

大型分布式地理信息系统的技术与发展

陈斌 方裕

(北京大学计算机系, 北京 100871)

摘要 随着地理信息系统和网络技术的发展,空间数据的处理正深入到各个专业应用领域,大型分布式地理信息系统对于空间数据的共享和互操作,以及专业领域应用系统集成具有重要的意义.本文通过对大型分布式地理信息系统的结构分析,介绍了硬件环境、系统软件和分布计算平台对大型分布式地理信息系统的影响,以及适应大型分布式地理信息系统空间数据共享和互操作发展要求的空间数据模型、海量空间数据库和空间数据操作服务技术,以及应用系统集成技术.

关键词 地理信息系统 分布式 OpenGIS 互操作 开放系统

中图法分类号: P208 TP316.4 文献标识码: A 文章编号: 1006-8961(2001)09-0861-04

Technology and Development of Massive Distributed Geographic Information System

CHEN Bin, FANG Yu

(Department of Computer Sci. & Tech., Peking Univ., Beijing, 100871)

Abstract With the developing of GIS and network technology, handling of geospatial data has been integrated into all sorts of special domains. The massive distributed GIS has great meaning to share and interoperation of geospatial data and integration of domain application. This paper analyses the structure of massive distributed GIS, and introduces impacts from development of hardware environment, system software and distributed computation platform. For the spatial data share and interoperation issues, several key technologies such as spatial data model, massive spatial database and spatial operation service are introduced. OpenGIS specification from OpenGIS Consortium is strongly recommended to the coming massive distributed GIS. At the end of the paper, application integration technology is emphasized.

Keywords GIS, Distribute, OpenGIS, interoperation, Open system

0 引言

随着计算机技术的普及,地理信息系统(GIS)的应用范围也越来越广泛,从最初地学意义上的测量和统计,发展到处理一切与地理空间位置相关的信息.地理信息系统技术与一些需要从地理空间获取数据和信息的专业领域结合,产生了许多基于空间信息分析和管理的信息系统.地理信息系统的应用覆盖了工业、农业、交通运输、国防等许多领域,几乎涉及人类生活的各个方面.

值得注意的是,随着网络技术的发展,一些大型

应用系统工作方式发生了改变,其中处理的海量空间数据也从原来的集中、独占走向分布、共享.如何通过标准的分布计算平台,对海量空间数据库进行高效、可靠的管理和使用成为当前地理信息系统技术面临的重大课题.

随着空间信息基础设施的进一步完善,空间数据的处理也逐步从科学计算型过渡到决策支持型,这就要求构建包括海量空间数据处理、各专业领域业务处理和跨专业领域综合业务处理在内的大型分布式地理信息系统.大型分布式地理信息系统在测绘、国防等大型关键性业务处理中将起举足轻重的

作用,同时也是构造“数字地球”、“数字城市”系统的基础性技术。目前对大型分布式地理信息系统关键技术的研究方兴未艾,也有许多利用当前的分布式计算技术来实现互操作和分布式管理的工作^[1]。

1 重要特征和系统结构

大型分布式地理信息系统的关键技术在于实现基于网络的无缝的海量空间数据共享和互操作,以及对地理操作服务的共享和分布处理。为此,在大型分布式地理信息系统中,开放性非常重要,它不仅要适应当前对海量空间数据处理需求,还要继承已有的空间数据和成果,同时也能够容纳将来空间数据处理的发展。所以,大型分布式地理信息系统应该具有如下的一些重要特征^[2]:

(1) 可互操作 系统提供了空间数据和空间操作的标准接口。这种接口既能够支持孤立结构,也能够支持网络访问,使得无论在系统内部的应用程序之间或是系统的内部应用程序和外部应用程序之间能够通过标准接口实现空间数据的共享和交换。同时,互操作能力还应该支持客户/服务器和对等实体两种分布计算模式。

