



# 网格环境下信息组织与管理的探讨

孙洁丽<sup>1,2,3</sup> 孙青云<sup>4</sup>

- (1. 中国科学院文献情报中心, 北京 100080; 2. 中国科学院研究生院, 北京 100080;  
3. 河北经贸大学信息技术学院, 石家庄 050061; 4. 南京林业大学机电工程学院, 南京 210000)

**【摘要】** 本文介绍了网格体系结构, 阐述了网格环境下实现信息组织与管理的原理和方法, 并重点分析了 OGSA-DAI 项目的框架模型, 最后介绍了基于网格的查询处理项目 OGSA-DQP 的功能和实现。

**【关键词】** 网格; 开放网格服务体系结构; OGSA-DAI; OGSA-DQP

**【Abstract】** This paper introduces OGSA technology, and analyzes its theory and method of information organization and management based on grid. Then it focuses on discussing the model framework of OGSA-DAI. At last the function and implementation method of OGSA-DQP and its effect to digital library are pointed out.

**【Key words】** Grid; OGSA; OGSA-DAI; OGSA-DQP

**【中图分类号】** G250.76 **【文献标识码】** A **【文章编号】** 1008-0821 (2007) 07-0051-04

网格专家 Forster 2002 年把网格计算描述为在多个虚拟社团机构间实现动态资源共享和协同解决问题的过程。网格为这种合作提供了各种控制策略与手段, 可以根据需要, 动态地与不同的组织或个人建立各种级别的工作关系。随着大量网格项目的展开和网格技术的发展, 可以将网格定义成一种系统, 该系统通过标准、开放的通用协议和接口来协调分布式的资源以提供最好的服务质量。

## 1 网格及其服务体系结构

应用网格技术的目的在于对位置分布、异构(使用不同硬件支撑、不同网络通信技术)和动态变化的虚拟机构的资源和服务可视化 and 进行集成与管理, 使一旦需要就能够对这些计算机、应用服务、数据和其他资源能够进行访问。要实现这一目标的关键是实现标准化, 从而将地理位置分散、属于不同机构的资源和服务当作单个虚拟系统进行管理, 去发现、访问、调度和监控。

网格资源管理包括资源发现、资源目录、故障隔离、资源供应、资源监控、各种自主能力以及服务水平管理活动。其中最重要的是根据服务水平的需求从网格资源池中挑选恰当的资源, 然后高效提供给用户以满足其要求。

按照采取的网络对象数据模型的不同, 网格的信息服务模式可分为四种, 四种模型在表示网格实体的关系上各不相同:

基于层次模型的网格信息服务, 基于层次模型的网格信息服务的代表性成果为 Globus 项目组设计和开发的 MDS (Metacomputing Directory Services/Monitoring and Discovering Services);

基于对象模型的网格信息服务, GGF (Global Grid Forum) 组织提出的网格对象描述 GOS (Grid Object Specification) 是网格信息服务的数据定义语言, 目前有待进一步的

研究和发展的;

基于关系模型的网格信息服务, 关系模型应用于网格环境的缺点是分布式管理能力较弱, 现在还没有重要的基于关系模型的网格信息服务软件包出现;

基于开放式网格服务体系结构 (Open Grid Services Architecture, OGSA) 的网格信息服务, 基于 OGSA 的网格信息服务系统继承了基于层次模型的网格信息服务系统和基于关系模型的网格信息服务系统的优点, 是理想的网格信息服务解决方案。

OGSA 是在 Globus 网格计算工具包和 web 服务技术融合的基础上提出的一套规范和标准, OGSA 包括两大关键技术, 即网格技术和 Web Service 技术。

OGSA 在原来 Web Service 服务概念的基础上, 提出了“网格服务 (Grid Service)”的概念, 用于解决服务发现、服务创建和服务生命周期管理等问题。

