ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.11, November 2006, pp.2224−2233
DOI: 10.1360/jos172224
© 2006 by *Journal of Software*. All rights reserved.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

网格环境中模式复用的异构数据库访问和集成方法*

陈小武+, 潘章晟, 赵沁平

(北京航空航天大学 计算机学院,北京 100083)

A Schema-Reusable Method on Heterogenous Databases Access and Integration in Grid Environment

CHEN Xiao-Wu+, PAN Zhang-Sheng, ZHAO Qin-Ping

(School of Computer Science and Engineering, BeiHang University, Beijing 100083, China)

+ Corresponding author: Phn: +86-10-82317610, Fax: +86-10-82317610, E-mail: chen@buaa.edu.cn, http://www.buaa.edu.cn

Chen XW, Pan ZS, Zhao QP. A schema-reusable method on heterogenous databases access and integration in grid environment. *Journal of Software*, 2006,17(11):2224−2233. http://www.jos.org.cn/1000-9825/17/2224.htm

Abstract: Recently the research about heterogeneous database access and integration is booming in grid technology and related applications. UDMGrid (university digital museum grid) is devoted to the integration and sharing of resources from 12 university digital museums with grid technology, and it is necessary for UDMGrid to investigate the mechanism and method for heterogeneous specimen-database access and integration. In this paper, an effective schema-reusable method on database access and integration in UDMGrid (UDMGrid-DAI) is presented. It includes a schema-reusable virtual database to define the global uniform or heterogeneous database resources, and to describe the concepts and their relations of one or one sort of scientific domain through virtual tables and virtual columns. It also includes the ways of database registering and query mapping to the virtual databases based on the above virtual database mechanism. Furthermore, This UDMGrid-DAI provides a consistent interface to the virtual databases through the registry processing and the virtual database services. Finally, the experimental results demonstrate the work of this paper, and show that this schema-reusable mechanism of virtual database is correct and useful on database access and integration, especially in UDMGrid.

Key words: grid; information grid; database access and integration; virtual database; schema-reusable

摘 要: 如何为网格环境的资源使用者提供异构数据库资源的全局统一视图和一致访问接口,是网格技术及其应用的关键问题和研究热点.由于大学数字博物馆网格(university digital museum grid,简称 UDMGrid)需要整合和共享分布在 8 个城市 12 个大学数字博物馆的异构数据库资源,因此提出了一种网格环境下模式复用的异构数据库访问和集成方法(database access and integration in UDMGrid,简称 UDMGrid-DAI).该方法首先提出了一种模式复用的虚拟数据库构建策略,定义了异构数据库资源的全局统一视图;在此基础上给出了数据库资源

* Supported by the National Natural Science Foundation of China under Grant No.60503066 (国家自然科学基金); the China Education and Research Grid Project; the China Next Generation Internet (CNGI) under Grant No.CNGI-04-15-7A (中国下一代互联网示范项目); the New Centaury Excellent Talent Foundation from the MOE of China under Grant No.NCET-05-0187 (国家教育部新世纪优秀人才支持计划)

Received 2006-06-13; Accepted 2006-08-25

© 中国科学院软件研究所 http://www.jos.org.cn

注册方式和虚拟数据库查询映射方法,为资源使用者提供了访问数据库的一致访问接口;最后,在 UDMGrid 的异构数据库资源环境下,实验了从数据库资源注册到使用的全过程,验证了 UDMGrid-DAI 方法的正确性和实用性.

关键词: 网格;信息网格;异构数据库访问和集成;虚拟数据库;模式复用
中图法分类号: TP311 文献标识码: A

根据现代远程教育公共资源中考古、人文、美术、地质、农业、生物、医药、航空航天等领域的 12 个大学数字博物馆,中国教育科研网格计划(China education and research grid,简称 ChinaGrid)的大学数字博物馆网格(university digital museum grid,简称 UDMGrid)将分散在中国 8 个城市的海量数字博物馆资源整合起来,实现多种资源的互连互通与有机共享,消除标本信息“孤岛”现象,并以数字、文字、图像、视频、音频、动画、三维模型等形式,通过网格门户为各种不同类型的用户提供“一站式”标本信息类型的网格服务[1−5].由于这些海量数字博物馆标本信息属于多种不同的学科或领域,各个大学数字博物馆的标本信息主要存储在各自的数据库中,例如 SQL Server,Oracle,DB2 等,所以,如何确定网格环境中异构数据库的访问和集成方法(database access and integration,简称 DAI),为用户提供异构数据库资源的全局统一视图和一致访问接口,是 UDMGrid 面临的关键问题[6].

