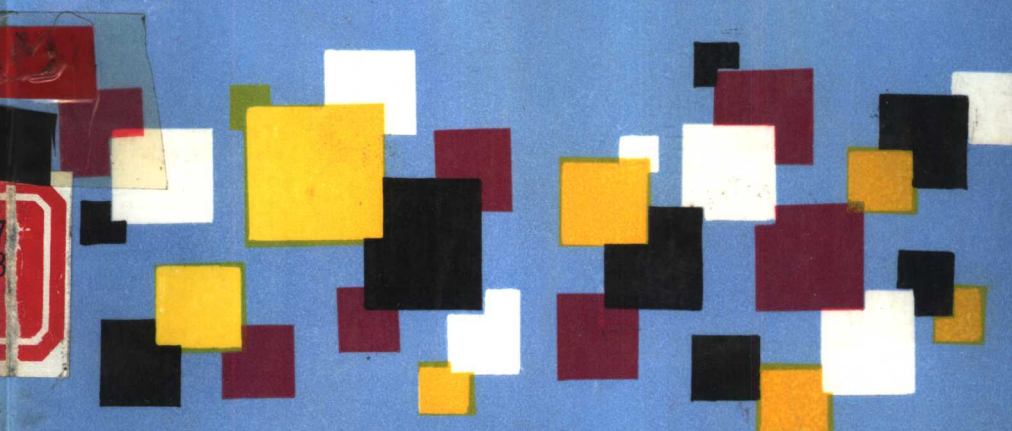


数据库技术

陈建新 叶飞跃 马玉书 仝兆岐 编著



石油工业出版社

数据库技术

陈建新 叶飞跃 马玉书 仝兆岐 编著

石油工业出版社

(京)新登字 082 号

内 容 提 要

本书系统而又全面地介绍了数据库技术的基本概念、基本理论、实现方法和技术,以及数据库技术的最新发展及新一代数据库。书中内容安排合理、选材适当,是一本通俗易懂、深入浅出的数据库技术普及读本。它可作为大学本科非计算机专业学生的教材和工程技术人员及管理干部的实用参考书,也可供各类数据库技术培训班、专科学校计算机专业作为数据库课程的教材使用。

图书在版编目(CIP)数据

数据库技术 / 陈建新等编著

—北京: 石油工业出版社, 1995.2

ISBN 7-5021-1302-9

I. 数…

II. 陈…

III. 数据库—基本知识

IV. TP311.43

石油工业出版社出版

(100011 北京安定门外安华里2区1号楼)

石油工业出版社印刷厂排版印刷

新华书店北京发行所发行

*

850×1168 毫米 32 开 $7\frac{1}{2}$ 印张 197 千字 印 1-4000

1995 年 2 月北京第 1 版 1995 年 2 月北京第 1 次印刷

定价: 10.00 元

前 言

数据库技术作为数据管理的最新技术，在我国各个行业中已得到了广泛的普及和应用。dBASE、Foxbase、Sybase、ORACLE 等数据库管理系统的普及几乎已到了家喻户晓，人人皆知的程度，它为社会的信息化、管理信息系统的开发、办公室自动化水平的提高、计算机辅助设计等许多方面作出了贡献。目前国内已有一些数据库教材出版，然而这些书籍，或是偏重某个具体的数据库管理系统的使用方法介绍和手册，缺乏系统的数据库理论知识；或是适用于计算机专业人员的著作和教材，内容过深，预备知识和先行课程要求较高，难以被一般人接受。本书填补了这两类教材之间的空白，是一本适合于非计算机专业的数据库设计 and 应用人员进一步提高数据库知识和理论水平，通俗易懂，深入浅出的数据库技术普及读本。

