维普资讯 http://www.cqvip.com

网格数据库服务中的需求和解决方案

张 凌[1] 王 康[2] 冯 欣[1]

(重庆大学计算机学院 重庆 400044)[1](重庆大学网络中心 重庆 400044)[2]

摘 要 网格计算是专门针对复杂科学计算的一种新型计算模式,而分布式数据库系统是数据库系统的发展趋势,本文定义了网格数据库服务的需求,并给出了一种概念框架模型。这种模型能够很好地屏蔽数据库的异构性,并且有标准的接口,易于扩充,具有良好的发展前景。

关键词 网格,数据库服务,数据库管理系统,外部资源模型,逻辑资源模型

1 引言

随着人们日常工作遇到的商业计算越来越复杂,人们越来越需要数据处理能力更强大的计算机,而超级计算机的价格显然很难进入普通人的工作领域。于是,人们开始寻找一种造价低廉而数据处理能力超强的计算模式,网格计算(Grid Computing)就是在这种情况下应运而生的。这种计算模式利用互联网把分散在不同地理位置的电脑组织成一个"虚拟的超级计算机",其中每一台参与计算的计算机就是一个"节点",而整个计算是由成千上万个"节点"组成的"一张网格"。这样组织起来的"虚拟的超级计算机"有两个优势,一个是数据处理能力超强;另一个是能充分利用网上的闲置处理能力。

简单地讲,网格是把整个网络整合成一个虚拟的巨大的超级计算环境,实现计算资源、存储资源、数据资源、信息资源、知识资源和专家资源的全面共享,目的是解决多机构虚拟组织中的资源共享和协同工作问题。在技术层面上,网格计算是分布式计算的一个新的阶段,它解决在动态的异构的虚拟组织中如何控制和协调对资源的共享。网格数据库服务是连接现有数据库和网格应用的桥梁,它是网格计算中的重要组成部分。随着网格技术的不断发展与成熟,数据库系统也将面对网格带来的巨大挑战。

2 研究现状

目前,很多网格应用使用了结构化的数据,例如生命科学研究和地球科学研究绝大多数的商业应用中已经部署了数据库系统[1],而网格本身用结构化的数据来实现其运行和管理,因此在网格上使用数据库系统是必需的。按照应用层次的不同可以把网格分为3种:计算网格,提供高性能计算机系统的共享存取;数据网格,提供数据库和文件系统的共享存取;信息服务,网格则支持应用软件和信息资源的共享存取。其中数据网格保证用户在存取数据时无须知道数据的存储类型(数据库,文档,XML)和位置。涉及的问题包括:如何联合不同的物理数据源,抽取源数据构成逻辑数据源集合;如何制定统一的异构数据访问的接口标准;如何虚拟化分布的数据源等。目前,数据网格研究的问题之一是:如何在网格环境下存取数据库,提供数据库层次的服务,因为数据库显然应该是网格中十分宝贵且巨大的数据资源。数据库网格服务不同于通常的数据库查询,也不同于传统的信息检索,需要将数据库提升为网格服务,把数据库查询技术和信息检索技术有机结合,提供统一的基于内容的 TOP-K 数据库检索机制和软件[2,3]。

网格数据库面临的主要问题:(1)数据库必须符合网格的标准,数据库应该成为网格中的一种资源并且提供相应的服务;(2)不同种类的数据库产品在功能和接口上也有很大的不同,在集成各种数据库系统到网格中时要尽可能保留这些系统的全部功能;(3)网格鼓励数据共享,因此需要中间件来完成异构数据的集成;(4)网格包含结构化数据、半结构化数据和无结构的数据。

针对上述问题,使用较少的集中控制,同时又要用最高质量的服务来实现跨众多虚拟组织共享的资源之间的高度交互,这是一项技术挑战。全球网格论坛(Global Grid Forum,GGF)必须设法通过一组软件体系结构标准和其它框架来使这一资源共享的

张 凌 硕士研究生,主要研究方向:Internet 网络技术、信息网络安全技术;王 康 教授,硕士生导师,主要研究方向:Internet 网络技术,信息网络安全技术;冯 欣 博士研究生,主要研究方向:计算机网络信息安全。

