

信息管理 数据库 管理系统

11

51-54

信息管理技术的最新进展

——参加国际智能信息管理系统会议总结

冯玲 冯玉才

(华中理工大学计算机系 武汉 430074)

TP391

TP311.13

1994年6月1日~3日我们出席了在美国华盛顿D.C.召开的国际智能信息管理系统学术会议,该会议由国际小型、微型计算机协会举办,来自美国、加拿大、德国、日本等的许多专家出席了会议。本届会议着重于研究信息的智能管理问题,录用的37篇论文中,涉及数据库领域的占14篇,网络领域的占13篇,并行领域的占4篇,应用领域的占6篇,现就以上领域,分述如下。

一、数据库技术

数据库是计算机数据管理的一门技术,具有数据结构化、低冗余度、高独立性、易扩充性和数据共享等特点,能有效地处理海量信息并高效地进行管理与维护。然而,今天市场上提供的数据库系统通常涉及到商务应用领域,其处理的事务都比较小。随着领域的拓广,这种传统的数据库管理系统难以应付新的应用需求,尤其自八十年代初、中期以来,出现了许多数据库的非传统应用,例如复杂对象、时态数据、多媒体数据、空间数据等。为了适应这些新应用的需求,提出了不少新的数据模型、系统结构和新的概念。几年来逐步出现了面向对象数据库、实时数据库、多媒体数据库、联邦数据库等。本届会议对上述各类数据库研究有不少论文作了阐述。

1. 面向对象数据库

近几年面向对象数据库已成为数据库发展的一个重要方向,之所以研制面向对象数据库,主要是因为需要数据库具有结构模拟和行为模拟的能力。这次会议上,面向对象数据库及面向对象方法是论文的重点之一,大致集中在性能提高、面向对象程序设计和实现技术三方面。这些论文已不再介绍一些面向对象数据库系统原型或面向对象数据模型的设计,而是比较深入地提出和解决面向对象数据库一些关键的实现技术问题和效率问题。

复杂对象的查询是面向对象数据库系统的核心技术之一,直接影响到系统的执行效率,是展现其强大生命力的关键,长期以来,人们尚未建立一个标准的面向对象查询语言。本届会议有两篇论文对此作了细致、深入的探讨。

在“面向对象数据库的智能查询处理”一文中,作者根据面向对象数据模型类层次概念所允许的类和实例,隐式地表达具有不同抽象级的查询结果,对于支持复杂的数据密集型应用,作者认为当前工作应集中在如何为面向对象数据库的查询提供一些方法和手段。目前的现状是,大多数方案以扩展的关系框架为基础,支持各种面向对象概念。典型地,一个查询的结果是满足特定谓词的对象集,这既不能提供一种丰富的表达技术,又非一种表达复杂信息的有效方法。事实上,人们可能对对象的特征及其所满足的条件感兴趣,即关心对象查询的内涵结果而非外延结果,因为内涵查询处理具有如下优点:

- 内涵结果通过公式集合来描述,它们独立于数据库的特定环境,因此较外延结果更稳定;
- 给定外延结果所需满足的条件,内涵结果提供了比实例集更简洁、更直观的形式;
- 内涵查询处理是对外延结果的解释;
- 在计算方面内涵查询处理优于外延查询处理,因为无需访问数据库中的数据,就可直接获得内涵结果,从而大大减少查询处理的代价。

在另一篇关于“面向对象数据库的查询预处理”一文中,作者通过大量实践工作,对面向对象数据库系统的研制,提出了一种全新的面向对象数据库系统的查询策略:经过查询预处理,任何一个面向对象数据库系统的关联查询均能转化为关系数据库中以单关系检索技术实现的查询。该文对查询预处理的过程进行了详细介绍,并给出了有关面向对象数据库查询的形式定义与描述。

尽管当今有些面向对象数据模型存在,且其查询语义较之传统数据库更为丰富、复杂,但它们均未达到充分的优化,尤其是涉及对象关联访问的应用,其执行效率远不如传统的数据库系统。作者认为,应将层次、网状数据库的导航式技术与关系数据库的描述式技术融合于面向对象数据库的查询处理之中,对任何一条 OSQL(扩展 SQL 支持面向对象数据库)查询进行以下预处理:首先将其分解成几个单类上的子查询,再通过 OSQL 的嵌套谓词“IN”引入子查询间的导航路径,从而无需联接运算,一个面向对象数据库的关联查询即可转化为一个完全能用关系数据库的成熟查询技术实现的查询。

