

数据库技术及其最新发展

罗 可*

长沙电力学院 (长沙 410077)

湖南大学 (长沙 410082)

摘 要 该文介绍了数据库技术的产生和发展,论述了数据库领域的研究成果和发展趋势,提出了今后数据库技术研究的若干热点。

关键词 数据库 数据仓库 面向对象 多媒体 DBMS

The Database Technology and Its Current Development

Luo Ke

(Hunan University, Changsha 410082)

Abstract: This paper introduces the birth and the development of a database technology. It states the researches and the tendency in the field of database. It predicts the popular problems in future.

Keywords: Database, Data warehouse, Object-oriented, Multimedia, DBMS

1 引言

计算机技术能迅速普及,数据库和计算机网络技术起了非常重要的作用。数据库技术是研究如何科学地组织和存储数据,如何高效地检索数据和进行数据处理的一门学科,它是当代信息系统的基础,有信息网络无数据库源,就好象有路无车一样。

从 50 年代中期开始,计算机应用由科学研究部门扩展到了企业、行政等部门,数据处理很快上升为计算机应用的重要方面,到 60 年代,磁盘技术取得了发展,计算机用于管理的规模更加庞大,数据量急剧增长,为了提高效率,人们着手开发和研制更加完美的数据管理方式,提出了数据库的概念。美国通用电气公司 Bachman 等人于 1964 年成功开发了 IDS (Integrated Data Store),IDS 奠定了网状数据库的基础,IBM 公司于 60 年代末推出了第一个商品化的层次数据库管理系统 IMS (Information Management System),标志着数据管理技术进入了数据库系统阶段。

层次和网状数据库是 60 年代技术条件下的合理产物。它们为数据库技术奠定了基础,打开了应用局面。但是数据库最有意义的成就是关系数据库的发展。1970 年, Codd 提出了关系数据模型,以关系 (relation) 或称表 (table) 作为描述数据的基础,在其后的 10 年中, Codd 又发表了一系列文章,奠定了关系数据库的理论基础。

关系数据模型一提出,便立即受到数据库界的重视,但是也有相当多的人认为关系数据模型仅仅是理想化的模型,尤其担心关系数据库的性能难以被用户接受。1974 年,数据库界开展了一场分别以 Codd 和 Bachman 为首的支持和反对关系数据库的大辩论,辩论促进了关系数据库的发

展。

进入 70 年代以来,数据库技术迅速发展,得到了广泛的应用,成为计算机科学中的一个重要分支。数据库技术的发展使数据管理上了新台阶,几乎所有的信息管理系统都以数据库为核心。到了 80 年代,关系数据库成为发展的主流,几乎所有新推出的 DBMS 都是关系型的。随着微机的出现和计算机网络的广泛应用,分布式处理系统在 80 年代得到很大的发展,与此相适应,分布式数据库系统也成为 80 年代数据库研究的重点,日趋成熟。80 年代且几乎所有分布式 DBMS 产品都是关系型的。80 年代是关系数据库的全盛时代。

总之,数据库技术自从 60 年代产生以来,无论是理论还是应用一直是热门课题,现已成为计算机科学的重要分支。现在,通常把 70 年代广泛流行的网状、层次数据库称为第一代数据库系统,而把 70 年代处于实验阶段 80 年代广泛流行的关系数据库系统称为第二代数据库系统,这两代数据库系统的应用领域主要是商务领域,所处理的事务一般比较小,比如财务管理、人事管理等等。随着计算机的广泛应用,新的应用又提出了新的要求。人们开始发现关系数据库的许多限制和不足,这又推动了数据库技术新一轮的研究,道路无非有两条;一是改造和扩充关系数据库,以适应新的应用要求;二是改用新的数据模型,例如面向对象数据模型、基于逻辑的数据模型等,研制新型的数据库。目前,这两方面都取得了可喜的成果,出现了下列新的数据库技术。

2 并行数据库技术

数据库的并行处理 (Parallel Processing) 是提高数据库

作者简介:湖南大学 97 级博士研究生(在职),长沙电力学院副教授、副系主任,曾独编或主编著作 5 本,发表论文约 20 篇,研究方向为数据库技术、计算机应用与系统工程。

