

网格 GIS 系统设计及其在遥感图像处理上的应用*

沈占锋¹, 骆剑承¹, 陈秋晓¹, 蔡少华^{1,2}, 郑江¹, 孙庆辉^{1,3}, 明冬萍¹

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室; 中国科学院研究生院 北京 100101; 2. 解放军测绘信息中心, 北京 100088; 3. 解放军信息工程大学 测绘学院, 河南 郑州 450052)

摘要: 随着地学数据特别是遥感数据与日剧增, 许多数据却因为数据量大、处理困难而得不到有效利用。因此, 如何更好地发展计算机网络技术, 使得海量数据能够得到更好的共享与应用成为了困惑广大计算机专业及地学工作者的一大难题, 网格计算提供了此问题的解决途径。以网格计算及其在 GIS 的应用为主线, 首先分析了网格技术的特点, 并进一步结合其在 GIS 中的应用给出了网格 GIS 的应用前景。同时以局域网为单位, 结合网格设计的思想, 借助中间件技术, 讨论了基于局域网络的网格 GIS 系统的实现技术, 并列举了所需的几项关键技术。特别就当前 GIS 所面临的共性问题进行了分析, 在分析现有网格技术发展的基础上, 指出了基于 GridGML 描述语言的 Web Services 技术的中间件分布式架构是 GIS 软件发展的一个主要方向及解决当前 GIS 领域中所存在的问题的最佳方式, 并就项目的图像分割任务完成情况讨论了其实现过程。

关键词: 数据共享; 网格计算; 网格 GIS; 网格中间件; GridGML; Web Services

中图分类号: TP391.41 文献标识码: A 文章编号: 1001-3695(2004)08-0122-04

System Design of Grid GIS and Its Application on Remote Sensing Image Processing

SHEN Zhan-feng¹, LUO Jian-cheng¹, CHEN Qiu-xiao¹, CAI Shao-hua^{1,2},
ZHENG Jiang¹, SUN Qing-hui^{1,3}, MING Dong-ping¹

(1. National Laboratory of Resources & Environmental Information System, Institute of Geographic Science & Natural Research, Chinese Academy of Sciences; Graduate School of the Chinese Academy of Sciences, Beijing 100101, China; 2. The PLA Surveying & Mapping of Information Center, Beijing 100088, China; 3. School of Surveying & Mapping, the PLA Information Engineering University, Zhengzhou Henan 450052, China)

Abstract: With the development of computer science and its relative subjects, Internet technology has entered a period of high-speed development, and geo-information science has developed itself with the development of computer science and Remote Sensing (RS). But with the data gained becoming larger and larger, many data can't be fully used for the difficulty of data duplicating and data storage. So how to develop Internet to fill the need of colossal data sharing becomes a difficult problem for many computer experts to solve, the idea of grid computing then appears. This paper centers on grid computing and its application on GIS, after analyzing the common problem and the development of grid GIS, this paper points out that the distributing architecture which is based on GML Internet language and Web Services technique and middleware is a main developing direction of current GIS software and the best way to solve the problem it faces, then its implementation process is discussed.

Key words: Data Sharing; Grid Computing; Grid GIS; Grid Middleware; GridGML; Web Services

1 引言

网格计算是近几年来计算机学者们为解决当前资源分配与利用不均而提出的一种先进的资源共享模型。随着计算机技术的发展, 计算机所面临的任务也越来越艰巨, 其中许多涉及大量的数据处理与网络计算的问题, 在人们试图改善现有基础设施的同时(如宽带网络、计算机硬件), 人们不禁考虑: 能不能有一种方法将现有的资源“共享”起来。这里的“共享”并不同于现有网络的共享, 而是按人们意愿将各种资源, 包括计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等