(2) 支持语义表达和转换 真正意义上的空间数据共享或者互操作,不仅在于通用的数据格式,更重要的是对空间数据中所包含的意义的共同理解。系统应该提供一种语义的表达和转换机制,使得具有不同应用背景的系统 and 用户能够方便地交换空间数据而不会造成语义上的误解和损失。

(3) 普遍适用性 系统通过定义标准的接口和协议,使得系统内外的应用程序,无论是属于什么专业领域,都可以方便地访问系统中的空间数据和空间操作;

(4) 可移植性 系统应该独立于具体的软件环境、硬件平台和网络体系;

(5) 可伸缩性 系统的可伸缩性表现在功能的增减和处理数据量的大小上,可以通过实现相对独立的功能组件和对系统应用组件进行重新配置来实现;

(6) 可扩展性 系统可以随着新的空间数据类型和处理方法的发展而发展,具有容纳新的空间数据处理技术和新的分布计算平台技术的能力;

(7) 兼容性:以往的地理信息系统产生了大量空间数据和处理结果,系统应该能够与旧有的空间数据保持兼容。

大型分布式地理信息系统具有如下的典型系统结构:

应用系统层:实现各种专业领域内和跨领域的通用业务处理		
领域服务层:面向专业领域的空间数据操作		
通用服务层:通用空间数据操作服务		
数据表示层:空间数据模型,海量空间数据存储与组织		
分布计算层:实现分布计算		
系统软件层:操作系统、数据库管理系统及其他系统软件		
人机交互设备	网络设备	计算硬件平台

2 硬件环境和系统软件

现实环境中存在着各种各样不同的软硬件平台环境,大型分布式地理信息系统必须能够适应这些环境。由于操作系统等系统软件已经对具体硬件设备进行了抽象和管理,可以认为硬件环境的变化将不会影响到大型分布式地理信息系统的结构和本质技术。

但是硬件环境的发展十分迅猛,人机交互技术和网络技术的发展也将对大型分布式地理信息系统提出新的要求。如随着各种虚拟现实设备的应用和普及,要求系统能够提供逼真的、三维的、可以实时漫游和交互的场景。而高速网的发展,则要求系统能够容纳更多的并发应用和更丰富的媒体表达形式。嵌入式环境和移动环境则要求系统具备强大的可伸缩性,能够依据应用环境进行剪裁,完成必要的空间数据处理功能。

3 分布计算平台

大型分布式地理信息系统应该能够适应不同的分布计算平台,包括面向过程的 DCE (Distributed Computing Environment) 规范以及面向对象的 CORBA (Common Object Request Broker Architecture) 规范和 DNA (Distributed interNetwork Architecture) 规范。这些分布计算平台提出了远端过程调用 RPC 的基础协议,同时,在面向对象的分布计算平台中,还提供了远程对象生成、调用、查询、代理、安全和事务处理等机制。

分布计算平台扩展了客户/服务器的体系结构,使得服务需求者不需要了解服务提供者的位置信息和配置信息,达到完全透明的计算分布。这对大型分布式地理信息系统提供了标准的分布计算基础。通过分布计算平台的接口,系统内部或者外部的应用就可以无需了解空间数据资源和空间操作计算资源的所在,而得到完整统一的空间数据服务。

4 空间数据模型

空间数据模型的选择和建立关系到大型分布式

地理信息系统内部以及内部和外部之间进行数据交换和数据共享的程度,所以空间数据模型的标准化非常重要,必须能够容纳旧有的空间数据,也能够具备很好的可扩展性,适应将来新的空间数据对象和数据操作。

目前广泛应用的各种地理信息系统中所采用的空间数据模型基本上都沿用了图库、图幅、图层、地理对象、几何对象这样的组织结构,从本质上说是面向纸质地图的数字化处理的,而不是面向客观存在的空间实体及其时空关系,具有比较突出的缺点^[3]:

(1) 以图层作为空间数据处理的基本单位,人为割裂地理空间,其处理能力,尤其是对大范围空间的数据处理能力较弱,跨图层、跨图幅的处理尤为困难;

