

网格环境下空间元数据目录服务

姜俊, 朱小冬, 杨崇俊, 任应超

(中国科学院遥感应用研究所 数字地球与导航实验室, 北京 100101)

摘要:针对网络环境中分布式空间元数据库整合问题,提出了采用网格技术来实现空间元数据集成的方法。利用网格中间件将已有的分布式空间元数据库节点构建成网格分节点,在管理中心节点上部署目录服务和管理模块,实现了分布式元数据存储、分布式元数据查询、统一的目录服务、元数据管理等功能,有效地整合了网络中分散的空间元数据,为用户提供无缝的一体化目录服务。测试结果表明,该方法简单易行、系统可扩展性强。

关键词:网格; 空间元数据; 目录服务; 空间信息资源; 元数据管理

中图法分类号: TP311 **文献标识码:** A **文章编号:** 1000-7024(2011)03-0901-04

Catalogue service of spatial metadata in grid

JIANG Jun, ZHU Xiao-dong, YANG Chong-jun, REN Ying-chao

(State Key Laboratory of Remote Sensing Information Sciences, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: Spatial metadata catalogue service is significant for the query, location and accessing of the spatial data. A grid spatial metadata catalogue service is presented to solve the integration of the distributed spatial metadata databases. The distributed metadata databases are built as the grid nodes, and the catalogue and management services on the federation node are deployed. The grid spatial metadata catalogue service supports the distributed queries on all the spatial metadata databases, and integrates the distributed spatial metadata on the internet efficiently. Therefore, a seamless integration metadata catalogue service is accessed to the users.

Key words: grid; spatial metadata; catalogue service; spatial information; metadata management

0 引言

空间元数据是“描述空间数据的数据”,是关于空间数据集和信息资源的描述性信息,是对空间特征的概括和抽取^[1]。空间数据具有跨行业、多部门、多数据类型、多尺度、分布式的特征,给空间数据的共享和应用服务带来了集成与定位的困难。空间元数据目录服务是以空间元数据为核心的目录查询,是元数据系统利用元数据技术提供信息服务的一种标准模式。它通过元数据标准的核心元素将信息以动态分类的形式展现给用户。用户通过浏览门户网站提供的元数据摘要(核心元数据)可以快速定位自己所需的信息范围,然后要求门户网站在该范围内进一步搜索^[2-3]。

随着计算机信息技术的快速发展和网络平台的普遍应用,各单位的元数据更加丰富与规范,元数据库建设也将逐步完善。元数据的丰富为查询、了解和应用空间信息资源提供了便利,各单位机构为整合、共享已有的空间信息资源相继开发了各自的元数据目录服务。但是,随着数据共享与空间信

息服务共享的需求进一步扩大,局限于各单位内部的元数据目录服务已经不能满足用户的需求。特别是网格空间信息服务的发展,对网格环境中的资源共享提出了更高的要求,如何整合已有的空间元数据资源,向用户提供一体化的元数据目录服务,是当前急需解决的一个问题。

本文针对网络环境中的分布式空间元数据,提出构建一个集合网格服务管理节点和多个网格分节点元数据库的目录服务体系,整合分散于各单位节点的空间元数据库,为用户提供无缝集成的元数据目录服务,从而为空间信息资源的定位与共享带来便利。

1 空间元数据的集成技术

1.1 空间数据目录

空间数据目录是以空间元数据为核心的应用系统,它是以一定规则排列的空间信息资源的属性集合,是一个存储着关于各种资源属性的特殊数据库,这些属性可以供访问和管理空间信息资源时使用^[4-6]。空间数据目录系统一般由空间元

收稿日期:2010-03-11;修订日期:2010-05-16。

基金项目:国家863高技术研究发展计划基金项目(2009AA12Z227、2009AA12Z215);科技部科技支撑基金项目(2008BAH23B04)。

作者简介:姜俊(1983-),女,四川内江人,博士,研究方向为网络地理信息系统;朱小冬(1980-),男,江苏东台人,硕士,工程师,研究方向为地理信息系统、遥感数据共享;杨崇俊(1954-),男,四川资中人,研究员,博士生导师,研究方向为地理信息系统;任应超(1980-),男,山西太原人,博士,助理研究员,研究方向为地理信息系统。E-mail:jiangjun224@gmail.com

