

基于 CSCW 的分布式数据库系统

沈 剑, 奚建清, 钟远明, 黄启荣

(华南理工大学 计算机工程与科学系, 广东 广州 510641)

摘要: 随着 CSCW 概念的引入, 计算机已经能给人们提供一个更智能的工作环境. 本文在分布式数据库系统的基础上提出了一种新的思想, 将 CSCW 技术与分布式技术相结合, 形成了一种新型的数据库系统, 即基于 CSCW 的分布式数据库系统. 它一方面解决了多用户对多类型数据库系统的协同操作问题, 另一方面对数据库中数据的利用率和访问速度也作了有效的改进.

关键词: CSCW; Agent; 分布式数据库系统

中图分类号: TP311

文献标识码: A

文章编号: 1007-7162(2000)04-0020-04

在信息化时代, 数据处理仍然是计算机应用的一个重要领域. 目前任何一个单位, 大至一个跨国公司, 小至一个工厂都往往由地理上分散的部门组成. 各部门地理上的分布造成了数据的分布. 为了有效的管理数据和共享数据, 我们在集中式数据库系统的基础上设计了分布式数据库系统(distributed database system, 简称 DDBS)^[1]. DDBS 是一个将一组有关的数据分布在计算机网中, 由分布式数据库管理系统(distributed database management system, 简称 DDBMS)统一管理的系统. 它比较有效地解决了集中式数据库系统通信开销大、可用性不高、可扩充性差及难以管理等问题. 但是, 随着数据库技术的不断发展, 数据库的类型变得越来越丰富. 不同用户对各种类型的数据库进行协同操作已经成为迫切的需要.

基于 CSCW 的分式布数据库系统就是试图通过将 CSCW 概念引入到分布式数据库系统中, 建立一个具有群体性、交互性、分布式和协作性的人机网络工作环境. 通过这个网络环境的支持, 地理上分布的不同用户对地理上分布的各种类型的数据库可以进行协同性数据库操作. 它是将 CSCW 技术与分布式技术相结合后在数据库领域中的典型应用^[2].

1 CSCW 环境的建立

1.1 CSCW 概念的提出

随着计算机技术和通信技术的飞速发展, 网络已经将不同地理位置的机器, 不同领域的人才紧密联系在一起. 为了实现一个较为复杂的设计, 我们需要不同领域的设计专家运用不同的设计方法和技术来达到设计目的. 例如机械产品的设计需要从产品的功能设计、产品的外观设计、产品的加工工艺设计、产品的费用、产品市场需求、产品的质量控制在以及产品的重用等几方面来并行地、整体地、综合地考虑某一产品的设计过程. 这一设计过程涉及到覆盖产品相关的各

收稿日期: 2000-03-31

基金项目: 广东省重点攻关项目资助(99B01702G).

作者简介: 沈剑(1977), 男, 硕士生; 主要研究方向: 面向对象数据库, 分布式数据库系统.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.>

个领域, 需要这些领域的专家运用相应的专业知识来进行设计. 每一领域的知识都是关键的, 但却又很难满足整个设计的总体要求. 我们需要不同领域的专家相互协调来共同完成设计工作. 在此背景下, 出现了一个崭新的领域, 计算机支持的协同工作 (compute supported cooperative work, 简称 CSCW)^[3].

CSCW 是综合研究计算机科学、心理学、人类工程学、认知科学、社会学等多个学科的前沿课题, 它通过先进的通信技术和多媒体技术来建立一个支持合作工作的平台, 该平台能使用户对同一产品进行共同开发、对同一专题进行共同讨论亦或对同一数据进行协作处理等. 由此可见, CSCW 的基础是以数据、图象和音频信息为特征的多媒体通信, 而其核心目标则是支持多个用户的协同工作.

1.2 CSCW 的基本分类

CSCW 在概念上是与人们合作方式相对应的, 而人们的日常合作往往是以同步或异步、同地或异地来进行划分的. 因此 CSCW 也可以分为四种不同的模式, 同地同步模式、同地异步模式、异地同步模式和异地异步模式. 其中最令人兴奋并且具有广泛的应用价值的是异地同步模式, 即在同一时间, 各个工作人员在不同地点共同协作完成某项任务, 现今较为成熟的异地同步模式的 CSCW 系统有视频会议系统, 该系统可以为分布于各地的会场实现视频、音频及数据的共享服务, 在其之上, 人们能够进行自由的讨论.

1.3 CSCW 的特点

CSCW 是指不同领域的专家在不同地理位置通过计算机网络进行协同工作. 相应地, CSCW 具有分布性、多主体性、协同性、共同性和灵活性等特点.

(1) 分布性: 不同领域的专家或用户在地理上是分布的, 因此他们享有的数据在物理上也是分布的.

(2) 多主体性: CSCW 技术的应用过程中都要求两个或两个以上设计专家参与, 这些设计专家通常是相互独立的, 并且各自具有领域知识、经验和一定的问题求解能力.

(3) 协同性: 各个设计专家的领域知识一般都是独立的, 而且相互之间有可能造成冲突, 因此我们需要具有一种协同机构来保证各个设计专家完成共同的设计目标.

