

# 空间信息网格中的元数据模型机理研究<sup>\*</sup>)

高劲松<sup>1,2</sup> 陈 思<sup>2</sup> 刘东升<sup>3</sup> 张 文<sup>2</sup>

(华中师范大学信息管理系 武汉 430079)<sup>1</sup> (武汉大学遥感信息工程学院 武汉 430079)<sup>2</sup>

(内蒙古师范大学计算机学院 呼和浩特 010022)<sup>3</sup>

**摘要** 网格元数据一直是网格研究的重点之一,直接影响到网格资源的组织、维护以及共享的安全控制等诸多问题。本文针对空间信息服务流程,充分考虑网格元数据服务的需求,提出了一种基于网格的空间信息元数据模型。它为实现一体化服务资源的透明访问、空间信息服务的高效获取以及知识的有效管理提供了有益的支持。最后介绍了这种元数据模型的实现。

**关键词** 元数据模型,空间信息网格,空间信息服务,一体化服务

## Research on Metadata Model of Spatial Information Grid

GAO Jing-Song<sup>1,2</sup> CHEN Si<sup>2</sup> LIU Dong-Sheng<sup>3</sup> ZHANG Wen<sup>2</sup>

(Department of Information management, Huazhong Normal University, Wuhan 430079)<sup>1</sup>

(School of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079)<sup>2</sup>

(School of Computer, Inner Mongolia Normal University, Huhhot 010022)<sup>3</sup>

**Abstract** Grid metadata, one of the keystones for grid research, have direct influences on the organization and maintenance of grid resources, the secure control of share and so on. Focusing on the spatial information service workflow, a spatial information grid metadata model is introduced with the needs of spatial information grid metadata services taken into account to realize the high efficiency integrated services, and to realize the transparent access to resources and the transparency of spatial handling workflow dynamic changes. Finally, a realization of the metadata model is introduced.

**Keywords** Metadata model, Spatial information grid, Spatial information service, Integrated service

## 1 引言

社会对空间信息的大量需求推动着空间信息理论与技术的不断发展,网格技术为空间信息服务领域的发展带来了新的契机。空间信息网格(Spatial Information Grid, SIG)是一种空间信息基础设施,汇聚了分布的海量空间信息资源,并对资源进行一体化组织处理,具有强大的空间数据处理能力。诸多空间信息服务皆为数据密集、计算密集或访问密集的信息,涉及海量空间数据的分布式存储访问和在线分析等。连接 Internet 上的空间信息服务站点、实现按需服务(Service On Demand)和一步到位服务(One Click Is Enough)是空间信息网格的最终目标<sup>[1~3]</sup>。

网格中的资源是复杂多样的,它们在地理上分散,可以动态地加入或移出网格。如何有效地调用这些动态可变的资源,是空间信息服务需要重点解决的问题。网格元数据可以将网格中的实体抽象化,为网格资源提供统一的逻辑视图,是高效透明地访问和使用网格资源的关键。空间信息网格元数据将不仅考虑网格资源的组织维护以及共享的安全控制等问题,还需构建更有针对性的元数据模型,为空间信息网格一步到位服务的实现提供高效合理的元数据支持。

本文从空间信息网格元数据需求出发,综合考虑网格环境的特点及空间信息的一体化服务目标,针对空间信息服务流程,探讨了如何建立更有效的空间信息网格元数据模型,提

出了空间信息网格元数据分类体系并构建了基于该体系的空间信息网格元数据模型。

## 2 空间信息网格元数据分类体系

空间信息服务是伴随空间信息的急剧增加、空间信息处理技术日趋丰富以及人们对空间信息需求日益旺盛的背景下产生的。围绕空间信息服务效率问题,产生了许多实用技术和方法,也借鉴了其他领域先进的技术方法,如组件技术、软件总线技术等。有些技术已在空间信息服务领域得到应用,如基于 Web Services 的空间信息服务,它推动了空间信息服务的进一步发展。但从实际的应用情况来看,对提供实时遥感数据分析、大型数据集交互分析和虚拟现实可视化等同时具备数据密集、计算密集和访问密集特点的空间信息服务来说,一体化服务目标的实现需要元数据服务考虑更多的因素:

