

文章编号: 1000-5013(2008)02-0218-04

OGSA-DAI 的异源空间数据访问技术及应用

樊明辉, 陈崇成, 吴小竹

(福州大学 空间数据挖掘与信息共享教育部重点实验室, 福建 福州 350002)

摘要: 采用开放式网格服务体系结构 数据访问与集成(OGSA-DAI) 中间件的数据访问模式, 利用空间数据引擎的空间数据管理能力, 构建分布式的多源网格数据服务(GDS). 基于 WKT(Well Known Text) 中间数据模型, 构建一个空间数据查询和访问的统一模式, 并将结果集转换为通用的空间数据格式. 实例表明, 通过 WKT 中间数据模型, 可实现自定义查询或跨数据服务资源的联合查询, 查询结果可以方便地转换为 XML, SVG, GML 文档.

关键词: 开放式网格服务体系结构 数据访问与集成; 网格数据服务; 空间数据引擎; WKT 格式

中图分类号: TP 393. 02; TP 311. 13

文献标识码: A

开放式网格服务体系结构 数据访问与集成(OGSA-DAI) 是网格环境中访问和集成独立数据源的数据访问中间件, 它支持关系数据库和 XML 数据库的访问和集成. 由于空间数据的非结构化特性, 目前的 OGSA-DAI 还不支持对空间数据库的直接操作. 因此, 网格环境下, 对空间数据的访问和集成一般采用改进 OGSA-DAI 的方式^[1]. 通过改进的 OGSA-DAI 访问空间数据库有许多不同的思路. 金宝轩^[2]利用 ESRI 公司 ArcSDE 的 Java APIs, 在 OGSA-DAI 的活动映射类中实现了对 SQL Server 和 Oracle 空间数据库的查询. 韩金华^[3-4]提出了利用空间数据引擎(ArcSDE 或 Oracle Spatial) 和 SQL, 分别访问空间数据和属性数据并共同构建网格数据服务服务(Grid Data Service, GDS) 的空间数据访问框架. Meyer 等^[5]提出一个网格环境下分布式空间数据访问框架, 结合 Globus 平台的 GRAM 和 MDS 模块实现空间数据资源的发现和监控. 本文提出在网格环境下, 以空间数据引擎为中间件层的空间数据访问框架, 利用 WKT 中间数据模型存储空间数据查询或访问操作的中间结果.

1 基于空间数据引擎的网格数据服务

OGSA-DAI 以网格服务的形式为用户提供数据访问和管理服务, 其核心功能以 3 种服务来体现^[6]. 即 GDS, GDSF(Grid Data Service Factory) 和 DAISGR(Data Access Interface Service Group Register). GDS 支持数据的访问、集成和发布, GDSF 用于创建 GDS 实例并访问特定的数据源, DAISGR 通过各种服务的状态信息(名称、生命周期、连接数据源等) 的注册以供应用程序动态访问. 三者之间的关系, 如图 1 所示.

考虑物理上分离的两个或者多个网格节点间异构空间数据库系统的数据访问和交互问题, 数据访问的总体结构可以分为 4 层^[7], 如图 2 所示. 底层为分布式存储的异构数据源, 可以包括各数据管理系统和文件, 进入网格的数据需注册资源信息, 网格中数据以服务的形式发布. 中间件层由各种空间数据引擎组成, 空间数据引擎采用一系列数据模型和空间索引机制, 对分布在各网格节点上的空间数据库或文件系统进行统一的有效管理, 为用户提供了一致的操作接口. 用户通过 API 接口统一调度并实现数据请求和服务, 兼容异构数据源. 网格服务层包括 OGSA-DAI 提供的 3 种服务, 用户通过与 DAISGR

收稿日期: 2007-12-05

作者简介: 樊明辉(1974), 男, 博士, 主要从事空间数据挖掘与网络信息服务的研究. E-mail: fanmh@fzu.edu.cn.