78 1999.5 计算机工程与应用

系统对事务快速响应能力的有效手段,从硬件上讲,并行处理是设置若干能同时工作的部件和设备;从软件上讲,并行处理是设置若干个可以同时运行的单位(如线程),这些并行单位可以完成相同或不同的数据处理功能。

并行数据库技术包括了对数据库的分区管理和并行查询。它通过将—个数据库任务分割成多个子任务的方法由多个处理机协同完成这个任务,从而极大地提高了事务处理能力,并且通过数据分区可以实现数据的并行 I/O 操作。

多线路技术和虚拟服务器技术是并行数据库技术实现中采用的重要技术。并行数据库系统以高性能、高可用性和高扩充性为目标,充分利用多处理器平台的并行能力,通过多种并行性,在联机事务处理与决策支持应用两种典型环境中提供快速的响应时间和较高的事务吞吐量。虽然并行数据库系统的许多关键技术(如:并行数据库的物理组织,并行数据操作算法的设计与实现等)仍需深入研究,但普遍认为并行数据库系统不久将成为高性能数据库系统的佼佼者。

3 分布式数据库系统

由于计算机网络技术的发展,才有可能把分散在各处的数据库系统通过网络通信技术连结起来,这样形成的系统称为分布式数据库系统(Distributed Database System)。

分布式数据库系统由多台计算机组成,每台计算机上配有各自的本地数据库,各计算机之间由通信网络连结。在这种系统中,大多数处理任务由本地计算机访问本地数据库来完成(这种应用称为本地应用),对于少量本地计算机不能单独胜任的处理任务,则通过数据通信网络与其它计算机相联系,并获得其它数据库中的数据(这种应用称为全局应用)。分布式数据库系统在系统结构上的真正含义是指物理上分布,逻辑上集中的分布式数据库结构,数据由系统统一管理,使用户不感到数据的分布,用户看到的似乎是一个全局数据模式的集中式数据库。

与集中式数据库比较,分布式数据库具有体系结构灵活性大、系统可靠性高、可用性好、可扩充性好等优点,但也存在下列不足:

由于数据库系统的应用通常是逐步发展的,最初是建立各种孤立的数据库,而管理这些数据库的计算机系统和 DBMS 可能是不同的,也就是异构的,当应用需要转向分布式数据处理时,就需要解决异构数据库的集成问题,这在技术上有一定的难度,目前,还很难用一个通用的 DBMS 来解决这样的问题。人们希望用一种新的数据库技术——联邦数据库系统来解决这一问题。

分布式数据库系统虽然有利于改善性能,但如果数据库设计不好,数据分布不合理,使远距离访问过多,特别是当分布连接操作过多时,会降低系统的性能。因此,对分布式数据进行有效的管理和方便的访问就成为分布式数据库技术必须解决的问题。

随着数据库技术的发展,分布式数据库系统也在向面向对象系统、智能库和知识库方向发展。

4 联邦数据库系统

分布式数据库系统不能很好解决异构数据库的集成问

题,一种较为有效的解决方法是通过建立联邦数据库系统来解决。

通常称相互独立运行的数据库系统为单元数据库系统(CDBS)。所谓联邦数据库系统(FDBS)是一组彼此协作且又相互独立的单元数据库系统的集合,它将单元数据库系统按不同程度进行集成,对该系统提供整体控制和协同操作的软件叫做联邦数据库管理系统。

联邦数据库主要是在分布的环境下实现数据的集成,比如,人类在进步过程中积累了浩瀚的数据并且采用了不同的 DBMS 分别进行管理,即使以后的数据库技术大大进步了,人们不可能也不应该放弃这笔巨大的财富。由于各个数据库的数据模型、DBMS 和计算机机型有很大的差别,完全用一种统一的数据库来改造它们几乎是不可能的,这时,就必须求助于联邦数据库,因此,联邦数据库实际上是结点异构型分布数据库系统的推广。联邦数据库不但要解决分布环境下的并发控制和数据一致性问题,而且要解决由于数据模型、语言和语义解释的不同以及操作系统的异构带来的各种问题。虽然联邦数据库的实现较困难,但人们仍在向这方面努力。

