

主 任: 刘国林

秘书长: 魏庆余 和 葵

委 员: (按姓氏笔画为序)

王寿彭	王晓东	白建新	任德山
刘国林	刘福源	刘振华	杨学军
李桂福	吴修书	宋士忠	张 丽
张进发	张其友	张彦民	张晓秦
张敬德	罗林平	封兆才	和 葵
金瑞英	郑春江	胡建华	侯 玲
祝立明	贾 斌	顾士熙	高文元
游铭钧	章宏伟	常汝吉	彭松建
韩永言	葛 君	鞠建泰	魏庆余

总 序

本世纪初叶，商务印书馆王云五先生得到胡适之、蔡子民、吴稚晖、杨杏佛、张菊生等 30 余位知名学者、社会贤达鼎力相助，编纂出版了《万有文库》丛书。是书行世，对于开拓知识视野，营造读书风气，影响甚巨，声名斐然，遗响至今不绝。

1000 多年以前，南朝齐梁学者钟嵘在《诗品》中以“照烛三才，晖丽万有”来指说天地人间的广博万物。今天，我们全国各地的数百家出版发行单位与数千名作者以高度的历史责任感，联袂推出《中华万有文库》，并向社会各界读者，特别是青少年读者做出承诺：

传播万物百科知识，营造有益成功文库。

我们之所以沿用《万有文库》旧名，并非意图掠美。首先，表明一个信念：承继中国出版界重视文化积累、造福社会、传播知识的优秀传统，为前贤旧事翻演新曲，把旧时代里已经非常出色的事情在新时代里再做出个锦上添花。其次，表明我们这套丛书体系与内容的鲜明特点。经过反复论证，我们决定针对中小学生在提倡素质教育的需要和农村、厂矿、部队基层青年在提高文化与科学修养的同时还要提高劳动技能的广泛需要，以当代社会科学与自然科学的基础知识为基本立足点，编纂一套相当于基层小型图书馆应该具备的图书品种数量与知识含量的百科知识丛书。万有的本意是万物。百科知识是人类从自然界万物与社会万象之中得到的最重要的收获。而为表示新旧区别，丛书之名冠以中华。这就是我们这套丛书的缘起与名称的由来。

《中华万有文库》基本按照学科划分卷次，各卷之下按照内容分为若干辑，每一辑大体相当于学科的一个二级分支，各卷辑次不等；各辑子目以类相从，每辑10至20种不等，每种约10万字左右，全书总计约300辑3000种。《中华万有文库》不仅有传统学科的基本知识，而且注意吸收与介绍相关交叉学科、新兴学科知识；不仅强调学科知识的基础性与系统性，而且注重针对读者的年龄特点、知识结构与阅读兴趣而保持通俗性和趣味性；不仅着眼于帮助读者提高文化素质与科学修养，而且还注重帮助读者提高社会生存能力与劳动技能。

每个时代，图书最大的读者群是10至20岁左右的青少年。每个时代能够影响深远的图书，是那些可以满足社会需要，具有时代特点，在最大的读者群中启蒙混沌、传播知识、陶冶情操、树立信念的优秀图书。我们相信，只要我们老实地做下去，经过几个甚至更为漫长的寒暑更迭，将会有数以百万计的青少年读者通过《中华万有文库》而打开眼界，获取知识；《中华万有文库》将会在他们成长的道路上留下鲜明的痕迹，伴随他们一同走向未来，抵达成功的彼岸。

天高鸟飞，海阔鱼跃。万物霜天，凭知识力量，竟取成功，争得自由。在现代社会中，任何人都没有任何理由拒绝为了获取力量而读书。这是《中华万有文库》编纂者送给每一位本书读者的忠告。

追求完美固然是我们的愿望，但是如同世间只有相对完善一样，《中华万有文库》卷帙庞大，子目繁多，难免萧兰并擷，珉玉杂陈。这些不如人意之处，尚盼大家幸以教之。我们虚心以待。是为序。

