

网格 GIS 体系结构研究及应用

蔡正林, 韩金华, 李梦琪

(河海大学 计算机及信息工程学院, 江苏 南京 210098)

摘要: 网格 GIS(地理信息系统)是 GIS 在网格技术支持下,用以构建空间信息服务体系的技术系统。在分析传统 GIS 技术和网格技术的基础上,提出了一种在网格平台下,借助于 OGSA-DAI 构建的跨区域的、异构的 GIS 模型,并初步实现了在 GLOBUS 平台上的一个 GIS 应用。实验表明该系统能够有效地集成异地 GIS 资源,提高了传统 GIS 运算速度。

关键词: 网格地理信息系统; 网格计算; 空间数据共享

中图分类号: TP393; P208

文献标识码: A

文章编号: 1673-629X(2006)07-0221-03

Research on Grid GIS Architecture and Its Application

CAI Zheng-lin, HAN Jin-hua, LI Meng-qi

(College of Computer Science and Engineering, Hohai University, Nanjing 210098, China)

Abstract: Grid GIS is the application of GIS on grid computing. It can help us construct the spatial database service for people. Based on the technology of OGSA-DAI and the grid computing plat, the paper presents the architecture of grid GIS, moreover, discuss the implementation process based on the grid-GIS. All the data tested proved the efficiency of the application.

Key words: grid GIS; grid computing; spatial data sharing

1 研究背景

随着 GIS 技术的迅速发展, GIS 处理的数据量也在以指数级增加。GIS 系统对数据传输带宽和处理速度提出了挑战,传统的 GIS 技术已经满足不了这种要求。同时随着因特网技术的快速发展,在 Internet 基础上发展起来的基于 Web Service 的 GRID(网格)平台有着以下 3 个特征^[1]:资源一体化和服务一体化;可保证网格中各服务资源协同工作,共同完成应用任务;经整合后的平台提供了强大的计算能力及高速传输带宽,更为开发 GIS 应用提供了平台。因此如果将网络强劲的运算功能、跨地域的信息资源共享功能运用到当前 GIS 中,不失为解决当前 GIS 信息化过程中的瓶颈的一条捷径。

2 网络计算与 GIS 技术

2.1 网络计算

网络的一个比较通用的定义是:“网络是一个集成的计算与资源环境,或者说是一个计算资源池”^[2]。网络计算指基于网络的问题求解方案,网络计算突破了计算能力大小的限制,可以让不同机器同时为一个任务协同工作,因此可以提供足够的计算能力,网络计算突破了地理位置的限制,可使分布在各处的,甚至不可复制的资源突破地理限制;网络计算的另一个支撑技术是标准的存取异构网

格的应用框架,即 Web Service,其核心是在大的异构网络上实现将各种应用连接起来,借助于 Web 标准(UDDI, WSDL, XML/SOAP)将 Internet 从一个通信网络发展到一个应用平台^[3]。同时网络突破了传统的共享和协作方面的限制,网络允许对其它资源进行直接控制而不是仅在数据文件传输层次上^[4]。

2.2 地理信息系统

地理信息系统(GIS, Geographic Information System)作为一门新兴科学,以其独特的空间分析功能和属性信息相结合的方式,已广泛应用于资源管理、环境监测、灾害评估、城市规划等众多领域。应用 GIS,可以完成许多传统方法无法进行的信息处理和分析工作,为管理提供直观、系统、科学的管理工具;可以规范管理数据,实现信息共享,共享的空间数据范围除简单的图形数据外,还包括卫星遥感影像数据、数字地面模型数据等,这给当前空间数据共享带来了很大的困难。另外,随着地理信息系统应用的不断发展,在各个部门,存在着将本部门的专业系统与地理信息系统相结合的问题,不同部门之间的数据协作也变得日益重要。

3 理论模型及解决方案

网络环境下的 GIS 涉及到的技术包括万维网服务、资源整合、计算网络、数据网络、GIS 等。由于网络本身就是一个新生事物,加之网格 GIS 都是建立在商业化网络计算及 GIS 系统之上,很少会去从底层开发,因此当前并