维普资讯 http://www.cqvip.com

过程标准化。GGF 已经着手进行一些体系结构标准化工作，以便提供更佳的软件互操作性、高级别的安全性、资源定义和发现、策略和可管理性。开放网格服务体系结构(Open Grid Service Architecture，OGSA)就是这样一个体系结构标准化过程。该体系结构的基本组件是开放网格服务基础结构(OGSI)[4,5]。DAIS 工作组(Database Access and Integration Services Working Group)隶属于 GGF 的数据领域，其主要研究内容是如何将数据库运用到网格中[6]。DAIS 工作组正在制定网格数据库服务的标准，该标准的草案可以从 GGF 的网站上下载。OGSA2DAI (Open Grid Services Architecture2Data Access and Integration)项目的目标是构造一个中间件，用于网格环境中对数据的存取和集成，是 DAIS 工作组制定的网格数据库服务标准草案的一个参考。

3 网格数据库服务

网格数据库服务的目标是将现存的数据库通过网格数据库服务(Grid Database Services)集成到网格环境下，使其能被网格应用访问。网格数据库服务是连接现有数据库和网格应用的桥梁，我们要做的工作是定义符合 OGSI 的网格数据库服务。目前并没有网格数据库这一概念。

首先构造一个网格环境下访问数据库的场景：一个用户想得到某霉菌中有特定功能的蛋白质的相关信息，假设没有单个的数据库能够提供这样的数据，那么他必须知哪个数据库能够提供这样的信息。用户获取这样的数据必须经历以下步骤，首先用户要查询 Information Service，得到相关数据库服务的元数据；得到服务之后用户必须选择服务；根据选取的服务来满足用户的需求，根据这个场景我们提出了网格数据库需求，并给出了一种解决方案的概念模型。

3.1 网格数据库的需求

网格数据库的需求总体上分为两个方面的需求：一个方面是针对网格数据库服务的要求，包括查询、数据库存取控制等；第二个方面是对网格数据库服务作为网格服务的要求，包括安全考虑、可扩展性、透明型，性能监控和可调度等，并且网格数据库的服务必须符合网格的标准，可以与其它网格服务顺利交流，具体一点可以有以下六个方面的需求。

1)元数据。数据库的元数据是数据的数据，元数据的作用在于表述每个数据库的位置和接口、其上允许的操作、存储数据的类型等等。网格应用通过元数据对数据库访问需要首先查询目录服务确定哪些数据库服务满足要求，然后将请求提交给相关的数据库服务。最好能够提供自动产生元数据的工具和定义元数据的标准。

2)历史元数据。提供关于一条数据的起源和历史使用情况的信息，是对数据的一种特殊形式的跟踪，记录了数据的创建和来源、拥有者、机密程度、移动情况、在数据上进行的处理以及产生的结果，包括“谁在什么时候做了什么”，一条记录的历史元数据的结构和内容可能会非常复杂，因为数据(特别是经过提取的数据)经常来自多个数据源，经过多阶段的处理、分析和整合。历史元数据十分重要，它用于确定数据的拥有者、质量、可信度和提供了关于数据创建的信息，对重复实验起到很重要的帮助作用，在网格数据库中必须提供记录历史元数据的解决方案。历史元数据应该尽量以自动的方式获得，而不是手动去填写。应该提供给用户工具，帮助用户以最小的努力建立关于现存数据的关键性的历史元数据，应该提供分析历史元数据以及检查一致性和有效性的工具。

3)数据发布和发现。当一组数据库服务发布到网格上时，就是关于这组数据库服务的元数据注册到一个目录服务中，网格数据库必须制定统一的标准来描述发布的数据。所提供的表示发布数据的方法必须十分灵活，所制定的标准必须能够支持所有类型数据的表示，无论数据的大小、内部结构和格式，也要支持用户以自己定义的格式来描述和组织数据；另外，应该尽量保持元数据的定义和发布过程是自动的，最好应该提供可交互的浏览工具，根据应用或用户自己的兴趣来发现数据。用户应该能够基于统一的语义和形式，以自己定义的条件和规则使用这些工具查询。