2. 联邦数据库

分布数据库留下来的一个难题,就是如何把已运行的分布式异构数据库集成起来,构成联邦数据库,使用户好像在一种共同的数据模型上访问各类数据库。本届会议有两篇关于联邦数据库模型研究的论文。一篇是“Cyrano:联邦数据库系统的元模型”,文中所提出的异质数据库系统的元模型与已往的元模型有以下几点不同:

- 完全支持各成员数据库的查询语言,而不仅限于支持等价模式;

- 与传统的对象分类不同,它强调预先存在的对象分类,即文中给出的三类:门类、基础类和导出类。门类指将成员数据库中的数据映射为元模型;基础类指供 Cyrano 内部实现用;导出类指包含标识类实例存在的规则,可以访问其它的类。Cyrano 正是利用导出类来解决数据库的异质问题的。

在另一篇关于“联邦数据库管理的面向对象方法”文章中,介绍了联邦数据库的面向对象执行模型,给出了联邦数据库的定义——互相合作、自治、异质型数据库系统的集合,以及联合级上操作的处理,统一用对象来描述:查询对象、事务对象、安全性对象、完整性对象、模式对象。

3. 多媒体数据库

信息系统发展到今天大有将计算机信息系统、通信系统和大众传播系统统一成一个综合一体化大信息系统之势,而多媒体技术将在这个进程中扮演一个极为重要的角色。对多媒体数据的管理与操作较一般的信息处理不太相同,除了查找、检索、排序等常规操作外,需要诸如识别、理解、转换、压缩、合成、创作等特殊操作。论文“基于时间的多媒体数据

库概念模型”提出了一种新的多媒体数据库概念模型用以结构模拟多媒体文本,它将一个媒体视为若干个表示行为的集合,每一个表示行为负责一个多媒体对象的表现,而各个表示行为又可分解为多个子表示行为,以此构成一个文本的树结构。该模型通过一个称为“同步消息传递模型”,表达各个表示行为之间的合作与同步关系。文章采用了面向对象方法,以对象描述各表示行为,以各种不同的聚集关系描述对象间的同步关系。

4. 实时数据库

实时系统对事务的处理时间有着严格要求,对于大多数应用,与其在限定的时间内得不到一个数据库查询的精确回答,不如得到一个近似的查询结果。“实时数据库中临时数据的近似查询处理”一文将近似查询处理与临时数据模型相结合,设计出一个近似查询处理器,使之既满足时间要求,又符合临时数据的精度要求。

5. 分布式数据库

随着微型计算机和计算机局域网的迅猛发展,分布式数据库系统在信息处理技术领域中的重要性越来越突出,但是,随之也带来了分布式数据库管理数据的复杂性、数据完整性、一致性维护等一系列困难问题。有一篇名为“分布式数据库全局约束的局部测试公式的编译”文章,专门讨论了分布式数据库中较困难的完整性约束问题。基于一阶谓词逻辑,作者提出了分布式数据库中完整性维护的一种方法:从一条全局约束来生成若干条局部约束的测试公式,通过局部约束公式的简单、高效的真假值判定,完成全局约束的一致性检查与维护。

6. 关系数据库

关系数据库技术发展至今,其理论与实现技术日臻完善,其应用已渗透到社会生活的各个领域,产生了巨大的社会效益和经济效益。论文“数据库专家:一个关系数据库的设计工具”为数据库管理者和终端用户介绍了一种关系数据库的设计工具。它实现了 Bernstein 算法,可以将多个用户视图转化为一个最小的界面;也实现了基于推理规则的 Carlson/Arord 命题搜索算法。该工具允许重新定义用户视图,依据产生的最小界面和视图定义,提供映射机制,产生商用关系数据库系统的标准语句输入(如 ANSI 中的 CREATE TABLE、CREATE VIEW 语句)。同样,该工具的解释机制可以接受类似自然的

