

# 分布式空间数据库集成访问技术\*

蒙智敏, 熊伟, 陈宏盛, 程果, 赵亮

(国防科学技术大学 电子科学与工程学院, 湖南 长沙 410073)

**摘要:** 提出了采用网格中间件技术解决面向分布式空间数据库的查询问题, 设计并实现了一个网格中间件系统 OGSA-SDQP。给出了该系统的设计思想, 重点研究了其中的空间数据类型转换、空间数据集成、空间操作函数扩展、空间数据查询流程等关键技术, 给出了系统实现及查询性能测试。实验结果表明, OGSA-SDQP 能够高效处理网格环境下的分布式空间数据查询。

**关键词:** 网格; 分布式; 空间数据查询; OGSA-SDQP

中图分类号: TP392

文献标识码: A

文章编号: 1674-7720(2010)24-0012-04

## Integrated access technology on distributed spatial database

MENG Zhi Min, XIONG Wei, CHEN Hong Sheng, CHENG Guo, ZHAO Liang

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** The grid middleware technology is adopted to solve the distributed spatial databases oriented spatial query problem, and a grid middleware system—OGSA-SDQP is designed and implemented. Firstly, the system design is presented. Secondly, research focuses are put on the key technologies of the system which include: the spatial data type transformation, spatial data integration, spatial operation extension, spatial query flow, etc. Finally, the implementation of the system is given together with some performance tests. The experimental results show that OGSA-SDQP can effectively process the distributed spatial queries under the grid environment.

**Key words:** grid; distributed; spatial data query; OGSA-SDQP

空间数据库用于存储空间和属性数据, 其数据模型和查询语言能支持空间数据类型和空间索引, 并且提供空间查询和其他空间分析的方法。现实环境中, 由于 GIS 隶属单位地域差异及异构 GIS 数据库的存在, 空间数据库系统不仅地域分散, 而且数据具有异构性<sup>[1]</sup>, 因而空间数据共享一直是 GIS 研究领域的一个很有挑战性的问题。

实现不同空间数据库的信息共享和集成, 必须解决不同的通信协议、不同的系统结构、不同的操作系统、不同的数据库和其他应用服务带来的异构性<sup>[2]</sup>。网格是在网络之上运行的以实现资源共享和协作为目标的软件基础设施, 提供了一种集成的资源和服务环境。网格技术具有高性能计算、资源共享、互操作性、服务管理、高可靠性、安全性等特性<sup>[3]</sup>, 它的迅猛发展为实现分布式

空间信息的访问和操作提供了新的思路。开放网格服务框架 OGSA(Open Grid Service Architecture)是以服务为中心的框架结构, 在网格体系结构中有广泛的影响力。开放网格服务框架下的分布式查询处理器 OGSA-DQP(Open Grid Service Architecture-Distributed Query Processor)符合 OGSA 框架, 解决了一般数据的分布式异构的问题, 但是它不能有效地集成空间数据, 因为空间数据种类繁多, 数据格式各异, 获取时段各不相同, 存在多种比例尺、多种空间参考系和多种投影类型<sup>[4]</sup>。为此, 针对网格环境下的分布式空间数据集成问题, 以 OGSA-DQP 为基础研究并开发了一个中间件系统, 开放网格服务框架下的分布式空间查询处理器 OGSA-SDQP(Spatial DQP), 能够高效地处理分布式空间数据查询。

### 1 OGSA-SDQP 设计思想

空间数据具有分布式异构的特点, OGSA-SDQP 首先屏蔽了其异构性, 然后在此基础上集成了不同的空间数据源。OGSA-SDQP 支持并扩展了空间操作函数。

\* 基金项目: 本项目获得国家高技术研究发展计划(863 计划)重点项目(NO.2007AA120400)和国家自然科学基金(NO.40801160 和 No.60902036)的支持