基金项目: 国家自然科学基金资助项目(60602052); 福建省自然科学基金资助项目(D0610008); 福建省科技计划项目(2005H028); 福建省教育厅重点项目(2005K4004)

交互直接发现已创建的 GDS, 也可查找 GDSF 以创建 GDS. 此时, 通过网格服务层将用户、GDS 和空间数据引擎进行绑定, GDS 支持多数据源的配置, 即 1 个 GDS 可以支持多个空间数据引擎的并行访问.

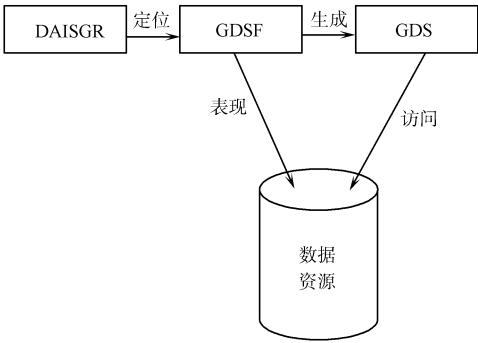


图 1 OGSA-DAI 的几种服务及其关系

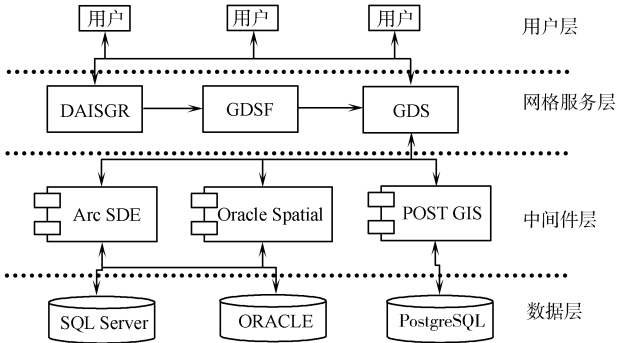


图 2 空间数据访问总体结构

用户通过发送执行文档与 GDS 进行交互, 执行文档封装了一系列的活动. 每个活动指定 1 个要执行的动作, 如查询数据源、转换数据格式, 或者导入和导出数据, 活动在活动映射文件中映射了具体的 Java 实现类. 因此, 通过扩展 Java 实现类可执行对应的活动^[8], 从而实现多类空间异构数据库的访问.

2 中间数据模型

在 GDS 利用多个空间数据引擎访问空间数据源的过程中, 当用户提交 1 个执行文档, 完成查询、转换, 或者导入、导出任务后, 数据层传回中间件层 1 个响应文档. 响应文档是 XML 格式, 内容包括客户请求的执行状态和处理后返回的结果. 由于空间数据数据库不仅存储了空间要素的属性信息, 而且还存储了空间位置、空间关系信息. XML 文档难于表述非结构化或半结构化的空间信息, 因此, 本文采用 1 种中间数据模型存储来自于空间数据库的查询信息, 确保空间位置和空间关系信息的完备性. 然后, 将其导入 XML 响应文档, 并返回给客户层, 如图 3 所示.

中间数据模型采用 OGC (Open Geospatial Consortium) 的 WKT 格式. WKT 是表述空间对象的标准方式之一, 以文本的形式表示要素 (图 4) 的几何对象的类型信息和坐标信息, 可以直接用于属性查询和空间查询 (GeoSQL), 而且还支持查询以 GML 或 SVG 格式返回查询结果.