收稿日期:2005-10-24

作者简介:蔡正林(1952-),男,江苏南京人,教授,研究方向为网络高性能计算、多媒体。

不存在基于网络协议的 GIS 软件,故网络环境下的 GIS 系统的实现是采用网络计算理念、基于当前现存的计算操作系统平台,通过部署网络服务来构建面向应用层的网络 GIS 软件来实现信息的共享和互操作。下面就网络环境下 GIS 体系结构作一简单探讨,并围绕这一体系结构实现了一个具体应用。

3.1 开放网络服务体系

Open Grid Services Architecture (OGSA) 是公认的网格体系结构,也是网格 GIS 的支撑平台,所有的 GIS 应用均在其基础上开发。具体 OGSA 请参考文献[5]。Globus Toolkit(GT)是 OGSA 实现的一个软件包,它是由一系列用来监测、发现、管理网格共享资源的服务、库以及关于安全和文件管理等软件构成的;在 GT 中并没有提供具体的应用层,这一层需要程序员根据各行业的需求作适应性开发,正因为此,就需要在其上架设 GIS 应用软件层。

3.2 网络环境下的 GIS 体系结构

网格 GIS 体系结构实际上是借用了网格的思想,在底层借助于 OGSA-DAI 将各种软硬件资源(包括 GIS 资源)进行整合,借助于 Globus Toolkit,在其应用层开发应用组件以调用网格 OGSA 服务[6]。

GIS 逻辑体系结构可分为五层,如图 1 所示。

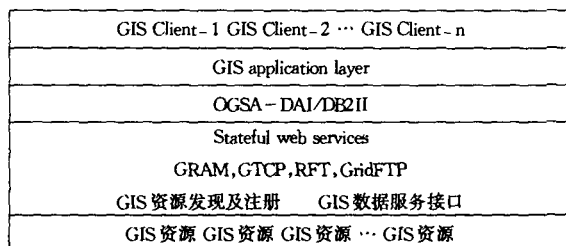


图 1 GRID GIS 体系结构图

①最底层为硬件及资源层,此处有本机(集群)的硬件资源和软件资源(包括 GIS 资源),依托 OS 实现本地的 GIS 资源管理,并向 GT4 环境注册信息;

②第二层实际上是目下 GT4 提供的功能,这也是本系统实现的一个基础环境。该层完成 GIS 资源注册及服务,用户可以提交跨部门 GIS 资源的作业与作业所需的各种资源脚本文件,一切操作对用户是全透明的,究竟取用了哪些资源,在什么机器上取的,由哪些机器同时参加了该作业的运算等相关细节都对程序员透明。这层主要涉及到 Web Services 等传统的网络协议标准,特别需要指出 Web Services 在提供底层连接时进行了扩展,实现了 Web Services 服务从无状态到有状态的服务,同时网络上一直使用的协议如 http,ftp,smtp,telnet 等都保留,只是根据需要进行了扩展[7],如 http 协议使得能够在 SSL(加密套接字协议)使用 GSI 等;

③第三层主要架设 OGSA-DAI 和 DB2II 完成各种数据库资源的整合,将多个集群或本地机器的 GIS 空间数据库加以集成,实现作业的提交、监视及删除等工作;

④第四层是 Globus Toolkit 中的应用层部分,程序员需在此写软件服务接口、服务的实现和部署描述器,从而

实现了网络环境中的 GIS 应用,对于下层资源及 GIS 服务的 API 调用,用户根据需要驱动软件调用相应的 API 函数,实现所需的服务。

⑤第五层为客户的浏览器,用户需根据需要下载相应的插件,当用户有 GIS 服务请求时只需以 HTTP 形式发送,经由 XML,SOAP,WSDL 层层包裹,去驱动 GT 为其服务,从而实现了 GRID GIS 的应用。