## 软件天地 Software Technology

数据库之间具有异构性,为了有效集成来自不同数据库的数据,需要有统一的数据类型。OGSA-SDQP 用 Java 语言实现,需要将 SQL 数据类型转换成统一的 Java 数据类型<sup>[5]</sup>。对于常规的数据类型,不同数据库产品之间的语义差别不大,例如:SQL 数据类型中的 VARCHAR、CHAR、LONGVARCHAR 对应 Java 数据类型中的 java.lang.String 和 char[];SQL 数据类型中的 INTEGER 对应 Java 数据类型中的 int。但是空间数据具有存储格式多样性的特点,各个数据库的空间数据类型不相同,Java 中也没有表示空间数据的类型。在 OGSA-SDQP 中,空间数据被转换成一致的数据类型,实现了数据类型的统一。

在面向分布式空间数据源的 SDQP 查询中,组合查询被解析为针对不同数据源的子查询,子查询通过数据服务获取查询结果,在 SDQP 中将不同数据源返回的结果转换成一致的数据类型后进行组装,并返回给请求用户。

OGSA-SDQP 支持空间操作函数。单数据源查询过程中,提交到数据服务的查询串被完整地推送到后端数据库执行查询,后端数据库支持的查询函数都可以嵌入在查询语句中提交给数据服务并获得查询结果。对于数据表中包含的空间信息,查询语句可以根据嵌入的查询函数经后端数据库处理后返回计算结果。而对于多源数据的分布式查询,需要在查询语句推送前对应不同数据源拆分为相应的查询子句。在查询子句返回数据结果之后,根据嵌入在 OGSA-SDQP 内部的基于标量函数拓展的空间分析函数,对结果进行处理、合并,返回给服务请求客户端。

## 2 OGSA-SDQP 关键技术

### 2.1 空间数据类型转换

系统使用的两个空间数据库 PostGIS 和 Beyondb 都支持空间数据。Beyondb 是项目组其他成员在开源项目 Ingres 的基础上进行空间扩展生成的支持空间数据的数据库系统。目前,OGSA-SDQP 仅支持矢量空间数据。

首先,将空间数据库的基本信息提交给 OGSA-SDQP,这些信息主要包括数据库的 URI、数据库的驱动程序 JDBC 等。由于每个空间数据库的空间数据类型各不相同,PostGIS 的空间数据类型是 PG\_Geometry, Beyondb 的空间数据类型是 ST\_Geometry,OGSA-SDQP 根据每个空间数据库的 JDBC,将 Geometry 类型解析为通用文本 WKT(Well-Known Text)类型。WKT 格式是对几何体编码的一个结构化文本格式,可以像操作字符串一样对 WKT 进行操作。通过将不同的空间数据类型转换为 WKT 的方式,屏蔽了不同空间数据类型的异构性。

空间数据的类型转换如图 1 所示,当用户提交查询操作时,OGSA-SDQP 主要利用两个行为 Activity 将数据转换成结果集 ResultSet 的形式,SQLQuery 和 TupleToWe-

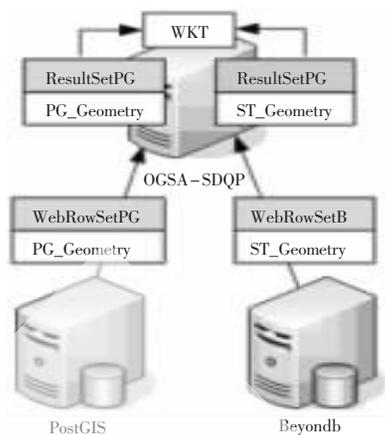


图 1 空间数据的类型转换

bRowSetCharArrays。它们处在一个工作流中,SQLQuery 的输出是 TupleToWebRowSetCharArrays 的输入。ResultSet 是 Java 对象,不适合网络传输,在传输前将其转换成 WebRowSet 格式<sup>[6]</sup>,OGSA-SDQP 接收到数据后将其转换回 ResultSet 格式。ResultSet 中的空间数据类型是空间数据库自己定义的类型,OGSA-SDQP 将空间数据转换成 WKT 格式,空间数据类型转换是空间数据集成的前提。

