

27-29, 17

时态数据库二十年

TDB模型

历史数据库

⑦

信息管理

计算机科学 1999 Vol. 26 No. 2

时态数据库的沿革、特色与代表人物*

The Background, Characteristics and Representative Researchers of Temporal Databases

——时态数据库二十年回顾之一

唐常杰

TP392

(四川联合大学计算机系 成都 610064)

Abstract The temporal database has been investigated for twenty years. Near hundred of TDB models are proposed. About 500 articles on TDB were published. This paper gives a simple Retrospection to the 20 years research of TDB, introduces the its background, characteristics and difficulties, representative researcher and articles. In the second part will discuss the main models and shortcoming in TDB research as well as the author's opinion about futures TDB investigation.

Keywords Temporal databases, Cronon, Time interval, Temporal information

1 时态数据库产生的背景

伴随着物理流的信息流包含着事件的时态信息(Temporal Information),其中有时刻信息(Instant Information),时间区间信息(Interval Information)和相对时间信息(之前、之后、重叠)等等。传统的数据管理系统把时态数据作为用户定义时间字段进行存贮和管理,只反应了一个对象在某一个时刻的状态(快照),不联系其过去和未来。

日益广泛的数据应用要求了管理被处理事件的历史性信息,和系统中元事件的时态信息,例如查删改事务的时刻,时态信息数据反映了事物发生发展的过程,有助于(揭示事物发展的本质规律和这些数据有助于提高数据库系统的可靠性和效率。

60年代末,人们开始认识到时态信息的重要性,并展开了研究。G. Wiederhold 和 J. F. Fries 在 1970 年研制的医疗系统在处理时态信息方面作了较早的尝试, Kahn Ketal 于 1977 年在 Artificial Intelligence 上发表的“Mechanizing Temporal Knowledge”反应了早期关于时态信息的基础性研究工作。

2 富有特色的开创性工作

70年代末、80年代初,数据库技术的日渐成熟和大容量高速存储设备的发展为时态技术和数据库

技术的融合创造了条件, J. Ben-Zvi, J. Clifford 和 S. Ginsburg 三位学者独自在时态数据库方面作了富有特色的先驱性工作。

J. Ben Zvi 于 1979—1982 年对时态数据库作了开创性研究,他的博士论文(加州大学洛杉矶分校, 1982)总结了他的一系列工作。Ben Zvi 的贡献突出在下列几点:(1)提出了时态数据库模型,引入了时间间隔(Time Period),后来被学术界改称为时间区间。(2)在 1979—1982 这个关系规范型研究的热点时期, Ben Zvi 突破思想禁区,提出并研究了非 1NF 的 TDB。(3)以时间区间作字段值,刷新了人们认为数据库字段值只能是一个数或串的概念。(4)引入了后来被称为双时态的概念,即用有效时间表示被管理对象在库中生命周期,用事务时间表示数据库本身的历史。(5)引入了时态索引结构。上述成果都是开创性的工作。影响了 TDB 领域中一代后来者的思路,仅管 1982 年以后, Ben Zvi 在 TDB 领域消声匿迹了,他对 TDB 的贡献功不可灭。

1982 年 J. Clifford 在纽约大学完成了博士论文“A logical frame work for the Temporal Semantics and Natural Language Querying of Historical Database”及相关的一组文章,对历史数据库的开创性作了重要贡献。它注意到了被管理对象的生命周期(Lifespan),研究了在关系、元组、字段值上加时

* 国家自然科学基金 69773051 号支持项目, 唐常杰 教授, 从 1983 年起从事时态数据库研究。

态信息的技术细节,引入了历史关系模型,历史关系代数,研究了历史数据库中投影、选择、连接的特殊要求和特殊规律,研究了历史关系模型与传统关系模型的兼容性,即当把区间缩小为一点(Now-Now)时历史数据库即退化为传统快照数据库,且相应的时态代数运算退化为传统快照关系运算。