本书系统全面地介绍了数据库技术有关的基本概念、基本原理和实现方法及技术。在内容上，它既具有传统数据库技术的知识，又介绍了数据库的最新发展。本书第一章简要地叙述了数据库系统的发展历史和数据库基本概念，对数据库的特点、系统组成及其在石油工业中的应用作了概括的介绍。第二章扼要介绍了如何用数据库来表达现实世界以及目前流行的几种数据库模型。通过这一部分的学习，读者将掌握把现实世界中所关心的对象转换为计算机中的数据结构的方法，以及在常用的网状、层次、关系三种数据库模型中如何组织数据，并对数据库结构有一个全面的了解。第三章介绍计算机中数据的存储结构，从存储介质，文件，到数据库中如何组织数据逐一加以介绍，使读者了解数据在计算机的各种存储设备上是如何存放的。鉴于关系数据库具有许多显著的优点并已日趋成熟普及，我们重点在第四章中讨论了关

系方法、关系数据库的基本结构、关系数据语言和关系数据库理论，这有助于对关系数据库的理解，且可为今后的数据库设计奠定理论基础。第五章全面介绍了数据库设计的技术、方法和一般步骤。第六章介绍了开发大型分布式数据库依赖的基础——计算机网络技术，以及分布式数据库产生的背景、概念、特点和主要实现技术。第七章介绍数据库技术的新发展和新一代数据库，指出了传统的数据库存在的不足和新一代数据库的发展途径，以及各种新型的专用数据库、智能数据库和面向对象数据库等各类新一代数据库，便于读者了解和跟踪当前数据库发展的最新动态和方向。

本书第一章由叶飞跃、陈建新编写，第二章至第五章由叶飞跃编写，第六章由仝兆岐、马玉书和叶飞跃编写，第七章由马玉书编写。全书由马玉书统稿和审校。

中国人民大学数据与知识工程研究所所长王珊教授和陈红讲师认真审阅了全部原稿，并提出了许多宝贵意见和修改建议，在此表示衷心的感谢。

由于作者水平有限，书中难免有许多不妥之处，恳请有关专家和读者批评、指正。

作者

1993.9.

目 录

第一章 概述	(1)
§ 1.1 数据库的产生和发展	(1)
§ 1.2 数据库系统的构成	(10)
§ 1.3 数据库系统的特点	(21)
§ 1.4 数据库技术在石油工业中的应用	(27)
第二章 现实世界的描述	(30)
§ 2.1 信息的三个世界	(30)
§ 2.2 数据模型	(32)
§ 2.3 E-R 模型	(35)
§ 2.4 层次模型	(40)
§ 2.5 网状模型	(48)
§ 2.6 关系模型	(54)
第三章 数据的存储结构	(59)
§ 3.1 存储介质	(59)
§ 3.2 文件的存储结构	(62)
§ 3.3 数据库中的存储结构	(74)
第四章 关系模型数据库	(85)
§ 4.1 关系的基本概念	(85)
§ 4.2 关系数据语言	(91)
§ 4.3 SQL 语言	(100)
§ 4.4 关系系统及 ORACLE 关系系统简介	(117)
§ 4.5 关系模式的规范化(1NF~3NF)	(123)
第五章 数据库设计	(141)
§ 5.1 数据库设计概述	(141)
§ 5.2 数据库设计的一般步骤	(144)

§ 5.3	需求分析	(148)
§ 5.4	概念设计	(151)
§ 5.5	逻辑设计	(156)
§ 5.6	物理设计	(162)
§ 5.7	数据库实施	(163)
§ 5.8	数据库的运行和维护	(165)
第六章	计算机网络和分布式数据库	(166)
§ 6.1	数据通信基础	(166)
§ 6.2	计算机网络及拓扑结构	(170)
§ 6.3	网络协议	(171)
§ 6.4	网络互连	(176)
§ 6.5	分布式数据库基本概念	(178)
§ 6.6	分布式数据库管理系统 DDBMS	(184)
§ 6.7	客户/服务器技术	(187)
第七章	数据库技术的新发展和新一代数据库	(189)
§ 7.1	数据库技术的发展与展望	(189)
§ 7.2	专用数据库	(194)
§ 7.3	智能数据库	(199)
§ 7.4	面向对象数据库	(219)
参考书目		(234)