完全共享^[1], 用于提高现有计算机在海量数据处理方面的能力, 同时也可以很大程度地减少数据的存储冗余, 网格计算便由此产生。

随着越来越多的专家学者致力于网格计算的研制与开发工作, 网格计算已从其概念阶段发展为现在的设计与实现阶段。目前国际上已经出现了一些网格计算的雏形, 其中一些已投入试运行阶段, 比较著名的有: 美国自 1997 年底的“分布式网格”研究项目开始, 先后有 NASA 的 IPG 网格, 美国能源部的 ASCI 网格, 美国国防部的 GIG 网格试验运行; 欧洲共同体的 Euro 网格和 Data 网格也投入了研究实验; 我国科技部在“九五”开展了国家高性能计算环境(网格)的建设和关键技术的研究。“十五”期间国家科技部加大了对网格技术研究和推广的力度, 目标是突破网格关键技术, 建立网格计算技术标准, 将网格计算技术应用到行业和企业应用中, 建立行业和企业应用

收稿日期: 2003-09-18; 修返日期: 2004-05-21

基金项目: 国家“863”基金资助项目(2002AA135230); 国家自然科学基金资助项目(40101021)

网格, 进一步加强全社会共享的国家高性能网格计算环境的建设, 推动我国网格产业的形成和发展^[2]。而网格在各领域中的应用也进入了探索性研究阶段。

Internet GIS 是利用 WWW, CGI, Java 等技术来提供开放的、基于标准的综合性服务计算机网络系统^[4], 它将处理储存通信能力移入网络, 具有 Frame 和 C/S 的优点; 同时信息高度分散, 通过 HTTP 可访问连接 URL 资源和应用, 共享程度高, 可伸缩扩展性强, 具有高度开放性和灵活性, 同时具备高速率和安全性。Web GIS 和本文讨论的网格 GIS 都是 Internet GIS 的不同表现形式。

网格 GIS 与 Web GIS 都强调共享, 强调 Internet GIS, 所以有很多人容易把网格 GIS 与 Web GIS 的概念混淆, 实际上两者的侧重点不同: 网格 GIS 是在有网格计算的架构基础上的 GIS 应用, 是在网络互连的基础上通过网格节点的普遍的、各种资源的共享, 而 Web GIS 主要侧重于利用现有网络来实现数据的共享及互操作; 网格 GIS 的基础架构是网格计算, 而 Web GIS 的基础是现有网络, 实现网格计算的全球互连则是比现有网络更复杂、功能更强大的“下一代网络”, 它的实现必须借助众多学科, 而其功能也必将更强大。本文后面部分将讨论其具体实现。

数字地球战略被认为是解决地球信息科学发展的海量数据处理的一个较好途径, 它试图通过地球科学、系统科学与信息技术的交叉融合, 以数字的形式来表示地球。也就是说, 数字地球战略与全球可持续发展战略是在计算机技术、网络技术、虚拟现实技术和 3S 技术等支持下, 对全球海量数据进行综合分析、管理和应用的超级系统, 它具有全球性、共享性、动态性与可视性等优点^[5]。而从实现技术来说, 其实现技术包括信息高速公路和计算机宽带高速网络技术、高分辨率卫星影像、空间信息技术、大容量数据处理与存储技术、科学计算以及可视化和虚拟现实技术等, 而实现这些技术都需要解决巨大的数据量和计算量。它力图把地球上一切与地理位置有关的信息, 用数字的形式进行描述并存储成为丰富的资源, 通过网络进行共享, 从而为全社会服务^[6]。数字地球概念实际上是网格技术在地球信息科学领域的一种体现形式。迄今为止, 只有借助网格计算技术才能够很好地实现这些关键技术, 同时, 网格计算被认为是解决数字地球问题的最好方法。

2 网格计算及其实现技术

2.1 网格计算及网格中间件

网格计算是在 20 世纪 90 年代由 Ian Foster 提出的, 其目标是将广域网上一些计算资源、数据资源和其他设备等互连, 形成一个大的可相互利用、合作的高性能计算网。网格计算环境的特点是可连接已存在的大计算中心, 对各种类型的设备进行分布式协同并行处理来为用户服务^[7]。目前美国 Globus 是全球致力于网格底层开发最著名、最有影响力的公司之一, 他联合了 IBM, Microsoft, HP 等公司, 构建了 GT2, GT3 网格开发环境, 并对网格环境中的各部分功能有了比较详细的规定, 形成了 Globus 的“五层沙漏结构”^[8]。在此基础上, 结合最新的 Web Services 技术, 提出了以服务为中心的下一代的网格体系结构(Open Grid Services Architecture, OGSA)。OGSA 包括两大关键技术, 即网格技术和 Web Services 技术。本次网格 GIS 架