网格服务是一种 Web Service, 该服务提供了一组接口, 这些接口的定义明确并且遵守特定的惯例, 解决服务发现、动态服务创建、生命周期管理、通知等问题。在 OGSA 中, 将一切都看作是网格服务, 因此网格就是可扩展的网格服务的集合。网格服务可以用不同的方式汇集起来满足虚拟组织的需要, 虚拟组织自身也可以部分地根据它们操作和共享的服务来定义。

由于网格环境中所有的组件都是虚拟化, 因此, 通过提供一组相对统一的核心接口, 所有的网格服务都基于这些接口实现, 就可以很容易地构造出具有层次结构的、更高级别的服务, 这些服务可以跨越不同的抽象层次, 以一种统一的方式来看待。

虚拟化也使得将多个逻辑资源实例映射到相同的物理资源上成为可能, 在对服务进行组合时不必考虑具体的实

收稿日期: 2007-03-27

作者简介: 孙洁丽 (1969—), 女, 北京航空航天大学计算机学院获硕士学位, 中国科学院文献情报中心博士生, 河北经贸大学信息技术学院副教授, 研究方向: 网格计算、信息组织和数字图书馆。

孙青云 (1971—), 女, 南京林业大学获硕士学位, 南京林业大学机电工程学院讲师, 研究方向: 计算机图形学和信息管理。

现,可以以底层资源组成为基础,在虚拟组织中进行资源管理。通过网格服务的虚拟化,可以将通用的服务语义和行为,无缝地映射到本地平台的基础设施上。

开放网格体系结构的意义,就在于它将网格从以科学与工程计算为中心的学术研究领域,扩展到更广泛的以分布式系统服务集成为主要特征的社会经济活动领域。

## 2 网格环境下的信息组织与管理

### 2.1 信息组织概述

信息组织是将处于无序状态的特定信息,根据一定的原则和方法,使其成为有序状态的过程,其目的是将无序信息变为有序信息,方便人们利用信息和有效地传递信息。信息组织是一个序化的过程,其目的在于方便人们对信息的利用。

网格技术关注大规模的资源共享,强调多机构之间的资源共享和合作使用,在解决异构平台兼容、集成已有系统方面有着独特的优势,提供了资源共享的基本方法。

### 2.2 网格环境下信息组织与管理

因为网格的主要目标是支持在共享资源上的协同工作,所以数据库集成是目前网格环境下数据管理的研究热点。网格环境中,数据资源是异构的,数据可能存储在不同的数据载体中,如各种关系型数据库,各种文件系统。为了在这些异构的数据资源之上实现资源共享和协同使用,网格中间件必须首先同构化这些异构的数据资源,即异构数据资源的统一访问技术是实现数据资源共享和调度的首要前提。通过统一访问技术,将存储、检索数据高层用户的请求映射为数据资源的底层操作,实现广域范围内对数据有效的统一访问和管理。

目前有三种网格数据库集成策略:

(1) 虚拟数据库。虚拟数据库只有一个数据库模式,所有的用户不会觉察到多个独立的数据库存在的事实。虚拟数据库在概念上是受欢迎的,但是难以实现。在构造虚拟数据库时,需要考虑以下几种透明性:异构透明性、命名透明性、属主和费用透明性、并行透明性和分布透明性。

(2) 定制集成。这种方式是指由应用程序自己完成数据库集成。例如在有的科学应用程序中,开发人员自己找到相关的数据源,然后将集成任务划分为查询、要执行的程序、中间数据源的构建、显式的数据传输和数据变换,以及存储结果等等。网格数据库管理系统应该提供对这种集成方式的支持,使之降低成本、减少时间耗费和错误的发生。

(3) 增量集成。虚拟数据库是一个理想目标,定制集成又过于注重细节,增量集成居于二者之间。在增量集成中,开发人员无须完成集成的每一个细节,高级的数据存取和集成组件(如 OGSA - DAI)可以自动完成一些后期的集成步骤。

GGF 主要目标是建立网格的技术标准。它由 7 个领域组成,每一个领域又分为若干个工作组和项目组。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。OGSA 对数据的访问和集成只提供了低层次