第一章 概 述

随着计算机在数据处理领域的普及和广泛应用，数据处理技术发展得越来越快，处理的数据量也在急剧增加。面对庞大的数据，如何在计算机中有效地组织、存储和利用它们，是数据管理的重要任务。数据库技术就是在 60 年代初期发展起来的一种最新的数据管理技术，目前它已成为计算机学科的一个重要分支。

§ 1.1 数据库的产生和发展

自从世界上第一台电子计算机问世以来，人们就开始利用计算机来进行各种数据处理。所谓数据处理 (data processing) 指的是对数据进行收集、组织、加工、储存、抽取和传播的过程。数据处理的目的是要从大量的、零乱的、难以理解的数据中，获得对某个特定的应用领域来说是有价值有意义的信息，作为管理、决策的依据。例如气象预报中的数据处理和石油地震资料的数据处理，都是先采集原始数据，然后按预先设计的数学模型进行处理，得到的结果提供给专业人员作为预报或预测的依据。

数据处理已经历了漫长的历程。早期人们利用各种初级的计算工具，如算盘、手摇计算机等来进行计算，这是手工数据处理的初级阶段。中期，到 1880 年美国进行人口统计，采用 Herman Hollerith 发明的卡片制表机编制人口普查表，利用穿孔卡片来存储信息，用机械方法进行数据处理，开始了机械数据处理的阶段。后期，本世纪 40 年代电子计算机的发明，则使数据处理进入了电子数据处理的新时代。由于电子计算机处理速度快、存储容量大、自动化程度高，使数据处理工作得到了飞速的发展。数据处理的信息量急剧上升，数据的储存形式也多样化，

有数字、文字、声音、图形、图像等等，数据的结构亦越来越复杂。如何管理这些数据就成为一个极其重要的问题。

数据管理 (data management) 是指对数据的组织、存储、检索、更新和维护等工作。它是数据处理的核心。高效的组织方式、存储结构、检索手段和安全措施是数据管理研究的主要内容。从 60 年代开始，由于人们对数据的要求越来越高，对数据管理技术提出了更高的要求。同时大容量直接存取介质磁盘的出现也为数据管理技术的发展奠定了物质基础。于是人们在文件系统的基础上，开发出了各种各样专用的数据管理软件包，并由此逐渐发展成为一种通用的系统软件——数据库系统。数据库就是 60 年代作为数据管理的最新技术登上数据处理舞台的。在讨论数据库的发展历史之前，先让我们回顾一下数据管理技术的发展历史。

§ 1.1.1 数据管理技术的发展

数据管理技术的发展经历了如下三个阶段。

1. 人工管理阶段

这是计算机用于数据处理的初期阶段 (50 年代及以前)。计算机出现之后，最早期的用户必须“软硬兼施”，计算一道题目需要重新搭接硬件线路。后来，计算机较普及之后，用户也需是专门的程序设计人员，他们在编写程序时，需要自己编制外部设备输入输出程序，亲自设计数据的存储结构及存储分配。此后又出现了简单的操作系统，可提供基本的外部设备输入输出程序和通用的存取数据的方法，使输入输出、存储分配标准化，这样，使用户再不必为外部设备和存储分配担心。这一阶段数据管理具有如下的共同特点：用户在编制数据处理程序时，需要考虑到数据的组织方式，自行设计存储结构；数据的逻辑组织和物理组织往往是一致的，程序和数据混为一体；在需要存取数据时，直接按物理存储位置进行访问；当数据的物理结构或存储介质改变时，其应用程序必须重新编写，不具备数据的独立性。在这一阶段，数据一般不长期保存，没有专门软件对数据进行管理，一组数据

