

doi: 10.3969/j.issn.1006-1576.2010.03.018

基于 OGSA-DAI 扩展的空间数据访问和集成

钟凯, 熊伟, 陈萃, 程果

(国防科学技术大学 电子科学与工程学院, 湖南 长沙 410073)

摘要: 针对分布式地理空间信息共享、资源管理和空间分析的需求, 结合网格技术的特点, 研究基于网格的分布式多源异构地理空间信息查询处理集成的关键技术, 提出了基于 EDIF 框架改进的集成框架。利用 OGSA-DAI 的扩展机制引入了对分布式空间数据的拓扑分析和互操作支持; 使用 JDBC Driver for OGSA-DAI 组件实现数据资源访问。最后实现了一个分布式地理空间信息集成访问和操作原型系统, 并通过实验对系统进行了测试和验证, 能为用户提供统一的服务访问接口。

关键词: 地理空间信息; 网格; OGSA-DAI; 分布式; 集成; 拓扑分析

中图分类号: TP311.3 **文献标识码:** A

Access and Integration of Geospatial Data Based on OGSA-DAI Extension

ZHONG Kai, XIONG Wei, CHEN Luo, CHENG Guo

(School of Electronic Science & Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Aiming at the demand of data sharing, resource management and spatial analysis of distributed geographic information, put forward a new method of integration framework based on EDIF improve on frame, through combining the characteristics of grid computing, studies on the key technology of distributed multi-resource heterogeneous geographic information based on grid computing. By taking advantage of the extension mechanism of OGSA-DAI, a new approach to topological analysis and special interoperation is presented. JDBC Driver for OGSA-DAI is used to implement the access to the data resource. Finally a distributed heterogeneous GIS integration and interoperation prototype is implemented. The prototype is able to provide a uniform service-based interface for users, and is validated by experiments.

Keywords: Geospatial information; Grid; OGSA-DAI; Distributed; Integration; Topological analysis

0 引言

随着地理空间信息海量增长及遥感和测量设备的广泛应用, 每天都在产生 TB 级的地理空间信息。如何对获取的空间信息进行存储和继承、分布式计算、空间分析、认知和表达等已成为地理信息科学的主要问题。目前, WebGIS 是解决分布式 GIS 应用的主要途径, 但 WebGIS 体系与实现海量时空信息共享和协同管理的要求还有很大距离, 其系统稳定性、负载均衡能力也有待提高^[1]。网格技术的研究目标是希望用户在使用网格计算能力时, 如同使用电力网一样方便高效^[2]。网格计算有超强的数据处理和资源共享能力, 能充分利用网络中闲置的计算资源^[3], 为分布式 GIS 提供新的解决思路。OGSA-DAI (Open Grid Services Architecture- Data Access and Integration)^[4]是在网格框架上发展而来, 符合开放式网格框架标准 (OGSA)^[5]的数据集成和访问中间件, 能把各种分布式异构的关系型数

据库、XML 数据库、文件数据呈现成网格服务, 并为用户提供统一的访问查询接口。但是, 目前单纯的基础网格技术并不能完全满足地理空间领域的需求, OGSA-DAI 应用于地理空间信息系统的技术还不成熟^[6]。中科院计算所的南凯针对 OGSA-DAI 在数据集成方面的不足, 在 OGSA-DAI 数据服务模型的基础上, 提出了以服务组合方式进行扩展的数据集成框架 (Extended Data Integration Framework, EDIF)^[7]。EDIF 框架能够满足现实中一些应用场合的需要, 同时也为语义层面的集成提供了基础。不过由于 EDIF 是在 OGSA-DAI 数据服务的基础上再做一次封装和数据解析处理, 再加上服务本身复杂性和冗余信息的开销, 使得这种框架的性能有所下降。而且 EDIF 是对一般性数据资源类型进行的集成, 没有考虑地理空间数据资源的特性, 集成方法只涉及了 UNION、AND 和 OR 这些一般性操作, 无法完成分布式空间数据的拓扑分析等互操作。故对 EDIF 框架进行改进, 构建分布式地理信息访问