的集成,网格应用更需要一种高层次的数据访问和集成方法。DAIS 工作组正在制定网格数据库服务的标准。DAIS 工作组制定的网格数据库服务标准草案的一个参考实现是 OGSA - DAI (Open Grid Services Architecture - Data Access and Integration) 项目。

英国 e - science 中心开发的项目 OGSA - DAI 是数据库网格研究领域的代表作。OGSA - DAI 是一种数据访问和集成的中间件平台,目标是构造一个中间件,这个中间件将有助于在网格环境中对数据的存取和集成,它主要实现了对多种数据库的服务化访问,使得数据库能够用网格服务的形式对外共享数据。

OGSA - DAI 是在 Globus Toolkit3 (GT3) 的基础上开发的专门用来访问数据库数据的中间件。GT3 是 OGSA 的一种参考实现,它重点解决了网格中的安全性、信息基础设施、资源管理、通信、故障检查等问题,为网格应用程序提供了一个良好的运行环境。OGSA - DAI 项目试图通过 Grid Services 实现对不同的数据库系统的统一访问。OGSA - DAI 对复杂数据的操作进行了封装,提供了一种统一的网格服务接口,使网格用户或服务在网格环境中通过服务接口,可方便地访问和集成网格中的各种分散的异构数据源。

OGSA - DAI 通过服务的方式获取数据,客户端通过这些服务间接同数据源交互,它可提供三个与 Web Service 和 Grid Service 方案兼容的服务,即 WS - I (Web Service Inter - operability), WS - RF (Web Service Resource Framework), OGSi (Open Grid Services Infrastructure), 其框架模型见图 1。

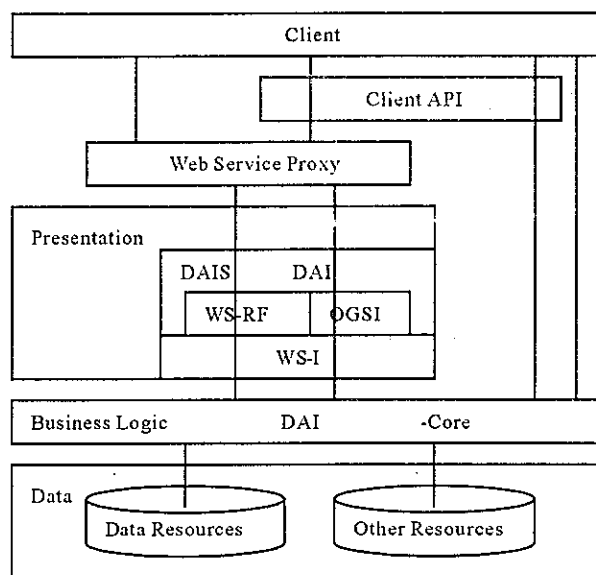


图 1 OGSA - DAI 框架模型图

(1) 数据层: OGSA - DAI 可以访问各种数据源,包括关系数据库(如: MySQL, SQL Server, Oracle 等)、XMI 数据库(如: Xindice)、文件(如: BinX 文件)。

(2) 业务逻辑层: 该层封装了 OGSA - DAI 的核心功能,可以接收客户端的请求(如查询、修改等),响应客户端的请求,管理数据传输和发布,以及与数据源的连接、管理、交互。

(3) 表示层: 该层封装了 OGSA - DAI 向网格展示的所



有服务,包括 DAISGR (Data Access Interface Service Group Register, 数据访问接口服务组注册), GDSF (Grid Data Service Factory, 网格数据服务工厂), GDS (Grid Data Service, 网格数据服务), 这些服务通过 Web Service 或 Grid Service 接口实现,采用 WSDL (Web Service Description Language, Web 服务描述语言) 和 XML Schema 来描述相应接口。