广义的异构数据库集成可将分散、异构的数据以统一、一致的形式表现出来,为用户屏蔽各种数据在来源和结构等方面的差异,它主要包括异构数据库的访问(access)与狭义的集成(integration).其中:异构数据库访问主要针对不同数据库之间的互连互通、数据交换与数据共享等问题;狭义的集成则侧重于屏蔽不同数据库系统在语义方面的异构性,研究如何向用户提供统一的模式和访问接口等问题.在异构数据库集成的数据仓库方法中,SQL Server 提供的数据导入、导出和数据传输服务,可以将不同数据库系统的数据复制到相同的数据库中,消除平台和系统层的异构性,但不能消除语义层的异构性.异构数据库集成的虚拟集中方法涉及有全局统一模式的多数据库系统和无全局统一模式的联邦式数据库系统.全局统一模式一般采用属性等价的方法,并结合 XML Schema 和本体(ontology)技术,在异构数据库集成方面具有较好的效果.联邦式数据库系统没有全局统一模式,但支持异构数据库之间的松散耦合关系,数据库可以通过输入、输出模式进行数据访问[7].

随着网格技术及其应用的不断发展,网格环境中异构数据库的访问和集成已经成为网格研究的重要组成部分[8,9].全球网格论坛(global grid forum,简称 GGF)的数据库访问和集成服务工作组(database access and integration services working group,简称 DAIS-WG)正致力于网格数据库服务的标准建设,其中包括异构数据库系统的一致访问规范、网格框架中数据库系统的协同机制等方面.目前,该工作组已经提出了网格环境下数据库服务接口的部分规范,其中涉及关系数据库、XML 数据库和索引文件等[10].英国 e-Science 网格计划的开放网格服务体系框架数据访问和集成(open grid services architecture data access and integration,简称 OGSA-DAI)是支持数据库访问和集成的网格中间件,能够支持多种数据库资源的查询、更新、传输和交付,为数据库资源生成配置文件(例如,数据库的类型、JDBC 驱动名、URL、用户名和口令等),并根据配置文件自动部署 GDSF 服务,为用户提供数据库的查询和访问服务[11−13].但是,OGSA-DAI 目前主要侧重于数据库访问服务,其数据库资源注册不包括其他数据库模式信息.中国科学院计算机网络信息中心提出了一种基于网格服务的数据库元数据管理框架,在网格服务(grid service)的基础上,使用轻量级目录访问协议(lightweight directory access protocol,简称 LDAP)存储各数据库的元数据,使用 JDBC(Java database connectivity)实现异构环境下数据库的统一访问接口,以网格服务方式提供元数据的访问功能[14].

上述异构数据库访问与集成方面的研究工作对大学数字博物馆网格建设有一定的启发和借鉴作用,但也存在一些需要进一步探索的问题:(1) 由于 UDMGrid 的异构数据库资源来自各个大学数字博物馆,数据复制方法可能会产生数字标本的版权保护等问题,也可能产生网格存储资源的浪费;(2) 由于 UDMGrid 的数据库资源涉及许多不同的学科和领域,基于数据库元数据方法的网格应用经常要求数据库资源的专业知识;(3) DAIS-WG 和 OGSA-DAI 的研究工作侧重于多种数据库的访问,难以支持标本信息类型相差较大的异构数

© 中国科学院软件研究所 http://www.jos.org.cn

据库集成工作,并且难以描述 UDMGrid 数据库资源的全局模式.

所以,在大学数字博物馆网格的研究与建设过程中,本文提出了一种网格环境下模式复用的异构数据库访问和集成方法(database access and integration in UDMGrid,简称 UDMGrid-DAI).该方法首先提出了一种模式复用的虚拟数据库构建策略,定义了异构数据库资源的全局统一视图;在此基础上,给出了数据库资源注册方式和虚拟数据库查询映射方法,为资源使用者提供了访问数据库的一致访问接口.如图 1 所示,UDMGrid-DAI 的构建过程是:根据标本信息的学科或领域类型[15]、信息资源的各种本体[16],利用模式复用的虚拟数据库描述各种数据库资源,通过虚拟数据库模式提供数据库资源的全局统一视图;采用资源注册的方式,资源提供者确定其数据库模式与虚拟数据库模式之间的映射关系(图中虚线箭头表示资源注册过程);通过查询映射方法,将基于虚拟数据库的查询请求转换为基于真实数据库资源的查询请求,并利用虚拟数据库服务提供访问数据库的一致接口.最后,在 UDMGrid 的异构数据库资源环境下,实验了从数据库资源注册到使用的全过程.