《中华万有文库》编委会

目 录

第一章 绪论	(1)
1. 信息、数据和数据处理	(2)
2. 数据管理技术的发展	(4)
第二章 数据库系统	(14)
1. 数据库系统	(14)
2. 数据模型	(21)
3. 网状数据库	(25)
4. 层次数据库	(27)
5. 数据库保护	(30)
第三章 关系数据库	(33)
1. 关系数据语言	(33)
2. 关系代数	(37)
3. 结构化查询语言——SQL	(43)
第四章 数据库设计	(49)
1. 规范化	(49)
2. 关系数据库的设计	(53)
第五章 分布式数据库	(59)
1. 数据通讯	(59)
2. 分布处理	(61)
3. 分布式数据库管理系统	(66)
第六章 面向对象数据库	(72)

1. 面向对象数据库	(72)
2. 面向对象数据库的现状与未来趋势	(83)
第七章 数据库技术的发展	(86)
1. 数据库技术发展综述	(86)
2. 多媒体数据库	(102)
3. 专用数据库技术——工程数据库	(108)
4. 专用数据库技术——科学数据库	(111)
5. 专用数据库技术——统计数据库	(117)
6. 数据仓库与数据挖掘	(124)

第一章 绪论

在远古时代,人们就从现实世界中抽象出了数的概念,从原始的结绳记事,到罗马人、阿拉伯人发明了今天仍在使用的数字,人们渐渐学会用数字来描述现实世界里的物,用数据运算表示现实世界的变化。随着人类文明的进步,社会活动的更加活跃,数据运算越来越频繁,越来越复杂,人类从而又有了利用机器实现运算的机械化、自动化的需求。中国古代的算盘和西方 19 世纪的机械计算装置反映了人类的这种追求。自上个世纪末以来,由于工业革命带来的对社会的巨大推动作用,社会呈现加速发展的趋势。本世纪 40 年代,逐步兴起的电子技术,以及为了应付二战战术数据处理的需要,人们开始了对电子自动工具的研究。1946 年,第一台电子计算机诞生了。从此,人类在对数据处理能力上,有了质的飞跃。50 年来,计算机以人们始料不及的速度发展着,以让人眼花缭乱的强大功能迅速渗透到社会的各行业各场合,并且仍在不断地以新面目面向世人。

计算机的计算能力和存储量,几十年来有了翻天覆地的变化,这些变化导致了计算机的作用从单纯的数字计算演变为对电子数据的处理上来。这种转变是一个划时代的转折,它使计算机从少数的天才科学家手中的怪物转变为广大科技人员和管理人员工作的有力工具和得力助手,它为推动信息化

社會轉變起到了決定性的作用。

1. 信息、數據和數據處理

計算機處理的電子數據，來自於現實世界的信息。信息是現實世界在人腦中的抽象反映，是通過人的感官感知出來並經過人腦的加工而形成的反映現實世界中事物的概念。這裡所說的“事物”不僅是那些看得見、摸得着的物體，而且也包括那些不可觸及的抽象概念，如產量、質量、速度等。因此信息可以看作是現實世界的真實反映。信息不僅為人們所認識和理解，而且能夠把它作為知識來進行推理、加工和傳播，從而達到認識世界、改造世界的目的。

在用計算機處理信息的時候，要將信息轉化為計算機可以識別的符號，也就是數據。數據是表示信息的一種手段。比如對一本書的認識——信息，可以在計算機中表示為標題、作者、開本、價格、出版社、摘要和全文等等符號化的數據。

事物——信息——數據，實際上貫穿了三個世界，即現實世界——信息世界——計算機世界。

現實世界存在着各種物體的集合，如某個班集體、書、零件等，這些可以稱為事物類，事物類也可以是某種抽象概念的集合，如成績。這是現實世界進行管理的基础。每一個事物類都有具體的事物組成，如某班集體的學生，可以是張三、李四等。每一個具體的事物又具有自己的內涵，如張三具有姓名、性別等內涵。事物類、事物、內涵構成三個層次，與事物、內涵相對應的是實體和屬體。

在数据世界，即计算机世界中，与三个层次对应的概念分别是文件、记录和字段。三个世界的类比关系可由表 1.1、1.2 表示。

表 1.1 三个世界的类比关系

现实世界	信息世界	数据世界
事物类	实体集	文件
事物	实体	记录
内涵	属性	字段

表 1.2 数据世界的事物表示形式

姓名	性别	出生日期	籍贯	地址
张三	男	1969.11	北京	北京师范大学
李四	男	1966.08	山西	北京大学
...	...	...	...	...