(4) 客户端: OGSA - DAI 支持与 OCSI, WS - RF, WS - I 兼容的客户端的访问,这主要依赖于服务器端表示层中提供的服务,通过客户端工具可以实现对底层数据源的检索、更新等操作。

该框架中还提供数据层与业务逻辑层接口,可以调用 JDBC 驱动器,XMLDB 驱动器等;表示层与业务逻辑层的接口负责在两层间的通信,支持对 OGSA - DAI 核心功能的调用。

目前使用 OGSA - DAI 的项目有: AstroGrid、Biogrid、BioSimGrid、Bridges、FirstDIG、GeneGrid、ODD - Genes、OGSA - WebDB 等。由于 OGSA - DAI 项目有着雄厚的资金支持和技术积累,随着版本的不断升级,其功能会更强大,包括更好的性能、支持更多的 DBMS、支持更多的 SQL 访问、DBMS 管理操作、文件访问、客户端类库等。

### 2.3 基于网格的数字图书馆信息查询

如果基于网格的数字图书馆在多个网格结点建立了数据库,则应用就需要做分布式查询。网格环境下的分布式查询处理和其他环境下的分布式查询处理的主要区别在于:

(1) 网格不仅提供了一套系统化的方法来安全地利用远程的数据资源,它还使得远程的计算资源同样可以安全使用;

(2) 网格提供动态资源发现、分配和监控机制;

(3) 网格提供监控网络连接状态的机制,这一点对于在广域网环境下执行查询很重要;

(4) 网格遵循 (正在开发的) 标准,在网络上各种形式的数据库可以以统一的形式存取。

OGSA - DAI 主要解决了在网格环境中对各种异构数据源的动态访问和数据在各网格服务间的传输问题。在数字图书馆应用中,需要一种高层次的数据集成方法。目前基于网格技术的分布式查询项目有 polar \* 和 OGSA - DQP (Open Grid Services Architecture - Distributed Query Processing) 等项目。

Polar \* 是英国 Manchester 和 Newcastle 大学的研究人员开发的用于网格环境下的分布式查询系统,Polar \* 项目研究网格环境下分布式查询处理涉及的问题,根据网格本身的特点实现高性能的分布式查询处理。

Polar \* 项目的研究者认为,网格环境下分布式查询处理最大的特点是它的自适应性,它必须能充分利用网格强大的并行计算能力和适应网格资源动态演化的特性。

Polar \* 设计了一种描述型查询语言,实现了对分布异构数据库的并行查询功能,并研究查询如何动态响应资源的改变。

Polar \* 项目已经并将完成以下工作:

(1) 设计一种描述型查询语言,用户使用它来表达数

据存取和数据分析任务;

(2) 考虑可用的网格资源、用户需求和代价约束,研究这种查询语言的查询优化方法和代价模型;

(3) 研究查询优化器生成的查询计划并行执行的方式;

(4) 研究长时间运行的任务如何利用在运行中变为可用的额外资源来提高性能;

(5) 研究查询如何动态响应资源的改变,资源的改变包括出现新的可用资源、分配的资源被回收了、资源失效和共享资源上负载的改变等等。

研究网格环境下分布式查询处理的还有 OGSA - DQP 项目,它建立在 OGSA - DAI 项目的基础之上。OGSA - DQP 是在 OGSA - DAI 的基础上开发的面向用户的基于网格服务的分布式查询系统,它采用并行数据库技术,对复杂的数据密集型请求,实现了具有并行功能的分布式查询。

OGSA - DQP 符合基于 OGSA 的网格标准,是建立在 GGF 的数据库访问和继承服务标准的参考执行 OGSA - DAI 上的,其底层组件采用了 Polar \* 的并行优化机制。OGSA - DQP 使用了 OGSA/OCSI 的参考实现即 Globus Toolkit 3, GT3 通过把资源虚拟化为网格服务 (Grid Services) 来实现了一种基于服务的架构。