### 2.2 空间数据集成

空间数据集成是将分布式数据库构成一个大的虚拟的数据库,屏蔽数据库之间的异构性,为用户提供统一、透明的访问接口。通过下面的语句可以从两个空间数据库中一次查询出铁路信息和公路信息并返回一个结果集:select rail.geom from beyondb\_railway\_pln rail union all select road.geom from postgis\_road\_pln road。

为了向用户提供统一的数据视图,屏蔽底层数据库的位置信息和平台差异,实现异构数据库的无缝连接。虚拟表与数据源的映射关系如图 2 所示,OGSA-SDQP 首先建立了一张虚拟表,虚拟表中定义了虚拟表名、虚拟字段名等属性。然后,又建立了一张虚拟关系表来维持虚拟表名与实际表名、虚拟字段名与实际属性之间的映射关系,虚拟表中的一个属性可以对应 0 个或多个数据源对象的属性<sup>[7]</sup>。OGSA-SDQP 循环每个数据源结果集的行,再循环每行的字段,根据映射关系查找该字段在虚拟表中的位置,然后根据位置信息,将数据插入虚拟表。

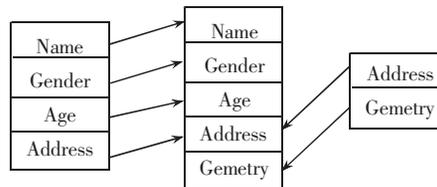


图 2 虚拟表与数据源映射关系

### 2.3 空间操作函数扩展

在数据访问与集成处理过程中,行为(Activity)是基本工作模块,每类行为负责完成一项特定的数据查询、

转换、传输等处理任务<sup>[8]</sup>。OGSA-SDQP 提供了行为的扩展接口,用户通过该接口可以定义自己的行为。

OGSA-SDQP 提供服务行为拓展机制,可供拓展的函数类型为标量函数与聚合函数两类。聚合函数从选择列数据中提取出代表一种趋势的值或者汇总值,并聚合列得到结果集,返回的数据类型对应于该列或者传递到函数中的值,如 COUNT、MAX、SUM 等。标量函数指的是基于输入值定义类型返回单个数据值。聚合函数针对通用数据类型进行操作,而空间分析过程需要针对空间数据表中每条记录包含的几何字段进行计算,因此拓展空间分析函数主要基于标量函数的形式拓展。

JTS Topology Suite 是一系列 Java API,它完成了 Java 对几何对象、空间拓扑的核心操作算法<sup>[9]</sup>。为了丰富用户的个性化操作,利用 OGSA-SDQP 的扩展功能和 JTS Topology Suite,添加了一些支持矢量数据的空间操作,包括缓冲区计算、凸外包框计算和空间相交计算等。

#### 2.4 空间数据查询流程

OGSA-SDQP 是一个基于服务的分布式查询处理器,它可以一次对多个数据服务进行并行查询。OGSA-SDQP 分布式查询的实现需要两个重要服务的支持:空间网格分布式查询服务 SGDQS (Spatial Grid Distributed Query Service) 和空间网格查询评估服务 SGQES (Spatial Grid Query Evaluator Service)<sup>[10]</sup>。

SGDQS 负责创建 SDQP 服务,生成数据资源实体的表结构 schema,对查询计划进行编译、优化、分解、制定查询计划,并将分解后的查询发送给 SGQES。另外 SGDQS 还负责与用户的交互,接受用户提交的请求,返回查询结果给用户。

SGQES 是查询计划的分支,是查询的执行节点,它在 SGDQS 制定好查询计划后动态生成,每个查询计划分支对应一个 SGQES 服务,SGQES 直接与数据资源和其他 SGQES 交互,并将执行结果返回给 SGDQS。SGQES 使用经 SGDQS 分解的 SQL 语句进行查询访问。图 3 说明了一次查询进行的数据流程,其中数据服务资源 SDSR (Spatial Data Service Resource) 标识和承载其数据库的 URI 地址,充分屏蔽了异构性,实现了透明访问<sup>[11]</sup>。