5 多媒体数据库

随着计算机的广泛应用,在办公自动化、生产管理和控制等领域,人们对用户界面和信息载体提出了越来越高的要求,希望有声有色,图文并茂等。通常,把能够管理数值、文字、表格、图形、图像、声音等多种媒体的数据库称为多媒体数据库 (Multimedia Database, MDB)。

由于不同媒体上的信息具有不同的性质与特性,如何组织存在于不同媒体上的信息,将不同媒体上的数据一体化,是多媒体数据库系统要解决的问题。

与传统 DBMS 一样,多媒体 DBMS 也要进行数据的处理、查询和事务处理的管理等,但有不同的用户接口和存储构造,在多媒体 DBMS 中特别强调“媒体独立性”,所谓媒体独立性是指不论管理的多媒体数据库的媒体如何变化,都不需改变 DBMS,因此,多媒体应能支持。

媒体的混合:在 DBMS 中,如遇到各种混合在一起的媒体,需对它们进行统一管理。

媒体的扩充:DBMS 能处理图形、图像、声音等信息,此外,它应该能根据要求处理新的媒体。

媒体的转换:在 DBMS 所遇到的各种媒体,应能在 DBMS 下相互转换,以保障媒体的互换性。

多媒体 DBMS 的结构主要有:

单一 DBMS 结构:它是指用一个多媒体 DBMS 管理各个媒体数据库,同时它也管理着结合媒体的数据库对象。

主从 DBMS 结构:它是指各个媒体数据库均有自己的从 DBMS 进行管理,而这些从 DBMS 只由一个主 DBMS 统一管理,用户只需面对主 DBMS,因此,用户看起来仍然只有唯一的、统一的多媒体数据库。

联合型(也称协作型)DBMS:它是指各个媒体数据库由分散的 DBMS 管理,各 DBMS 间通过相互通信进行问答和相应的更新,用户可面对任何一个 DBMS,而把全体 DBMS 看成一个多媒体数据库。

6 面向对象数据库

面向对象(Object-Oriented,OO)数据库系统是数据库技术与面向对象技术相结合的产物。它应满足两个标准:它既是一个数据库系统,而且也是一个面向对象系统。

对象是面向对象数据库中进行存储、修改和提取的元素。对象可实现多种类型的数据,包括图形、图像和声音等,在面向对象数据库中,现实世界中的所有实体都能表示为对象。对象除了能存放数据外,还能存放关系和行为。其中关系在内部表示为到其他对象的链;行为在内部表示为方法。与传统数据模型比较,面向对象数据模型具有下列优势:

- (1)具有表示和构造复杂对象的能力。
- (2)通过封装和消息隐藏技术提供了程序的模块化机制。

- (3)继承和类层次技术提供了软件的重用机制。

- (4)通过滞后联编等概念提供系统扩展能力。

但是 OODB 至今没有统一标准,其原因是:

- (1)缺乏通用的数据模型。
- (2)缺乏坚实的形式化理论基础。
- (3)具有较强的实践性。

至今,面向对象数据库的核心技术问题已基本解决,但正是因为面向对象模型的丰富表达能力,模型相对复杂,系统要处理的事情多,实现起来较困难,所以尽管面向对象系统多,但大多是实验型的,尚未商品化。

7 数据采掘与数据仓库

无论是商品企业、科研机构或者政府部门,在过去若干年的时间里都积累了海量的、以不同形式存储的数据资料。由于这些资料十分繁杂,要从中发现有价值的信息或知识,达到为决策服务的目的,成为非常艰巨的任务。数据采掘就是解决上述问题而提出的。

数据采掘(Data Mining)指的是从大型数据库或数据仓库中取提人们感兴趣的知识,这些知识是隐含的、事先未知的潜在信息。提取的知识表示为概念(Concepts)、规则(Rules)、规律(Regulation)、模式(Patterns)等形式。数据采掘是当前数据库和信息决策领域的最前沿研究方向之一。数据采掘技术涉及数据库、模糊逻辑、人工智能、机器学习和统计分析等多种技术。数据采掘过程一般由3个主要的阶段组成:数据准备、采掘操作、结果表达和解释。知识的发现可以描述为这3个阶段的反复过程。