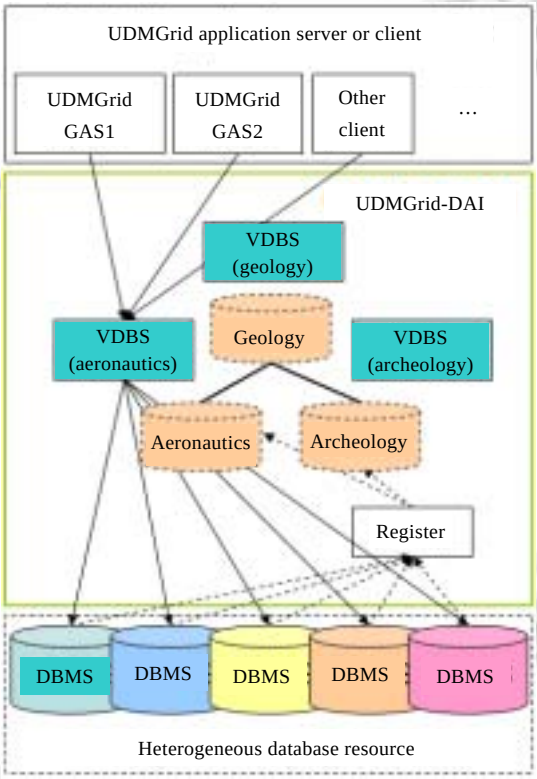

Fig.1 Constructing process of UDMGrid-DAI

图 1 UDMGrid-DAI 构建过程

1 模式复用的虚拟数据库构建策略

UDMGrid 的标本信息来源于 12 个不同的大学数字博物馆,隶属于多种不同的学科或领域,且主要存储于各个大学数字博物馆的数据库中.为了将这些异构数据库整合成按学科领域划分的虚拟数据库,同时避免虚拟数据库之间的相对孤立性,本文提出了一种模式复用的虚拟数据库构建策略.

其中的模式复用,主要是在虚拟数据库的构建过程中,依据大学数字博物馆隶属学科或领域之间的内在联系[15],描述网格信息资源的全局本体与局部本体[16],借鉴面向对象设计方法的继承特性,利用树状结构表现虚拟数据库之间的继承关系,从而减少虚拟数据库构建的重复性工作,并保留虚拟数据库之间的内在联系,也体现了学科或领域之间的内在联系.如图 2 所示,在 UDMGrid 的虚拟数据库树林中,“人文与社会学科”可以继承“博

© 中国科学院软件研究所 http://www.jos.org.cn

物馆”的虚拟数据库模式,“考古”与“艺术”又可以继承“人文与社会学科”的虚拟数据库模式.这种继承关系,使得子虚拟数据库能够复用父虚拟数据库的模式.

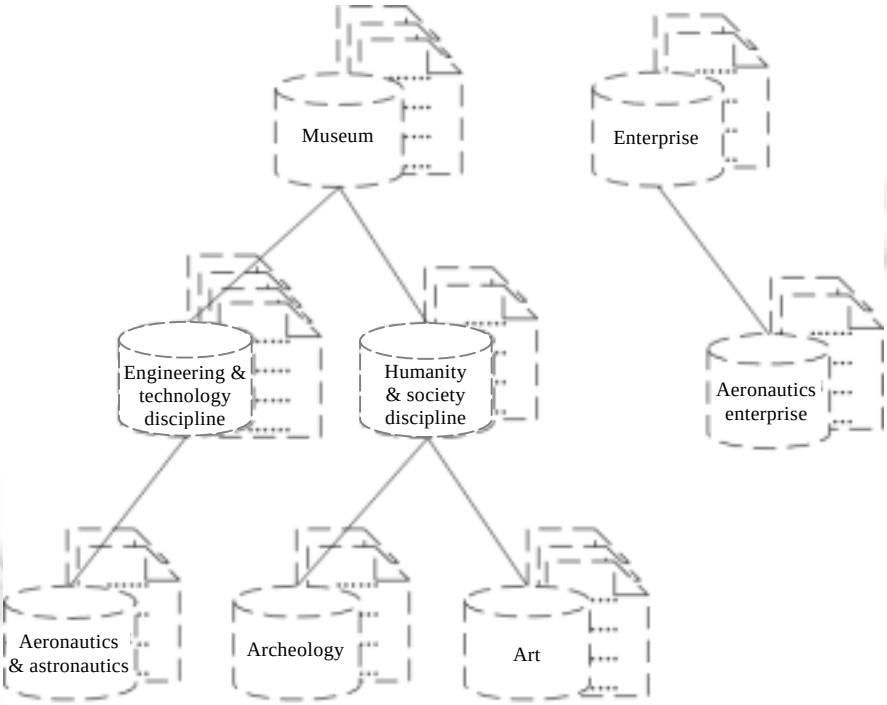

Fig.2 Forest structure of virtual databases

图 2 虚拟数据库的树林结构

一般地,虚拟数据库可以包含一个或多个虚拟表,虚拟表可以包含一个或多个虚拟字段,虚拟数据库、虚拟表和虚拟字段的概念分别类似于关系数据库理论中的数据库(database)、表(table)和字段(column)的概念.具体而言,本文提出的模式复用的虚拟数据库构建策略主要包括以下内容:

(1) 在创建子节点虚拟数据库的虚拟表时,既可以继承父节点虚拟数据库的虚拟表,也可以不继承任何虚拟表而创建自身的新虚拟表,并且这种继承关系具有递推性,可以一直上溯到根节点的虚拟数据库.

(2) 在创建子节点虚拟表的虚拟字段时,如果子节点的虚拟表继承了父节点的虚拟表,则称前者为子虚拟表,后者为父虚拟表.子虚拟表自动具有父虚拟表的所有虚拟字段,其中包括父虚拟表自身创建的虚拟字段和继承祖父的虚拟字段,并且子虚拟表也可以创建其自身的虚拟字段.

(3) 根据标本信息的学科或领域类型,相近或者相同学科的虚拟数据库位于同一棵子树中,相差较远学科的虚拟数据库位于不同的子树中.

