

基于分布式地质图空间数据库的数据交换策略

杨林¹, 杜小平², 周顺平²

1. 中国地质大学研究生院, 湖北武汉 430074

2. 中国地质大学信息工程学院, 湖北武汉 430074

摘要: 为了解决分布式地质图空间数据库之间的数据共享和一致性维护问题, 在 MAPGIS7.0 面向地理实体的空间数据模型的基础上, 根据地质图空间数据组织和分布的特点设计了一种“横向网格、纵向多级”的分布式多级服务器组网方案, 结合该方案提出了一种以地质空间要素类为单位, 面向服务的数据交换策略。该策略在国家地质调查空间数据库管理系统中的应用表明: 以要素类为基本数据交换单位可以有效地维护分布式地质数据的空间、拓扑和语义的一致性; 以服务的方式操作使数据交换能够跨平台、跨操作系统以及适合异构数据库。实践证明, 这种策略是适合于分布式地质图空间数据库的一体化组织与管理的。

关键词: 地质图空间数据库; 分布式; 数据交换; 发布; 订阅; MAPGIS7.0.

中图分类号: TP311.13

文章编号: 1000-2383(2006)05-0659-04

收稿日期: 2006-05-30

Data Exchange Strategy of Distributed Geological Spatial Database

YANG Lin¹, DU Xiao-ping², ZHOU Shun-ping²

1. Graduate School, China University of Geosciences, Wuhan 430074, China

2. Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China

Abstract: The main purpose of this study is to solve the problem of data sharing and consistent maintenance in distributed geological spatial database. Based upon the spatial data model of MAPGIS7.0 and the characteristics of data distribution, this paper designs a distributed multi-level server network architecture called “grid horizontal and multi-levels vertical”. Under the architecture, a data exchange strategy is presented which regards geological feature class as elementary unit and it is service-oriented. The application of the strategy to nation's geological spatial database system demonstrates that: the spatial, topological and semantic consistence can be maintained effectively by means of exchanging data with geological feature class, and service-oriented operation can deal with data exchange between different operating systems and databases. In conclusion this strategy is suitable for the integrated management of a distributed heterogeneous geological spatial database.

Key words: geological map database; distributed; data exchange; publish; subscribe; MAPGIS7.0.

0 引言

地质调查工作是采集、处理地质数据并向社会提供信息服务的活动, 自从我国开展系统的地质调查工作以来, 国家对各种地下、地面的资源数据和信息进行采集和获取, 获得了可观的纸质数据和电子文档。地质调查获取的空间数据主要包括经过遥感、区域地质调查、矿产资源勘查等获得的空间数据及

其属性信息(中国地质调查局发展研究中心, 2005)。地质图空间数据库的建设即将这些地质调查空间数据按照一定的数据组织方式存储于文件或商用数据库的过程。

这些地质图空间数据库分布在不同的地理位置, 具有同构或者异构的数据库资源。维护这些分布式多源异构的地质空间数据的一致性就变得越来越重要。分布式地质空间数据库不仅要求提供对任意

地理位置上的任意数据库资源完全透明的访问,即具有一定的通用性,而且由于地质空间数据的特殊性,还需要对地质数据的空间性、拓扑信息以及语义信息进行一致性维护。因而急需一种数据交换策略来解决分布式地质图空间数据库群中空间数据的交换问题。

1 地质图空间数据库的特点

1.1 面向地质实体的空间数据组织

MAPGIS7.0 采用面向地理实体的空间数据模型,充分考虑了“类”在空间数据组织中的作用,采用面向“对象—关系”的思想进行整体的数据描述与组织,突破了传统的水平分幅、垂直分层的组织模式,实现了以独立、完整、具有地理意义的实体及空间关系为基本单位进行数据组织和表达(MAPGIS7.0 技术白皮书,2003)。

国家地质图空间数据库的数据组织以 MAPGIS7.0 的地理空间数据库为基础,在地理数据库的概念下,充分利用地理数据库中的新对象进行地质数据的组织和管理。以地质空间要素为对象组织数据,按照“地理数据库—数据集—类”这几个层次组织数据,以满足对不同专题地质空间数据的组织和管理需要。在地理数据库下,地质空间要素类是同一类地质空间要素的集合,是基本的数据对象。