美国南加利福尼亚大学的 S. Ginsburg 教授原是形式语言,特别是上下文无关文法研究的先驱者。他于 1983 年提出对象历史模型(Object History),该模型以可计算对象如银行帐目、支票帐户的历史性数据为研究背景,其特色在于:(1)深刻地分析了基于可计算元组的对象历史的特殊要求和特殊规律,提出了基于记录、事件驱动的代数结构模型。(2)研究内容(包括提出的问题,分析建模和成果表现)与其他 TDB 模型完全不交叉,强调了历史+现状→未来的机制。(3)揭示了一系列出人意料的结果,包括关于对象历史中的若干判定性问题等难度很大的结果。(4)问题提出,分析和模型都接近金融系统的各种应用。

3 百家争鸣、竞争研究

80 年代初期 TDB 的研究开始繁荣,根据 E. Mckenzie 的报告,其中 1982—1986 年有 80 多篇关于时态数据库的论文发表。80 年代中期,学术界提出了近百种为数据库管理系统增加时态信息处理能力的方案,经过十多年竞争性研究,逐渐归并为十多种模型,其中有 13 种被收入世界第一本关于时态数据库专著《Temporal Database—Theory, Design and Implementation》^[1],下面列出这 13 种模型的代表性人物、及相应代表性著作的发表时间:

(1) Time Relational Model, 该模型是 Ben Zvi 在 1979—1982 在博士论文中提出,是 TDB 领域的先驱性研究工作。它开创了对时态数据库、时态查询语言的研究。

(2) HRDM (Historical Relational Data Model), J. Clifford, 1982.

(3) TempSQL, Sharshi. K. Gadia & Sunil, S. Nair, 1985.

(4) IXRM (Interval-Extended Relational Model), Nikos A. Lorentzos, 1987.

(5) TRM 及 TSQL (Temporal Extensions to the Relational Model) K. B. Navathe, 1987.

(6) HSQL (Historical Query Language), N. L. Sarda 1987.

(7) TQuel, R. Snodgrass, 1985

(8) TRC (Temporal Relational Calculus), Abdullah Tausel, 1992

(9) TEER, (Temporal Query Language for Enhanced Entity Relationship Model), R. Fimasri, 1985.

(10) TDM (Temporal Data Model Based On Time Sequence), Arie Segev & Arie Shoshanu, 1988.

(11) OODAplex (Object Oriented Aplex), U. Dayal, 1989.

(12) Object History, S. Ginsburg, Tanaka, 唐常杰等, 1983.

(13) Temporal Deductive Databases, Marianne Bandmet 等, 1989.

上述 13 个 TDB 模型从不同的需求、观点出发,各自独立地建立了一套术语、概念、数学模型,并形成了一套各自独立的理论体系。

在文献[1]中列出二十年来时态数据库研究的 500 余篇的目录文献,这些文献的总篇幅近万页,文献作者遍及美国、欧洲、中国、日本、印度、土耳其。最主要的研究集体集中在美国纽约大学、依阿华州立大学、加州大学伯克利分校、南加利福尼亚大学、亚利桑那大学。

4 特殊功能、特殊困难

4.1 时态数据库的功能性分类

按 Spipada 和 Snodgrass 的意见,时态数据库按功能可分为三类:(1)历史数据库,被管理对象的生命周期称为有效时间(Valid Time),对象历史由 DBMS 内部机制处理。(2)事务数据库,其中,数据库本身被查删改的时间称为事务时间(Transaction Time),其历史由 DBMS 内部机制处理。(3)双时态数据库,既能管理对象历史,又能管理数据库本身的历史。

系	系主任
信息系	李 军
电力系	王 林

图 1

4.2 从实例看历史数据库的时间维结构

传统的关系数据库是{属性}×{元组}的二维结构。图 1 中的关系记录了大学中系和系主任姓名信息,它只是一个快照,不能反应被管理对象的历史。

以“年”为时间粒度,将94年、97年、98年(Now)的三个快照放在一起,如图2,构成一个有时间维的三维数据库。这在一定程度上解决了管理历史的问题,但又引出了新问题。

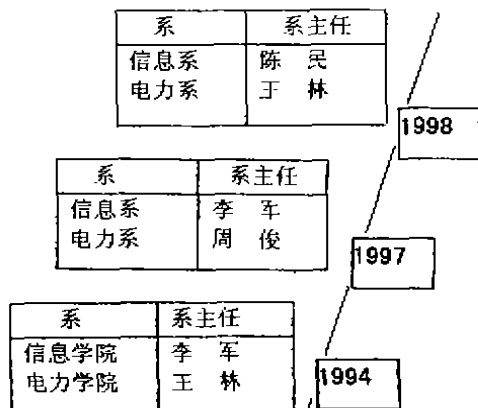


图2 带时间维的三维数据库

4.3 时态管理引出的特殊困难

图2中把多个快照(即在不同时刻保存的数据库的备份)放在一起,构成了历史。长期以来,在没有使用时态数据库的年代里,企业及机关管理历史性数据都是这样作的。管理人员为三个问题感到困惑:

(1)取多大的时间间隔保存快照?如果间隔太大,则不足以保证数据的准确翔实,如果间隔太小,则数据冗余多,占存贮空间大。以一个城市银行为例,每天作一备份,每份占磁盘(或磁带)空间1000Mb。当某一事件涉及三个月帐务历史时,搜索空间将达100Gb。

(2)传统关系数据库中,一个表的多个快照,不能简单地同时装入内存,不能简单地使用传统的选择、投影、连接运算进行操作和查询,因为同一元组同一属性下的值在不同快照中可能不同,必须作较复杂的、非通用的编程。

(3)传统数据库中,对数据库本身历史的维护支持不足,一般只有供恢复回退用的事务日志(Transaction log),缺乏相应的事务查询命令(例如快速地查出谁对某一字段作了超过十次的修改)。

为了解决这些问题,人们提出了各种时态数据库模型。

5 各种TDB模型的共性

各种TDB模型有一个共同的管理时态信息的目的,从而决定了它们在概念、研究方法和机制上有

一些共同的特点。

(1)时间量子和时间区间概念 时间量子,早期文献称Time quantum,近年逐渐统一为Cronon,被定义为时态系统所支持的最小的、不可分解的时间间隔。在微机Windows平台上,时态系统通常取时间量子为0.01秒(高精度系统)至1秒(低精度系统)。系统支持的时间起点记为0,系统时间论域是有限、离散、有序的时间量子的集合,一般记为 $Sys-T = \{0, 1, 2, \dots, Now, \dots, MaxSysTime\}$,时间区间是 $Sys-T$ 中以两个时间量子为上下限的时间量子的集合,例如 $\{t, t+1, t+2, \dots, T\}$,简记为 $[t, T]$ 。

(2)扩展关系数据库 TDB模型一般是传统关系数据库的扩展,而传统关系数据库是它的特例,当生命周期 $[t1, t2]$ 缩小为一个时刻(Now)时,它是被管理对象的一个快照,即传统关系数据库,传统关系数据库的操作如选择、投影、连接都在这里有所扩展,而当生命周期缩为 $[Now, Now]$ 时,这些运算与传统RDB运算兼容。TDB中增加一些传统RDB没有,但对历史数据库有意义的运算,如时态连接、时间态选择、AFTER、BEFORE和OVERLAP等等,并引出出人意料的深刻的结果(命题或定理)。

(3)形式化的描述和原型 TDB模型一般都有形式化的描述,对时间、区间、时态元组、时态关系下一批定义,引入一批运算或扩展传统关系数据库的运算,例如时态选择,时态投影,时态连接等等。在模型和概念基础上深入下去,取得一批有意义的结果。例如某性质在关系数据库上成立,在时态数据库上也成立(平移,扩展);某性质在关系数据库上成立,在时态数据库上加条件才成立,等等。在条件允许时,研究者都要实现个原型,检验理论结果和运行效率,分析模型的优缺点及可实现性。

在这些原则的指导下,TDB研究者从不同的观点,根据不同的需求,提出了各种TDB模型。在下篇“TDB二十年回顾之二”中,我们将介绍TDB代表性成果,分析目前在TDB领域研究工作的不足,并对TDB未来发展策略提出作者的看法。

参考文献

- 1 Tansel A. et al 编. Temporal Databases. --Theory, Design and Implementation. The Benjamin/Cummings Publishing Company, 1993
- 2 Zvi J B. The Time Relational Model [Ph. D. Thesis]. Computer Science Dept. UCLA, 1982