数据准备 这个阶段又可进一步分成3个子阶段:数据集成、数据选择、数据预处理。数据集成将多文件或数据库运行环境中的数据进行合并处理。数据选择的目的是辨别出需要分析的数据集合,缩小处理范围。预处理是为了克服目前数据采掘工具的局限性。

数据采掘 这个阶段进行实际的采掘操作。包括:决定如何产生假设;选择合适的工具;发掘知识的操作;证实发现的知识。

结果表达和解释 根据最终用户的决策目的对提取的信息进行分析,把最有价值的信息区分出来,并且通过决策支持工具提交给决策者。

数据仓库也是近年来提出的新概念。数据仓库(Data Warehouse)是指这样一种数据的存储地,来自于异地、异构

的数据源或数据库的数据经加工后在数据仓库中存储、提取和维护。数据仓库提供来自种类不同的应用系统的集成化和历史化的数据,使用户拥有任意提取数据的自由,而不干扰业务数据的正常运行。数据仓库不是单一的产品,而是由软、硬件技术组成的环境。它把各种数据源(源数据库)集成为一个统一的数据仓库,便于最终用户访问并能从历史的角度进行分析,最后作出决策。作为一个系统,数据仓库至今至少包括3个基本的功能部分:

数据获取:这个部分负责从外部数据源获取数据。

数据存储和管理:这个部分负责数据仓库的内部维护和管理。

信息访问:这个部分属于数据仓库的前端,面向不同种类的最终用户,这里主要由桌面系统的各种工具组成。

目前,数据采掘的研究已和数据仓库的研究结合起来。数据仓库的产生和发展为数据采掘技术开辟了新的战场,同时也提出了新的要求和挑战,当前的研究主要着眼于数据仓库的构建和维护的基本理论、方法上,下一步将把重点放在数据仓库的有效应用上,为高级的决策支持服务是数据仓库的最终目的。

8 演绎数据库

演绎数据库是一种基于逻辑推理的数据库,它将数据库看成一个演绎系统,即一个数据库可与一个演绎理论相一致,由一些公理组成,通过公理中的演绎规则导出定理,或者说演绎数据库是一种具有演绎推理能力的数据库系统。

演绎数据库发展较早,1969年,Green成功开发了QA问题解答系统,它被认为是演绎数据库领域中创始性的工作。到今天,演绎数据库的发展可分为3个阶段:

孕育形成阶段(1969~1978年)。在这个阶段,关系演算查询语言和逻辑程序设计语言PROLOG的产生。对演绎数据库的形成有重大影响。1978年出版的《逻辑与数据库》一书标志着该领域的诞生。

理论发展与基础研究阶段(1979~1987年)。这个阶段是演绎数据库与逻辑程序设计语言的理论研究以及与实现有关的基础研究迅速、全面地发展与完善的时期,1987年,由Minken教授编辑的会议记录《演绎数据库与逻辑程序设计基础》一书,标志着该理论的成熟。

实现与应用阶段(1988年至今)。虽然有了许多实验性系统,但尚未有成功的演绎数据库系统投入市场。人们已认识到,演绎DBS比关系DBS有许多优越性,但实现起来比较复杂,为此,人们希望在90年代,演绎数据库能像关系数据库一样成为广泛应用的数据系统。演绎数据库的基于逻辑的规划管理部分已被结合进第5代计算机系统和知识库系统。

演绎DBMS和OODBMS是继关系DBMS以后的两个重要发展方向。

9 模糊数据库

传统的数据库仅允许对精确的数据库进行存储和处理,而客观世界中许多事物是不精确的。比如:“明天有可能下雨”、“这个人个子高”等等,这种不确定性可分为两类,

一类是随机性所产生的不确定性,像“明天有可能下雨”,而另一类则是由概念的外延本身的模糊而产生的不确定性,像“个子高”,到底多高算个子高,没有明确定义。模糊数据库的研究和实践就是为了解决模糊数据的表达和处理问题,使得数据库描述的模型能更自然、更贴切地反映客观世界。传统的 DBMS 对此是无能为力的。国际上对模糊数据库的研究是从 80 年代初开始的,旨在克服传统数据库难以表达和处理模糊信息的弱点,开拓更新、更广的应用领域。10 多年来,大量的研究工作集中在模糊关系数据库方面,即对关系数据库进行模糊扩展,比如:模糊数据表达、模糊查询、模糊数据库设计等。近年来,亦有许多工作是对关系模型之外的其它数据库模型进行模糊扩展,如模糊 ER (实体-关系)、模糊面向对象数据库等等。