对应一个应用程序，其相互关系可由图 1.1 所示。应用程序负责数据的管理，数据在不同的应用之间无法共享，因而数据存在着严重的冗余（重复存储）现象。

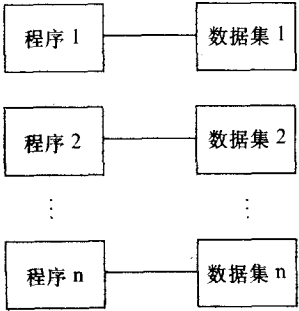


图 1.1 人工管理阶段

2. 文件系统阶段

人工管理阶段，其应用程序依赖于数据的物理组织，数据独立性差，不同应用之间的共同数据重复地多次存储，造成很高的数据冗余，给数据维护带来了许多困难。到 50 年代后期，随着操作系统的逐步完善，出现了文件系统，有了专门负责数据管理的软件。它承担各种文件中数据的存取，可用简单的描述性语言进行数据处理，用户不再需要与存储结构和数据的类型打交道。特别是 60 年代初期，直接存取介质的出现，使得索引、散列 (Hash) 等技术充分发展，文件系统发展得相当成熟。用户在编制应用程序时，只需涉及数据的逻辑结构，不必再与物理设备打交道，文件系统负责逻辑结构与物理结构之间的转换。文件按不同的组织方式分为顺序文件、索引文件和随机文件，从而提供了一些标准的顺序存取或直接存取的方法。一个应用程序可以建立、维护和处理一个或多个文件。按不同组织方式的文件所提供

的存取方法，可顺序地或直接地按记录为单位来存取数据，而无需关心使用什么外部设备或具体的物理地址。这一阶段数据与程序之间的对应关系如图 1.2 所示。

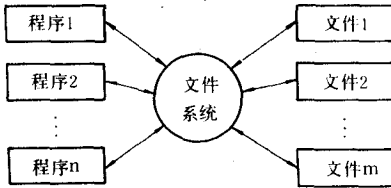


图 1.2 文件系统阶段

在这一阶段，文件系统作为应用程序和数据文件之间的一个接口，起到了将数据的逻辑结构与物理结构独立开的作用；并提供了统一的存取方法。当物理结构或者外部设备变化时，应用程序不需修改照常可以运行，这就提高了数据的物理独立性，并实现了以文件

为单位的数据共享。但这一阶段仍然存在许多问题，例如数据共享的级别还比较低，未实现记录或数据项级别的共享，数据的冗余还相当严重，数据的逻辑结构还是面向应用进行设计的，即一个文件基本上对应于一个应用程序，数据的独立性不高，而且文件中记录结构单一，文件与文件之间缺乏联系，表示复杂数据结构的能力有限。不过文件系统在数据管理中曾起过重要的作用，至今它仍然是一种使用广泛、相当有用的一种数据管理方法。

3. 数据库管理阶段

数据库管理技术出现于 60 年代。这个时期软件上索引、散列技术已发展得相当成熟，硬件上出现了大容量的直接存取设备磁盘。由于文件系统的局限，用户对数据管理提出了更高的要求，要求具有更高的数据共享、更高的数据独立性、表示更复杂的数据结构，因而导致了数据库技术的诞生。数据库实际上是一个存储在计算机内的所有相关数据构成的集合。其基本思想是对所有用户数据实行统一的、集中的管理、操作和维护。数据独立于程序而存在，并可提供给各类不同应用共享使用。如图 1.3 所示。

数据库技术有别于文件系统
系统的最大特点是数据的共享。
数据不再属于某个特定的应用，
而是面向整体来组织数据。虽然
在操作系统下，数据仍是以多个
文件的形式存在，但这些文件内
部的组织已经与文件系统中的文
件组织不同，文件之间也已不
再是毫无关联

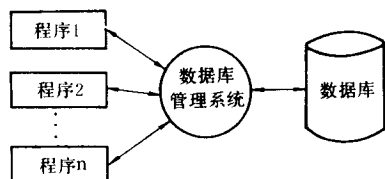


图 1.3 数据库系统阶段

的，常采用某种数据模型将全部数据文件组织为一个结构化的整体，数据库系统提供它们之间的交叉访问的手段，这就大大减少了数据的冗余度。数据库技术还提供了较强的数据独立性、安全性和完整性，使得数据受应用程序的影响大大减少，数据的可靠性得到了保证。另外，数据库系统还提供统一的数据操纵手段，为用户提供了方便的统一的用户接口，从而使数据的应用更为有效。