### 3 系统实现及实验

#### 3.1 系统环境

部署了 PostGIS 和 Beyondb 两个数据库服务器。在 WindowsXP 操作系统上的 Beyondb 数据库中存放了长沙市的铁路数据,在虚拟机 Ubuntu 操作系统上的 PostGIS 中部署了长沙市的公路数据。系统的服务器部分 OGSA-SDQP 部署在 Ubuntu 下。客户端分别部署在 WindowsXP 和 Ubuntu 下。

#### 3.2 系统实现方式

该系统分为服务器和客户端两个部分。服务器是空间扩展后的 OGSA-SDQP,支持空间数据类型和空间操

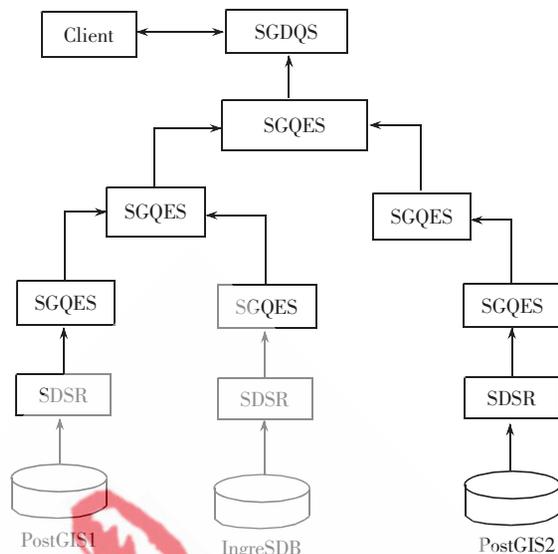


图 3 OGSA-SDQP 查询流程

作;客户端包括注册、查询、空间渲染、导出四个模块,这四个模块为用户提供了接口,用户可以根据需要选择其中几个模块建立自己的图形化系统。其中空间渲染模块和导出模块的实现上使用了开源项目 Geotools。Geotools 是 Java 语言编写的开源 GIS 工具包,提供各种 GIS 算法、各种数据格式的读写和显示,可以实现地图的可视化。

#### 3.3 实验结果

通过 OGSA-SDQP 访问空间数据库,既可以像以前一样访问单个数据库来获取空间信息,如图 4(a)、图 4(b)所示,也可以通过集成查询,将铁路和公路一次读取并显示出来,看起来就像是对一个数据库进行操作一样,如图 4(c)所示;还可以对获取的空间数据进行操作,图 4(d)是对铁路的外包框操作,图 4(e)是对公路做缓冲区操作,图 4(f)求铁路和公路的交点。