通过这种方式构建的模式复用型虚拟数据库,表现了异构数据库资源的全局统一视图.如图 3 所示,以 UDMGrid 中所构建的虚拟数据库模式为例,虚拟数据库“博物馆”拥有一个名为“标本”的虚拟表,“标本”虚拟表有“名称”、“类型”、“时间”、“说明”、“图片路径”、“二进制图片”等虚拟字段;虚拟数据库“工程与技术学科”是“博物馆”的子虚拟数据库;虚拟数据库“航空航天”是“工程与技术学科”的子虚拟数据库,并有两个虚拟表“飞机”和“发动机”,都继承了祖父的虚拟表“标本”.同时,虚拟表“飞机”和“发动机”不但继承了“名称”、“类型”等虚拟字段,而且“飞机”创建了“产地”、“速度”和“用途”等自身虚拟字段,“发动机”创建了“推力”、“转速”等自身虚拟字段.

所以,模式复用的虚拟数据库构建策略支持两种继承关系:其一是虚拟数据库之间的继承关系,这是一种比较“弱”的继承关系,这种继承关系使得子虚拟数据库的虚拟表可以继承父虚拟数据库的所有虚拟表的某个虚

© 中国科学院软件研究所 http://www.jos.org.cn

拟表,如果子虚拟数据库的所有虚拟表都没有继承父虚拟数据库中的虚拟表,则这种继承关系没有产生作用;其二是虚拟表之间的继承关系,这是一种比较“强”的继承关系,这种继承关系会使子虚拟表自动拥有父虚拟表的所有虚拟字段.也正是这种继承关系,使得父虚拟数据库的模式能够被子虚拟数据库所复用,并且这种复用的方式具有较大的灵活度.第 1 种继承关系是第 2 种继承关系的基础,第 2 种继承关系是第 1 种继承关系的体现.

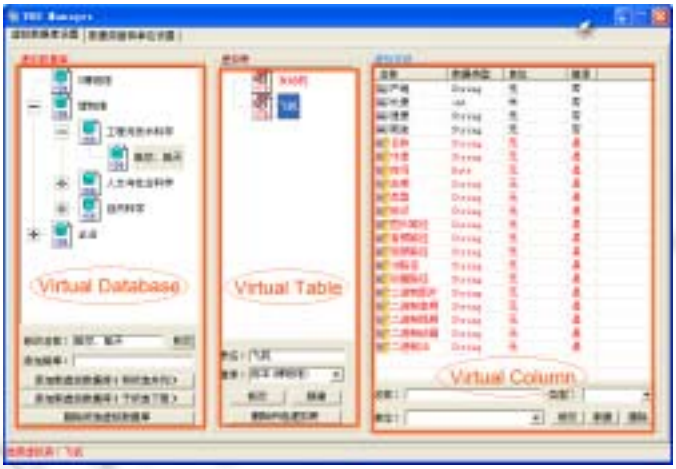

Fig.3 Virtual database manager

图 3 虚拟数据库管理器

该策略具有的继承特性有利于网格环境下异构数据库资源的注册与映射.数据库资源提供者在注册数据库资源时可以选择一个与自身最相近的虚拟数据库,设定其资源与该虚拟数据库之间的映射关系.例如,在注册航空航天数字博物馆数据库资源时,如果选择“航空航天”虚拟数据库,在设定航空航天数据库与子虚拟表“飞机”、“发动机”之间的映射关系后,由于“飞机”和“发动机”继承了“博物馆”的“标本”虚拟表,这使得航空航天数据库与父虚拟表“标本”之间自动建立起同样的映射关系,从而减少了数字博物馆资源注册的重复性工作.所以,网格应用服务在查询虚拟数据库时,无论是查询“博物馆”还是查询“航空航天”的“数据”,都能根据这些映射关系获得来自航空航天数字博物馆后台数据库的真实数据.

2 数据库资源注册

在上述模式复用的虚拟数据库构建策略基础上,数据库资源提供者注册其所持有的数据库资源时,需要设定 3 方面的信息:本地数据库访问参数、本地数据库模式、本地数据库与虚拟数据库之间的映射信息.

(1) 外部网络访问本地数据库所需要的参数.如果本地数据库可直接被外网访问,则这些参数包括数据库的类型、服务器 IP、端口号、数据库名称、只读用户与口令等.如果外网不能直接访问本地数据库,可以使用一个部署在 Web Services 容器中的 Web 服务,作为本地数据库访问的代理服务,提供者只需提供该 Web 服务的地址和访问参数.

(2) 本地数据库表和视图的模式信息.在资源提供者设置上述的访问参数之后,客户端软件能够自动注册本地数据库表和视图的模式信息.