构设计是在 OGSA 的架构基础上, 应用 Web Services 技术, 基于局域网的原型设计。

中间件一般是指运行在客户机或服务器系统上的一种独立的系统软件或服务程序, 是一种新型的软件设计模式。在实际应用中, 它可以实现多种功能, 如提供远程进程管理、空间信息资源分配、信息存储与访问、系统安全登录和认证、系统安全或服务质量监测等^[6,9,10]。网格中间件就是在网格计算环境下的中间件。网格中间件除具有以上中间件所具有的功能外, 还有其特定的特点和功能。在不同的网格计算环境中, 由网格中间件所承担的任务的不同, 可将网格中间件划分为消息中间件、对象中间件、安全中间件等, 再进一步可分为注册中间件、复制中间件、任务管理中间件、错误处理中间件等。

2.2 网格实现技术

2.2.1 Web Services 与 WSDL

Web Services 就是建立可互操作的分布式应用程序的新平台, 即在进行网络间通信时, Web Services 通过向外界提供的 API 来完成网络间的各种操作^[11]。应用 Web Services, 我们能够使用业务描述进行描述, 发布到业务注册库, 通过标准机制进行查询, 对网络中的 API 进行调用, 与其他服务结合等^[12]。

WSDL(Web Service Description Language) 是基于 XML 的语言, 用于描述 Web Services 及其函数、参数及返回值等的语言。它通过在服务提供者和服务请求者之间交换的消息来描述 Web 服务^[12]。应用 WSDL 描述语言, 用户可以动态地发布、查找及绑定 Web Services。

Web Services 符合网格计算中应用要求的平台无关性、位置无关性、简单易用性, 且为其提高安全性提供了底层保证。可以说, 在本次基于 OGSA 的网格架构设计中, Web Services 技术是解决网格计算问题的最好途径。

2.2.2 XML 与 GridGML

XML(eXtensible Markup Language) 是 W3C 组织于 1998 年 2 月发布的标准, 是 SGML(Standard for General Markup Language) 的子集, 是目前用于 Internet 网络间通信的一种通用语言。XML 语言具有灵活的可扩展性、自描述性、开放性、简单易学等特点, 同时 XML 把数据和形式进行了分离, 更利于数据的传输和快速定位。

GML(Geography Markup Language) 是 XML 语言的扩展, 是由 OGC 公司组织并发行的可用于 GIS 网络间通信的一种高效语言之一。GML 语言可用于存储和传输空间地理特征的属性信息和几何信息, 它强调的是空间信息的传输, 而不强调地理特征的可视化。GML 语言符合 OpenGIS 模型标准, 其内定义了大量的空间体在网络间通信的协议与范式, 并对元数据和空间实体模型(矢量与栅格)及各种坐标系统有了精确的定义^[13]。GML 也具有 XML 的特点, 可扩展性强, 用户可以根据实际情况规定自己的 GML 规范。

基于 GML 语言的特点, 结合网格通信的需要, 规定了适合网格环境的 GML 语言通信语言, 即为 GridGML 语言。GridGML 语言尽可能地保持 GML 的优点, 同时进行相应地扩展, 以利于 GIS 中各种资源的描述和传输。

2.2.3 SOAP 与 UDDI

SOAP(Simple Object Access Protocol) 是网际通信中用于

交换 XML 编码信息的轻量级协议。通过 SOAP 的封装(Envelope)、编码规则及远程调用(RPC),结合 GridGML 语言,即可实现此网络 GIS 中的各种网络通信及中间件间通信的任务。

UDDI 提供了在 Web 上描述并发现服务的框架。应用网络 GIS 中的 UDDI,可以快速、准确地查找、发布及绑定服务^[14]。

2.3 网络 GIS

网络 GIS 指应用网络计算的环境与思想来解决 GIS 的问题,其侧重点在于 GIS 领域中涉及大量数据共享与交换及海量信息处理的应用。本次应用以大图像分割为例,列出了网络 GIS 要解决的问题及途径。