OGSA - DQP 本质上是一个高吞吐量的分布式数据流引擎,它依赖于对网格资源的面向服务的提取并假定数据资源可以通过基于服务的接口进行访问。OGSA - DQP 不仅可以对网格数据服务 (GDSs) 查询,而且还可以调用网格中的 Web 服务,因此它可以把数据访问与数据分析功能结合在一起,具有很高的使用价值。

OGSA - DQP 设计了一个框架:支持通过网格数据库服务 (GDSs) 和其他网络上可用的 Web 服务来支持公开的查询,并因此可以分析合成数据访问;借鉴并行数据库的技术来为复杂的密集型数据请求提供潜在的并行机制,这样查询的各个部分就可以在不同的主机上高效地并行执行;通过查询优化来对复杂、繁重、专门的配置和资源利用实现自动化决策;使用 GDSs 已形成的标准来提供对数据库元数据的一致性访问和与网格数据库的交互;使用 OGSA 的工具来动态获得必要的资源,以达到有效评估分布式查询的目的。

OGSA - DQP 提供了两个服务来实现它的功能:网格分布式查询服务 (GDQS) 和网格查询评估服务 (GQES)。随着实际应用中这种中间件需求的增加,OGSA - DQP 必将向着功能更加强大、使用更加方便的方向发展。

### 3 结束语

以 e - science、e - learning、e - business 和 e - government 为代表的信息环境正带来新的用户需求、用户行为和用户信息应用机制。当前数字图书馆以用户为核心来聚合信息资源、信息服务、信息利用活动,其最终目标就是着力于支持用户利用信息、提炼知识、解决问题,成为用户工作、学习环境的有机结合。

目前,网格服务技术尚不成熟,还在不断的发展中,许多新技术、新规范正在不断地被采用,如数据流语言等。但网格服务技术可以解决不同系统间数据复制的异构性、



互操作性、分布式查询等问题,是数字图书馆发展的方向。国外虽然已经有相关项目的开展,但具体标准正在制定,并且项目实用性方面还有待进一步加强,所以我们国内的相关研究应尽快展开,或者能够取得有价值的成果。

参 考 文 献

[1] Ian Foster, Carl Kesselman. The Grid: Blueprint for a New Computing Infrastructure, second edition [M]. Morgan Kaufmann Publishers, San Francisco, CA, USA, 2003.

[2] Francine Berman, Geoffrey C.Fox, Tony Hey. Grid Computing: Making the Global Infrastructure a Reality [M]. Wiley, New York, 2003.

[3] Globus [EB/OL]. <http://www.globus.org>

[4] GGF [EB/OL]. <http://www.ggf.org>

[5] DAIS [EB/OL]. [http://www.gridforum.org/6\\_DATA/dais.htm](http://www.gridforum.org/6_DATA/dais.htm)

[6] Ingo Frommholz, Predrag, et al. Supporting Information Access in Next Generation Digital Library Architectures [A]. Pre-proceedings of the Sixth Thematic Workshop of the EU Network of Excellence DELOS [C]. 2004.

[7] Agosti, M., Ferro, N., Frommholz, I., Thiel, U. Annotations in digital libraries and collaboratories facets, models and usage [A]. In: Proc. 8th European Conference on Research and Advanced Technology for Digital Libraries (ECDL) [C]. 2004.

[8] Neuhold, E.J., Nieder'ee, C., Stewart, A. Personalization in digital libraries: An extended view [A]. In: Proceedings of ICADL 2003 [C]. 2003.

[9] 戴维民. 信息组织 [M]. 北京: 高等教育出版社, 2004.

[10] 周宁. 信息组织 (第二版) [M]. 武汉: 武汉大学出版社, 2006.

[11] 桂小林. 网格技术导论 [M]. 北京: 北京邮电大学出版社, 2005.

[12] (美) Joshy Joseph, Craig Fellenstein. 网格计算 [M]. 战晓苏, 张少华, 译. 北京: 清华大学出版社, 2005.

[13] 徐志伟, 冯百明, 李伟. 网格计算技术 [M]. 北京: 电子工业出版社, 2004.