(1)在资源获取过程中,用户在不了解存储系统的存储管理结构和访问路径等细节的前提下,能高效地定位和访问资源。就是说,资源获取的细节对于高层用户应该是透明的。

(2)在资源使用过程中,由于网格资源的拥有或失去是不可预测的,网格资源的管理和使用需要更高的动态适应性。在资源动态减少或出现故障等问题时,网格应能自行采取一定的措施,最大限度地保证任务的完成,减少用户的损失。

(3)在分布式环境中,面对海量空间数据的实时处理、高频率的访问以及空间计算资源的高消耗等问题,空间信息服

<sup>\*</sup> 本研究得到湖北省自然科学基金(2004ABA013)和武汉大学创新基金(2003-2005)资助。高劲松 副教授,博士生,主要研究方向:网格计算、地理信息系统及计算机图像处理。陈 思,张 文 硕士生,主要研究方向:网格计算、地理信息系统。刘东升 教授,主要研究方向:多媒体技术。

务响应速度应有根本性的提高。

综合考虑空间信息一体化服务的发展需求,这里将空间信息网格元数据模型中的元数据分为三类:资源元数据、空间处理程式元数据和知识元数据。

### 2.1 资源元数据

资源元数据为解决网格中资源的组织和维护以及共享安全控制等提供了基本保障。根据每类资源在访问频率、更新机制及与其他网格实体的逻辑关系等方面的特性可将资源元数据进一步细分为:

① 应用元数据:包括空间数据及与特定应用领域相关的业务信息的描述。标志了数据的结构,指明了数据的领域特性、内在构成、依存关系和转换方式等。

② 计算元数据:对分析工具(算法和软件)和空间信息模型等空间计算资源的描述。

③ 系统元数据:网格中各主机的系统描述信息。包括动态和静态的网络互联情况、存储系统容量、计算机空闲情况以及当前主机运行的线程等信息,是网络活动首要考虑的因素。

④ 用户元数据:网格中用户的管理信息(如用户名),它是可控制的资源共享的基础。

⑤ 副本元数据:包括各种资源的副本信息,也包括元数据副本的信息(如文件与具体存储系统之间的映射信息)。

### 2.2 空间处理程式元数据

空间信息网格中的空间数据分析是根据用户的需求,选择一系列合适的资源,由相应的应用程序处理执行的。这就构成了一套空间信息处理执行方案。这里将空间信息网格中的数据执行方案定义为空间处理程式(以下简称程式),它是一系列空间数据、空间分析工具、空间算法模型的逻辑集合,是对资源的综合集成。在其中可定义一套应用服务中所必需的输入数据、处理工具、作业方式及输出模型等,是提供一体化服务的核心。通过抽象空间处理程式,可以从空间处理程式实体的逻辑特征中生成空间处理程式元数据(以下简称程式元数据)。

### 2.3 空间知识元数据

知识是经过对空间数据进行分析处理、挖掘等深加工后所产生的满足某种特定应用需求的信息。为便于空间知识的维护和理解而产生的元数据为空间知识元数据<sup>[4~6]</sup>。

## 3 空间信息网格元数据模型结构

空间信息网格元数据模型以空间信息服务流程为核心,按照上节所提出的分类体系给出的空间信息网格元数据模型结构如图1所示。

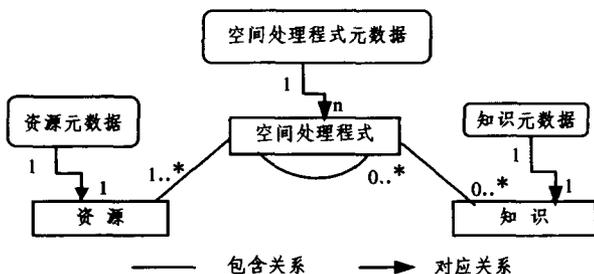


图1 空间信息网格元数据模型

图中,一套空间处理程式至少由一项资源构成,也可以包括知识以及其他的空间处理程式。

网格环境中的用户对空间信息领域的认知程度决定了他们

们选择怎样的处理方案,而采用不同的分析手段产生的结果往往会有很大差异,同时网格的动态性使得空间处理程式缺乏一定的稳定性。基于以上原因,模型中包含了空间处理程式元数据。它抽象了空间处理程式的逻辑特征。一个空间处理程式元数据可以对应多个空间处理程式,它们之间是一对多的关系。资源元数据和知识元数据是为了满足其定位、再利用和维护等需求而产生的,它们惟一确定了相应的资源和知识,它们之间是一一对一的关系。