收稿日期: 2009-11-12; 修回日期: 2009-12-04

基金项目: 国家科技支撑计划《区域地理空间信息网格技术服务关键技术研究、开发与应用示范》(2007BAH16B03); 国家 863 计划《高可信地理空间数据库管理系统及其重大应用》(2007AA120402)。

作者简介: 钟凯 (1982-), 男, 福建人, 国防科学技术大学硕士, 从事地理信息系统和空间数据库研究。

和集成框架，实现了分布式地理空间信息集成访问和操作原型系统，为用户提供统一的服务访问接口。

1 基于 OGSA-DAI 的地理空间访问模型

定义 1 空间数据服务 (Spatial Data Service, SDS): 空间数据服务是承载空间数据服务资源的容器，是 OGSA-DAI 服务的提供点，用户根据 URL 来定位空间数据服务。一个空间数据服务用来配置多个空间数据服务资源 (Spatial Data Service Resources)。SDS 与 SDSR 的关系可以表示为 $sds_i \xrightarrow{list Resource} \{dsr_1, dsr_2, \dots, dsr_n\}$ 。

定义 2 空间数据服务资源 (Spatial Data Service Resources, SDSR): 空间数据服务资源承担着空间数据集成访问的核心功能。它从空间数据服务 (SDS) 上接受执行文档 (Perform Documents)，解析并使执行文档生效，继而顺序运行执行文档中“承载”的每个行为指令 (Activity)，生成文档 (Response Document)。一个空间数据服务资源 SDSR 代表空间数据资源 SDR，即 $sdsr_i \leftrightarrow sdr_i$ 。

定义 3 空间数据资源 (Spatial Data Resource, SDR): 空间数据资源 (SDR) 主要包括关系地理空间数据库 (如 PostGIS, SQLServer, Oracle Spatial 等)、文件形式存储的空间数据等等各种实体。

空间数据服务 SDS、空间数据服务资源 SDSR、空间数据资源 SDR 之间的关系如图 1。

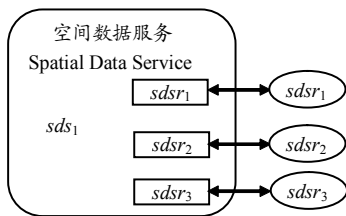


图 1 SDS、SDSR、SDR 之间的关系

用户要对一个地理空间数据资源实体进行访问，需要指定这个实体 SDR 所在数据服务 SDS 的 URI 地址和对应的 DSR 的 ID，然后提交包含查询请求的执行文档 Perform Document (PD)，执行完查询后返回给用户一个包含查询结果的响应文档 Response Document (RD)，即 $RD=request\{sds_i, sdsr_i, PD\}$ 。可见，在这种模式下用户并不需要了解 SDR 实体的相关信息，如空间数据资源类型、操作系统、硬件类型、IP 地址等，只需要知道该 DR 对应的 DSR 标识和承载其的 DS 的 URI 地址，因此充分屏蔽了异构性，实现了透明访问，而且拥

有统一的服务接口。

2 基于 OGSA-DAI 扩展的地理空间集成访问和操作模型

2.1 EDIF 框架

以上模式，用户一次只能对一个空间数据实体资源进行访问，还不能达到数据集成的效果。EDIF 框架对数据服务资源 DSR 进行扩展建立虚拟数据资源 (Virtual Data Resource)，一个 VDR 并不代表一个实体的数据资源，而是代表一个由 OGSA-DAI 封装之后以服务方式展现的数据资源。DS 和 DSR 是定义一个 VDR 的关键属性，即： $vdr_i \leftrightarrow (ds_j, dsr_k)$ 。EDIF 框架还定义了虚拟数据服务 (Virtual Data Service) 用来对 m 个 VDR 进行包装，即 $vds_i \rightarrow \{vdr_1, vdr_2, \dots, vdr_m\}$ ，向 VDS 提出的服务请求是在该 VDS 包含的全体 VDR 上实施的，返回的结果是对所有 VDR 操作结果的综合，这时， $RD_{vds_i} = request\{vds_i, PD\}$ 。VDS, VDR 以及 DS 的关系如图 2。