数据库是一项新型的数据管理技术，由于它的出现，使得数据处理以程序为中心开始向以数据为中心转变。传统的程序设计语言都是以程序为中心的，数据或文件只是程序加工的对象，但在数据库系统中，它是以数据为中心的，围绕着数据完成查询、更新等操作，使得数据处理的效率大大提高。数据库系统现在已成为现代管理信息系统不可缺少的强有力的工具。

§ 1.1.2 数据库技术的发展

60 年代后期发展起来的数据库技术，最主要的有三种模型：网状模型、层次模型和关系模型。这三种模型以如下三个事件为标志，这也是数据库发展史上的三个里程碑：

· 1969 年美国数据系统语言协商委员会 CODASYL (Conference On Data System Language) 的下属组织数据库工作小组 (DBTG) 发表了“DBTG 报告”，提出了以网状模型为基础的一个数据库系统方案，并形成了数据库的一个规范。

· 1968 年美国 IBM 公司推出了 IMS 系统，这是一个以层次模型为基础的数据库管理系统。

· 从 1970 年起，美国 IBM 公司的 E.F.科德发表了一系列关系数据库的论文，为新一代关系数据库的诞生奠定了基础。自此，数据库技术的发展逐渐以关系模型为主导地位。

下面就这三个事件的发生，分别把三种数据库系统的产生过程简介如下：

1. DBTG 报告

“DBTG 报告”是一种网状模型数据库系统方案。它的产生可以追溯到 1960 年。当时通用电气公司的 C.W.贝奇曼（世界公认的网状数据库奠基人之一）在他发表的两篇论文中提出了用链式结构将不同文件中不同类型的相关联的记录连接在一起的思想，并在 1963 年用这种方法开发了著名的 IDS 系统(Integrated Data Storage System)，这就是后来的网状数据的最早雏形。后来匹兹堡的 W.西蒙斯接触到了 IDS 系统，在他的积极鼓动下，美国数据系统语言协商委员会 (CODASYL) 于 1965 年 10 月成立了一个下属组织“表处理任务分队”(LPTF —— List Processing Task Force)。CODASYL 原是一个由制造商和用户集团组成的自发组织，早先从事 COBOL 语言的研究、开发和规范。它成立的表处理分队专门从事数据库的研究，并采纳了贝奇曼的方法。表处理分队 1967 年改名为数据库工作小组 DBTG(Data Base Task Group)，于 1968 年 1 月发表了“用扩展的 COBOL 处理数据库的报告”，1969 年 10 月编制了一套语言规范——“数据库建议书”，并于 1971 年 4 月将其修改名为“Report”的修订报告，为网状数据库提出一个完整的系统方案和语言规范。这就是著名的“DBTG 报告”，也称 CODASYL 报告。这个报告标志着以 C.W.贝奇曼为首的网状数据库系统的发展达到了巅峰阶段，贝奇曼因此而获得了 1973 年计算机界的最高奖——图灵奖，他的图灵奖的演讲题目是“作为导航员的程序员”。随后，DBTG 组织就解散了。需要指出的是，自 1971 年

“DBTG 报告”发表以后，CODASYL 已将这个文本作了多次修改，特别是 1975 年根据 ANSI / SPARC 研究小组内部报告的建议采用三级数据库系统结构，将概念模式和内模式的定义明确地分开。后来，到 70 年代中期和末期，有人提出在 CODASYL 方案的基础上开发标准数据库，但由于关系数据库的兴起，这个建议未能被采纳。