(下转第17页)

$$\partial^2 f(x, y) / \partial x^2 + \partial^2 f(x, y) / \partial y^2 = 0$$

我们定义一个矩形网络(grid)来求偏微分方程的近似解。其 Jacobi 迭代的高散形式为:

$$f_{max}(i, j) = (f(i-1, j) + f(i+1, j) + f(i, j-1) + f(i, j+1)) / 4$$

计算函数点的新值时用周围函数点的旧值。并行程序在给定边界的情况下,对其进行 Jacobi 迭代,然后判定其是否收敛。

在对泊松问题进行并行计算时,我们对该矩阵进行一维划分。每个结点机分得一个子阵。对于阵的边缘结点进行迭代时,需要与相邻的结点机进行通信。

在结点机数目为 4 的情况下,实际的计算时间与使用性能预测工具得到的预测时间的比较如图 2 所示。

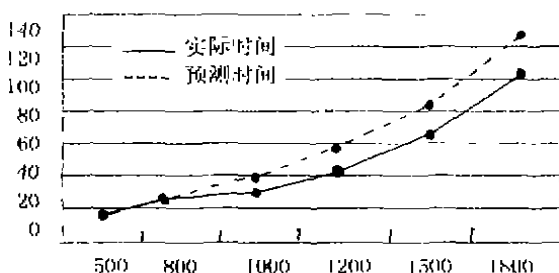


图 2

我们发现随着规模的增加,预测时间与实际执行时间的偏差增大,这是因为误差的累积。在实际运算中,并行计算的速度还受到系统进程、内存容量等其他因素的影响。另外一个因素是同步函数 Barrier 和 AllReduce 等集合操作的影响,这些状态在预测

模型中没有考虑,数据规模越大,这些因素的影响越大。

结束语 目前并行性能分析和预测工具主体已经完成。并行性能预测工具 TH-PP 尚需进一步完善。问题规模的扩大和系统资源利用率的关系、可伸缩性分析是将来需重点探讨的两个问题。

参考文献

- 1 Miller B P, et al. *The Paradyn Parallel Performance Measurement Tools*. IEEE Computer, Special issue on performance evaluation tools for parallel and distributed computer systems, 1995, 28(11)
- 2 Martonosi M, et al. *Tuning Memory Performance of Sequential and Parallel Programs*. IEEE Computer, 1995(4): 32~39
- 3 Calzarosea M, et al. *Medea: A Tool for Workload Characterization of Parallel Systems*. IEEE Parallel & Distributed Technology, 1995 Winter: 72~83
- 4 Boden N J, et al. *Myrinet: A Gigabit per Second Local Area Network*. IEEE Micro, 1995, 15(1): 29~36
- 5 Crovella M E, LeBlane T J. *The Search for Lost Cycles, A New Approach to Parallel Program Performance Evaluation*. [Technical Report 477] The University of Rochester, Computer Science Department, Rochester, New York 14627, 1993
- 6 Sarukkai S R, Mehra P. *Automated Scalability Analysis of Message-Passing Parallel Programs*. Same to [3]: 21~32
- 7 Borgeest R, et al. *A Trace Based Performance Evaluation Tool for Parallel Real Time Systems*. Parallel Computing, 1995, 21: 551~564
- 8 Clifford J. *A Model for Historical Databases*. In: Proc. of Workshop On Logical Bases for Database. Toulouse, France, 1982
- 9 Gadia S K. *Temporal element as a primitive for time in temporal databases and its application in query optimization*. In: Proc. of 1986. ACM Computer Science Conf.
- 10 唐常杰, 熊明. *The Temporal Mechanisms in HBase*. J. of Computer Science and Technology, 1996, 11(4): 365~371
- 11 唐常杰, 于中华, 等. *时态数据库的分史存储技术*. 第 15 届全国数据库会议论文集, 1998

(上接第 29 页)

- 3 Ginsburg S, Tang Changjie (唐常杰). *Projection of Object History*. Theoretical Computer Science, 1986, 48: 297~325
- 4 Ginsburg S, Tang Changjie (唐常杰).
- 5 Novathe S B, Ahmed R. *A Temporal Model and Query Languages*. [UF-CIS Technical Report TR-8516-Computer and Information Science]. Department Univ. of Florida, April 1986
- 6 Snodgrass R, Ahn I. *Temporal Databases*. IEEE Computer, 1986, 19(9): 35~42
- 7 Tansel A U. *Adding time dimension to relational model and extending relational algebra*. Information Systems, 1986, 11(4): 343~353