(3) 本地数据库模式和虚拟数据库模式之间的映射关系信息.具体而言,这是指本地数据库中某一个字段到虚拟数据库中某一个虚拟字段之间的映射关系,其映射关系可以定义为:映射关系信息(MappingInformation)是映射规则(MappingRule)集合的子集,一条映射规则包括字段映射(ColumnMapping)子集和表间关系(TableRelation)子集,并且字段映射子集的所有元素必须针对同一虚拟表(VirtualTable).可以表示如下:

$$MappingInformation=\{mr|mr\in MappingRules\}$$

$$MappingRules=\{\langle CMs,TRs\rangle|CMs\subset ColumnMappings,TRs\subset TableRelations,\forall cm1,cm2,cm1\in CMs,$$
$$cm2\in CMs\rightarrow cm1.virtualColumn.virtualTable=cm2.virtualColumn.virtualTable\}$$

© 中国科学院软件研究所 http://www.jos.org.cn

$ColumnMappings=\{\langle column,virtualColumn\rangle | column\in Columns, virtualColumn\in VirtualColumns\}$

$TableRelations=\{\langle (column1,column2),relation\rangle | column1,column2\in Columns, relation\in Relations\}$

$Relations=\{“=”, “<”, “>”, “<>”, “>=”, “<=”, “LIKE”\}$

其中,*MappingRule* 为映射规则的集合,*ColumnMapping* 为字段映射的集合,*TableRelation* 为表间关系的集合,*Columns* 为本地数据库中字段的集合,*VirtualColumns* 为虚拟数据库中虚拟字段的集合,*Relations* 为逻辑关系的集合.

3 虚拟数据库的查询映射

在虚拟数据库的查询映射之前,资源使用者首先要确定与网格应用相关的一个或多个虚拟数据库.由于父虚拟数据库能够自动具有子虚拟数据库中所设定的部分映射关系,所以,网格应用开发者可以选择虚拟数据库树林中的任何一个虚拟数据库节点进行访问.如果资源使用者期望获得比较“专业”的数据,则可以选择虚拟数据库中靠近下层的虚拟数据库进行访问;如果资源使用者期望获得更广泛的数据,则可以选择虚拟数据库中靠近上层的虚拟数据库,这样可以获得来自更多真实数据库中的数据,但也损失了一部分数据的“精度”.例如,如果资源使用者访问“博物馆”虚拟数据库中的“标本”虚拟表,就肯定不能获取到“推力”这一虚拟字段所对应的数据,因为“推力”是“航空航天”虚拟数据库中“发动机”虚拟表所特有的.另一方面,如果资源使用者访问“发动机”虚拟表,则可以顺利获得“推力”所对应的数据,但不能获得“考古”虚拟数据库中的“出土地”虚拟字段对应的数据.所以,在虚拟数据库的选择过程中,“精度”和“广度”是一对矛盾,资源使用者可以根据具体应用的需求特点选择合适的虚拟数据库,从而在“精度”和“广度”之间获得取舍点.模式复用的虚拟数据库构建策略为资源使用者提供了选择虚拟数据库的灵活性,同时也没有增加系统构建者和资源提供者在初始化工作过程中的负担.

根据模式复用的虚拟数据库构建策略,在数据库资源注册与虚拟数据库选择的基础上,需要给出虚拟数据库的查询映射算法.本文的虚拟数据库查询映射算法主要包括:

(1) 虚拟查询语句的转换.根据数据库资源注册确定的映射关系,将资源使用者对虚拟数据库的查询语句(即虚拟查询语句)转换为多个真实数据库的真实查询语句.

(2) 数据库查询结果的合并.将来自各真实数据库的查询响应合并为一个结果集,该结果集的元数据结构能够匹配所访问的虚拟数据库模式.

4 UDMGrid-DAI 的总体结构与应用服务

在大学数字博物馆网格 UDMGrid 的建设环境中,根据本文模式复用的虚拟数据库构建策略以及数据库资源注册、虚拟数据库选择和虚拟数据库查询映射等方式,在中国教育科研网格环境下设计并实现了异构数据库数据库访问和集成方法(database access and integration in UDMGrid,简称 UDMGrid-DAI).如图 4 所示,UDMGrid-DAI 总体结构主要包括 8 个部分:(1) 虚拟数据库管理器(VDB management).在应用系统构建过程中,领域专家或系统构建者用它来维护虚拟数据库的模式(VDB schema).(2) 数据库资源注册服务与客户端(ResInit service,ResInit client).数据库资源提供者在向应用系统注册其数据库资源时,使用该客户端自动注册其持有的数据库模式,并设定其数据库资源与虚拟数据库之间的映射关系.(3) 数据库访问代理(access Agent),为 UDMGrid-DAI 中其他模块(如资源监测服务、查询与响应转换服务等)提供数据库访问服务.该代理可以支持对各数据库资源的并行访问.(4) 虚拟数据库查询服务(VDB SGR).网格应用服务(GAS)或其他客户端通过虚拟数据库查询服务,获取应用系统中的虚拟数据库模式信息.(5) 虚拟数据库服务(VDBS).网格应用服务或其他客户端通过访问虚拟数据库服务获得来自各异构数据库中的信息.为了尽量避免因虚拟数据库集中处理应用查询请求而降低系统效率,可以在 UDMGrid 中部署多个虚拟数据库服务,以分流查询请求.(6) 查询与响应转换服务(Req/Res TranServ),将资源使用者对虚拟数据库的访问请求转换为对真实数据库的访问请求,并将来自多个真实数据库的响应合并为针对一个虚拟数据库的响应.(7) 资源监测服务(ResMonitor),定期通过数据库访问代理监测已注册数据库资源的状态.(8) 元数据目录,用来存储虚拟数据库模式、数据库资源元数据和数据库资源到

© 中国科学院软件研究所 http://www.jos.org.cn

虚拟数据库之间的映射信息.其中,数据库资源注册服务、虚拟数据库查询服务、虚拟数据库服务等部署在网格服务容器中,分别被数据库资源注册客户端、网格应用服务(grid application service,简称 GAS)或其他客户端调用.