(2) 面向平面二维结构,对于三维空间数据的处理仅限于表面造型,没有实质上处理三维空间数据的能力;

(3) 静态空间数据的处理模式不能够适应动态时空关系的处理;

(4) 对影像数据的管理和处理能力较弱,不能适应目前大规模快速数据获取技术的发展;

(5) 空间数据模型标准化工作较少。

这样,在大型分布式地理信息系统中的空间数据模型就需要突破现有空间数据模型的限制,具备描述连续地理空间、真三维空间、动态时空关系、较强的影像处理能力和标准化程度高的特征。

Open GIS 协会(Open GIS Consortium, OGC)从1997年开始,提出和倡导了开放地理数据互操作规范(Open Geodata Interoperability Specification, OpenGIS), OpenGIS 规范中提供了一个与程序设计语言、硬件设备和网络环境无关的开放的地理数据模型(Open Geodata Model, OGM)^[4]。这个地理数据模型以易于理解的地理要素(Feature)为核心,以面向对象的分析方法规定了地理要素的定义、空间属性、语义属性和与其他地理要素的关系,以及地理数据模式等元数据的定义和表示。OpenGIS 规范经过历年的发展,已经得到包括有微软、IBM 等多家计算机厂商及主要 GIS 厂商、政府机构、大学、集成商的广泛支持^[5]。

开放地理数据模型 OGM 将会得到广泛的应用,大型分布式地理信息系统应该将 OGM 作为空间数据模型设计的基础和重要参照。

5 海量空间数据库

海量空间数据库为大型分布式地理信息系统中

的空间数据提供了存储、组织、检索的软件基础,目前的主要商用关系数据库管理系统(如 Oracle, Informix, DB2, SQL Server 等)相继实现了对空间数据的管理^[3],空间数据的管理手段有了明显的提高。但是由于关系型数据库本身的限制,对空间数据的操纵手段还比较原始和初等。

目前也有系统实现了采用商用 RDBMS 来实现空间对象中几何数据和属性数据的一体化存储和初步的一体化查询,提高了空间数据的操纵能力。但是这种一体化查询和存储还比较初等,表达能力也比较弱。

大型分布式地理信息系统中的海量空间数据库不仅要实现空间数据的一体化存储,解决海量空间数据的索引、组织和分布管理的问题,还需要有一体化的空间查询语言作为操纵接口。由于结构化查询语言 SQL 已经非常成熟,具有面向问题和接近自然语言的良好特征,目前已经有许多对关系型 SQL 语言进行空间扩展的 GSQL 语言的研究。一般来说,空间查询语言必须具有以下特征:

(1) 采用标准、一致的空间数据模型 能够清晰地表达空间数据的层次和组织,以及容纳各种空间对象。

(2) 具备完备的数据操纵能力 包括了数据定义、操纵、查询、事务处理等各种语句。

(3) 具备完备的空间运算能力:能够作为空间数据的运算函数和关系判定表达式,满足空间数据操作和条件表达的需要,如拓扑关系谓词包括相等、相离、相交、相接、包含与被包含和交叠,空间函数包括取坐标、计算长度、计算面积、计算周长、计算外接矩形、计算质心、取得构造对象、计算缓冲区、计算交叠结果等^[6]。

(4) 具备严密的语法和无二义的语义结果 对于相同语句的语义解释应该是唯一的。

(5) 具有可实现性。

6 空间数据操作服务

空间数据操作服务提供了应用系统访问和处理空间数据的接口,大型分布式地理信息系统中的空间数据操作服务可以分成通用的空间数据操作服务和面向每个专业领域的空间数据操作服务。一般来说通用的空间数据操作包括了如下几个类别^[6]:

(1) 空间数据访问 在空间数据库中定位、检索、获取和更新空间数据,可以通过空间对象访问的接口或者结构化空间查询语言(GSQL)来访问。

(2) 地理要素综合:基于指定的规则简化地理

要素的表现形式,使得在简化的同时保持地理要素的必要特征.

(3) 空间信息提取:从遥感影象或者扫描影象中提取地理要素和地形信息.

