

数据库 管理系统 维也纳 拉古纳海

19-22

数据库管理系统的未来研究方向 ——从拉古纳海滩到维也纳

周龙骧

(中国科学院数学研究所 北京100080)

TP311.13

A

摘要 1993年8月在爱尔兰首都都柏林举行了第19届国际超大型数据库会议(VLDB93)。会议的一个特邀报告是 M. Stonebraker 所作的“处于十字路口的 DBMS 研究:维也纳修改意见”。该报告回顾了1988年2月美国加州拉古纳海滩讨论会以来 DBMS 研究方向的情况和变化,提出了今后最有希望的研究方向和被认为是误导的研究方向。该文表达了一批美、德资深学者对 DBMS 研究方向的看法,显然对我国学者的研究工作具有重要的参考价值。本文是对该文的综述和评论。

1 引言

1993年4月23日在维也纳举行了 IEEE 数据工程国际会议的一次专题讨论会,有五位数据库研究界的学者参加,他们是美国的 M. Stonebraker, R. Agrawal, U. Dayal 和德国的 E. J. Neuhold 及 A. Reuter。讨论的主题是 DBMS 的未来研究方向。这次讨论会的主题与组织方式和上一次会(即1988年2月4日至5日在美国加州拉古纳海滩举行的讨论会)相似^[1]。拉古纳海滩会议由前西德国家计算机科学研究中心(GMD)在美国加州所建的国际计算机科学研究所(ICSI)发起,有九位美方学者和七位德方学者参加,他们是 P. Bernstein, U. Dayal, D. Dewitt, D. Gawlick, J. Gray, M. Jarke, B. Lindsay, P. Lockemann, D. Maier, E. Neuhold, A. Reuter, L. Rowe, H. Schek, J. Schmidt, M. Schrefl, 和 M. Stonebraker。讨论会由 GMD 的 E. J. Neuhold 和 Berkeley 的 M. Stonebraker 主持。

会议首先由每位学者介绍四项他认为是有意义的 DBMS 研究方向(即他本人并未在做但他是愿意做的)和两项他认为意思不大的方向。然后进行投票,选出最有希望最应予以重视的研究方向和最没有意思不应再做的研究方向。拉古纳海滩的投票结果如表1所示。

从表1可以看出一共投了144张赞成票和64张否决票。其中前十个课题集中了144张赞成票中的101票,而最后六个课题集中了64张否决票中的42票。这说明拉古纳海滩讨论会的与会者们的意见还是比较

一致的。

表1 拉古纳海滩表决结果

课 题	赞成票	否决票
End User Interfaces	14	0
Active Data Bases	15	1
Parallelism	11	0
New Transaction Models	10	0
CIM, Image, IR Appl.	10	0
CASE Applications	9	0
Security & High Availability	9	0
Large Dist. Data Bases	9	1
DB/OS Interaction	7	0
Transaction Design Toods	7	1
Large System Admin.	5	1
Real Time DBMS	3	0
DBMS Impl. Blkhd Paradigm	3	0
IMS-style Joins	3	0
Automatic DB Design	4	2
Tool-kit DBMS Systems	5	3
Data Translation	2	1
OODB	7	6
Dependency Theory	0	3
Interface btwn DBMS & Prolog	0	5
New Data Model	0	5
Common OO Data Model	2	7
Tradi. Concurr. Control	0	7
Hardware DB Machines	0	8
General Recursion Queries	0	10

周龙骧 研究员。*)见笔者所译《数据库管理系统研究的未来方向》,本刊1988年第5期。

维也纳会议采用了与拉古纳海滩讨论会相似的方式,五位学者中有四位参加过拉古纳海滩的会议。五位学者中的每一位介绍四项他认为很有意义而他本人并未在做的课题和四项他认为没有意义不应再做的课题。然后进行投票选出最有意义应予以高度重视的课题和意思不大不应再做的课题。维也纳讨论会表决的结果如表2所示:

表2 维也纳讨论会表决结果

课 题	赞成票	否决票
End User Interfaces	4.5	0
Big Systems	4.5	0
Legacy Applications	4	0
Multimedia Applications	4	0
Data Mining	4	0
Mobile Data Bases	2	0
Method Optimization	2	0
Embedded DBMS	2	0
Object Repositories	1	0
OO Data Base Design	1	0
Constraints	1	0
Relational Extensions	0	1
Synchronization Theory	0	1
Simplistic Data Integration	0	1
Less than ACID Transactions	0	2
Persistent C++	0	2
Multi-Data Base Transactions	0	3
New OO Data Models	0	3
Replication Algorithms	0	4
Bizarre Performance Studies	0	4
Traditional Engines	0	4
General Recursion	0	5

从表2可以看出,前五个课题受到最大限度的重视而最后六个课题则受到大多数与会专家的反对。

2 拉古纳海滩会议分析

拉古纳海滩会议选出了十个备受关注的重点课题。第一个是 DBS 的最终用户接口,它得到与会学者的完全一致的支持,被认为是一个极端重要的领域。1988年以前的十年中用户接口取得了极大的进展,他们指出了诸如 Macintosh 的图形用户接口、电子报表、WYSIWYG 接口、超极卡等,这些技术和成就导致了在人机接口方面的一次革命。

与会专家们注意到的一个明显事实是在用户接口这一如此充满活力并已取得意义重大进展的领域里研究界很少参与、贡献甚少,基本的成就和进展都是工业界做出的。专家们经过讨论对这一现象做了一些解释。首先是发表关于用户接口的论文有其固有的困难。正如一位与会专家所指出:“最好的用户接口是不需要用户手册的。”故而要写出抓住其本质的论文难度很大。其次是学术界特别是学术会议的程序委员会反对用户接口的文章,因为这类文章缺乏数学或其它“深奥”的内容。第三是用户接口必须做成之后才能显示其精妙之处。为做到这一点必须进行大量的“基本建设”,费时费力代价高昂。一旦不成功,“基本建设”的大量低级支持代码就等于白做,因此使研究者视为畏途,以这种“示范或死亡”(Demo or Die)来证明一种想法(idea)的事不是轻易能让人下决心的。但是考虑到用户接口的极端重要性以及工具环境等“基本建设”基础的逐步出现,占拉古纳海滩讨论会1/3的专家表示愿在这领域做研究。会议还建议 SIGMOD 和 VLDB 等国际会议接受录像记录(Video Track)以代替提交论文,从而便于用户接口研究者阐述他们思想。

第二个课题是积极或主动数据库。将触发器,报警器和约束作为 DBMS 的服务来支持,与会者一致认为这是一个富于成果的研究领域,提供这种服务将是未来系统的一大成功。

1994年9月7日到9日在希腊雅典召开的第五届国际数据库和专家系统应用会议的第一个综述报告(Tutorial)就是 S. Gatzia 所作的“主动数据库系统”。

第三个课题是并行性,诸如在单 CPU 系统中保持多个盘忙碌,共享主存中使用多 CPU 以及紧密耦合的分布式系统等。它提供用户以实时响应的幻觉能力和持有较短时间区间的封锁能力。这一课题得到的支持仅次于用户接口和积极数据库。

第四个课题是新事务模型。1988年以前的十五年中事务的概念没有发生过变化,如原子性,可串行性等。但新的应用如 CIM, CASE 等需要新的事务模型,如:(1)带有版本控制的 Check-in, Check-out 协议;(2)比可串行性更弱的语义;(3)在 COMMIT 之后能 UNDO 的事务等。这一研究领域受到专家们的大力支持。

第五和第六个受到大力支持的课题是新的应用,如计算机辅助软件工作(CASE)、计算机集成制造(CIM)、图象 DB(资源卫星数据,地理信息系统,医学

系统)、信息检索(IR)等。

第七个课题是安全和高可用性更好的故障管理。专家们认为这一课题并不容易,且隐藏在重要结果之后的是大量的低级代码工作。

第八个课题是大型分布式 DB。DDBMS 在 1988 年以前的十年中已做过了充分的研究。现在的关键问题是规模问题,例如 50000 个终端、上千个结点的分布式管理。会议认为研究界很少具有这种大规模系统的条件和经验。