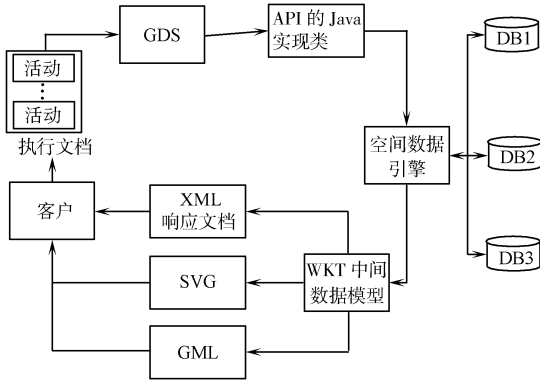


图 3 基于中间数据模型 WKT 的任务流程

Fig. 3 The task flow based on WKE middle data model

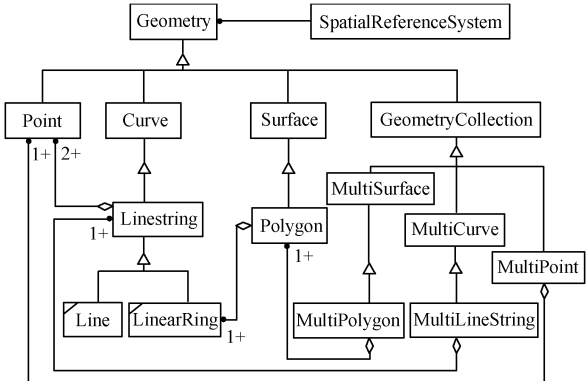


图 4 用于 SQL 查询的 OGC 简单要素规范

Fig. 4 Simple features specification of OGC used for SQL

用字符来描述要素的空间对象的实例, 如点: POINT (0 0); 线: LINESTRING (0 0, 1 1, 1 2); 多边形: POLYGON ((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1)).

以下是空间点要素对象的 SQL 查询的例子, 分别返回 WKT 文本、SVG 和 GML.

(1) Select astext(the_geom) from City where city_name= '福州'. 其中, the_geom 字段是要素空间对象 GEOMETRY 类型. 返回的查询结果:

POINT(119°18' 26°05');

(2) Select assvg(the_geom) from City where city_name= '福州'. 返回的查询结果:
cx="119°18'cy="26°05";

(3) Select asgml(the_geom) from City where city_name= '福州'. 返回的查询结果:
< gml: Point>
< gml: coordinates> 119°18, 26°05< /gml: coordinates>
< /gml: Point>

基于上述原理, 可以完成面向 WKT 的自定义查询或者跨数据服务资源(数据库)的联合查询. 上述 3 种查询结果可以方便地转换为 XML, SVG, GML 文档.

3 实例

(1) 搭建网络环境. 以局域网某网段(211.80.198.*)内 5 台 PC 机和 1 台 SGI 小型机组成两个虚拟组织(VO), 其节点的 IP 分别为 136, 145, 176, 177, 178(SGI), 179, 如图 5 所示. 操作系统均为 Linux RedHat 9, 每台机器均安装 GT 4.0.2 All Source, 局域网以带安全模式的网络传输协议 HTTPS 连接, 以 SGI 小型机连接两个 VO, 并作为 VO 内的 Web 服务器成为 2 个 VO 向外提供服务的出入口, 而接受并执行从 VO 外 Web 服务器 183 的指令. 在节点 176 上安装 Oracle 9i 数据库和 Oracle Spatial 空间数据引擎, 节点 177 和 179 上分别安装 Postgre SQL 数据库和 PostGIS 空间数据引擎.

(2) 部署网格数据服务. 以在节点 176 上部署基于 Postfgr SQL 的数据服务为例. OGSA-DAI WSRF 安装包的主目录下运行下列命令, 发布数据服务: \$ ant deployService -Ddai.container= d:\tomcat -Ddai.service.name= pgdb/sharedb; 可通过 URL 地址(http://211.80.198.176:8080/wsrfl/services/pgdb/sharedb) 访问该服务, 发布数据服务资源: 修改数据服务资源属性文件, 配置数据库的访问地址、访问用户名、密码、驱动信息等. 1 个数据服务可以配置多个数据服务资源, 1 个数据服务资源被分配 1 个资源 ID. 因篇幅所限, 这里不赘述.