测试 OGSA-SDQP 的性能如图 5 所示。横轴表示查

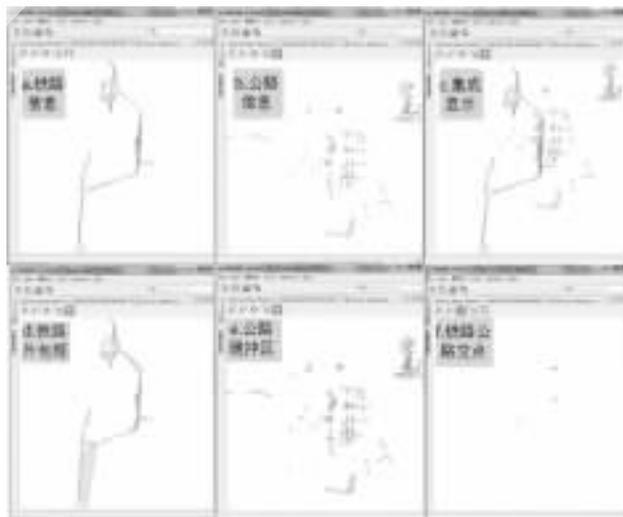


图 4 查询结果

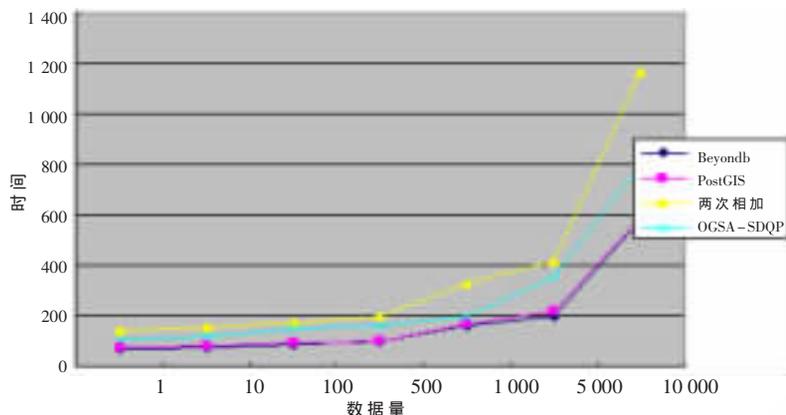


图 5 集成查询性能测试

询的数据量,纵轴表示查询时间。先分别对两个空间数据库进行查询,将它们的查询时间相加,再与 OGSA-SDQP 的查询时间做比较。实验发现,由于采用并行查询,OGSA-SDQP 在查询记录较小时,基本上与单独查询的时间一致。随着数据量的增大,耗费的资源也随之增大,因此在查询数据量大时集成查询所需时间会比单独查询时间更长,但总体上还是小于累计的时间。OGSA-SDQP 的性能取决于网格服务器的性能和各个数据库服务器的性能。

针对空间数据分布式异构的特点,使用网格技术,构建了分布式多源异构空间信息的集成访问环境 OGSA-SDQP。在该环境下,用户可以采用并发的方式同时访问多个空间数据源;扩展了空间操作功能,支持对查询结果的处理。最后,通过实验证明了该系统在集成查询中的性能优势。

#### 参考文献

- [1] 龚强. 地理空间信息网格计算环境研究[J]. 信息技术, 2008,32(7):1-4.
- [2] 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.

[3] JOSHY J, FELLESTEIN F 著. 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] 张丰. 面向网格的海量时空数据访问、集成于互操作研究[D]. 浙江: 浙江大学, 2007.4.

[5] HOARAU W, TIXEUIL S, RODRIGUES N. Dependability evaluation of the ogsa-dai middleware [A]. Achievements in European Research on Grid Systems[C]. 2008:223-236.

[6] LYNDEN S, MUKHERJEE A, HUME A C. The design and implementation of OGSA-DQP: A service-based distributed query processor[J].

FutureGeneration Computer Systems, 2009,25(3):

224-236.

[7] 刘会强. 网格环境下数据集成和查询处理的研究与实现[D]. 北京: 北京工业大学, 2009.4.

[8] CHUE Hong N, ANTONIOLETTI M, KARASAVVAS K. Accessing data in grids using OGSA-DAI[A]. Knowledge and Data Management in GRIDS[C]. 2007.3.

[9] 高昂. 面向空间数据的分布式服务计算研究与实现[D]. 北京: 中国科学院地理科学与资源研究所, 2010.3.

[10] 张建兵. 基于网格的空间信息服务关键技术研究[D]. 北京: 中国科学院遥感应用研究所, 2006.5.

[11] 庞雄文, 齐德昱, 李拥军. Globus 下的数据访问方法[J]. 计算机应用研究, 2006,27(4):68-70.

(收稿日期: 2010-09-20)

#### 作者简介:

蒙智敏, 男, 1985 年生, 硕士研究生, 主要研究方向: 地理信息系统与空间数据库技术。

熊伟, 男, 1976 年生, 讲师, 主要研究方向: 空间数据库技术。

陈宏盛, 男, 1958 年生, 副教授, 主要研究方向: 地理信息系统和数据库技术。