1.2 分布式多级

分布式空间数据库系统是物理上分散而逻辑上集中的空间数据库系统。它使用计算机网络将地理位置分散而管理和控制又需不同程度集中的多个逻辑单位联系起来,共同组成一个统一的空间数据库系统。在分布式空间数据库系统中,被计算机网络连接的每个逻辑单位称为站点或节点(杜小平等,2004)。

从横向上来看,国家地质调查空间数据库由物理上分布在全国各个省、市、县的地质调查空间数据库组成,是一个典型的分布式空间数据库。从纵向上来看,国家地质调查数据库按“中国地质调查局—大区地质调查中心—省级地质调查院”3 个层次分布,又具有清楚明晰的多级特征。

根据地质图空间数据库的分布式多级特征,提出了“横向网格、纵向多级”的分布式多级服务器组网方案(图 1)。

“横向网格”是指在每个层次上都可以构建一个

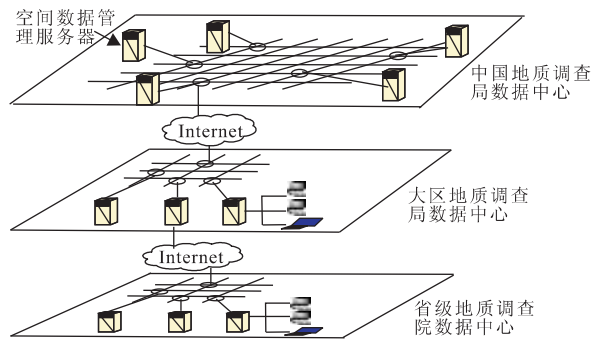


图 1 分布式地质图空间数据库组网方案

Fig. 1 Organization of distributed spatial database

网格,网格中的每个结点对应一台空间数据管理服务器,与下级的一个地质调查空间数据库相对应;“纵向多级”是指 3 级地质调查空间数据库,最高层为中国地质调查局,中间层为大区地质调查中心,最低层为省级地质调查院。如在中国地质调查局数据中心的横向网格中设置多台服务器,每个服务器负责连接中间层的一个大区地质调查中心,大区地质调查中心的横向网格中的每个结点又对应一个省级地调院,通过这种方式将全国的地质图空间数据库组织成一个巨大的分布式地质图空间数据库群。

每个节点为实现本地管理都拥有自己本地数据库,中国地质调查局高层数据库的数据来源于各个大区地质调查中心,各个大区地质调查中心的数据又来源于自己所辖属的各个省级地调院。当底层数据库发生变化时,存放在异地的高层数据库应实现同步修改。

2 数据交换策略

在“横向网格、纵向多级”的组网方案下,为了实时地同步各级地质图数据库中的数据,需要研究一种能够实现多级异构的服务器数据交换策略来实现数据同步,来保证各级数据具有极大的可恢复性,当上级服务器出现故障时,上级服务器可以从所有的下级服务器中取得最新的数据,来建立全局数据,当下级服务器出现故障时,下级服务器可以从上级服务器中下载最新的数据,实现多级服务器之间数据一致性和有效性的维护。

2.1 基于数据库自身的同步策略

目前流行的一些商业数据库的厂商也提供了针对分布式数据库的数据同步策略,这些策略提供了多级服务器间在表一级上的数据复制方案。其数据

复制方案完全利用数据库提供的组件完成系统的部署,不需要其他的技术支持,实现非常简单。但对于地质图空间数据库的多级服务器结构,存在着明显的缺点:

(1)系统扩展不够灵活,如果一台作为订阅服务器,又作为另外一级的发布服务器,则系统层次复杂;

(2)所有环节全靠数据库维护,错误解决和跟踪复杂;

(3)只能实现表级的复制,不能实现面向地质空间要素的空间实体间的复制;

(4)无法实现异构的数据库资源间的数据复制。例如 SQL Server 使用 SQL-DMO(SQL distributed management objects)来进行分布式数据管理以维护多个数据库之间的数据一致性,但是这种方案要求在各成员数据库之上建立一个全局模式,若成员数据库是异构的,则很难建立。