查询语言。一旦用户对数据库进行了修改,该工具即刻产生 ALTER TABLE 和 ALTER VIEW 语句反映其更新操作。总之,作为关系数据库设计的辅助工具,DB-EXPERT 可以保证设计的正确性,减轻设计负担。

二、并行技术

VLSI 技术正刺激着计算机系统结构朝着多处理机系统发展;许多大型计算需要的速度使目前超大型机难以胜任,如图象实时处理需要 13 亿次/秒。所以多处理机并行系统已成为当今及今后计算机发展的主要趋势,对此,本届大会即做出了反响。

目前,并行计算方法分为两大类:

- 具有多并行处理器的系统,如 CM-5, Intel Paragon。这些计算机内的节点由特殊硬件构成,大部分为高速微处理器,处理单元(PE)通过高带宽的通讯网络互联,这类网络可具有不同的拓扑结构。由于这类并行处理计算机使用了特殊的硬件和互联网络,因此每个目标系统都有特定的软件。

- 工作站间的并行。这种基于网络的工作站聚簇分布式计算在科研、工业领域中已为越来越多的人所接受。不同于第一种方法,这种方法由于只有一个通讯通道,其工作站聚簇的通讯带宽很低,而且网络不随工作站数目的增多而增多,所以在工作站增多时只会带来系统性能的下降。但是,大多数工作站都运行 UNIX 操作系统,如果设计者使用其通讯设施,则工作站聚簇上并行应用的设计与开发可完全独立于使用的硬件平台,当今已开发出多种工具为系统程序员提供高级的程序界面。

基于上述两种方法,在“工作站聚簇中使用并发网络体系结构的并行计算”一文提出了工作站聚簇的又一新型体系结构——并发网络结构,即使用二维或更高维数的网络结构改善传统聚簇的通讯带宽。该思想吸收了并行处理器系统与传统工作站聚簇的优点,填补了两者计算方法上的差异,聚簇结构被划分成多个子聚簇,形成一种新的并行计算,可以合作并发执行,因此较传统的聚簇,更加有效、更加灵活。

但是新的体系结构需要相应的新网络管理支持,除传统的网络管理功能,如用户管理、安全性管理或 IP 管理之外,还需要提供其它的新设施,如何将资源分配给不同的应用/用户,如何将聚簇划分成

多个子聚簇,如何有效利用每个节点上的资源(如具有协处理器的节点)。

在另一篇关于“ASTRA 多处理器通讯系统的体系结构”文章中,作者介绍了由 ASTRA (ASsociative RISC 结构)处理器构成的多处理器系统的通讯单元体系结构。

近年来微处理机设计中出现的 RISC(简化指令系统计算机)革命,使单个处理机速度提高十倍,到 1995 年将生产出每秒执行 5 亿条指令的微处理机。这些高速处理机芯片正在构成更快的并行计算机系统的部件。为了进一步提高工作站性能,设计新处理器体系结构势在必行,ASTRA 即为一种革新的体系结构,融合了支持协处理器和多处理器的新型体系思想,其核心包括现代微处理器中支持协处理器的重要构件,该处理器可以作为独立的处理器,或者作为支持通用处理器的主动/被动型协处理器,或者作为中心网、CAM 处理器的控制器,它可实现 SIMD 或 MIMD 域中新型相关多处理器体系结构。基于 ASTRA,人们可设计出直接网络包转换多处理器系统,类似于自治协处理器,ASTRA 的通讯单元可以处理多处理器系统中一个节点上的所有数据传输工作。

但是 ASTRA 通讯单元需要网络的额外虚拟通道,需要设计新的链接单元,且 ASTRA 处理器核心与通讯单元缓冲区之间还要求有更优化的耦合方式。

三、网络技术

本届会议所涉及的网络领域论文大部分着重于网络资源的研究,对资源的需求预测与均衡分配,并介绍了一个回溯型网络分析工具。

在智能通讯网络中,及时、准确、可靠地预测资源的需求对于分配和控制带宽模式十分重要,从以下三个级别我们可以看出预测的重要性:1)长范围的预测(如,月、周、日、时)用于网络资源的规划与测定;2)短范围的预测(如,日、时、分、秒)用于决定何时分派和重新分派资源,以及是否为外带访问公共网络的自动控制模式;3)秒范围的预测用于资源的动态转换以随时满足非预先安排的需求,或者根据用户即席输入,决定何时何地加紧或放松频率的控制机制。智能网络必须提供秒范围预测设施,能够预测到带宽要求,并能预测何时能够达到主要的命令阈值。