## 4 空间处理程式元数据

空间处理程式元数据针对特定的应用需求,把能满足一定精度要求、提供相同空间处理能力的处理程式抽象为一类。系统根据用户请求中指明的服务特征,如输出模型、精度大小等就可以获得相应的程式元数据,然后按照程式所定义的资源实体间的引用关系确定访问次序,动态地调度任务,从而缓解空间信息服务存在的应用执行割裂情况,提高任务运行的效率,这一点对周期性的服务请求的处理表现得尤为明显。同时,程式元数据对程式的抽象使得空间处理程式更能适应网格的动态变化。当物理资源发生变更时,可以通过改变程式元数据和程式中对应的映射参数加以调整,毋须直接修改应用程序。正是空间处理程式元数据到程式的这层映射隐藏了程式的动态变更,保证了网格中空间处理程式所引用资源的逻辑独立性和较高的物理独立性,从而提高了空间信息服务的动态承受力。因此,通过空间处理程式元数据的描述,可为高可用性的一体化服务提供有益的支持。表1是空间处理程式元数据的内容设计。

表1 空间处理程式元数据的内容设计

内容	描述
名称	空间处理程式元数据的名称
最终服务	描述的是用户最终需要获得的空间信息服务,有固定的阈值范围,这样就可以由用户请求定位到惟一元数据文件
描述对象	记录可以获得同一应用服务的不同空间处理程式

从实现的角度来看,空间处理程式描述的是构成程式的资源元数据的组合表达式。在应用程序获得确定的处理程式后,通过资源元数据进行资源定位,可以实现物理资源的透明访问<sup>[7~9]</sup>。描述空间处理程式的内容由四个元素组成,内容设计见表2。

表2 空间处理程式的内容设计

内容	描述
名称	空间处理程式的名称
引用资源	记录程式所引用的资源和其它空间处理程式
组成方式	组成空间处理程式的运算式,是分解获得所引用资源的重要依据
存储系统	存储系统的标识信息,以供完成对处理程式的访问获取

图2是用户请求获得惟一确定的空间处理程式的流程。从图示可知,首先网格解析用户的请求,经由元数据目录访问后,通过元数据中所描述的最终服务,定位获得惟一的空间处理程式元数据,再根据空间处理程式元数据中的描述对象映射到多个实例化的空间处理程式。由程式处理的海量空间数据的存取往往需要考虑当前的网络状态、资源的成本、数据传输时间等多项因素,而通过程式优化算法,集成各种资源元数据(如当前CPU占用最少的主机,目前可用的空间数据源等动态信息),将得出最优化的空间处理程式。

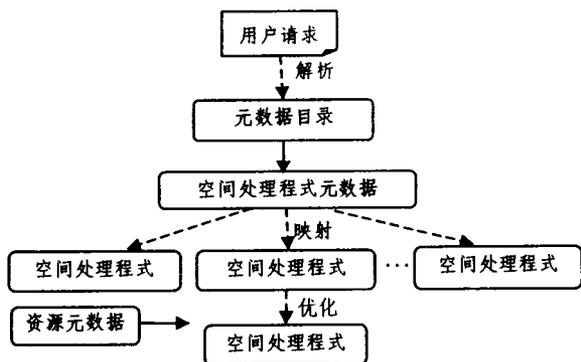


图2 最优空间处理程式流程

## 5 空间信息网格元数据模型的实现

### 5.1 实验平台

为了验证网格元数据模型的可行性,这里搭建了实验平台,对该模型进行初步的实现。用户通过 GUI 界面提交请求,空间信息网格服务器使用 UNIX 操作系统,该环境中配置了 Globus Toolkit 3.0<sup>[10]</sup>,其中包括 MDS (Metacomputing Directory Service) 2.4.1,同时采用 LDAP (Light Directory Access Protocol) 目录服务技术构建元数据目录服务。Globus 提供了基于 LDAP 的网格信息服务 MDS 及相关的客户端 API。针对 MDS 对静态信息支持不足的特点<sup>[11]</sup>,对 MDS 进行扩充,采用了关系数据库存储静态信息。

实验中,将网格中的元数据分为两部分:静态元数据和动态元数据。LDAP 目录中的元数据是通过人工或程序注册的静态信息,即应用元数据、用户元数据、副本元数据等。MDS 采集动态信息,由 Information provider 程序动态生成。