模糊数据库技术在管理信息系统、专家系统、决策支持系统中占有重要的位置,有人预言,第 6 代计算机将是模糊计算机。

10 模糊演绎数据库

模糊演绎数据库是模糊数据库与模糊演绎推理技术相结合的产物,其目的要求在数据库上进行模糊推理的功能,这是知识库管理系统、专家系统开发工具、决策支持工具的主要功能之一,实现模糊演绎数据库的关键技术是下列 3 个方面:

- (1)模糊推理规则的表达。
- (2)模糊数据库与模糊演绎推理功能的结合方法。
- (3)模糊推理方法。

11 主动数据库

主动数据库是近年来出现的一个重要而活跃的研究领域。在传统数据库中,当人们需要改变、增加或删除数据库中某些满足一定条件的数据时,用户可以通过相应的命令或操作来实现其目的,但是,这些功能都有一个重要的特征,就是“数据库本身是被动的”,用户给什么命令,它就做什么动作,此数据库系统仅在用户或应用程序直接请求时才执行事务,它不会主动根据数据库的状态提出做些什么,这种数据库被称为“被动式数据库”。在主动数据库中,则打破了这种规则,主动数据库除了具有传统数据库的一切被动服务功能外,还具有主动进行服务的功能,这主要是通过将一些规则预先嵌入数据库系统的办法来实现。例如,在系统中预先嵌入如下形式的规则

```
WHEN <事件>
IF <条件> THEN <动作>
(或后跟一组 IF-THEN 规则)
```

系统提供一个“自动监视”机构(一般可以是一个直接由操作系统控制的独立进程或某种硬件设施等),它主动地不时地检查着这些规则中包含的各种事件是否已发生,一旦某事件被发现,就主动触发执行相应的 IF-THEN 规则(或规则组)。显然,此时 DBMS 本身可主动履行一些预先由用户设定的动作,可把诸如完整性约束、存取控制、例外处理、触发警告、主动服务、状态开关切换,乃至复杂的演绎推理功能等以一种统一的机制而实现。

实现主动数据库的关键在于实现一种有效的事件自动监视装置,使得事件一旦发生便很快被发现,从而触发执行相应的规则,这往往需要软、硬件结合来解决。

12 结束语

限于篇幅,还有一些新数据库,比如时态数据库、集成的工程数据库等的研究和发展就不一一表述了,读者可参阅其它文献。可以预见,在未来的若干年中,数据库技术仍将是热门课题,下列等课题将是数据库技术研究的热点:

- (1)关系数据库理论的深入研究。
- (2)数据库技术与人工智能技术相结合。
- (3)数据库技术与计算机网络技术相结合。
- (4)新的数据模型的研究。(定稿日期:1998 年 7 月)

参考文献

1. 王能斌. 数据库系统. 北京:电子工业出版社,1995
2. 张健沛等. 数据库系统教程. 北京:中国铁道出版社,1995
3. 黄上腾等. 数据库原理. 上海:上海交通大学出版社,1996
4. 罗可. FoxPro 教程. 北京:电子工业出版社,1997
5. 科学数据库与信息技术论文集. 北京:中国科学技术出版社,1994
6. 周立柱等. 数据库研究的发展与方向. 微电脑世界,1997;3:36~38
7. 陈楠. 刍议数据库技术的最新发展. 微电脑世界,1997;7:51~54
8. 胡侃等. 基于大型数据仓库的数据采掘. 研究综述. 软件学报,1998;1:53~62
9. 袁占亭等. 大型数据库信息系统设计的关键技术. 计算机工程与应用,1998;4:1~5
10. 王浩等. 演绎的面向对象数据库研究. 计算机研究与发展,1998;4:289~295