3.3 在网格 GIS 环境中涉及到的几个重要内容

3.3.1 XML Schema 与 GML

XML 是 SGML 的子集,XML 由于易学已经成为网络间通信的通用语言,GML 是 XML 语言的扩展,主要用于传输空间地理信息数据;为实现互操作性,网格服务 GIS 系统平台使用扩展的 GML 语言来沟通不同平台、编程语言和组件模型中的不同类型系统[8]。为了更好地达到这一目的,将不同层的定义放到相互独立的 GML 命名空间中,这些名字必须能够被用户和程序员的服务模式直接引用,同时要将尽可能多的逻辑放到 GML 类型定义中去,以更好地方便用户对服务的调用。这样不管是 GIS 标准还是网格标准均以 GML 方式来定义,可由 SOAP 协议封装计算请求,通过 GML 语义解析,跨平台完成网格 GIS 服务的互操作,以便于各部件间相互传递消息。

3.3.2 简单对象访问协议(SOAP)

简单对象访问协议(SOAP)提供了标准的 RPC 方法来调用网格服务组件。用户请求服务时按照 WSDL,SOAP,XML,HTTP 的顺序进行封装,而用户在接收到信息时再按照相反的顺序剥离出所需的信息[9]。

3.3.3 网格服务描述语言(WSDL)

网格服务描述语言是基于 XML 的组件描述,服务描述为调用 Web Services 提供了具体的方法。WSDL 是一个基于 XML 格式的定义服务的实现和接口的基础标准。

3.3.4 网格服务发布(UDDI)

在这一层次,服务提供者能够直接向服务客户端发送本地关于 GIS 的 WSDL 文档,同时,服务提供者也可以选择将 WSDL 文档发布到本地的 WSDL 注册库,或是公共/私有的 UDDI 注册中心。服务客户端可以通过这些注册库来获得相关的关于 GIS 的 WSDL 描述文档。

3.3.5 对 Web Service 的扩展

Web Service 最有用特点之一就是其能够自我描述。这意味着一旦你定位了 Web Service,你能要求其“描述自己”并告诉你如何操作和使用它。

Web Service 由于使用标准的 XML 语言因而是平台无关、语言无关的;通常在网络上传递的数据都是无状态的,在 GIS 系统中信息交互需要的是有状态的服务,因此在 GT 中对 Web Service 进行了扩充,将无状态的 Web Services 扩展为有状态的,因此 Globus Toolkit 4 的核心构件是 GT4 WSRF Core。在扩展的 Web Services 中为实现数据传输的质量,最通常的 QoS 要求网格服务能够在服务循环中恢复到正常状态,因此需要一种服务恢复机制,

故在 GIS 网格服务中采用在 Web Service 应用描述中设置状态检查点, 遇到故障点及时恢复即可实现高质量的传输要求, 在 GT4 中通常也采用 DPS(Date Replication Server)。

3.3.6 安全保障

网格环境下的 GIS 体系结构的安全保障主要依靠 Globus Toolkit4 的安全机制来保证其安全; 默认配置中, 所有用户和服务实体都拥有一个 X.509 公钥证书。

4 应用实例

实验系统中使用了四台机器, 均为 Dell PowerEdge 2800, 其中一台机器作为 CA Server(certification authority)作为权限认证服务器, 另外三台机器作为不同的 Grid Server, 操作系统全部安装 Red Hat9, 其上分别安装了 SGE 和 Globus Toolkit4。为演示方便, 取一幅图像的 3 个图层, 分别为各个地域边界图层、公交线路信息图层和大型建筑物图层信息, 分别定义数据结构, 来自一幅图像的三图层用相同的 ID 作为一幅图像的唯一识别标志, 并分别存储在 3 个 GRID 服务器上, 各 GRID 服务器向网格信息中心注册相应的信息。系统的目标是将三台 GIS 服务器上的图像数据的相关信息取过来, 然后叠加还原成原始图像。