认识到了现实世界中的事物可以表示成计算机世界的的数据，计算机就有了进行数据处理的基础。数据处理正是对各种形式的数据进行收集、储存、加工和传播的一系列活动的总和。其目的是从大量的、原始的数据中抽取、推导出对人们有价值的信息，作为行动和决策的依据；是为了借助计算机科学地保存和管理复杂的大量的数据，以便人们能方便而充分地利用这些宝贵的信息资源。比如说，对天文观测的数据处理，可以预报天气的变化；对经济运行中产生的数据处理，可以帮助制定科学的经济政策；当计算机保存并处理了图书馆的书刊数据时，就可以迅速从浩瀚的书海中迅速找到所需的资料。借助于计算机网络，数据处理可以使人们在社会活动中如虎添翼。如在北京可以迅速得知上海的商业行情；从中国可以得到美国的某篇文献的原文。这一切均来自于计

算机的对电子数据的处理功能。

现在让我们来思考一下，上文中所述及的计算机完成的种种功能如何才能实现呢？要想对数据进行处理，首先要作的是对数据的分类、组织、储存、索引和维护等。这些工作是数据处理的基本环节，这些工作统称为数据管理。

2. 数据管理技术的发展

数据管理的水平是和计算机硬件、软件的发展相适应的，是随着计算机技术的发展而发展的。人们的数据管理技术经历了三个阶段的发展：

- 人工管理阶段；
- 文件系统阶段；
- 数据库系统阶段。

1. 人工管理阶段

这一阶段，大致是在 50 年代中期之前。此时计算机技术相对落后。这时的计算机主要用于科学计算。硬件方面，计算机的外存只有磁带、卡片、纸带，没有磁盘等直接存取的存储设备，存储量非常小；软件方面，没有操作系统，没有高级语言，数据处理的方式是批处理，也即机器一次处理一批数据，直到运算完成为止，然后才能进行另外一批数据的处理，中间不能被打断，原因是此时的外存如磁带、卡片等只能顺序输入。

这一阶段数据管理的特点是：

- (1) 数据不保存。在需要计算时，利用卡片、纸带等将

数据输入，经过运算得到运算结果，数据处理的过程就结束了。

(2) 数据不能独立。数据是作为输入程序的组成部分，即程序和数据是一个不可分隔的整体，数据和程序同时提供给计算机运算使用。对数据进行管理，就像现在的操作系统可以以目录、文件的形式管理数据。程序员不仅要知道数据的逻辑结构，也要规定数据的物理结构，程序员对存储结构，存取方法及输入输出的格式有绝对的控制权，要修改数据必须修改程序。要对 100 组数据进行同样的运算，就要给计算机输入 100 个独立的程序，因为数据无法独立存在。

(3) 这一时期，尚没有文件的概念。数据的组织完全由程序员自行设计。即使人们发现了这样作的弊病，也无可奈何。因为此时计算机的外存能力是很弱的。

(4) 数据是面向应用的。一组数据对应一个程序。不同应用的数据之间是相互独立、彼此无关的，即使两个不同应用涉及到相同的数据，也必须各自定义，无法相互利用，互相参照。数据不但高度冗余，而且不能共享。

综上所述，所以有人也称这一数据管理阶段为无管理阶段。

2. 文件系统阶段

从 50 年代后期到 60 年代中期，数据管理发展到文件系统阶段。此时的计算机不仅用于科学计算，还大量用于管理。外存储器有了磁盘等直接存取的存储设备。在软件方面，操作系统中已有了专门的管理数据软件，称为文件系统。从处