数据的管理、查询和访问组成。其中,管理模块负责对元数据库中的数据进行注册、删除和修改,整理后的元数据通过发布服务供用户使用。查询模块是通过查询元数据从而对空间信息资源进行定位、浏览,用户可以通过这个服务了解空间信息资源的基本信息。访问模块是可选模块,提供对空间信息资源的访问服务^[7-8]。

1.2 基于网格技术的空间元数据集成

空间元数据的集成是将分散于网络环境中的空间元数据整合,为用户提供无缝集成的统一服务。目前,关于空间元数据的集成,比较通用的方法是建立一个中央全局元数据目录服务器,对各个局部元数据目录服务进行索引和数据缓冲。用户向中央全局元数据目录服务的门户提交查询,中央服务器将这个查询分析调整后,分配到各个局部元数据目录服务器,查询分布式元数据库后,将结果传输到中央服务器进行整合,最后返回给用户。

在上述集成方法中,新的节点加入时,必须依照整个系统的服务接口先建立局部的元数据目录服务,并且当系统需要提供新的服务接口时,各个局部节点都需要增加相应的服务,可扩展性不强。针对这个问题,本文采用网格技术来集成空间元数据,将各个局部节点构建成网格分节点,网格分节点上不需要部署目录服务,而只需在管理节点上部署统一的目录服务,目录服务门户在接到用户查询请求后,直接在各网格分节点的空间元数据库中进行分布式查询,最后再整合结果,返回给用户。该方法避免了目录服务的重复建设,并且各网格分节点可以很容易地加入或退出系统,因此系统可扩展性强。

2 网格元数据目录服务

2.1 网格元数据目录服务结构

网格元数据目录系统是由管理节点和网格分节点组成的分布式系统,如图1所示。管理节点上部署3个功能模块:节点管理模块、元数据管理模块和目录服务模块。元数据管理模块和目录服务模块都需要和各网格分节点交互,所以系统构建了服务引擎模块来完成管理节点的任务分配和分发等任务。

(1)节点管理模块:节点管理模块控制网格节点的加入和退出,这个功能通过配置系统全局资源文件来实现。网格分节点以资源的形式注册到节点管理中心,每个节点资源包括网络地址、名字、数据库版本、数据范围等属性信息。由于节

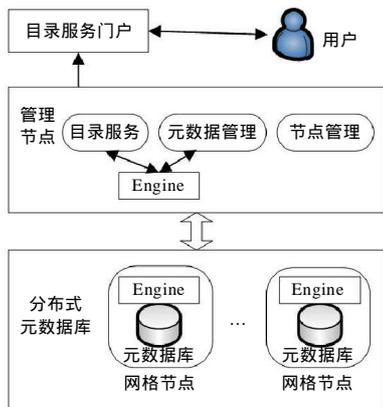


图1 系统结构

点管理模块可以实时地添加和删除网格分节点,所以网格元数据目录系统具有很好的可扩展性。

(2)元数据管理模块:元数据管理模块具有添加、删除和修改分布式元数据库记录的功能,为元数据目录的管理者使用。这个模块包装了对数据记录的操作功能,它将数据操作整合成服务提供给管理者。管理者在发出修改数据的请求后,管理模块将其整合成SQL语句发送给引擎模块,然后引擎模块将任务分发到各个分节点执行。修改操作完成后,分节点将执行成功的消息返回到引擎模块,最后返还给用户。

(3)目录服务模块:目录服务模块为用户提供元数据的查询和定位功能。用户在目录服务的网络门户上提交需要获取的数据属性,例如数据的空间范围、地名、关键字等。目录服务模块在获取这些查询条件后,将其整合成查询任务,传送给引擎模块。查询任务完成后,引擎模块将查询结果集传递给目录服务模块,最后展示给用户。