(3) 数据的查询与转换. (i) 自定义查询功能. 通过用户自定义的查询语句查询数据库, 查询结果集为 WKT 文本型的中间格式, 如图 6(a). (ii) 跨资源的联合查询. 一个数据服务可配置多个数据服务资源, 即可对应多个空间数据引擎, 实现多个资源的“Join”查询; 在查询界面中设定相应的资源 ID, 即可对资源加以引用, 如图 6(b).

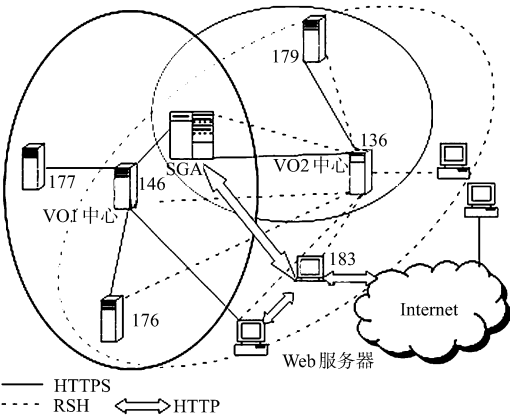


图 5 网络试验环境结构图

Fig.5 The structure of test environment of grid



(a) 自定义查询及其结果



(b) 跨资源的联合查询

图 6 数据查询与转换

Fig.6 Data query and translation

(4) 数据的存储与发布. 上述查询结果以 XML, GML, SVG 格式存储, 其中 GML 和 SVG 格式可

直接作为网络地图来发布.

4 结束语

网格环境下, OGSA-DAI 可通过空间数据引擎作为空间件对空间数据进行访问, 这为数据访问提供了统一的模式. 通过将扩展的空间数据引擎的 Java 实现类封装为 GDS 执行文档中的 Activities, 实现对多源空间数据库的访问. 通过 WKT 中间数据模型实现了自定义查询或者跨资源(数据库)的联合查询, 查询结果可以以 XML, GML 和 SVG 3 种方式返回给客户. 当前对空间数据的存取和转换是以矢量数据的 OGC 简单要素模型为基础的, 有关对栅格数据的存取还需进一步研究.

参考文献:

- [1] SHU Yair feng, ZHANG J F, ZHOU Xiao fang. A grid enabled architecture for geospatial data sharing[C]// IEEE Asia Pacific Conference on Services Computing (APSCC'06), 2006: 369-375.
- [2] 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.
- [3] 韩金华. 基于改进 OGSA-DAI 的空间数据集成研究[J]. 高性能计算技术, 2006, 182: 9-23.
- [4] 韩金华. 基于网格的空间数据库集成研究与应用[D]. 南京: 河海大学, 2006.
- [5] MEYER W S, SOUZA J M, RAMIREZ M R. Second grid: An infrastructure to study spatial databases in computational grids[C]// VII Simpósio Brasileiro de Geoinformática, 2005: 191-209.
- [6] 刘 瑰, 陈 琳, 何红旗. 利用 OGSA-DAI 实现异构数据的透明访问[J]. 高性能计算技术, 2006(180): 11-14.
- [7] 高 勇, 秦 适, 王晓明. 空间数据访问中间件技术[J]. 地理信息世界, 2004, 2(4): 45-48.
- [8] 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.

Access Techniques and Implement of Different Source Spatial Data Based on OGSA-DAI

FAN Ming-hui, CHEN Chong-cheng, WU Xiao-zhu

(Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education,
Fuzhou University, Fuzhou 350002, China)

Abstract: On basis of the data accessing ability of spatial data engine (SDE), the open grid services architecture data access and integration (OGSA-DAI) middleware was adopted to construct grid data services (GDS), which integrated distributed multi source spatial data under grid environment. The well known text (WKT) was introduced to form an unified spatial data querying and accessing mode, the querying result sets could be transformed into general data formats. The instance illustrated that the WKT data mode was fit for custom query and joint query across resources, query result sets were conveniently transformed into XML, GML and SVG.

Keywords: open grid services architecture data access and integration; grid data service; spatial data engine; well known text

(责任编辑: 黄仲一 英文审校: 吴逢铁)