2.2 基于 Web Service 的数据交换策略

由于基于数据库的复制策略不能适应地质图空间数据库的需求,根据地质图空间数据库组织的特点,提出的数据交换方案以面向“服务”思想为指导,以地质空间要素类对象为单位来实现数据交换,采用基于 SOAP 的 Web Service 技术完成地质调查空间数据库多级服务器间数据的一致性和协同维护。

其实现策略是:数据交换通过对地理数据库中地质空间要素类的发布和订阅操作来实现。发布和订阅的实现基于 Web 服务,它们通过部署在广域网中的 Web Service 站点——分发代理和接收代理来完成。其数据交换流程主要分 3 个阶段:首先是发布数据,将数据库服务器需要和其他级别数据中心保持同步更新的数据以地质空间要素类为单位,在分发代理进行注册;其次是订阅数据,用户在广域网上可以查看当前已经发布的地理数据库以及地质空间要素类,并选择所需内容进行订阅;最后是同步数据,当发布地理数据库中的数据发生改变时,相应的发布引擎将负责通知相应的分发代理同步数据,从而维护数据的一致(图 2)。

这种数据交换策略有以下优点:

(1)面向“服务”操作。分布式数据的存取操作采取面向“服务”方式进行,就是把“进行数据存取操作”变为“请求数据存取服务”,“数据存取服务”是所有服务的特例,充分体现“面向服务”的最新设计思想。地质空间数据分布式管理是跨平台的,按照“面向服务”的思想,每个节点上的数据“管理者”必须提

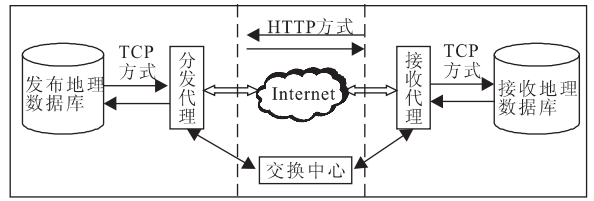


图 2 数据交换机制

Fig. 2 Procedure of data exchange

供“服务”,在“谁管数据谁提供服务”的原则基础上,可解决网格节点之间、父节点与子节点之间、不同平台不同系统之间数据不通问题。应用端只需请求服务而不需直接操作数据,服务端只需提供服务而不需提供数据内部结构。

(2)以地质空间要素类为交换单位。地质空间要素类是地质图空间数据库中的基本数据组织单位,因而发布地理数据库以地质空间要素类为单位。当发布类的地质图信息改变时,订阅端可以以地质空间要素类为单位来更新数据。这种数据交换机制实现了类地质实体一级的语义一致性维护,这有区别于商业数据库间提供的表一级的一致性维护。而且解决了不同商用数据库之间,数据库与文件存储载体之间无法进行同步更新的难题。

(3)采用基于 SOAP 协议 Web Service 技术实现。SOAP 协议是一个轻型分布式计算协议,允许在一个分散、分布的环境下交换结构化的信息,可以实现异构程序和平台之间的互操作性,使应用能够被广泛的用户所访问。分发代理和接收代理为部署于广域网的 Web Service 站点,提供 SOAP 服务接口给广域网。客户机应用程序使用 SOAP 在 HTTP 上远程过程调用发布的服务,从而实现远程空间数据的数据交换。

(4)更加安全。系统的部署充分考虑到了地理数据库的安全性,只将分发代理和接收代理提供的 Web 服务直接暴露于广域网环境中,而发布地理数据库、接收地理数据库、分发代理和接收代理则部署在局域网环境中,采用基于 TCP/IP 协议的 Socket 接口实现局域网内部地理数据库和分发接收代理之间的数据通讯,确保地质图空间数据库的安全。