(4)服务引擎模块:服务引擎模块是管理节点与网格分节点交互的关键,负责查询任务的分发和查询结果的汇集。下一节将重点介绍。

(5)网格分节点:网格分节点上通过部署网格中间件 Globus 搭建网格环境。每一个网格节点上都有空间元数据库和服务引擎。服务引擎与管理节点交互,负责接收任务和返回结果。当系统的功能需要修改和扩展时,只需要修改管理节点的模块功能,各网格分节点不需要改动。

2.2 网格元数据目录服务的关键技术

本文的网格元数据目录系统将所有服务功能配置在管理节点,如何将管理节点与网格分节点有机结合,使得查询、管理任务能够有效的传送到分节点、查询结果精确地返回到管理节点,这是系统需要解决的关键技术。本文提出在管理节点配置任务配置模块和结果整合模块,网格分节点配置服务引擎,通过模块间的消息传递来解决管理节点与分节点的通信问题,如图2所示。

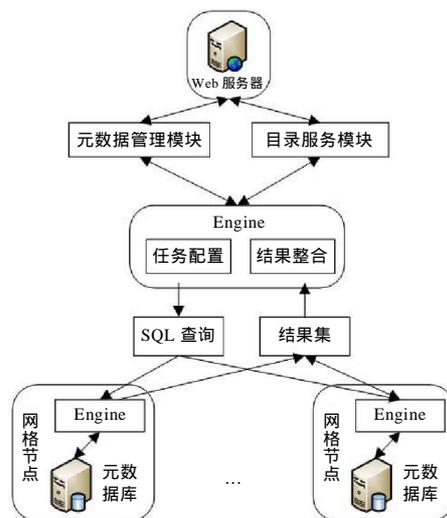


图2 系统关键技术模块

管理节点上部署了服务引擎,包括任务配置模块和结果整合模块。服务引擎接收来自目录服务和元数据管理模块的

SQL 查询。任务配置模块中存储了一个包含各节点数据库模式的虚拟数据库,任务配置模块接收到查询任务后,并不直接将 SQL 查询传递到各个网格分节点执行,而是首先利用已有的语法分析组件分析 SQL 语句,并结合虚拟数据库构建调适后的查询任务。新的任务不会传递到每个网格节点,只会发送给符合某些条件的元数据库,该方法提高了总体查询的效率。结果整合模块的功能是将网格分节点传回的查询结果进行处理,融合成一个统一的结果返回给用户。

每个网格分节点都部署了分节点服务引擎,负责网格分节点与管理节点服务引擎的联系。它将接收的 SQL 语句传递给元数据库进行操作,并将处理结果返回给管理节点服务引擎。所有网格分节点的查询任务是同时进行的。

此外,各网格分节点对于管理节点都是平等的。管理节点分发的任务需要在分节点上并行实施,因此分布式的元数据库需要具有统一的数据结构。通用的元数据标准是建立统一的元数据目录服务的前提,本文采用国家元数据通用标准构建元数据库,不仅可以全面地描述空间数据集,而且方便今后对元数据库进行整合共享等。

2.3 网格元数据目录服务的技术特点

网格空间元数据存储于物理分散的网格分节点上,目录服务系统整合了这些分布式资源,为用户提供了统一的无缝集成的目录服务,具有以下技术特点:

(1) 统一的元数据目录服务门户:空间元数据的搜索、添加、删除等实际操作任务是在网格分节点上完成的,网格管理节点部署了目录服务门户,向用户提供统一的查询与管理功能。网格管理节点与分节点的服务引擎中进行 SQL 数据流的交互,管理节点具有分发 SQL 任务和整合各分节点结果集的功能,为用户屏蔽了数据的分布特点,提供透明的、逻辑一体化的元数据库资源。

(2) 统一的元数据标准:网格分节点同时进行分布式元数据查询的特点要求各元数据库采取统一的元数据标准。本文实验采用的元数据标准是依据国家标准 GB/T19710-2005 所改进的,由于实验规模较小,本文的元数据标准较为简单。