4)数据的存取控制。网格数据库必须提供能保持数据机密性的机制和提供灵活的授权机制，包括数据拥有者必须能够赋予和回收其他用户对数据的访问权限，把这种权力赋给一个可信任的第三方以及动态赋予和回收，可以设定时间限制等。应该满足对数据不同粒度的访问需求，即可以从对整个数据库的访问到对数据的子集的一部分值的访问 ，而且各种权利应该可以融合，即在同一粒度上对已经拥有读权力的用户再赋予插入、更新、删除等权力，除了以上的需求以外也必须能够基于用户的角色来限制对数据的访问。

5)分布式查询处理。首先，网格数据库能够将查询请求分布到不同的 DBMS 管理的数据库系统中，并且所有的交互都要符合网格的标准，然后必须制定标准协议和接口，包括用户和网格数据库服务之间以及网格数据库服务和不同的数据库系统之间，当然还需要涉及到比如网络、处理器的资源调度和与其它网格服务的合作。

维普资讯 http://www.cqvip.com

6)数据操作。网格数据库必须能够将以元数据形式表达的提取目标、输出和提取条件转换成实际可获取的数据源和可执行的数据结构;必须能够根据所提供的参数(如是否是关系数据库、何种查询语言等)构建符合查询语言语法和语义的查询条件。当查询涉及到两个或两个以上的数据源时;必须有能力把查询分布到这些数据源上,并且以统一的形式将结果返回;数据操作必须能够以和用户交互的方式运行;传统的数据库查询采用批处理的方式,即用户先将数据操作构建完毕然后再提交给数据库,处理的过程中用户不能干涉。网格环境下,用户希望能够用高级的可视化工具来检查、分析和翻译正在处理的数据,甚至可以改变或追加查询条件。

3.2 网格数据库的解决方案

根据以上的需求,本文在网格数据最初的框架(CGF4)基础上,提出了一种新的网格数据库的概念模型,较好地解决了以上的问题。早期的数据库联合的框架图如图 1 所示,可以看出,早期的 CGF4 数据库联合有如下的缺点:1)不符合 OGSI 的规范;2)太底层;3)这里服务重视功能而不是接口的标准。

本文提出的新的网格数据库的模型是网格数据的一部分,一直以来,对于网格数据的研究都是先于网格数据库的研究,这种新的网格数据的资源分为外部资源模型和逻辑资源模型两个部分。