理方式上讲,不仅有了文件批处理,而且能够联机实时处理,联机实时处理是指在需要的时候随时从存储设备中查询、修改或更新,因为操作系统的文件管理功能提供了这种可能。这一时期的特点是:

(1) 数据长期保留。数据可以长期保留在外存上反复处理,即可以经常有查询、修改和删除等操作。所以计算机大量用于数据处理。

(2) 数据的独立性。由于有了操作系统,利用文件系统进行专门的数据管理,使得程序员可以集中精力在算法设计上,而不必过多地考虑细节。比如要保存数据时,只需给出保存指令,而不必所有的程序员都还要精心设计一套程序,控制计算机物理地实现保存数据。在读取数据时,只要给出文件名,而不必知道文件的具体的存放地址。文件的逻辑结构和物理存储结构由系统进行转换,程序与数据有了一定的独立性。数据的改变不一定要引起程序的改变。保存的文件中有 100 条记录,使用某一个查询程序。当文件中有 1000 条记录时,仍然使用保留的这一个查询程序。

(3) 可以实时处理。由于有了直接存取设备,也有了索引文件、链接存取文件、直接存取文件等,所以既可以采用顺序批处理,也可以采用实时处理方式。数据的存取以记录为基本单位。

上述各点都比第一阶段有了很大的改进。但这种方法仍有很多缺点,主要是:

(1) 数据冗余大。当不同的应用程序所需的数据有部分

相同时，仍需建立各自的独立数据文件，而不能共享相同的数据。因此，数据冗余大，空间浪费严重。并且相同的数据重复存放，各自管理，当相同部分的数据需要修改时比较麻烦，稍有不慎，就造成数据的不一致。比如，学籍管理需要建立包括学生的姓名、班级、学号等数据的文件。这种逻辑结构和学生成绩管理所需的数据结构是不同的。在学生成绩管理系统中，进行学生成绩排列和统计，程序需要建立自己的文件，除了特有的语文成绩、数学成绩、平均成绩等数据外，还要有姓名、班级等与学籍管理系统的数据文件相同的数据。数据冗余是显而易见的，此外当有学生转学走或转来时，两个文件都要修改。否则，就会出现有某个学生的成绩，却没有该学生的学籍的情况，反之亦然。如果系统庞大，则会牵一发而动全身，一个微小的变动引起一连串的变动，利用计算机管理的规模越大，问题就越多。常常发生实际情况是这样，而从计算机中得到的信息却是另一回事的事件。

(2) 数据和程序缺乏足够的独立性。文件中的数据是面向特定的应用的，文件之间是孤立的。不能反映现实世界事物之间的内在联系。在上面的学籍文件与成绩文件之间没有任何的联系，计算机无法知道两个文件中的哪两条记录是针对同一个人的。要对系统进行功能的改变是很困难的。如在上面的例子中，要将学籍管理和成绩管理从两个应用合并成一个应用中，则需要修改原来的某一个数据文件的结构，增加新的字段，还需要修改程序，后果就是浪费时间和重复工作。此外，应用程序所用的高级语言的改变，也将影响到文

件的数据结构。比如 BASIC 语言生成的文件, COBOL 语言就无法如同是自己的语言生成的文件一样顺利地使用。总之数据和程序之间缺乏足够的独立性是文件系统的-一个大问题。

文件管理系统在数据量相当庞大的情况下, 已经不能满足需要。美国在 60 年代进行阿波罗计划的研究。阿波罗飞船由约 200 万个零部件组成。分散在世界各地制造。为了掌握计划进度及协调工程进展, 阿波罗计划的主要合约者罗克威尔 (Rockwell) 公司曾研制了一个计算机零件管理系统。系统共用了 18 盘磁带, 虽然可以工作, 但效率极低, 维护困难。18 盘磁带中 60% 是冗余数据。这个系统一度成为实现阿波罗计划的严重障碍。应用的需要推动了技术的发展。文件管理系统面对大量数据时的困境促使人们去研究新的数据管理技术, 数据库技术应运而生了! 例如, 最早的数据库管理系统之一 IMS 就是上述的罗克威尔 (rockwell) 公司在实现阿波罗计划中与 IBM 公司合作开发的, 从而保证了阿波罗飞船 1969 年顺利登月。