(4) 共同性: 各个设计专家的设计目标是共同的, 其设计环境和上下文也是一致的.

(5) 灵活性: 参与设计的专家数目, 可以动态的增加或减少, 而且协同式设计的体系结构也是灵活的、可变的.

2 基于 CSCW 的分布式数据库系统的基本体系结构

基于 CSCW 的分布式数据库系统是在原有的分布式数据库系统基础上引入 CSCW 概念后实现的, 相应地, 其体系结构也是原有分布式数据库系统的体系结构的改进.

2.1 agent 概念的引入和多 agent 系统

agent 的概念源于分布式人工智能领域中的分布式问题求解与多 agent 系统 (multi-agent system, 简称 MAS) 等. MAS 的研究重点在于如何协同系统中的多个 agent 一起工作. 其中 agent 是指能在相应环境中连续自动运行的实体, 它能感知周围环境的变化, 并作用于环境. Agent 还有一定的精神状态, 这状态影响它的决策行为. 这种精神状态可由信念、许诺、能力等来描述. Agent 具有自学习能力, 它能够进行相互学习、询问与协作, 从而可以解决一共同问题.

2.2 多 agent 模型

基于 CSCW 的分布式数据库系统是将地理上分布的许多不同类型的数据库通过计算机网络来统一协调管理, 它的主要目的是在数据库分布的情况下, 保证用户对数据库操作的正确性和透明性. 正确性指的是对于各种数据库操作, 如查询、插入、删除等要保证数据正确. 而透明性指的是用户觉察不到数据库物理位置上的分布以及类型上的不同. 为了实现这些基于 CSCW 的分布式数据库系统的设计目的, 我们需要协同管理此系统中的各个数据库, 这一点与多 agent 系统有很大相似之处. 因此, 在基于 CSCW 的分布式数据库系统的体系结构的设计中, 我们提出了多 agent 模型^[4, 5].

2.2.1 总体设计

物理上分布的每一个数据库相当于多 agent 模型中的一个 agent, 当然我们需要将数据库封装起来, 使它满足多 agent 系统的设计需求. 在分布式数据库系统中, 数据分布一般有三种方式: 划分式、全重复式和部分重复式, 其中部分重复式是指部分结点拥有全部数据的复本, 它是前两种方式的折衷, 比较充分地体现了分布式数据库系统的特点. 在本系统的多 agent 模型中, 我们对于数据分布也采用部分重复式, 不过拥有全部数据复本的结点, 我们只设定为一个, 也就是说, 多 agent 系统是由一个主 agent 和多个次 agent 组成, 如图 1 所示. 主 agent 与次 agent, 次 agent 与次 agent 之间部可以相互通信, 在通信的过程中相互协调, 共同完成某一个数据库操作.

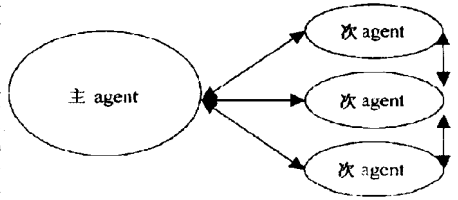


图 1 多 Agent 系统

2.2.2 agent 的内部结构

为了实现多 agent 系统之间的协调工作, 我们必须把原有的每一个结点上的数据库系统加以封装, 满足多 agent 系统的设计需求. 在本系统中, agent 由 5 大模块组成, 用户模块、数据模块、通信模块、控制模块和 cache 模块. 如图 2 所示.

a) 用户模块: 用户要对数据库进行各种操作, 当然首先要提出对数据库的操作申请. 该模块就是处理用户提出的各种申请, 并相应地转化为系统能够识别的语言.

b) 数据模块: 数据库中存放各种重要的数据信息, 数据库系统的重要目的就是有效地获取数据和处理数据.

因此, 数据模块是 agent 中最核心的模块, 它包括数据库和屏蔽数据库类型的接口.

c) 控制模块: 这是整个系统的主控模块, 它决定了系统是如何有效地协调工作.

d) 通信模块: agent 之间的通信渠道. 本系统中, 各个 agent 之间需要协调工作, 于是 agent 与 agent 之间的通信是必需的. 一般对于分布式数据库系统来说, 通信部分都是整个系统的瓶颈部分, 因此通信模块设计的好坏决定了整个系统的效率.

e) cache 模块: 在 agent 引入 cache 模块的主要目的是为了提高数据利用率和减少通讯流量. 在传统的数据库系统中, cache 技术主要运用于服务器端(相当于本系统中的主 agent), 目的是为了减少 I/O 操作. 在本系统中, 我们在主 agent 和次 agent 中都设计了 cache 模块, 目的是加快信息检索时间, 减少数据传输和减少主 agent 的负载.