“智能通讯网络中资源需求的自适应预测”一文提出了预测带宽要求的一个通用启发式方法,其预测的有效性基于有效的时间段分析方法。具体体现在:i)启发式;ii)模块化;iii)使用有效的时间段方法;iv)牺牲少许的预测精确度以换取速度,确保统计数据的有效性;v)对于通讯量源于异质资源的非稳定流这一知识,所作断言与假设不能减弱或影响对其理解;vi)可计量的时间间隔;vii)适用于瞬间条件。

文中的启发式计算可以并行地运行于不太昂贵的微处理器阵列上,而且该方法可扩展为对其它类型资源的预测(如缓冲区需求)和度量(如最大、最小需求量)。

在“局域网互连网中测定通讯量的一个框架”一文中,作者认为对计算机通讯网络效率的研究有三种方法:定性分析、模拟与流量测量。过去主要运用前二种方法,近年来人们已普遍接受“测定通讯量是研究网络效率的最精确方法”这一事实。通过通讯量的测定,可导出通讯模式,可对操作网络实行更佳的管理,但是测量通讯量较传统方法需要更多的资源与机制。虽然目前已有多种用于网络通讯量研究的工具和程序,但大部分研究基于即席方式,且通讯量测定过程只适用于特定研究中的局部要求。因此,需要有一个模型对通讯量的测定过程进行描述,并给出关于通讯量精确度的定义。

作者在文中首次引入了测定通讯量的一个框架,其过程如下图所示:

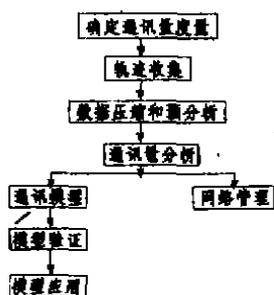


图1 通讯量测定框架

本文的工作对今后通讯量模型的研究及其网络

的管理具有指导意义。

一篇关于“多域网络管理环境下的资源划分”文章中,作者提出了多域环境下网络的管理问题。

随着开放式网络产品的开发与完善,如何对网络资源实行有效管理已成为当今亟待解决的问题。今天许多网络管理人员将精力花在维护网络的正常运行方面,忽略了网络资源的可视性,随着网络不断扩展,对客户虚拟网络资源的引入已成为必需。

另一篇“分布环境的自适应调度器”论文介绍了异质工作站网络上一种嵌入式的自适应分配体系结构,以解决分布环境下各种均衡问题。作者概述了自适应算法及其特点,通过一个工作站自动分派的简单例子,说明了算法的简单性与通用性;同时针对某些特定应用进行了种种优化。

“回溯型网络分析工具”一文介绍了一个运行于 SUN UNIX 操作系统下的回溯网络分析工具。随着基于 UNIX 的网络迅速普及,网络管理人员迫切需要有效的工具来对计算机网络实行最优管理。目前虽有一些工具问世,但绝大多数属实时分析工具,用于监控网络的运行,但能帮助管理和设计人员调整网络结构、确定最优资源分配的工具却不多。本文的网络分析工具运行于 Ethernet 上,能够帮助用户收集统计信息,分析各子网上的流量,并提供服务器上信息使用的各种情况,以辅助网络管理者实行网络的最佳规划。该工具已安装在 Martin Marriata、Syracuse 机上,能够诊断和分析网络结构方面的某些难题。

在本届大会上宣读的唯一一篇来自中国的论文“面向对象数据库的查询预处理”(冯铃 冯玉才)受到了与会者的高度评价,国际同行称赞我们做了件非常出色的工作,并充分肯定了我们在面向对象数据库领域所作的大量深入研究,这也为我国关于面向对象数据库的研究走向世界迈出了可喜的一步。

参考文献 Proceedings of the International Conference on Intelligent Information Management Systems, Washington D. C. 1994. 6