3 基于OGSA 的网络GIS 体系结构设计及实现

3.1 网络 GIS 架构设计

网络 GIS 架构是指网络 GIS 的运行体系。在网络系统中,其架构体系决定了整个网络的运行稳定性与可扩展性。网络架构定义了网格内及网格节点间的各种协议及 API,用于指导网络系统及相应应用程序间的操作^[15]。本网络 GIS 架构是以 OGSA 架构为基础,结合 GIS 专业项目的特点,应用中间件技术构建的基于局域网的网络架构。图 1 所示是网络 GIS 运行体系的架构。



图 1 网络 GIS 体系架构

整个系统由五层组成,层与层之间有着明显的层次关系。从底层向上分别是基础层、资源层、控制层、实现层及应用层。基础层包括现有网络基础结构,同时需在此层规定适合网络 GIS 体系的特定协议,用于保证其正常运行;资源层指当前系统可用的各种资源,包括本地资源及异地已注册可利用的共享资源,通过此层实现了资源的共享;控制层是整个系统的核心,它控制着本地系统及远程系统的正确运行,并有协调资源层与实现层的作用;实现层是系统的具体的实现部分,由各种中间件来协助完成,各中间件通过可扩展的特定接口与系统连接;最上层是应用层,由具体的用户应用界面组成,网络 GIS 项目的开发与应用就构架在此层。

在整个系统的设计与实现过程中,应始终保证系统的平台无关性、位置无关性、安全性、易用性与稳健性。

3.2 网络 GIS 架构实现

由图 1 可知,控制层是整个网络系统的核心,它控制着整个系统正确、安全地运行。而在控制层中,任务管理中间件与资源管理中间件又是此层的重要组成部分,可以说,网络系统中负责任务管理与资源管理的部分是决定着整个网络成败与否的关键。下面就本系统中这两部分给予详细的介绍。

3.2.1 任务管理中间件

任务管理中间件负责整个系统中的任务管理与调度,作业分类、分发与合并以及报告生成等任务。图 2 示意了任务管理中间件在网格中的作用。

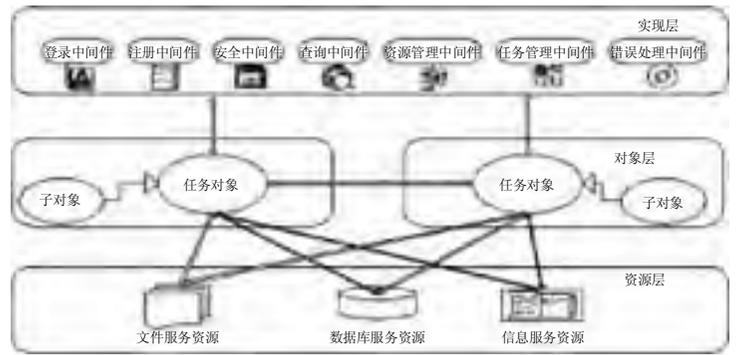


图 2 任务管理中间件与任务对象

从图 2 可以看出,此系统中任务管理中间件的任务是由任务对象来完成的。任务管理中间件在接收到用户的请求后,可由任务管理中间件派出不同的任务对象来完成不同的任务,每个任务对象具有面向对象语言中对象的特点,既可具有继承性、封装性与多态性,每个对象又可派生出他自己的子对象;用户不可以直接对资源进行操作,而必须通过任务对象对资源进行操作;不同的任务对象可完成不同的任务,在任务对象生成时即已规定了此任务对象的任务,如任务与资源的查询,中间结果的显示,暂停、重启、中止任务等;不同的任务对象亦可具有不同的性质,如具有智能性和可移动性的 Agent 等。

任务管理中间件对象与网格节点工作原理如图 3 所示。当进行图像分割时,首先由任务管理中间件对象生成相应的任务管理子对象,并将它们分别绑定到对应的网格节点上;任务管理子对象通过调用核心函数 GetGridInformation(...)来获取各已注册可利用网格节点的信息(包括计算机状态、CPU 利用率、存储状态等),再通过核心函数 CSegmentation(...)来绑定远程的连接,并传递图像数据、分割信息及要求;节点计算机会在相应要求下(设定为 $\frac{2}{3} \times \text{initInterval}$ 时间内)向任务管理处返回消息 SendGridReport(...),报告任务的进展状况;任务管理中间件对象根据其任务完成状态及连接状况(若与其失去连接则没有相应消息的返回)决定任务的进一步管理,并发送消息 SendGridMessage(...)进行处理。