### 5.2 模型的实现

空间信息网格需提供空间数据处理流程中所必需的安全

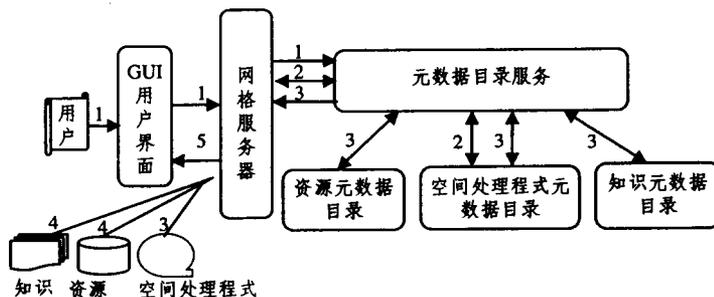


图3 空间信息网格元数据模型的实现

**结束语** 适应网格环境和空间信息一步到位服务目标的元数据模型的建立是空间信息网格建设尤为关键的一步。本文通过对空间信息服务流程的分析,结合已有的网格元数据服务的构建思想,提出了一种空间信息网格元数据模型,旨在实现资源和空间处理程式的透明和高效访问,为实现空间信息网格的一体化服务提供支持。

### 参考文献

- 1 金江军. 网格技术在地球信息科学中的应用. <http://tech.ccidnet.com/pub/article/c322-a92292-pl.html#4>
- 2 金江军. 网格计算及在数字城市中的应用. <http://www.csdn.net/subject/327/14964.shtml>
- 3 毕建涛, 王雷, 池天河, 等. 基于 Web Service 的地理信息服务研究. 计算机科学, 2004, 31(1): 69~71
- 4 Rajasekar A, Moore R. Data and metadata collections for scientific applications. In: European High Performance Computing Conf.

资源访问、资源和任务的分配、错误管理等服务。其中,元数据目录服务是资源定位、空间处理程式分解等服务的访问接口,主要完成的是由空间处理程式元数据到空间处理程式的映射、空间处理程式到资源的分解。正是这样的工作实现了资源的透明访问,为空间信息的一体化服务提供了基础支持。当用户向空间信息网格发出服务请求时,实现流程如图 3 所示。

(1)用户通过 GUI 界面向空间信息网格提交服务请求,网格解析用户请求,并将解析结果传给元数据目录服务。

(2)元数据目录服务在空间处理程式元数据目录中获取符合条件的空间处理程式元数据,并将处理程式的访问信息发送到空间信息网格。

系统根据元数据目录中资源的可用信息惟一确定最终的空间处理程式;分解空间处理程式后可获得处理程式中组合的资源元数据、知识元数据以及其他空间处理程式元数据的访问信息。

(3)元数据目录服务定位到相应的资源元数据、知识元数据及空间处理程式元数据。根据空间元数据搜索到符合要求的 GIS 数据,并对用户使用该数据的权限进行验证。

(4)由空间信息网格非元数据目录服务的网格服务进行程式和数据的协同分配。

网格服务把可执行代码和数据传送到指定的各个计算机上,依据优化的空间处理程式,确定资源的访问顺序,动态地调度任务,进行空间数据的处理。网格服务将负责任务调度和进程间的通信。

系统将知识获取过程中的信息收集与集成,对知识标引,以生成知识元数据。同时,将新产生的知识、知识元数据、资源、资源元数据存储并更新。

(5)网格服务将最终结果返回给终端用户。

- Amsterdam, Holland, 2001
- 5 Rajasekar A, Wan M, Moore R. My SRB & SRB components of a data grid. In: Proc. of the 11th IEEE Int'l Symp on High Performance Distributed Computing (HPDC11). Los Alamitos: IEEE Computer Society, 2002. 301~310
- 6 OGSA-DAI HOME. <http://www.ogsadai.org>
- 7 Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid; Enabling Scalable Virtual Organizations. International Journal of High Performance Computing Applications, 2001, 15(3): 200~222
- 8 Foster I, Kesselman C. The Grid; Blueprint for a New Computing Infrastructure. Morgan Kaufmann publishers, Inc., San Francisco, California, 1999
- 9 Declman E, Singh G, Atkinson M P, et al. Grid-Based Metadata Services. In: Proc. of the 16th Intl. Conf. on Scientific and Statistical Database Management (SSDBM'04). IEEE Computer Society, 2004. 1099~3371
- 10 The Globus Project. <http://www.globus.org>
- 11 张非, 阎保平. 一种基于网格服务的数据库元数据管理框架. 计算机工程与应用, 2004, 40(29): 209~212