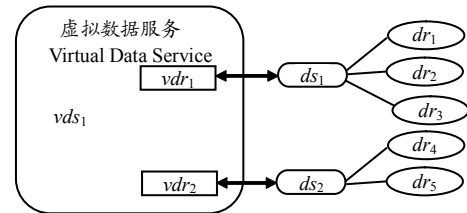


图 2 EDIF 框架下 VDS、VDR、DS 关系图

前述 EDIF 框架对数据服务 DS 做封装增加了数据的冗余和服务的复杂性，造成性能下降，另外缺乏地理空间信息操作和处理的功能，接下来对 EDIF 在服务封装上进行改进和简化，然后加入空间数据操作的扩展和支持。

2.2 基于 EDIF 改进的地理空间信息集成访问框架

定义 4 虚拟空间数据服务资源 (Virtual Spatial Data Service Resource, VS DSR): 用一个虚拟数据服务资源 VDSR 对 SDS 下具有相同语义结构的地理空间数据资源实体 SDR 对应的 SDSR 包装，形成虚拟空间数据服务资源 VS DSR，每个 VS DSR 是包含多个地理空间数据服务资源 SDSR (Spatial Data Service Resource) 的集群，即 $vdsr_i \leftrightarrow \{sdsr_1, sdsr_2, \dots, sdsr_n\}$ ；SDSR 和 SDR 还是一一对应的关系，即 $sdsr_i \leftrightarrow sdr_i$ 。访问这个 VS DSR 即是同时对多个 SDSR 进行访问，将返回的结果做集成处

理或拓扑分析之后返回给用户。地理空间数据服务 SDS、VSDSR、SDSR、SDR 之间的关系如图 3。

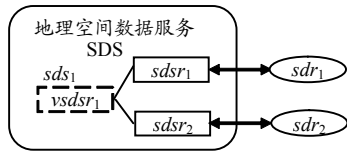


图 3 SDS、VSDSR、SDSR、SDR 之间的关系

一个虚拟空间数据服务资源包含多个空间数据资源实体, 即 $vdsr_i \rightarrow \{sdr_1, sdr_2, \dots, sdr_n\}$, 用户提交一个执行文档 PD 给 VSDSR, 可对这个 VSDSR 包含的所有 SDSR 进行访问, 即 $request\{vdsr_i, PD\} \rightarrow request\{sdr_1, \dots, sdr_n\}$ 。可对 SDSR 返回的结果进行 AND、OR 和 UNION 操作或者各种空间几何拓扑分析, 如求空间关系、叠加交叉、缓冲区等。设操作为 Operation, 则最后返回给用户的响应文档为:

$$RD = Operation (request \{vdsr_i, PD\})$$

2.3 基于 JTS 的空间数据操作框架

JTS Topology Suite^[8]是一套开源 Java API, 为兼容 OGC 标准的空间对象模型的几何操作提供空间谓词和空间操作函数。JTS 能创建和使用几何 (geometry) 对象, 计算空间关系, 处理叠加操作, 缓冲区处理, 多边形的联合等等多种几何操作, 功能强大, 性能比较稳定。JTS 遵循了 OpenGIS Simple Features Specification 规范, 广泛应用于各种 GIS 相关的开源平台, 如 Geotools、Udig、Open-JUMP 等。JTS 提供了丰富的空间谓词, 如 contains(), unions(), intersection() 等。JTS 对这些函数的实现非常强壮, 稳定性高, 面对奇异的几何对象或者临时的坐标系运算也不会轻易出错, 效率比较高。