2.3 数据交换模块设计

数据交换模块分别由地理数据库发布引擎、分发代理、接收代理、交换中心等子模块组成。

(1)地理数据库发布引擎模块:是在发布地理数据库上运行的一个功能模块,负责更新数据操作的记录和管理(包括类中添加、删除、更新的数据),并

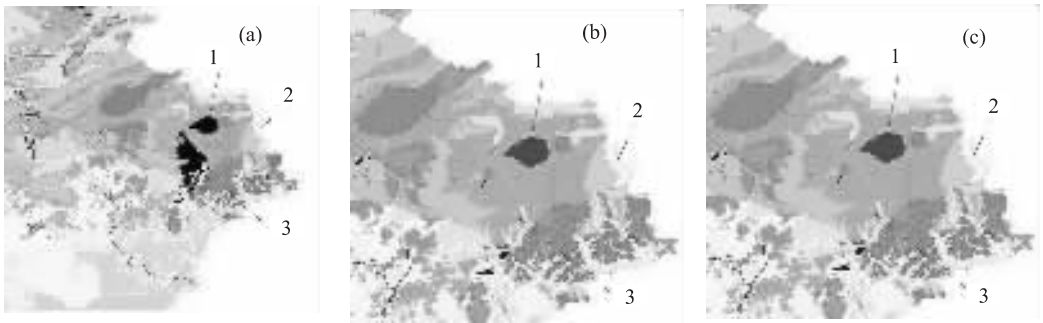


图 4 天津节点和北京节点上的同步更新结果示意(据中国地质调查局发展研究中心, 2005)

Fig. 4 Synchronous updating results between Tianjin and Beijing

通知分发代理改变。

(2)分发代理模块: Web 服务功能模块, 订阅端通过该模块获取地理数据库的相关信息和数据; 发布引擎可以将数据操作信息发送到该模块; 该模块还负责接收代理与交换中心交互; 是运行在发布地理数据库端和接收地理数据库间通讯的一个关键模块。

(3)接收代理模块: Web 服务功能模块, 通过该模块, 订阅地理数据库可以从发布端获取数据操作记录; 接收来自分发代理的数据; 也可以主动去分发代理获取; 是运行在接收地理数据库端和发布地理数据库间通讯的一个关键模块。

(4)交换中心模块: 记录发布服务器注册信息和发布信息, 并对外显示的服务器, 起到辅助发布的功能, 用户可以在分发代理上选择订阅, 也可以在交换中心上选择订阅。

3 方案验证

为测试地质图空间数据库之间的更新效果, 特选中在天津节点与北京发展研究中心节点之间、西安节点与北京发展研究中心节点之间进行地质图数

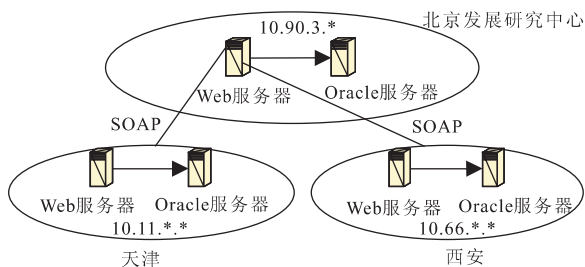


图 3 测试部署示意(据中国地质调查局发展研究中心, 2005)

Fig. 3 Deployment of test

据的更新。部署示意如图 3 所示。

在天津节点和北京发展研究中心节点同时部署有相同的河北和山东的 1:50 万地质图数据(图 4), 在天津节点上进行河北和山东的 1:50 万地质图数据的发布, 在北京发展研究中心节点上进行地质图数据订阅, 然后在天津节点上对发布的地质图数据进行更新, 图 4a 是未进行更新的山东省 1:50 万地质图, 图中有标号的地质体将进行修改。

在天津节点上对发布的山东省地质图进行模拟修改, 为明显起见, 将地质图的海岸线进行改动, 如图 4b. 在北京发展研究中心节点进行更新, 结果如图 4c, 可以很清楚地看到, 北京发展研究中心节点的山东省地质图也得到了更新, 更新结果和天津节点的更新结果相比较, 完全一致。实验证明, 通过这种数据交换策略, 能够有效地保证分布式异构地质图空间数据库之间的数据共享和一致性维护。

4 结论

本文以 MAPGIS 7.0 为支撑平台, 针对地质图空间数据库分布式、多源异构的特点, 以面向服务思想为指导, 提出了以地质图数据库中的地质空间要素类为单位进行发布和订阅的数据交换策略。按该策略实现的数据交换模块已在国家地质图空间数据库管理系统中良好地运行, 证明了这种数据交换机制可实现不同数据源间的数据一致性维护, 可解决异构数据源间数据一致性维护的难题。

References

China Geological Survey Development and Research Center, 2005. General report of geological survey spatial data

(下转 692 页)