

联邦数据库系统

数据库

互操作性

⑫

计算机科学1994Vol. 21No. 6

55-5)

联邦数据库系统研究

肖永桥 姚卿达

(中山大学计算机软件研究所 广州510275)

TP311.13

摘要 In the heterogenous mulit-databases environment, the interoperability of different database systems is becoming more and more important. Now there are four main ways to handle it, such as gateway, mulit-database system, distributed database system, federated database system. In this paper, we compare and analyze them by networking method, and we think federated database system is a better choice. In addition, we introduce the design structure of the federated database system-LNFDBS being developed by us.

关键词 Federated Database System Component Database System Multi-Database System Distributed Database System Special Database System Gateway

1. 引言

网络的发展使不同数据库系统之间的互操作在物理层成为可能,但要真正实现,还得依赖于逻辑层的处理,即数据库系统之间的交互。由于不同数据库系统之间的异质性和封闭性,使得这一技术实施起来很困难。

当前解决的方法主要有网关、多数据库系统、分布式数据库系统、联邦数据库系统等。本文用网络方法对它们进行了分析和比较,认为联邦数据库系统是一种较好的选择,并提出了我们正在开发的联邦数据库系统——LNFDBS的设计思想。

2. 分析和比较——网络方法

通常,多个不同的数据库组成一个异质多数据库环境,其中的数据库系统称为成员数据库系统。为了便于分析,我们采用网络方法,即把一个数据库系统看成一个结点,不同的数据库系统在物理上通过网络连接起来。这样,一个多数据库系统环境就可以看成是一个网络 (N, E) ,其中 N 是结点集, E 是边集,不同的数据库系统之间的互操作就可以看成是网络上不同结点之间的通信,也可看成网络互连(一个结点上的数据库系统可能是一个网络)。

2.1 网关(Gateway)

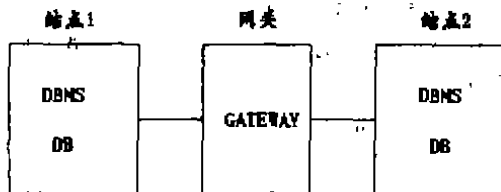


图1

网关是一个网络术语,用于把不同的网络连接起来,这里,只不过是用于不同数据库系统之间的连接^[1]。其结构如图1。

网关的主要作用是转换和通信,其中转换包括模式转换、操作转换等,通信可以看成是点对点通信。通过网关,一个结点的用户可以访问另一个结点的数据库,这种方法实现起来较为简单,一般的数据库系统本身就提供这样的工具,如 ORACLE 的 SQL * NET,但这种方法一次只能访问一个数据库,故称它为单数据库系统。

2.2 多数据库系统

多数据库系统是指一个结点可同时访问多个结点的数据库或同一结点的多个数据库(同质环境),而这些数据库系统都是局部自治的,并且没有集成的全局模式^[2,3]。其结构如图2:

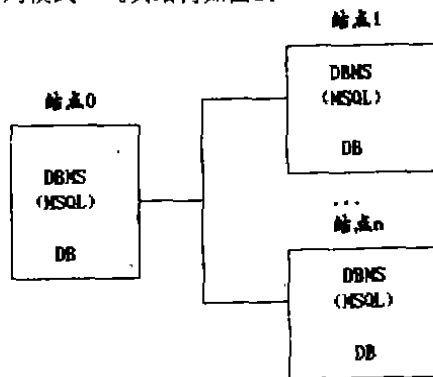


图2

多数据库系统由于允许一个结点同时访问多个结点的数据库,这样就得扩充数据库系统的操作能

力,一般要求数据库系统本身提供多数据库查询语言—MSQL,使用这种方法,用户对多个数据库的访问一般是非透明的,即用户得显式地指明所要访问的数据库。

2.3 分布式数据库系统

分布式数据库系统是把分布在不同结点上的数据库集成为一个统一的数据库模式(全局模式),从而提供给每一个结点以透明的分布式查询^[1,2],其结构如图3:

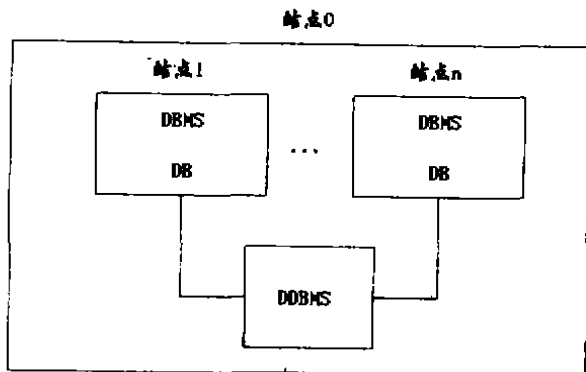


图3

分布式数据库系统集成多个结点的数据库为一个统一的全局模式,各个结点只能通过这个全局模式访问其它结点的数据库,它在逻辑上可看成一个结点,使用这种方法,用户对多个数据库的访问是透明的,即用户不必显式地指明所要访问的数据库,但是,由于不同结点的异质性,构造一个全局的数据库模式是不容易的,有时也没有必要。

2.4 联邦数据库系统(FDBS)

联邦数据库系统最早由 Hammer 和 Mcleod 提出时是作为一组松耦合的部件(如对象、记录、类型)的联邦^[3],这个概念后来由 Heimbigner 和 Mcleod 引伸为没有全局模式的松耦合库的联邦^[7],它相当于上面讲的多数据库系统,只是一些术语不一样罢了。后来 Sheth 和 Larson 运用这一术语来描述具有更广泛意义的系统^[8],其结构如图4:

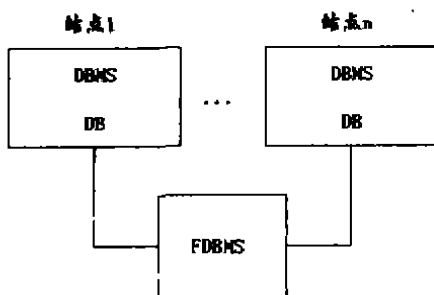


图4

Sheth 和 Larson 在文[8]中提出了联邦数据库系统的参考结构,其中包括五级模式:局部模式、成员模式、输出模式、全局模式(联邦模式)、外模式。局部模式是指各个成员 DBS 自治管理的概念模式,不同的局部模式可以用不同的数据模型。成员模式是指用 FDBS 的公共数据模型表示的局部模式,采用公共数据模型可以把用不同数据模型表示的局部模式转换为统一的数据模型。输出模式是成员结点向联邦提供的可供其它成员共享的数据库模式,它可以是成员模式的子集。全局模式是由各成员结点的输出模式集成的,成员数据库只有通过全局模式才能共享其它成员的数据库。外模式是指为用户或应用提供的视图。FDBS 在逻辑上不可看成一个结点,因为其全局模式不一定是各成员数据库系统的局部模式的全集,这一点与分布式数据库系统不同。使用这种方法,用户对多个数据库的访问也是透明的。此外,因其全局模式是由各个成员数据库系统的输出模式集成的,全局模式的构造可能比分布式数据库系统容易。

2.5 评论

上面我们用网络方法分析比较了四种方法,但实际上这四种方法(尤其是后三种)的区别越来越不明显,一些学者分别从不同的角度进行了扩充,如 Ram^[9]等。

3. 一个联邦数据库系统——LNFDBS

通过上面的分析比较,我们认为,联邦数据库系统是一种较好的选择,下面介绍我们正在研制的一个联邦数据库系统——LNFDBS,它综合了上述四种方法的优点。

3.1 系统特点

LNFDBS 有下列特点:

(1)客户/服务器体系结构。LNFDBS 采用这种结构不但可较容易地实现分布式处理,而且可减少网络传输费用,从而提高系统的效率。

(2)自治性。LNFDBS 允许局部结点具有自治性,如它有权选择自己的数据库模式和事务处理策略。

(3)分布性。LNFDBS 提供一个全局模式,各个结点通过这个全局模式可透明地访问其它结点的数据库。

(4)协商功能。LNFDBS 的全局模式是由各个结点的输出模式集成的,而各个结点的输出模式是由联邦 DBA 和局部 DBA 协商决定的。

(5)点-点通信。LNFDBS 除了允许结点操作局部模式和联邦模式以外,也允许结点之间点到点通信,这相当于上面网关的功能。

3.2 体系结构

LNFDDBS 的体系结构如图5:

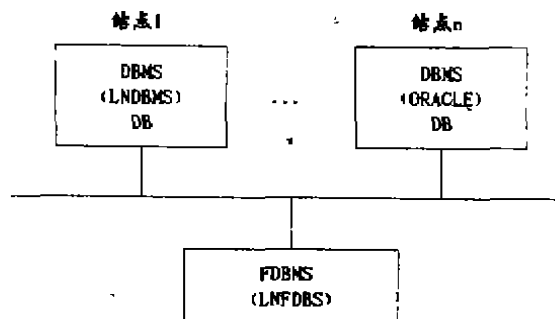


图5

LNFDDBS 采用类似于总线形网络的体系结构,各个结点相当于工作站,FDBMS 相当于服务器,各个结点的 DBMS 具有自治性,同时也可通过 FDBMS 访问其它结点。当然,更重要的是,它也允许任意结点之间的直接通信。这是必要的,因为 FDBMS 结点有可能成为瓶颈,影响系统效率,此外这种结构也更加适用于企业的组织模式。

目前,我们以 ORACLE、FOXBASE (FOX-PRO)、LNDBMS 作为 LNFDDBS 中的三个结点,ORACLE 是一个功能很强的通用数据库系统,它的安全性和完整性都很强,但它的效率在一般的微机资源下不很高^[10],FOXBASE (FOXPRO) 的效率较高,但安全性却很差^[10],LNDBMS 是我们正在开发的面向数据采集的专用数据库系统,它具有很高的效率和较好的安全性^[10,11]。我们选择它们就是想综合它们的优点,使 LNFDDBS 既有很高的效率和较好的安全性,又有很强的功能。

本文介绍了 LNFDDBS 的特点和体系结构,有关

其进一步的技术问题,将另文发表。

参考文献

- [1] 唐雪飞,数据库的互操作性,计算机科学,Vol. 20, No. 3, 1993
- [2] W. Litwin, et al., Interoperability of Multiple Autonomous Database, ACM computing Surveys, Vol. 22, No. 3, 1990
- [3] W. Litwin, et al., Multi-database interoperability, IEEE Comput. Vol. 19, No. 12, 1986
- [4] S. Ceri, et al., Distributed Databases-Principles and System, McGraw-Hill, New York, 1984
- [5] W. Litwin, et al., SIRIUS System for Distributed Data Management, in Distributed Data Bases, H.-J. Schneiders, Ed. 1982
- [6] K. Hammer, et al., On database management system architecture, Tech. Rep. MIT/LCS/TM-144, 1979
- [7] D. Heimbigner, et al., A Federated Architecture for Information Management, ACM Trans. on Inf. Syst. Vol. 3, No. 3, 1985
- [8] A. P. Sheth, et al., Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, ACM Computing Surveys, Vol. 22, No. 3, 1990
- [9] S. Ram, et al., Architecture of Distributed DataBase System, the Journal of System and Software, Vol. 10, No. 2, 1989
- [10] 肖永桥,姚卿达,单一式和客户/服务器数据库系统:测试和分析,计算机科学,Vol. 21, No. 5, 1994
- [11] 姚卿达,肖永桥,面向数据采集的专用数据库管理系统—LNDBMS 的设计与实现(待发表)

(上接第54页)

建立高层数据视图。

数据库封装的缺点主要表现在对厂商所提供封装软件的依赖上,因为,对数据库进行封装要求专门的技术和编程技能,通常厂商对数据库封装框架是全封闭的,不支持第三方厂家提供新的数据库封装手段;同时,厂商还必须提供自己的 API 和前端应用,允许客户机上的应用可以对数据库对象发出请求,往往这种 API 较难掌握,与 SAG 或者 Microsoft 的 API 相比,显得有此曲高和寡。尽管还存在一定的缺点,但随着面向对象技术的逐渐成熟,数据库封装将成为现阶段解决数据库互操作性,实现可互操作数据库平台最有希望、最有前景的体系结构。目前主要的数据库封装软件厂家有 Constellation Software 公司, OpenBooks Software 公司及 Ontos 公司。

6 结论

通过对上述五种体系结构的讨论,我们可以得出这样的结论:开放式信关和数据库封装是目前较理想的互操作数据库平台的体系结构,开放式信关以 SQL 语言为基础,对用户来说,实现现有的各种数据库互操作性是可以获得的,事实上 SYBASE 在这方面已取得了很大的成功,尽管基于数据库封装方法的互操作数据库平台产品还不成熟,但是所采用的面向对象的技术本身就意味着数据库封装将是更灵活、更长期的解决数据库互操作的途径。由于能够较好地解决一种数据库系统与多数据库系统的透明访问,带常规信关的 RDBMS 体系结构仍将继续为人们所接受。至于数据库访问 API,笔者认为,很有希望与开放式信关及数据库封装的客户 API 相结合,成为一种互操作数据库平台的应用界面的标准。(参考文献共14篇略)