图 3 任务管理中间件对象与网格节点工作原理

3.2.2 资源管理中间件

资源管理中间件的任务包括资源的描述方式,资源的注册、查找、应用绑定等。

图 4 示意了资源管理中间件的实现及流程。不同用户可拥有不同的权限。用户在登录后,可以将其资源声明为共享资源(分为局部性共享资源与全局性共享资源),同时在资源中心处进行资源注册,其资源描述语言 RDL 基于 GridGML 语言,此资源描述语言又是资源查询的重要机制与依据;资源注册中心的资源是按层次结构目录组织的基于数据库与 LDAP 形式的资源管理机制与索引服务结合的形式,即两者相结合以管理分布式的任务、计算资源及其他的各种资源,这样可以方便对任务的查询和控制,方便不同应用的信息交流,同时为不同地理位置的计算机系统之间的任务迁移、任务调度、负载均衡提

供技术支持。

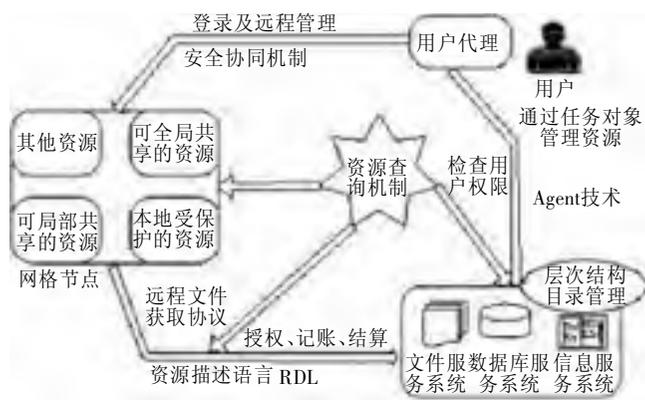


图 4 资源管理中间件的实现及流程

信息资源的注册机制可采用管理员授权的形式向网格服务器进行资源注册,资源描述方式统一为 GridGML 语言形式,以利于资源的描述、查找与定位;语言描述方法采用基于 GridGML 语言的信息交换形式,可满足异构性、分布性和动态性等因素存在的要求。用基于 GridGML 扩展语言可实现网格计算系统内部的信息查询、信息交流和信息控制,而且在不同的系统之间通过 XML 进行资源信息的交流和控制,可有效地解决信息表示、信息交换和信息控制的问题。资源信息的交换可通过 Web 页间统一格式的 SOAP 消息进行传递。