2. IMS 系统

IMS (Information Management System) 系统是 IBM 公司开发的一种层次数据库管理系统，是商品化的数据库系统中使用得较普遍的一种。该系统早期影响较大，用户较多，层次模型数据库的重要性至少部分是受该系统存在的影响。

在 60 年代初，当各公司都在开发自己的数据库管理系统的时候，IBM 公司在 IMS 系统上下了很大的投资，于 1968 年率先推出了第一版本 IMS / 360 Version I，1971 年推出第二版本 IMS / 360 Version II，1974 年推出第三版本 IMS / VS (IMS / Virtual Storage)。IMS 是在 OS / VS 操作系统下运行的，最早的第一版本是一种纯层次的数据库管理系统，即只能表示现实世界的层次结构，这就使数据库的应用受到很大的限制。而且当时由于 DBTG 的工作，IBM 公司发觉在数据库标准化道路上推行的是另一种数据模型，这是与他们的商业利益相抵触的，因而当“DBTG 报告”于 1969 年和 1971 年两次提交 CODASYL 组织并得到通过时，都受到了来自 IBM 公司一些人联合一致的反对。针对 IMS 只能表示层次结构的缺陷，IMS 的第二版本增加了逻辑数据库的概念，利用不同物理数据库片段型之间的逻辑双亲来表示现实世界中的网状结构。第三版本的 IMS 系统，在原来的基础上增加了数据库恢复、并发控制和辅助索引等功能，并提供了相应的一组软件产品，如数据字典、查询接口、应用生成器和设计辅助工具等，使性能进一步完善。目前 IMS 系统仍然还是一种使用较广泛的数据库管理系统之一。

3. 关系数据库

说来也许是历史的巧合，正是 DBTG 解散，CODASYL 完成了网状数据库语言规范的时候，IBM 公司的 E.F.科德从 1970 年起陆续发表了一系列论文，提出关系方法在数据库管理系统上的应用，开始了数据库发展史上的又一个进程。

关系方法是建立在集合论的基础上的。最早将这种方法用于数据处理的是 1962 年 CODASYL 发表的一篇文章“信息代数”，尔后于 1968 年 David Child 在 7090 计算机上实现了这种方法，称之为“集合论的数据结构”(set-theoretic data structure)，直到 1970 年科德才开始比较系统而又严格地对数据的关系模型作出了分析，提出基于关系演算的数据操纵语言 ALPHA 和一整套完备的关系数据理论，奠定了关系数据库的理论基础。科德因此而获得了 1981 年的图灵奖，他接受图灵奖时的演讲题目是“提高生产率的现实基础”，指出了关系方法使用关系数据结构，提供了高度的数据独立性，采用非过程式的数据语言，使程序员完全摆脱了数据管理，数据库系统自动完成诸如存储、检索、修改等操作的具体实现，从而使用户可以专心于他的应用问题，这就大大地提高了生产率。

关系方法虽然理论上早在 70 年代就发展成熟，但关系方法在实现上却遇到较多的困难。直到 80 年代超大规模集成电路的出现，计算机的运算速度和内存容量都得到大大提高，大容量外存也开始进入应用，所有这些才使得关系数据库迅速发展起来。System R 是关系数据库中的典型代表，它是 IBM 公司圣约瑟研究所 1974 年开始设计与研制的一种“实验关系数据库系统”，充分体现了关系方法的特点，采用了许多新颖的设计思想，对用户隐藏了数据的内部结构和存取路径，达到高度的数据独立性。经过多年的努力，System R 于 1979 年终于完成。由于 System R 是第一个公开的关系模型实验数据库系统，因而它的设计思想对其它各种关系数据库影响很大。此后，各家公司陆续推出了各种关系数据库系统，较著名的有 1979 年美国 ORACLE 公司推出的第一个商品化关系数据库管理系统 ORACLE；1981 年由关系

