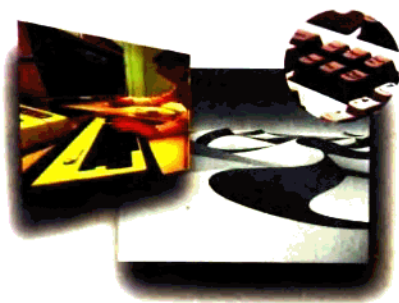




勘探开发工程 数据库系统理论 与 实践

叶飞跃 著



石油大学出版社

北京)

0.9

1

书号	124048
分类号	P618.130.9
种次号	004

勘探开发工程 数据库系统理论与实践

SU17/05

叶飞跃 著



石油0117138



石油大学出版社

图书在版编目(C I P)数据

勘探开发工程数据库系统理论与实践 / 叶飞跃. — 东营: 石油大学出版社, 1998.10

ISBN 7 - 5636 - 1150 - 9

I . 堪… II . 叶… III . 油气勘探 - 数据库系统 IV . TE17

中国版本图书馆CIP数据库核字 (98) 第26055号

勘探开发工程数据库系统理论与实践
叶飞跃 著

出版者: 石油大学出版社 (山东 东营, 邮编 257062)

印刷者: 山东东营新华印刷厂印刷

发行者: 石油大学出版社 (电话 0546-8392563)

开 本: 850×1168 1/32 印张: 5.75 字数: 150 千字

版 次: 1998年9月第1版 1998年9月第1次印刷

印 数: 1—600册

定 价: 9.80元

前 言

数据库作为一门实用而成熟的技术，已经广泛应用于石油勘探开发的各个行业。目前，国内外流行的各种数据库系统几乎全部是基于关系模型的。关系数据库具有统一的数据结构、友好的用户界面和一套完备的关系数据理论，被用户广泛接受。但是，由于关系数据库必须遵守关系规范化理论的要求，在石油勘探开发的特殊工程应用领域有一定的限制。其一，关系中的任一属性必须是不可分割的，对于勘探开发所涉及的种类复杂的数据，例如矢量、矩阵数组等数据，关系系统不能够直接支持这类数据，其数据结构的组织能力不强；其二，关系模型中不允许对象的标识，各个元组均不能相同，而工程应用中的数据有可能两组相同的值代表的却是不同的内容，例如测井曲线数据，虽然不同的曲线数据值不一样，但不可以以值的大小来区分，而应该有一种系统的标识方法，这在关系数据库中是无法表示的；其三，关系中的数据无先后次序，而勘探开发应用有时希望按顺序存放数据，关系模型却不便实现；此外，关系数据操作不支持工程数据的操作，也不具备版本控制的能力，从而使应用范围受到限制。

针对以上问题，结合石油勘探开发工程数据的特点，我们研究了在关系模型基础上建立勘探开发工程数据模型，改造了关系数据库系统的 1NF 限制，扩充关系为嵌套关系结构，用指针的方法建立关系之间的联系来实现嵌套关系的存储，使得原关系模型中的一个属性可定义为另一个关系。利用嵌套关系解决矢量、数组、文本、图形等数据的存储问题，而且可定义动态数据类型，提供了灵活组织数据的手段；关系中的各个元组亦可进行标识，使相同的对象用标识的方法在勘探开发工程数据库中方便地进行

存储；在扩展的嵌套关系数据模型下研究并定义了一套与 SQL 语言兼容的数据定义、数据查询、数据更新和数据控制语言，此外，为适应勘探开发工程数据多版本的存储问题，还研究了一种行之有效的数据库版本控制策略和方法。解决了大量数据重复存储的问题，提高了存储设备的利用效益。为石油勘探开发工程应用提供了方便、高效、统一的数据存储结构和数据存储方法，解决了常规的关系数据库系统下无法解决的一些工程数据的数据管理问题，为石油工业勘探开发数据库的建库和应用研究提供了有效的方法和手段。

Triebase是用C语言，基于SQL语言，在 ORACLE 数据库上，用扩充的嵌套关系数据模型开发的一套完善的工程数据库系统。提供了16种适合于勘探开发工程应用的新的数据类型，设计了一套在新的嵌套关系模型下的数据操作语言，规定了标准的数据操作方法，开发了交互和预编译用户界面，并为勘探开发应用提供了版本控制的手段。它具有与 ORACLE 系统 SQL 语言完全兼容、统一的数据操作方式、对用户透明性、较高的数据独立性、高效的存储空间利用和速度、可移植性强等特点。