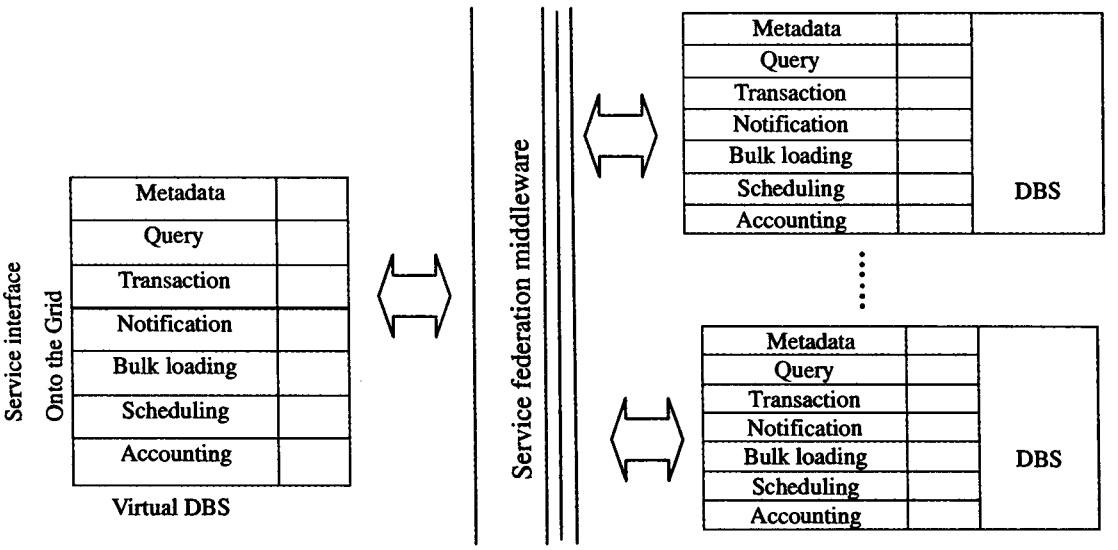

图 1 数据库联合

外部资源模型如图 2 所示,包括资源管理(EDRS),资源数据(EDR)和数据的结果集(EDS),这三个部分中资源管理(EDRS)是最外层的访问接口,它实现对数据库的访问,通常情况下,它包括多个数据库的访问实例(如 Oracle 9i,DB2 等);数据库的资源(EDR)包括多个物理数据库的集合,而多个数据库都又会被数据库管理系统(DBMS)管理;数据的结果集(EDS)是独立于数据库管理系统的。

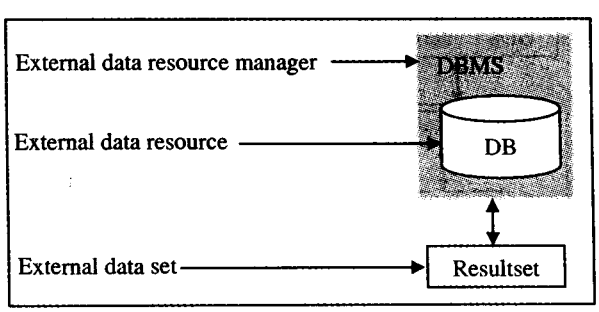

图 2 外部资源模型

与外部资源模型相比较,逻辑资源模型图 3 所示,它在外部资源的上一层,包括资源管理(DRM)数据活动会话,数据请求和数据集几个部分。数据资源管理器(DRM)是网格服务的集合,网格服务用来实现对数据集的管理,可以像获得任何其他网格服务一样来获得 DRM,即取得 DRM 的网格服务句柄(Grid Service Handle)和网格服务引用(Grid Service Reference);也可以通过工厂方法(factory)创建一个 DRM 服务,这个 DRM 通过外部资源管理器的名字绑定到事先已经存在的外部资源管理器上,逻辑资源管理器和外部资源管理器是一对多的关系;一个数据资源(DR)也是一个网格服务,主要是提供数据元的服务,它与一个外部数据资源是一一对应的关系。数据集提供数据集的服务,数据集被创建时可以绑定到一个外部数据管理器,或者初始化空;数据活动会话(DAS)提供了一次数据请求交互所需要的服务,网格应用通过数据资源(DR)创建一个数据活动会话(DAS)服务。一个数据活动会话(DAS)可以对创建它的数据资源(DR)进行操作。数据活动会话(DAS)和数据资源(DR)是多对一的关系,数据活动会话(DAS)和数据集(DS)一对多的

(下转第 100 页)

维普资讯 http://www.cqvip.com

2）针对用户要求和系统必需的功能，度量小组根据每个功能的实现难易、重要程度等因素，为各个功能赋予不同的权值；

3）利用系统运行情况、用户满意度等参数，由专门的测试人员对集成系统的应用状况进行黑盒式的综合测评，并根据测评结果，为系统的各个功能给出一个相应的分数；

4）用给出的测评分数乘上第 2 步给定的功能权值，就得到整个系统的等级，并可以查到其对应的若干改进建议。

通过集成度量，HMM 不但用具体的分数、翔实的数据给每个具体的集成系统分出了等级，还给出了完整、有针对性的改进建议，以指导企业完善自身系统。

结束语 在对电子商务-ERP 集成系统分层剖析的基础上，本文引入了一种层次度量模型 HMM。HMM 的前一阶段是对集成系统的各个层次分别进行结构度量——先利用软件能力成熟度模型等技术设计了针对 ERP 核心层的 ERP 满意度模型，再评估中间传接层的集成程度，最后对电子商务外层进行面向客户的性能测评；HMM 的后一阶段是对整个系统进行功能度量，分出具体系统的优劣，并给出相应的改进建议。

参 考 文 献

1 Gattiker T F, Goodhue D L. Understanding the Local - level Costs and Benefits of ERP through Organizational Information Processing Theory. Information & Management, 2004,41: 431～443