4 系统实例分析

图 5 示例了此次基于局域网的网格 GIS 架构的设计与实现。

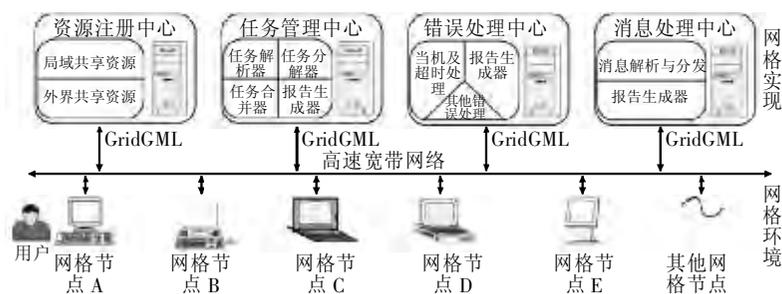


图 5 基于局域网的网格 GIS 的实现

在局域网内利用某些计算机做成资源注册中心、任务管理中心、错误处理中心、消息处理中心等,并按图 5 构建好网格环境后,用户即可登录到网格环境进行任务提交。首先各网格节点将自己的可共享资源送到资源注册中心处注册(并定期更新),此过程中消息描述与资源描述方法基于 GridGML 语言格式。假设用户从节点 A 登录到网格环境,并向系统提交一大图像分割任务,图像存储在节点 B 处,分割过程由节点 C、D、E 应用相同的分割算法来共同完成。任务管理中间件接收到用户提交的任務后,进行任务的解析、分解与分配,查找可利用资源及网格节点后,向节点 C、D、E 发送任务。节点 C、D、E 根据任务的要求进行分割,并发送消息及任务完成情况给任务管理中间件,任务管理中间件根据其任务完成情况进行相应的处理。如果任务比上次完成得多(没有当机),则向其发送“Keep Alive”消息,并使其计数器归零;否则进行其他错误处理。按此方式任务完成后,节点 C、D、E 将任务结果连同任务完成报告返回,再由任务管理中间件将结果合并提交给节点 B(按用户任务描述中要求的将结果提交到某处,此例中为节点 B),并将任务完成报告提交回节点 A,同时中断与节点 C、D、E 机的任务连接。

5 结束语

网格 GIS 是当前 GIS 研究领域的一个新的研究问题。在

给 GIS 领域提供了很好的应用前景的同时,网格 GIS 也给我们提出了更多的挑战。网格 GIS 是一项跨地区、跨平台、跨专业的研究项目,需要付诸很大的投入。针对 GIS 专业的特点及当前网格计算的发展,我们提出了基于 GridGML 语言的资源描述语言作为网格 GIS 网络间通信的基本语言,并应用 Web Services 技术,以中间件结构为软件架构来组织整个网格 GIS 体系,根据 GIS 所处理的具体问题及所需的数据,组织规定相应的协议体系,以实现信息共享和资源共享,为 GIS 行业服务。

此网格 GIS 架构是在 OGSA 架构的基础上,应用 dotNET 及 Web Services 技术而构建的网格环境。针对 GIS 中不同的任务,均有着不同的任务描述方法,因此,任务描述方式就显得尤为重要。如何采用先进的方法手段,使得在任务提交时的任务描述“统一化”,是我们仍需努力解决的问题之一,同时,也将是衡量网格 GIS 系统的智能性、自主性的重要指标,它同时也关系着网格 GIS 系统的发展速度问题。

参考文献:

- [1] Foster, C Kesselman. The Grid: Blueprint for a New Computing Infrastructure[M]. Morgan Kaufmann Publishers Inc., San Francisco, CA, 1999.
- [2] 肖依, 卢锡城, 王怀民. 网络计算的四种形式[N]. 计算机世界报, 2002(40, B1~B4).
- [3] 陈述彭, 鲁学军, 周成虎. 地球信息科学导论[M]. 北京: 科学出版社, 1999.
- [4] 修文群. 网络地理信息系统[J]. 中国图像图形学报, 2002, (6): 610-617.
- [5] 黄解军, 潘和平, 万幼川. 数字地球战略与数字城市建设[J]. 中国软科学, 2002, (7): 85-88.
- [6] 骆剑承, 周成虎, 蔡少华. 基于中间件技术的网格 GIS 体系结构[J]. 地球信息科学, 2002, (3): 17-25.
- [7] an Foster. What is the Grid? A Three Point Checklist[EB/OL]. <http://www.gridtoday.com/02/0722/100136.html>.
- [8] an Foster, Carl Kesselman, Jeffrey M Nick. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration[EB/OL]. <http://www.globus.org/research/papers/ogsa.pdf>.
- [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, (4): 36-41.
- [10] 张志檀. 中间件——技术·产品·应用[M]. 北京: 中国石化出版社, 2002.
- [11] asser Shohoud. Web Service 简介[J]. 程序员, 2002, 28-30.
- [12] 李安渝, 等. Web Services 技术与实现[M]. 北京: 国防工业出版社, 2003.
- [13] pen GIS Consortium Inc. OpenGIS Geography Markup Language (GML) Implementation Specification, version 2.1.1[EB/OL]. <http://www.opengis.net/gml/02-009/GML2-11.html>, 2002.
- [14] arsten Januszewski, Ed Mooney. UDDI version 3 Features List[EB/OL]. http://uddi.org/pubs/uddi_v3_features.htm.
- [15] he Globus Project™ Argonne National Laboratory USC Information Sciences Institute. Grid Architecture[EB/OL]. <http://www.globus.org>.

作者简介:

沈占锋(1977-), 男, 黑龙江大庆人, 博士生, 主要从事网格计算与分布式计算、网格 GIS、空间计算与图像理解等方面研究工作。