技术公司推出了 INGRES，它原先是美国加州大学伯克莱分校 70 年代初首创的关系数据库，后配置在 VAX 计算机上进入市场；后来 1982 年 IBM 公司推出了 SQL/DS，1985 年又推出 DB2。此外，自 1980 年以来，在微机环境下，dBASE、Foxbase 等很多关系数据库系统也陆续研制成功，虽然这些系统不如前面所提的大型关系数据库系统在存储能力和安全性、完整性、数据恢复及并发控制等方面的优良特性，但却具有成本低廉、易于非计算机专业人员掌握使用的特点。其中最著名的是由 Ashton-Tate 公司 1980 年推出的 dBASE II，后经扩充修改于 1984 年以 dBASE III 的名字发行，这是目前我国国内使用得相当广泛的一种微机数据库系统。

在关系数据库的发展中，值得一提的是用于关系数据库操作的关系数据语言的发展过程。前面提到，科德在提出关系方法时，设计了一种语言——ALPHA，这种语言最终并没有在计算机上真正实现，后来美国加州大学伯克莱分校在研制 INGRES 系统时，以 ALPHA 为母体，修订后作为 INGRES 系统的数据操纵语言，称之为 QUEL，这是一种功能很强，基于一阶元组关系演算的语言。不过 QUEL 最终没有成为关系数据库的主流语言，目前广为流行的关系语言称之为 SQL (Structured Query Language)。SQL 的前身为 SQUARE (Specifying Queries As Relational Expression)，是由 IBM 公司的 Boyce 和 Chamberlin 提出的用于 System R 上使用的语言，后几经发展演化为 SQL，并于 1976 年予以正式发表。SQL 语言由于其优越的特性而被许多系统所采纳，如 ORACLE、DB2 等。SQL 功能丰富、使用方便灵活、语言简洁易学，1986 年 10 月美国国家标准局 (ANSI) 批准了 SQL 作为关系数据库的标准语言，此后在 1987 年国际标准化组织 (ISO) 也作出了同样的选择。使 SQL 成为数据库领域的主流语言，对数据库技术的发展产生了重大的影响。自 SQL 成为国际标准后，各数据库厂家纷纷推出支持 SQL 的软件。如果有一日所有的计算机都采用 SQL 作为共同的数据

存取语言，作为其标准的用户接口，就有可能“使未来的数据库世界连接成为一个统一的整体”，这是有重大意义和十分诱人的，难怪有人把 SQL 采纳为国际标准称之为是一场革命，也有人把这叫做“SQL 一边倒”或“SQL 的时代”。

综观数据库的发展过程，60 年代是数据库技术孕育和诞生的时代，70 年代是层次、网状数据库的时代，80 年代则以关系数据库为主导地位。可以说自 80 年代之后，新研究的数据库系统几乎都是基于关系方法的。目前，数据库技术正向着演绎数据库、工程数据库、面向对象数据库、多媒体数据库、分布式数据库等方向发展，数据库这一领域尚有着广阔的沃土有待去开垦、去耕耘、去发现、去收获。

§ 1.2 数据库系统的构成

§ 1.2.1 数据库系统的基本概念

数据库 (Data Base) 广义地讲可以是任何数据的集合：书里的书，档案馆里的资料，一个地区的地质资料，磁盘、磁带上的数据和程序、歌曲、录像，甚至山洞里的史前壁画等都可以看成是数据库。不过我们这里讨论的数据库含义并没有这样广泛，通常是指计算机中用于数据处理的一种数据管理技术。简单地说，数据库是在计算机中按一定组织方式存储在一起的、相互有关的、为用户共同关心的全部数据的集合。这些数据具有最少的重复（冗余度小），能够同时为多个用户服务（数据共享），数据的存储独立于应用程序（数据独立性），对数据库进行更新和检索等操作时，由系统提供统一的控制方法，并能够保证数据的安全性、完整性和并发一致性。

由于数据库中存储的数据量大，又为多个用户共享使用，因此必须有一套专门的软件来管理数据库，同时负责数据库的建立、数据结构的定义、数据库中数据的更新和查询、多个用户并发访问数据库时的事务调度，并进行安全性和完整性检查，以及