第九个课题是 DB 和 OS 的交互作用。专家们认为由于 OS 自己的事情很多,期望 OS 对 DBMS 的需求作更多的响应是不现实的。但是对不同数据管理器或一个数据管理器同一个网络管理器的事务进行管理必须由建在 OS 核心中的事务支持来处理。OS 对 DBMS 的支持如快速大量存取数据,高性能的文件系统等研究应予重视和支持。

第十个课题是关于帮助用户设计事务的工具系统。

拉古纳海滩讨论会否定的研究课题集中为六项。受到一致否决的第一个课题是一般递归查询。会议认为不存在求解一般递归查询的客户,对于这种不存在需求的研究课题专家们抱怨各种学术会议接受了太多的这类论文。它重蹈了几年前没有多少实际意义的依赖性研究(Dependency Theory)的覆辙。

第二个被否决的课题是硬件数据库机。专家们认为运行并行软件的传统多处理器的纯软件数据库机更具竞争力,硬件数据库机无大指望。

不受欢迎的第三个课题是传统的并发控制。八十年代中发表了大量论文,但它们只有次要的细微差别。

否决的第四和第五个课题是一般的 OO 数据模型或其它新数据模型,理由也是文章太多而彼此差别不大。

第六个被否决的课题是 Prolog 和 DBMS 接口。会议认为太多的研究组在做着同一件事情,这种研究影响有限,建议不予支持。

3 维也纳会议分析

被肯定的第一个课题仍是用户接口,新的进展如工作流语言和合作工具(Workflow Lang. & Collaboration Tools)。遗憾的是进展仍主要由工业界做出而研究界贡献甚微。

第二个受到最大支持的是对超大数据库(10^{12} , 数百万兆或数万亿字节)和超大系统(10^3 , 数千客户)进行管理的 DBMS 规模的研究。这导致多级层次式的存贮管理和执行非常智能化的缓冲(Caching)策略,并要涉及数百万或数十亿的对象。大量的客户就要求系统在安装新应用程序拷贝时避免系统停机并保持不同客户应用的踪迹,因为不同客户可能运行同一应用的不同版本。

第三个课题是软件遗产的继承和大公司软件或信息投资的保护。大公司的商务活动所运行的应用系统一般都在十年以上或甚至更老。这些系统结构很差,使用的是不成熟的 DBMS 技术或甚至不是 DBMS。这些公司希望将这些遗产退休换上现代的 DBMS 和客户/服务器技术。有两种途径可以采取,一是推倒重写,一是增量式迁移。前一种办法工作量大,易错,很不现实,因此研究并提供可行的增量迁移策略是急迫需要的。

第四个课题是多媒体数据库,如存贮大的多媒体对象,如何在这些对象的内容上建索引,如何提供诸如有保证的递交等服务,等等。

受肯定的第五个课题是数据开采或数据发掘(Data Mining)。动力来自诸如大型连锁店的决策支持问题。零售时的每一项货品都会在 DB 中及时记录下来。主管要利用此数据库进行库存、销售、采购等决策。对 DB 的询问可能是:“告诉我令人感兴趣的问题”。系统应能“发掘”出有用的信息并提供出来。

以下三个课题亦受到一定的支持,尽管较以上五个课题的支持为少。第一个是移动式数据库,亦即某些客户有手持式设备,可断续地和 DBMS 服务器连接。故涉及到断开方式的系统操作然后在下游完成合并以及省电的查询优化等。

第二个课题是可扩充的及面向对象的数据库系统允许用户将用户定义的函数加进 DBMS 去。这种函数可用询问语言来写也可用通用语言如 C 混以 SQL 来写。对于后者其执行代价不易估计,处理含有这种函数的询问的智能询问优化器是重要的研究题目。

第三个课题是嵌入式 DBMS,例如电话交换应用的软硬件是在工厂做好的封闭世界,不会有其它用户干扰。但发生的问题是极高的可用性要求,如安装软件新版时系统不能停机以及极高的性能需求。

维也纳讨论会的专家们否决的第一个课题仍是一般递归查询。他们提出自拉古纳海滩会议以来的五

年中仍未发现有关该项技术的任何用户。

否决的第二个课题是应用于商业数据处理的传统单节点 DBMS 的优化。专家们说这一课题已经解决了。拉古纳海滩会议宣布传统的并发控制是一死亡课题。现在维也纳会议将这一宣布扩展到单节点 DBMS。