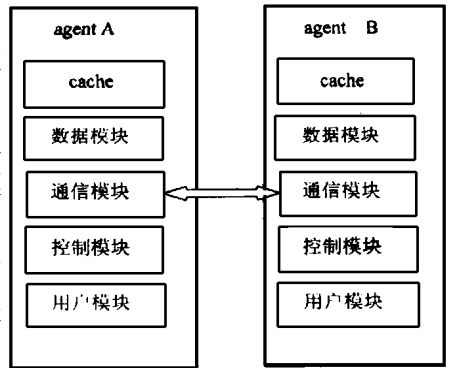


图 2 Agent 内部组成

2.2.3 agent 之间的通信机制

本系统中的每一个 agent 都需要不断与其他 agent 通信, 了解其他 agent 的最新数据信息. 只有这样, 整个系统才能及时准确地进行协调工作. 为了让 agent 之间更好地通信, 我们需要制定一个高效可行的标准来保证 agent 之间的通信. 我们通过实际的数据库操作来说明 agent 之间的通信机制.

a) 当用户第一次提出申请时, 首先询问主 agent, 由主 agent 指定满足用户需要而且离用户物理位置最近的次 agent. 这样, 用户所需的数据将从该次 agent 中得到.

b) 在系统运行过程中, 次 agent 按照一定的时间间隔与其他 agent 通信, 了解各种数据的实际位置. 这样对于用户的下一次申请, 用户就不需要先询问主 agent, 而直接通过上次访问的次 agent 来定位将要访问的次 agent. 这样就加快了数据访问的速度.

c) 对于各种数据库更新操作, 为了保证数据之间不冲突, 每一时刻我们只允许一个 agent 提出数据库的更新操作, 也就是获得了发言权. 发言权的获得与否由主 agent 控制. 数据库的更新操作流程如图 3 所示.

Agent 应该首先获得发言权, 然后系统判断其是否具有写的权限. 如果权限审核通过, agent 先将更新的数据置于本地的 cache, 而不作数据库插入操作. 等到把数据传送给其他 agent 并确定无冲突后, 数据库才进行插入操作, 并同时 will 各个 agent 进行更新. 否则就交出发言权, 进行下一步工作.

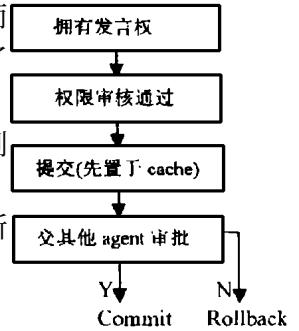


图3 数据库更新操作的流程图

3 CSCW 环境下分布式数据库系统的性能分析

将 CSCW 概念引入分布式数据库系统后, 该系统除了具有分布式数据库系统的一些基本功能外, 还具有 CSCW 系统的一些特别功能. CSCW 技术与分布式技术的结合, 使得本系统获得了比传统数据库更好的性能.

(1) CSCW 技术的引入, 使得各个物理上分散的数据库更好地协调工作, 提高了数据访问的局部性和数据的利用率.

(2) 主、次 cache 的设计加快了数据的检索时间, 减少了数据传输和减少主 agent 的负载.

(3) 多 agent 技术的运用, 使整个系统变得更加智能.

(4) 数据更新时用户权限的审核以及数据交其他 agent 审批的过程, 提高了数据的利用率, 避免了许多冗余的数据库操作.

4 总结

随着多媒体技术的发展与网络速度的提高, 基于 CSCW 的分布式数据库系统必将有着越来越大的应用前景. 本文仅介绍了基于 CSCW 的分布式数据库系统的实现构想, 但这是个新兴的课题, 其中还有许多值得深入和研究的地方: 如怎样有效利用原有的数据库, 如何在 agent 之间传送数据进行审批, 如何更好对用户权限进行统一管理, 这些都有待于进一步的研究.

Database in Developing Pitch Pug Mill' s Control System Under Kingview

GUO Xiu-jiang, WANG Ren-huang

(Faculty of Automation, GDUT, Guangzhou 510090, China)

Abstract: A brief instruction is given to the technology of operating database of the control system of asphalt concrete mixer plant. Kingview database interface and the ODBC data source of Windows98 operating system are used to operate Access database by SQL.

Key words: asphalt concrete mixer plant; asphalt concrete; control system database

(上接第 23 页)

参考文献:

- [1] 王能斌. 数据库系统[M]. 北京: 电子工业出版社, 1995. 175-181.
- [2] Rodden T. Distributed system support for computer supported cooperative work[J]. Computer Communications, 1992, 12 (8): 527-538.
- [3] 潘云鹤. 智能 CAD 方法与模型[M]. 北京: 科学出版社, 1997. 228-235.
- [4] Sycara K, Pannu A. Distributed intelligent agents[C]. IEEE Expert, Dec 1996, 36-46.
- [5] 李红臣, 史美林. Agent 在工作流管理系统中应用研究[J]. 通信学报, 1999, (9): 16-22.

Distributed Database System based on CSCW

SHEN Jian, XI Jian-qing, ZHONG Yuan-ming, HUANG Qi-ying

(Computer Science Department of South China University of Technology, Guangzhou 510641, China)

Abstract: A new idea based on the distributed database system was proposed which combines CSCW technique and distribution technique to form a new kind of database system. This system allow multi-user to cooperate in multi-kind database, and increase the access speed and the data utilization efficiency in database as well.

Key words: CSCW; Agent; DDBS