本书全面地介绍了勘探开发工程数据库系统及其原理和实现方法，对关系模型的特征、在石油勘探开发特殊工程应用领域的局限性进行了深入的讨论，并提出了解决的可行办法。书中还设计了在关系数据库基础上进行模型扩充，以嵌套关系为基础的包括数据结构、数据操作和数据约束控制的一套完整的工程数据模型，并讨论了嵌套关系下的数据操作语义和算法，也为用户提供了在此工程数据库上进行查询、修改、录入和删除的数据语言 EOSQL 语言。最后一章还介绍了以此原理为原型开发的一个实际的工程数据库系统 Triebase，该系统在实际应用中得到了验证和考验，系统是完善和可靠的，并能解决实际问题。

本书是石油天然气总公司“八五”重点科研攻关项目“石油勘探开发数据库建库和应用研究”子课题“基于ORACLE的勘探开发

工程数据库”的研究成果之一，得到了总公司的支持和资助。项目的研究和开发人员还包括高延文、段友祥、董国卿、张文东和刘志宏等同志，他们的工作对本书的形成起到了很大影响，在此向他们表示衷心的感谢。

中国科学院计算所的倪光南院士和石油大学石油工程系的张琪教授认真仔细地阅读了本书的手稿，并提出了许多中肯的修改意见，在此一并向他们表示衷心的感谢。

作者

1998年5月

目 录

第一章 数据库概述	1
§1.1 数据库的产生与发展.....	2
§1.2 现实世界的描述.....	11
§1.3 三种数据模型.....	20
§1.4 数据库系统结构.....	26
第二章 关系数据库	38
§2.1 关系的基本概念.....	38
§2.2 关系数据语言.....	45
§2.3 SQL 语言.....	52
第三章 传统数据库技术的局限	69
§3.1 数据库系统及对石油工业的影响.....	70
§3.2 关系方法的局限.....	73
§3.3 解决途径.....	77
第四章 嵌套关系数据模型	83
§4.1 勘探开发工程数据模型.....	83
§4.2 结构化数据类型及实现.....	86
§4.3 图形数据的存储结构.....	89
§4.4 版本管理.....	94
§4.5 数据字典.....	97
第五章 嵌套关系的数据操作	101
§5.1 并运算.....	102

§5.2	交运算	103
§5.3	筛选	105
§5.4	展接	107
第六章	EOSQL 语言	120
§6.1	EOSQL 语言语法	120
§6.2	表的定义及控制命令	124
§6.3	表与视图的授权命令	129
§6.4	数据插入、修改语句	132
§6.5	数据查询语句	136
§6.6	图形操作语句	140
第七章	TRIEBASE 系统介绍	150
§7.1	系统结构	150
§7.2	系统特点	155
§7.3	系统的安装与启动	159
§7.4	用户界面	160
§7.5	TRIEBASE 系统演示实例	162

第一章 数据库概述

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。由于数据库技术的出现，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。

数据库是计算机中按一定组织方式存储在一起、相互有关、为用户共同关心的全部数据的集合。由于数据库中存储的数据量大，持续时间长又为多个用户共享使用，因此有必要配备一套专门的软件来管理数据库，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，以及系统性能的监测、数据库的转储和故障后的恢复等，完成这些任务的软件称之为数据库管理系统（Database Management System）。

§1.1 数据库的产生与发展

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、存取和传播的过程。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，其相互关系如图 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。

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。

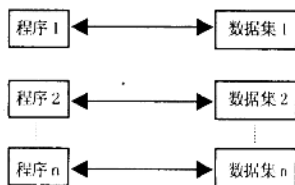


图 1.1 人工管理阶段

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。

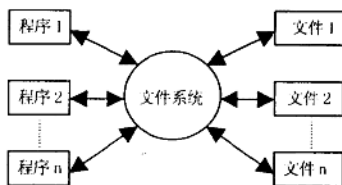


图 1.2 文件系统阶段

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，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且集中的管理、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，为用户提供了方便统一的用户接口，从而使数据的应用更为有效。

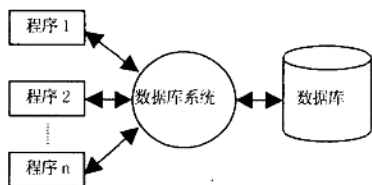


图 1.3 数据库系统阶段

1.1.1 数据库技术的发展

60 年代后期发展起来的数据库技术，主要有三种模型：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 C·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织的自发组织,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顶峰阶段,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 年两次提交给 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 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。

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; 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