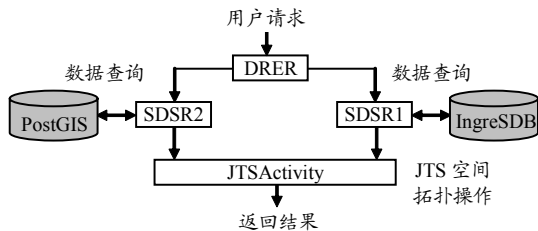


图 4 JTSActivity 数据操作流程

在上节建立的地理空间信息集成访问框架中对空间数据的操作进行基于 JTS 扩展实现。OGSA-DAI 的活动 Activity 是 OGSA-DAI 的扩展点, 可以创建和部署新的 Activity 来实现 OGSA-DAI 对空间数据操作的功能扩展。创建一个 JTSActivity 空间操作模块, 用来进行分布式的空间操作。该 Activity

实现了对 JTS API 的调用, 负责将空间数据资源服务 SDSR 上查询的地理空间结果集根据需求进行空间谓词或拓扑分析, 数据流程如图 4。

用户提交的执行文档包括 5 个参数, 分别是要访问的 SDSR 的 ID 和 SQL 查询语句, 以及空间操作符号 op: $PD = \{SQL_1, sdsr_1, SQL_2, sdsr_2, op\}$ 。

2.4 JDBC Driver for OGSA-DAI

经过对 OGSA-DAI 的了解和使用, 发现其易用性还有待提高, 无论是从框架搭建, 服务部署, 还是客户端的访问, 步骤都比较繁琐。从 Java 编程的角度来看, 用户要使用 OGSA-DAI 服务, 就必须写一段繁杂的客户端代码, 给维护和扩展带来了困难。本系统采用了 OGSA-DAI 的 JDBC 驱动: JDBC Driver for OGSA-DAI^[9] 开源工具, 使一般的 Java 应用也能方便地访问发布在网格环境的数据服务, 相当于为用户定制了数据层, 用户可把更多精力投入到业务逻辑层的实现。

JDBC Driver for OGSA-DAI 实际上是一个 Java 编写的 Activity, 使用时将其部署在网格数据服务 (DS) 之上。它包含有 3 个重要的类: Driver、DAIConnection、DAIStatement。

其中, Driver 用于接收 JDBC URL 连接串;

DAIConnection 使用 OGSA-DAI 客户端工具, 根据接收到的 JDBC URL 创建 stub 以访问数据服务工厂 (Data Service Factory), 并根据请求创建 DAIStatement;

DAIStatement 创建 data service, 设置数据服务和查询请求的超时时间, 发送查询语句到 data service, 并把结果转化成 ResultSet。

3 系统实现

3.1 分布式地理空间信息集成访问和操作系统

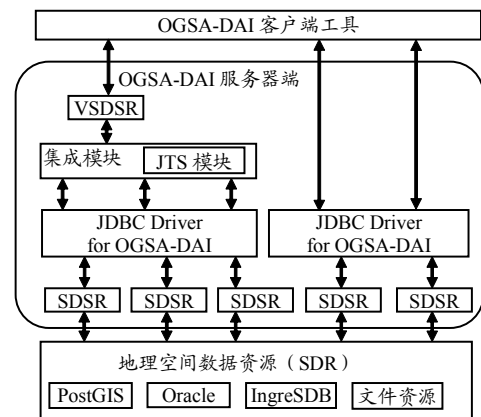


图 5 系统的体系结构

基于以上技术,构建基于 OGSA-DAI 扩展的分布式地理空间信息集成访问和操作平台框架,把各种分布式异构的空间数据资源呈现在网格服务上,用户在统一的接口上根据需要灵活的对分布式空间数据资源进行访问和操作,该框架体系结构如图 5。

用户在该系统中可根据需求对单个数据资源进行访问,也可对多个数据资源进行集成访问,并提交一个空间操作符或集成类型,如 Union、Intersection、Difference 等。系统将在网格服务中进行空间拓扑运算,并将最终得到结果返回给客户端。