2 陈启申. ERP — 从内部集成起步. 北京:电子工业出版社,2004

3 王东迪. ERP 原理、应用、实践[M]. 北京：人民邮电出版社，2004

4 托马斯·F·华莱士著. 陈德民译. 企业资源规划成功指南[M]. 上海：上海交通大学出版社，2003

5 卢向华,黄丽华. 信息系统的项目投资决策评估过程研究[J]. 计算机集成制造系统,2004，10(6)：651～655

6 Wei C C. An AHP-based approach to ERP system selection [J]. International Journal of Production Economics, 2005, 96: 47～62

（上接第 77 页）

关系. 之间的关系如图 4 所示。

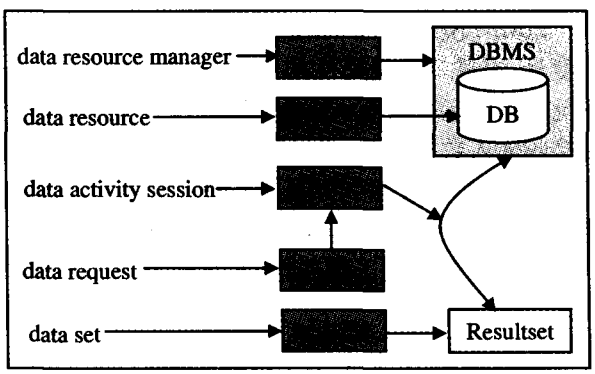

图 3 逻辑资源结构模型

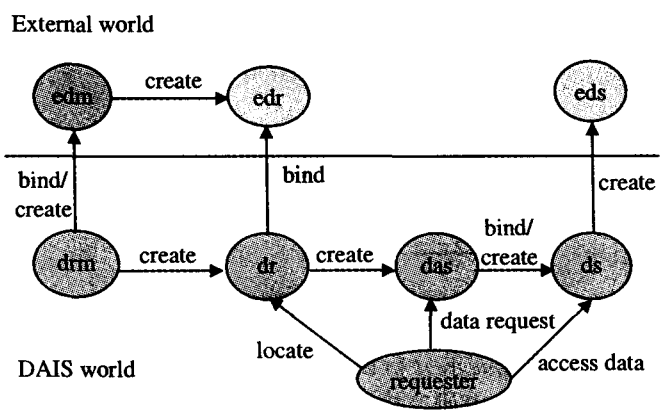

图 4 网格数据库概念模型各个组成部分的关系图

总结 本文首先对网格技术的发展作了介绍，并对网格技术如何用于数据库领域的当前的研究现状进行了深入的分析，然后对网格数据库服务六个方面的需求进行了确定，并提出了一种新的网格数据库服务的概念模型，从外部数据模型和逻辑数据模型两个方面对这种新的模型进行了描述和界定，这种模型能够很好地屏蔽数据库的异构性，并且有标准的接口，易于扩充，但这种模型只是一个框架，距离实际的应用需求还有一段距离，比如元数据标准的制定、历史元数据提取、数据发布、查询处理等标准都要进一步的实现。

参 考 文 献

1 王珊,张坤龙. 网格环境下的数据库系统[J]. 计算机应用,2004,24(10):1～3

2 孟小峰,周龙骧,王珊. 数据库技术发展趋势[J]. 软件学报,2004,15(12):1822～1836

3 Hristidis V, Gravano L, Papakonstantinou Y. Efficient IR-Style keyword search over relational databases. In: Freytag JC, Lockemann PC, Abiteboul S, Carey MJ, Selinger PG, Heuer A, eds. Proc. of the 29th Int'l Conf. on Very Large Data Bases (VLDB). Berlin: Morgan Kaufmann, 2003. 850～861

4 Foster I, Kesselman C. The physilolgy of the grid-An open grid services architecture for distributed systems integration [EB/OL]. http :// www. globus. org/research/papers/ogsa. pdf , 2002202211

5 Joseph J. 开发人员对于 OGSI 和基于 OGSI 的网格计算的概述. http://www-128. ibm. com/developerworks/cn/grid/gr-ogsi/, 2003,5

6 Watson P. Database and the Grid[R]:[Technical Report CS-TR-755]. University of Newcastle ,2001