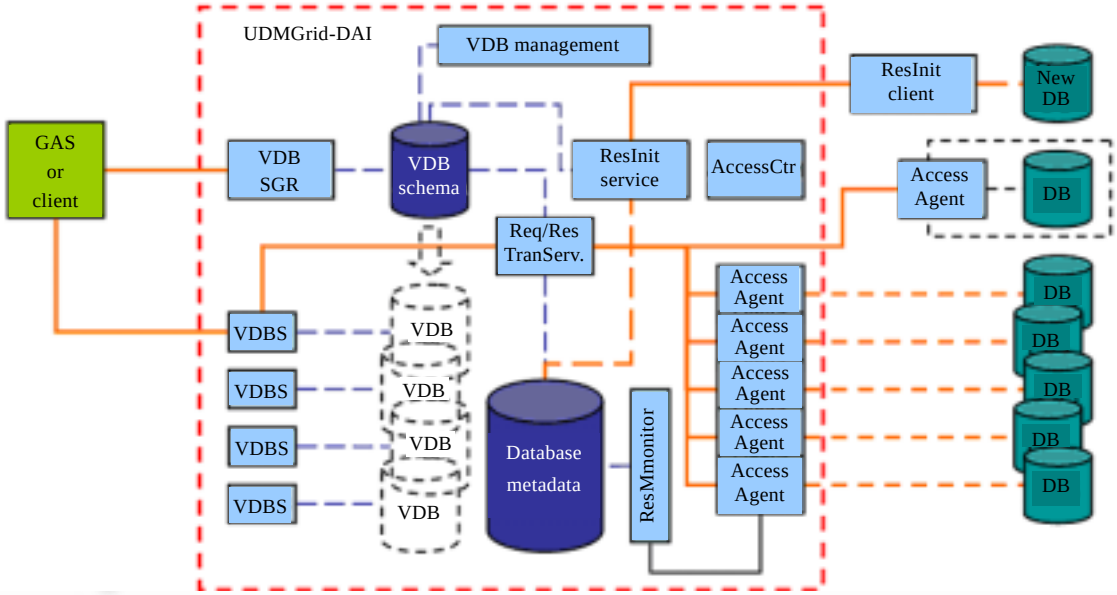

Fig.4 Architecture of UDMGrid-DAI

图 4 UDMGrid-DAI 总体结构

如图 5 所示,利用 UDMGrid-DAI,UDMGrid 的网格应用服务访问各异构数据库资源的过程如下:(1) 网格应用服务向 UDMGrid-DAI 的虚拟数据库提交访问请求;(2) 虚拟数据库服务将查询请求递交给查询映射服务;(3) 查询映射服务访问数据库元数据目录,结合虚拟数据库模式,将对虚拟数据库的访问请求转换为对相应真实数据库的访问请求;(4) 查询映射服务通过访问代理,顺次或并行访问真实的数据库资源;(5) 查询映射服务将合并后的查询响应返回给虚拟数据库服务;(6) 虚拟数据库服务将查询响应返回给网格应用服务,完成本次访问过程.