第三个课题是人为环境的性能研究。一个例子是具有小内存的基于盘的“玩具”DB 研究。由于当代计算机都有大内存。这种人为的环境是脱离现实的。因而没有意义。另一个例子是最近发表了的大量文章。其中设定等待封锁请求的概率大于 0.1。这也是十分脱离现实的。实际的 DB 事务极少等待。J. Gray 建立的定律是：一个良定义的 DB，等待封锁释放的概率小于 0.01。

第四个无吸引力的课题是 DDB 中数据对象多重副本的更新算法。对这一课题已写了大量论文且专家们感到还会再写许多论文。专家们认为这些文章与 1985 年发表的 El Abbadi, A. 等的论文所提出的算法少有本质不同。

受否决的第五个课题是新 OO 数据模型。已经发明和将要发明太多的数据模型。它们其实彼此差别不大。

其它获多数否决票的课题也都是由于离实际的应用有较大的距离。

4 总结

相隔五年举行了两次资深数据库学者的专题讨论会。他们有一定的代表性和权威性。但他们均来自系统领域。没有理论研究领域的专家与会。甚至连一位演绎数据库的代表也没有。因此他们的意见更倾向于实际。正如他们自己所提出的：应该面向应用而非面向技术！应该找一位顾客解决他的问题而不是找一技术问题并加以解决。这种观点也许同当今世界的计算机界的情况相吻合。著名的 COMPAQ 公司以前立足于高技术高质量。现在则更注重市场。这种情况当然会反映到学术界。笔者认为如果不是太走极端。则这种观点无可厚非。计算机科学和计算机软件不是数学。正如物理理论最终要以实验来检验那样。计算机

领域的研究自然最终要以应用来检验。

拉古纳海滩会议否决的课题这几年基本上停止了研究。唯一的例外是一般递归查询。该讨论会支持的重要研究课题在五年中广泛地受到重视。而能留在维也纳表上的则仅剩用户接口一个课题。这说明 DB 界对拉古纳海滩会议的呼声是响应的。在该会上有争论的 OODB 现在这种争议已不复存在。

对于维也纳会议结果的表 2。93 年的数据工程会有大量的文章是工作在其否决的课题领域。而几乎没有文章工作在表 2 支持的领域。这是不是反映了 DBMS 研究界多数是在错误的方向上工作？这就难怪维也纳讨论会的专家们把他们的报告定名为：“处于十字路口的 DBMS 研究”。他们认为诸如缓冲区管理、并发控制和询问优化等传统课题应该声明是已经解决了的课题。继续在这些熟悉了课题上“耕作”只能算炒冷饭。固然这么干较为安全。研究新课题则要承担更多的风险。不仅更为困难（例如数据采拙显然是 AI 完全的）且远离当前的 DBMS 活动中心。维也纳会议的专家还抱怨程序委员会出于“保险”原因而偏爱传统课题。但对新思想的论文特别是不够完整、不够形式化的文章过于苛求。呼吁改变这种状态。

笔者有幸作为程序委员出席了 VLDB93 都柏林会议。曾与个别专家进行过讨论。笔者认为用户接口确实非常重要。但它是数据库吗？用户接口在图形学会议上出现得更多。因此现在的到了一个学科交融的时代。传统的 DB 问题基本上已经解决。而新的课题已不是“纯”的数据库课题。从 VLDB 的发展也可以略窥端倪。早期的 VLDB 会正值 DB 研究顶峰。每次会议参加者逾千人。而 93 年的第 19 届 VLDB 会则不过三、四百人（作为对比。随后在法国举行的人工智能会议与会者超过千人。美国的多媒体会人数达两万多人。据说图形学方面的会议人数达五万多人）。认识到这种情势。那么不拘泥于传统的课题就很自然了。备受重视的用户接口除了它涉及数据的表示和使用之外显然与人工智能、多媒体有着密切的关系。几门学科的交汇必将产生崭新的前途无量的应用和研究领域。对这一点应充满信心。