3. 数据库系统阶段

从 60 年代后期开始, 数据管理进入数据库系统阶段。这一时期用计算机管理的规模日益庞大, 应用越来越广泛, 数据量急剧增长, 数据要求共享的呼声越来越强。这种共享的含义是多种应用、多种语言互相覆盖地共享数据集合。此时的计算机有了大容量磁盘, 计算能力也非常强。硬件价格下降, 编制软件和维护软件的费用相对在增加。联机实时处理的要求更多, 并开始提出和考虑并行处理。

在这样的背景下，数据管理技术进入数据库系统阶段。

现实世界是复杂的，反映现实世界的各类数据之间必然存在错综复杂的联系。为反映这种复杂的数据结构，让数据资源能为多种应用需要服务，并为多个用户所共享，同时为让用户能更方便地使用这些数据资源，在计算机科学中，逐渐形成了数据库技术这一独立分支。计算机中的数据及数据的管理统一由数据库系统来完成。

数据库系统的目标是解决数据冗余问题，实现数据独立性，实现数据共享并解决由于数据共享而带来的数据完整性、安全性及并发控制等一系列问题。为实现这一目标，数据库的运行必须有一个软件系统来控制，这个系统软件称为数据库管理系统 (Database Management System, DBMS)。数据库管理系统将程序员进一步解脱出来，就像当初操作系统将程序员从直接控制物理读写中解脱出来一样。程序员此时不需要再考虑数据中的数据是不是因为改动而造成不一致，也不用担心由于应用功能的扩充，而导致程序重写，数据结构重新变动。在这一阶段，数据管理具有下面的特点，这些特点正是数据库的改进之处：

(1) 数据结构不是面向单一的应用，而是面向全组织。仍以学校管理为例，要想避免数据冗余和数据程序之间的依赖性，就要将学生学籍及成绩两类不同的数据之间彼此建立关系。如图 1.1。

当需要增加新的应用，比如学生的体质状况管理，则只要再增加新的联系。

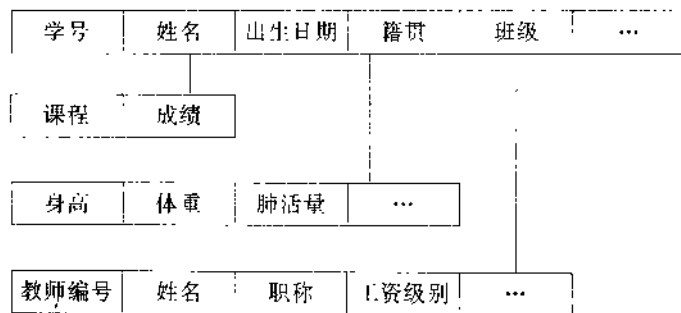


图1.1 关系化的数据

这种思想只是数据库方法的雏形，它从文件内部的记录的结构化，扩大到不同的文件记录之间建立一种联系。但是它还有局限性，因为它还是从应用的角度去看待数据，还应进一步从整个组织的数据结构考虑。假设所考虑的这个组织——学校，就还应该有教师人事信息、教务信息、教学关系等。不同应用考虑的是整个数据集合的某个有用的子集。整个组织的数据是结构化的。这样描述数据时不仅描述数据本身，还有描述数据之间的联系。

数据的结构化是数据库主要特征之一。这是数据库与文件系统的根本区别。至于这种结构化是如何实现的，则与数据库系统采用的数据模型有关，后面会有较详细的描述。

2. 数据冗余小，易扩充。数据库从整体的观点来看待和描述数据，数据不再是面向某一应用，而是面向整个系统。这样就减小了数据的冗余，节约存储空间，缩短存取时间，避免数据之间的不相容和不一致。对数据库的应用可以很灵活，面向不同的应用，存取相应的数据库的子集。当应用需求改