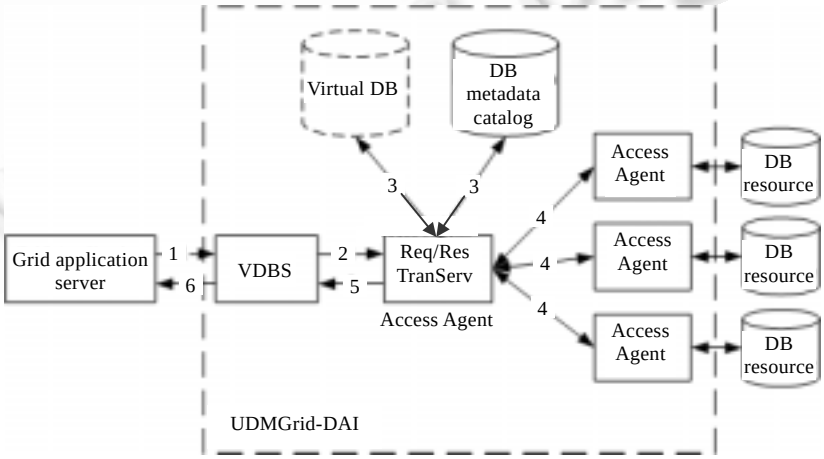

Fig.5 Process of access to heterogeneous database resource

图 5 访问异构数据库资源的过程

© 中国科学院软件研究所 http://www.jos.org.cn

5 实验与分析

为了验证 UDMGrid-DAI 在异构数据库资源整合方面的功能能否满足 UDMGrid 的需求,首先由系统构建者和领域专家创建 UDMGrid 数据库资源的全局模式;然后,由数据库资源提供者在 UDMGrid 注册其提供的数据库资源;最后,由网格应用服务的开发者使用虚拟数据库服务提供的接口获取来自异构数据库的数据.图 6 是虚拟数据库的构建过程,根据领域专家的建议,构建了“博物馆”、“航空航天”、“考古”等虚拟数据库.图 7 描述了航空航天数字博物馆数据库资源的注册过程,将资源数据库的“Model”等表映射为“飞机”和“发动机”等虚拟表.

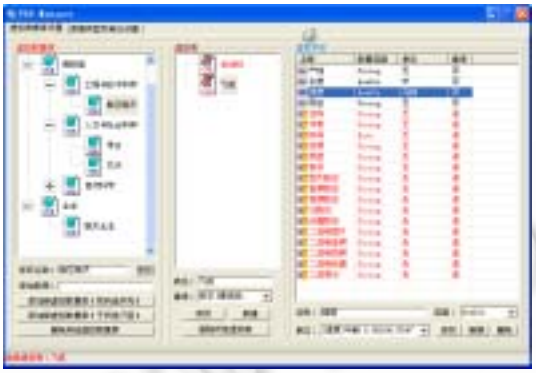

Fig.6 Creation and maintaining of virtual database

图 6 虚拟数据库的创建与维护

Fig.7 Register database resource of aeronautic & astronautic digital museum

图 7 航空航天数字博物馆后台数据库注册

UDMGrid 利用“博物馆”虚拟数据库开发的“一站式”标本信息服务,支持用户通过 UDMGrid 门户获得来自多个相关数字博物馆资源的标本信息.例如,用户输入标本名称“马”,设定查询范围为已注册数字博物馆资源,在提交任务之后,获得的信息服务结果如图 8 所示(由于输出内容非常多,移动了部分图片):左上角为来自中山大学生物数字博物馆中的“斑马”;右上角为来自中国农业大学昆虫数字博物馆的石雕“马上蝇”;左下角为来自航空航天数字博物馆的“野马”战斗机;右下角为来自山东大学考古数字博物馆的西汉时期文物“车马器”.本次服务从清华大学美术数字博物馆、西安建筑科技大学建筑数字博物馆、航空航天数字博物馆等 12 个数字博物馆中共获得 878 条数据库标本信息和 1 249 条 Web 标本信息,用时 5 分 22 秒.

UDMGrid-DAI 在网格环境下的异构数据库整合过程中提供了灵活的虚拟数据库机制,无论是对网格系统构建者,还是对资源提供者,或是对资源使用者,都具有较好的实用性.资源提供者只需要使用客户端程序,即可自动完成大部分的注册工作;资源使用者通过虚拟数据库查询服务,即可获得 UDMGrid 数据库资源的全局统一视图,不需要关心真实数据库资源的地理分布、数据库模式、语义信息等,并且可以通过虚拟数据库服务的一致接口获取其所需资源.与 OGSA-DAI 和 OGSA-DQP 相比,UDMGrid-DAI 更侧重于对异构数据库的集成,主要针对的是数据库资源的统一视图问题;OGSA-DAI 和 OGSA-DQP 虽然提供了统一访问接口并支持分布式查询评估,但并不提供数据库资源的全局视图,见表 1.

Table 1 Comparison between UDMGrid-DAI and OGSA-DAI

表 **1** UDMGrid-DAI 和 OGSA-DAI 的比较

Items	OGSA-DAI	UDMGrid-DAI
Focus on	Data access	Data integration
Global view of resources	No	Customizable/Flexible
Register of resources	Described roughly by system constructor	Described accurately by resource provider
Resource utilization	Learn metadata of resources indirectly	Visit VDB directly
Performance	High, support distributed query evaluate	Normal, simple technology of database connection pool
Popularization	Widely	UDMGrid only till now

© 中国科学院软件研究所 http://www.jos.org.cn

Fig.8 Result of “one-stop” specimen service (from 12 DB resources)

图 8 “一站式”标本信息服务结果(来自 12 个数据库资源)

本文在研制大学数字博物馆网格的过程中,提出了一种网格环境下的异构数据库访问和集成方法.该方法首先提出了一种模式复用的虚拟数据库构建策略,定义了异构数据库资源的全局统一视图;在此基础上,给出了数据库资源注册方式和虚拟数据库查询映射方法,为资源使用者提供了访问数据库的一致访问接口;最后,在 UDMGrid 的异构数据库资源环境下,对从数据库资源注册到使用的全部过程进行实验,验证了 UDMGrid-DAI 方法的正确性和实用性.今后可以在提高虚拟数据库服务的效率、非数据库资源的访问与集成等方面进一步开展研究工作.

致谢 感谢陈明基硕士研究生对本文实验部分给予的帮助.