(4) 空间坐标转换:将地理坐标在不同的空间参照系之间转换.

(5) 空间注记管理:在影象或者地理要素中添加额外的辅助信息,如标记、链接等.

(6) 图符管理:对符号库进行管理.

(7) 影象处理:提供对栅格影象的处理,如:改变大小、亮度对比度、修改颜色、进行各种滤波、增强、处理分辨率等,以及对影象信息的统计,如:计算直方图、计算卷积等.

(8) 地理要素处理:针对用户的需求,对地理要素的创建、质量控制方法、分析处理和显示的服务.

(9) 空间分析:对地理要素所作的空间运算,如缓冲区分析、路径分析、覆盖分析等.

(10) 影象配准:利用数学模型将栅格影象对应到实际的地理坐标中.

(11) 影象地图生成:将影象进行校正和拼接,生成影象地图或者作为其他用途.

(12) 影象比对:检测图象变化、图象之间差异及其差异的显著性分析、基于区域或者模型的差异分析.

(13) 空间信息可视化:提供地理数据到输出设备的接口,包括显示器等临时性显示和打印机等永久显示.

而面向各个专业领域的空间数据处理服务的种类则是由专业领域的空间信息需求来决定的,这些空间数据处理服务构成了大型分布式地理信息系统中的应用平台.应用平台可以用来构成面向领域的应用系统,也可以进行组合构造跨领域的通用系统.

另外,大型分布式地理信息系统还应该提供一些通用的服务,使得应用系统能够具备更加灵活和强大的处理功能,这些通用的服务可以包括:自动化和宏处理、数据压缩、国际化和时间处理、数据交换、工作流管理、安全控制、元数据管理等.

7 应用系统集成

大型分布式地理信息系统的最终目的是能够将空间数据以及空间操作服务无缝地集成到各个专业领域的应用系统中,达到空间数据和空间操作服务的最广泛最深入的应用.依据所应用的专业领域,应用系统可以分成面向专业领域业务处理的专业系统

和适应跨领域业务处理的通用系统两大类.随着地理信息系统的深入应用,越来越多的专业领域业务处理集成了空间数据和对空间数据的操作,如交通、运输、金融、电信、医疗、制造业、仿真等等专业领域.而象字处理、电子表格、报表生成、电子编著、分析设计等跨领域通用的应用系统也对空间信息的表达和操作提出越来越高的要求.

大型分布式地理信息系统应该能够支持各个专业领域应用系统对空间信息处理的集成和访问,并且提供各种层次上的二次开发接口.

参考文献

- 1 韩海洋,龚建雅,袁相儒. Internet 环境下用 Java/JDBC 实现地理信息的互操作与分布式管理及处理[J]. 测绘学报,1999,28(2):177~185.
- 2 方裕,陈斌,薛巍巍. 开放式地理信息系统应用开发平台的理论与实现[J]. 中国图象图形学报,1998,3(10):858~870.
- 3 方裕. 新一代 GIS 软件技术研究[A]. 见:教育部 21 世纪高校 GIS 发展照略研讨会论文集[C],武汉:中国地质大学,2001,1~10.
- 4 Open GIS Consortium, The OpenGIS(tm) Abstract Specification [EB/OL], www.opengis.org, March-1999.
- 5 张家庆. 新一代 GIS 的走向[J]. 遥感信息,1998,S1:6~12.
- 6 方裕,楚放,陈斌. 空间结构化查询语言——G/SQL[J]. 中国图象图形学报,1999,A4(11):901~910.

陈斌 1995 年,1998 年先后获北京大学计算机科学技术系学士、硕士学位,现在北大集软公司工作,并在读北京大学计算机科学技术系博士研究生.主要从事地理信息系统和虚拟现实等方面的研究和开发工作.

方裕 1946 年生,1968 年毕业于北京大学,现为北京大学计算机系教授.长期从事计算机软件工程、地理信息系统、计算机集成制造系统等方面的研究工作.