(3) 元数据的分布、异构存储:分布式空间元数据存在海量、多源、异构的特点,并且其存储量、访问量、更新频率都有不同的要求,因此元数据的存储方式与存储介质也各不相同。不同组织对自己拥有的空间元数据具有最高的使用权限,同时需要更新、维护自己的元数据,集中式的数据存储方式难以满足这个需求。网格空间元数据采取分步存储的方式,不改变数据本身的独立性,在充分尊重元数据自治性的基础上将其有机整合在一起。分布式存储提高了系统的运行效率、改善了性能,局部节点的故障不会影响整个网格的运行。网格分节点可以选择不同的数据库存储空间元数据,如 Oracle、PostgreSQL、MySQL 等,元数据目录服务屏蔽了数据的异构存储特征。

(4) 元数据互操作:网格管理节点与分节点之间能够安全有效地进行查询任务与元数据结果集的交互,要求管理节点与分节点有一致的元数据规范,能够进行元数据的互操作。一致的元数据规范能够保证各个节点之间相互理解各自的信息内容,管理节点根据用户的查询请求生成查询任务,然后派发

到分节点执行。管理节点与分节点对执行任务的理解相同,因此管理节点可以准确获得各个元数据库的查询结果,返回给用户。本文采用了网格中间件 OGSA-DAI 进行任务的分发与结果的汇集,主节点与分节点之间进行元数据互操作的基础是统一的元数据标准和 SQL 操作数据流。

3 系统流程

3.1 网格节点资源配置

在系统实现的流程中,需要将网格分节点加入到资源配置文件。资源配置文件以 XML 格式存储,它记录了网格分节点的域名、资源 ID 和是否是本地资源。文件片段如下所示:

```
<resource url=http://jj.digitalearth.cn:8080/wsrf/services/dai
resourceID="MySQLDataResource2"
isLocal="true" />
<resource url="http://chinagis:8080/wsrf/services/dai"
resourceID="MySQLDataResource"
isLocal="false" />
```

3.2 服务调用流程

网格元数据目录服务系统中,目录服务的调用流程主要分为以下几个步骤(如图 3 所示):

- (1) 用户登录目录服务门户,选择查询条件(如名称、坐标范围等)输入查询的内容,或是选择元数据管理操作(如添加、删除、修改),提交给服务器;
- (2) 管理节点的服务引擎接收请求后,提取有效值,组成 SQL 查询;
- (3) 服务引擎的任务配置模块查询资源配置文件,获取网格分节点信息;
- (4) 任务配置模块将 SQL 查询分发到网格分节点的服务引擎;
- (5) 网格分节点的引擎接收 SQL 查询,并传递到元数据库进行处理;
- (6) 元数据库将查询结果返回网格分节点的引擎;
- (7) 网格分节点引擎将查询结果传送到管理节点的结果整合模块进行整合;

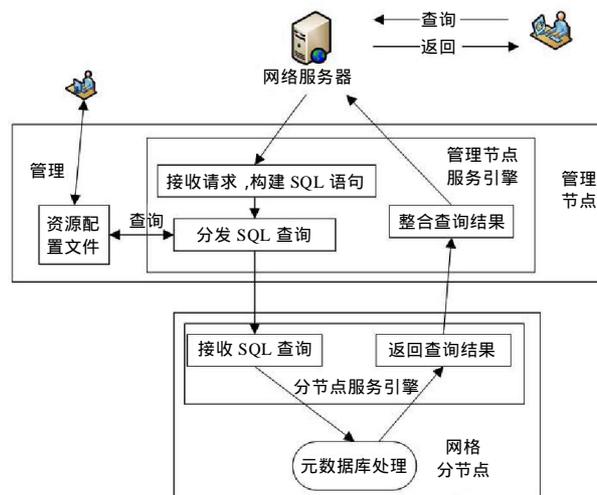


图 3 目录服务流程