References:

[1] ChinaGrid (China education and scientific research grid) portal (in Chinese). http://www.chinagrid.edu.cn

[2] Jin H. ChinaGrid: Making grid computing a reality. In: Chen ZN, Chen HC, Miao QH, Fu YX, Edward F, Ee-peng L, eds. Digital Libraries: Int’l Collaboration and Cross-Fertilization (ICADL 2004). LNCS 3334, Berlin, Heidelberg: Springer-Verlag, 2004. 13–24.

[3] UDMGrid (university digital museum grid) portal (in Chinese). http://www.udmgrid.net

[4] Chen XW, Luo XX, Pan ZS, Zhao QP. A CGSP-based grid application for university digital museums. In: Chen GH, Pan Y, Guo MY, Lu J, eds. Proc. of the 3rd Int’l Symp. on Parallel and Distributed Processing and Applications—ISPA 2005 Workshops. LNCS 3759, Berlin, Heidelberg: Springer-Verlag, 2005. 286–296.

© 中国科学院软件研究所 http://www.jos.org.cn

[5] Chen XW, Xu Z, Pan ZS, Luo XX. UDMGrid: A grid application for university digital museums. In: Jin H, Pan Y, Xiao N, Sun JH, eds. Grid and Cooperative Computing (GCC 2004). LNCS 3251, Berlin, Heidelberg: Springer-Verlag, 2004. 720−728.

[6] Pan ZS, Chen XW, Ji XY. Research on database access and integration in UDMGrid. In: Chen GH, Pan Y, Guo MY, Lu J, eds. Proc. of the 3rd Int'l Symp. on Parallel and Distributed Processing and Applications—ISPA 2005 Workshops. LNCS 3759, Berlin, Heidelberg: Springer-Verlag, 2005. 496−505.

[7] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: Buneman P, Jeds S, eds. Proc. of the '93 ACM SIGMOD Int'l Conf. Management of Data. Washington: ACM Press, 1993. 207−216.

[8] Foster I. The anatomy of the grid: Enabling scalable virtual organizations. In: Rajkumar B, et al., eds. Cluster Computing and the Grid, Proc. of the 1st IEEE/ACM Int'l Symp. IEEE Computer Society, 2001. 6−7. http://ieeexplore.ieee.org/iel5/7358/19961/00923162.pdf?tp=&arnumber=923162&isnumber=19961

[9] Nemeth Z, Sunderam V. Characterizing grids: Attributes, definitions, and formalisms. Journal of Grid Computing, 2003,1(1):9−23.

[10] Database access and integration services working group (DAIS-WG in GGF). https://forge.gridforum.org/projects/dais-wg

[11] OGSA-DAI. http://www.ogsadai.org.uk/

[12] Antonioletti M, Atkinson M, Baxter R, Borley A, Chue Hong N, Collins B, Hardman N, Hume A, Knox A, Jackson M, Krause A, Laws S, Magowan J, Paton N, Pearson D, Sugden T, Watson P, Westhead M. The design and implementation of grid database services in OGSA-DAI. Concurrency and Computation: Practice and Experience, 2005,17(2):357−376.

[13] Atkinson M, Karasavvas K, Antonioletti M, Baxter R, Borley A, Chue Hong N, Hume A, Jackson M, Krause A, Laws S, Paton N, Schopf JM, Sugden T, Tourlas K, Watson P. A new architecture for OGSA-DAI. In: Simon C, David W, eds. Proc. of the UK e-Science All Hands Meeting. 2005. http://www.allhands.org.uk/2005/proceedings/papers/432.pdf

[14] Zhang F, Yan BP. A database metadata management framework based on grid service. Computer Engineering and Applications, 2004,40(29):209−212 (in Chinese with English abstract).

[15] China State Bureau of Technical Supervision. National standards of P.R. China—Classification and code of disciplines (GB/T13745-92). 1992 (in Chinese).

[16] Luo XX, Chen XW. OOML-Based ontologies and its services for information retrieval in UDMGrid. In: Cao JN, Wolfgang N, Xu M, eds. Proc. of the 6th Int'l Workshop on Advanced Parallel Processing Technologies. LNCS 3756, Hong Kong: Springer-Verlag, 2005. 342−352.

附中文参考文献:

[1] 中国教育科研网格 ChinaGrid 网站. http://www.chinagrid.edu.cn

[3] 大学数字博物馆网格 UDMGrid 网站. http://www.udmgrid.net

[14] 张非,阎保平.一种基于网格服务的数据库元数据管理框架.计算机工程与应用,2004,40(29):209−212.

[15] 国家技术监督局.中华人民共和国国家标准——学科分类与代码(GB/T13745-92). 1992.

陈小武(1972－),男,湖南安仁人,博士,副教授,CCF 高级会员,主要研究领域为虚拟现实,增强现实,网格技术.

赵沁平(1948－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为虚拟现实,分布式系统,人工智能.

潘章晟(1979－),男,硕士,主要研究领域为网格技术及应用.

© 中国科学院软件研究所 http://www.jos.org.cn