3.2 实验与分析

为验证系统在分布式异构空间数据资源的集成访问和操作上是否满足要求,在 2 台主机上分别安装 2 个空间数据库管理系统 PostGIS 和 IngreSDB。PostGIS 上有一张表 roads 存储了某地区道路的空间信息,另一个数据库 IngreSDB 上有一张 rivers 表存储了同一地区河流的空间信息。将这 2 个数据资源都部署到系统服务中。假设用户需要查询该地区桥梁的位置坐标,虽然 2 张表中没有存储桥梁相关信息,但是可采用拓扑分析 Intersection,求河流与桥梁的交点得出。系统以网格服务的方式对这 2 个分布在不同主机的数据库图层进行集成访问,并获取道路和河流交点的空间信息。图 6 为系统界面。



图 6 系统界面

实验结果图 7(d)为河流与道路交点信息,达到了预期分布式空间信息集成访问和操作的效果。

4 结束语

网格 GIS 技术的日益成熟为地理空间信息共享和集成提供了新的解决方案。首先对 EDIF 框架运

用于分布式 GIS 系统的优劣进行了探讨,其次利用 OGSA-DAI 的扩展机制改进了 EDIF 框架,创建新的功能模块,实现对空间拓扑分析模块 JTS 的调用,最后构建了分布式多源异构地理空间信息的集成访问和互操作服务平台,系统在网格服务中充分利用网格资源进行具有数据密集和计算密集特性的空间拓扑运算,减小了客户端的计算压力和维护管理成本。网格 GIS 体系还有发展空间,研究基于网格的 GIS 信息处理标准和框架具有学术和商业意义。

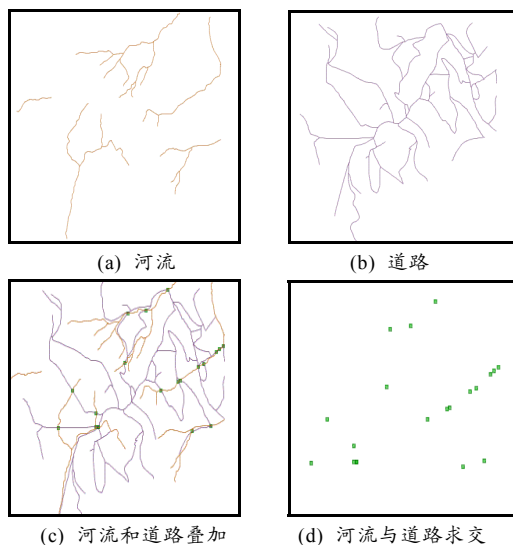


图 7 实验结果图

参考文献:

- [1] 金宝轩, 边馥苓. 基于 OGSA-DAI 的空间数据访问和集成研究[J]. 测绘信息与工程, 2005, 30(3): 20-22.
- [2] Foster, I., and Kesselman, C. The Grid: Blueprint for a New Computing Infrastructure[M]. San Francisco: Morgan Kaufmann Publishers, 1998.
- [3] Joshy, J. & Fellenstein, F. 网格计算(Grid Computing)[M]. 占晓苏, 张少华 译. 北京: 清华大学出版社, 2005.
- [4] The OGSA-DAI Group. OGSA-DAI User Documentation [EB/OL]. (2009-05-07) [2009-10-22]. <https://sourceforge.net/apps/trac/ogsa-dai/wiki/UserDocumentation>.
- [5] 都志辉, 陈渝. 以服务为中心的网格体系结构 OGSA[J]. 计算机科学, 2003, 30(7): 26-30.
- [6] 刘建新, 阎保平. OGSA-DAI 体系结构及其关键技术研究[J]. 计算机应用, 2004, 24(11): 81-83.
- [7] 南凯, 阎保平. 扩展 OGSA-DAI 的数据集成框架及原型[J]. 计算机工程, 2007, 33(10): 55-57.
- [8] Jonathan Aquino. JTS Topology Suite Developer's Guide[EB/OL]. (2003-05-06) [2009-10-22]. <http://www.vividolutions.com/jts/bin/JTS%20Developer%20Guide.pdf>.
- [9] Brito, M & Sato, L.M.. Extending OGSA-DAI Possibilities with a JDBC Driver[C]. Proceedings of the 2008 11th IEEE International Conference on Computational Science and Engineering IEEE International Conference on Volume, 2008. Washington, DC: IEEE Computer Society.