[14] OGSADAI [EB/OL]. <http://www.ogsadai.org.uk/>

[15] Polar\* [EB/OL]. <http://www.ncl.ac.uk/polar/>

[16] M Nedim Alpdemir, Arijit Mukherjee, Norman W Paton et al. OGSADAI - DQP: A Service for Distributed Querying on the Grid [A]. In: EDBT2004 [C]. 2004.

[17] M.Nedim Alpdemir, Arijit Mukherjee, Anastasios Gounaris, Alvaro A.A.Fernandes, Norman W.Paton, Paul Watson. An Experience Report on Designing and Building OGSADAI - DQP: A Service Based Distributed Query Processor for the Grid [EB/OL]. <http://www.ogsadai.org.uk/documentation/publications/alpdemir.pdf>

[18] 张晓林. 从数字图书馆到 E-Knowledge 机制 [J]. 中国图书馆学报, 2005, 31 (4): 5-10.

[19] 张晓林. 构建数字化知识化的信息服务模式 [J]. 津图学刊, 2003, (6): 13-16.

[20] 李广建. 数字图书馆技术的发展趋势 [J]. 图书情报工作, 2005, 49 (1): 6-7.

(上接第 50 页)

导。同时,行业协会信息功能的发挥,必须要有一支稳定的、高素质的科技信息人才队伍为基础,尤其应重视基层企业科技信息人才队伍的培养。

4.5 加强对企业领导信息意识的培养和行业协会信息功能的宣传

一些企业领导信息意识淡薄,是制约行业协会信息功能发挥的主要因素。实践证明,只要企业领导重视信息工作,企业的科技进步和管理创新就会因此受益,从而实现良性循环。同时,行业协会的领导和有关部门应大力宣传信息受益企业的案例和行业协会的信息功能。

4.6 重视对企业的个性化信息服务

针对不同企业在生产、销售、管理、技术开发、对外投资等方面对信息的需求,提供特定信息服务,满足企业的个性化信息需要,个性化信息服务是行业协会信息功能最具有发展潜力的信息服务方式。

参 考 文 献

[1] 刘凤军, 庞晓鹏. 美日德行业协会的职能、网络与自身建设 [J]. 首都经济贸易大学学报, 2004, (1): 29-33.

[2] 石宏英. 日本行业协会对企业的信息情报功能与经济功能 [J]. 日本问题研究, 2000, (1): 69-70.

[3] 张新文, 谢煥文. 西方发达国家行业协会的角色功能及运行机制 [J]. 广西民族学院学报: 哲学社会科学版, 2004, (9): 176-180.

[4] 谢媛. 从国外经验看我国行业协会发展模式 [J]. 理论导刊, 2004, (1): 29-30.

[5] 中国机械工业联合会. 发挥协会作用, 振兴机械工业 [EB]. <http://www.ica.gov.cn/news2006/zxdt/2005huibian/15jixigongye.htm>, 2007-02-26.

[6] 中国商业联合会. 提供服务, 为政府与企业构架信息桥梁; 加强交流, 为国内外企业搭建合作平台 [EB]. <http://www.cac.gov.cn/news2006/zxdt/2005huibian/8shangyietianhe.htm>, 2007-01-02.

[7] 中国印刷及设备器材工业协会. 中国印刷及设备器材工业协会情况 [EB]. <http://www.ica.gov.cn/news2006/zxdt/2005huibian/17yinshuashebei.htm>, 2007-02-23.

[8] 李丕仕. 基于行业协会的专业数字图书馆建设模式 [J]. 图书馆建设, 2004, (6): 62-64.

[9] 李丕仕. 基于行业协会的信息共建共享模式研究 [J]. 情报杂志, 2005, (12): 145-146.

[10] 2005年行业协会工作交流材料汇编 [EB]. <http://www.ica.gov.cn/news2006/zxdt/2005huibian/2005huibian.htm>, 2006-12-25.

信 息 化 与 网 络 建 设