由于 JAVA 具有平台可移植性、面向对象等优点, 且为 GT4 支持语言^[10], 故选用 JAVA 作为开发平台, 将编写好的 GIS 部署成网格服务(此处没采用任何 GIS 开发工具作二次开发), 以 HTML 方式提交就可以得到结果。系统几乎是同时从三台机器上分别取得 3 个相关的图像信息, 最后描绘成一幅图像; 如果图层间存在依赖关系则程序员必须要自己写服务, 否则会运算失败, 但是如果要求的信息之间没有直接的依赖关系, 此时将是网格发挥其最佳性能的时候, 所以在写网格服务时务必注意。该服务中的网格 GIS 接口定义如下^[11]:

```
package gt3tutorial.core.factory.impl;
public interface ComImages
{
    public Image GetImage(InetAddress ip, int pictID);
    public Image ComposedImages(Image[] images);
}
```

该服务接口的作用是根据系统提供参数获取一个图层, 而下一个接口的作用是实现将各图层有机的合成, 生成一幅完整的图像信息, 明显这两个接口的图像数据的取得及运算均已经分配到这三台机器上完成了。

只要按照生成 WSDL 描述接口, 产生 stubs, 实现服务, 部署描述器, 编译, 运行即可得到结果, 具体的运行部署方法可参考文献[11]。

5 总结比较

与普通 GIS 各种性能比较, 典型的特点就是系统需要的数据是来自异构的 GIS 服务器, 并且随着处理数据量的增加, 网格处理速度会呈现出快速的增长。在开始的时间由于 GT4 及各机器间通信的开销, 其强大的运算功能并不明显, 这时单机 GIS 却比网格 GIS 要快些, 而且随着时间的推移其运算速度还在下降。但随着时间的推移, 网格 GIS 多台机器均参与运算, 且强大的运算功能及宽通道的数据通信量优势完全发挥出来, 其处理的数据量也比单机要快很多倍。

参考文献:

- [1] 王 铮, 吴 兵. GridGIS—基于网格计算的地理信息系统[J]. 计算机工程, 2003(4): 37-40.
- [2] Foster I. WHAT IS THE GRID A THREE POINT CHECKLIST [EB/OL]. <http://www.gridtoday.com/02/0722/100136.html>, 2002-07-22.
- [3] Segal B. The European DataGrid project, First Presentation At Data Mining Workshop—CSC Scientific Computing—Otanienmi [EB/OL]. <http://web.datagrid.cnr.it/pls/portal30/docs/902.PDF>, 2005-04-01.
- [4] 郁志辉, 陈 渝, 刘 鹏. 网格计算[M]. 北京: 清华大学出版社, 2002.
- [5] 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.
- [6] 任健伍. GRID GIS 关键技术研究[D]. 南京: 南京师范大学, 2004.
- [7] 张月卓. GT3 体系结构介绍[D]. 北京: 北京信息工程大学, 2003.
- [8] Shen Zhanfen. Architecture design of grid GIS and its applications on image processing based on LAN[J]. Information Science, 2004, 166: 1-17.
- [9] 方金云, 何建邦. 网格 GIS 体系结构及其实现技术[J]. 地球信息科学, 2002(3): 37-40.
- [10] Grid Computing with the IBM Grid Toolbox REDBOOKS [Z]. 2003.
- [11] Foster, Kesselman C. The GRID: Blueprint for a New Computing Infrastructure [DB/OL]. <http://www.gridforum.org>. 2003-12-06.

(上接第 220 页)

- 业职业技术学院学报, 2005, 4(1): 49-52.
- [3] 张郑擎. SAN 技术及其在电信网络中的应用[J]. 现代通信, 2005(2): 23-25.
- [4] 刘福明, 李 明. NAS 和 SAN 技术的特点及发展[J]. 重庆教育学院学报, 2003, 16(3): 23-27.
- [5] 陈学锋, 陈颖行. 大型企业 SAN 数据存储方案设计[J]. 微机发展, 2003, 13(5): 88-92.