

# 一种基于数据网格技术的 MODIS 空间数据服务模型研究

曾怡<sup>①</sup>, 李国庆<sup>②</sup>

(<sup>①</sup> 北京林业大学信息学院, 北京 100083;

<sup>②</sup> 中国科学院对地观测与数字地球科学中心, 北京 100086)

**摘要:**介绍了 MODIS 数据共享与资源服务现状, 研究了一种基于 3 层结构的 MODIS 数据服务模型, 详细介绍了数据服务模型的管理和分发机制。为了确保数据服务模型的可扩展性和平台无关性, 设计了基于 XML 的可扩展性数据访问语言和数据目录服务。最后给出了在空间信息网格试验床上的示范验证, 并对实验结果进行分析。

**关键词:**数据网格; 空间数据管理; MODIS; 数据服务

**doi:** 10.3969/j.issn.1000-3177.2011.06.019

**中图分类号:** TP751      **文献标识码:** A      **文章编号:** 1000-3177(2011)118-0092-06

## A MODIS Data Management and Service Model Based on Data Grid Technology

ZENG Yi<sup>①</sup>, LI Guo-qing<sup>②</sup>

(<sup>①</sup> *Information School, Beijing Forestry University, Beijing 100083;*

<sup>②</sup> *Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing 100086*)

**Abstract:** This paper presents a research on utilizing the data grid technology to integrate heterogeneous distributed MODIS data sources. We studied and implemented a three-layer structure for data interoperation, and designed an extensible XML-based spatial data access language to support the cross-border data computation, and developed a data catalogue service to facilitate users. The system deployed on Spatial Information Grid platform, and the performance is introduced finally.

**Key words:** data grid; spatial data management; MODIS; data service

### 1 引言

作为世界上先进的对地观测系统, EOS-MODIS 可提供中等分辨率的遥感数据, 共有 44 种产品, 包括了陆地、海洋、大气、资源环境等多个方面, 是目前研究地球科学最佳的可免费接收的数据源<sup>[1~2]</sup>。美国是卫星制造大国, 但截止到 2006 年仅建设了 16 座 MODIS 卫星接收站, 覆盖全美。俄罗斯有 8 套, 英国、法国、德国和意大利等欧洲大部分国家只有 1 套, 各空间机构共享数据, 满足科研和应

用的数据获取需求<sup>[3]</sup>。

由于不同接收站在数据源组织方式与管理模式上的异构性, 如果缺乏对 MODIS 数据共享相关技术的研究, 将难以实现 MODIS 数据的跨域资源协同和使用。目前, 我国已建立的 MODIS 接收站有 50 座, 计划建站 80 座, 大大超过了美国和欧洲等国家, 存在着严重的重复建站问题<sup>[4]</sup>。建设接收站的每套设备如全套引进美国产品约 100 万美元左右<sup>[5]</sup>, 这种重复建设带来了巨大的资金浪费, 尽管如此, 也难以实现有效覆盖和共享。

收稿日期: 2011-03-23      修订日期: 2011-05-16

基金项目: 北京林业大学青年科技启动基金 2010BLx15。

作者简介: 曾怡(1980~), 女, 博士, 讲师, 主要研究方向为空间数据服务。

E-mail: zeng.bjfu@yahoo.com.cn

遥感数据作为一种资源,可以不断地反复使用,利用数据网格(Data Grid)技术搭建 MODIS 空间数据分发管理系统,可以将分布的 MODIS 数据资源进行无缝集成<sup>[6~7]</sup>,在保留数据源原有数据组织方式和运行前提下,在合理的层次进行多源异构 MODIS 数据的跨域共享,盘活现有的空间数据存量,有效减少重复建站的开支和成本,避免浪费。同时,方便用户从一个统一的接口进行数据的发现和访问,使用丰富的 MODIS 数据产品,缓解数据提供与数据应用的矛盾,并为地学计算及相关行业应用发展提供基础信息保障<sup>[8]</sup>。

本文介绍了 MODIS 数据的特点和资源服务现状,利用数据网格技术,研究并实现了一种由数据源层、空间信息资源 UDDI 注册中心、数据代理层和数据分发接口层组成的 MODIS 数据管理分发模型。设计基于 XML 的可扩展数据源访问语言 XSDAL,确保了数据服务模型的可扩展性和平台无关性。同时,利用该模型在空间信息网格 SIG(Spatial Information Grid)试验床进行了相关示范验证,对数据服务性能进行了实验分析。

## 2 数据网格技术及相关工作

面对重复建站的巨额成本和大量现存数据无法大范围共享使用的矛盾,研究建立一个高效稳定的 MODIS 数据共享平台是迫在眉睫的。利用数据网格技术,可以将异构分布的数据资源进行整合和共享,大大减少重复建站的成本开支,缓解数据提供和数据应用的矛盾,为地学计算和应用提供有利的支撑。

空间数据的管理与分发是联系地学计算和信息应用的关键环节,起到中介和桥梁的作用。数据服务模型的质量和效率直接影响到数据的利用率和应用程序的性能<sup>[9]</sup>。传统的空间数据管理分发模式大都采用基于 web 的界面集成,向用户提供服务,由于 HTML 语言面向表现和语义缺乏的特性,并不适合对不同来源和不同类型的数据进行管理和分发,一般只能分发同一个数据来源和同一种数据组织方式的图像数据,对嵌入集成的分布式遥应用更是无能为力。

在克服空间数据的分布和异构所带来的障碍方面,虽然 OGC 组织针对各种空间数据服务系统进行总结<sup>[10]</sup>,通过 W3S(WCS、WFS 和 WMS)3 个主要的空间数据服务类型制定了基于 web 服务的空间信息互操作系统框架和标准体系模型,但是由于

基于 OGC 的 W3S 只提供了标准和规范,并没有在技术手段层面提供解决策略。同时,基于它实现的空间数据分发系统对图像的查找支持不力,接口调用也比较复杂,难以得到有效使用。

数据网格技术是网格环境中管理和共享异构分布式数据资源的技术框架,对于数据密集型的地学计算和应用领域,借助数据网格技术能够实现空间数据快速、便利地搜索与整合,建立高效的资源使用和管理平台,为用户提供透明访问分布式异构数据源的机制<sup>[11]</sup>。

## 3 MODIS 数据服务模型组件分析

对于运行性 MODIS 接收站,每天的原始归档数据量约为 4G 左右,一个月约 120G。面对庞大的数据量,灵活的数据管理模型和高效的数据传输机制是研究的重点。基于数据网格技术的 MODIS 数据服务模型主要由数据源提供层、数据代理层和数据管理分发接口层组成。

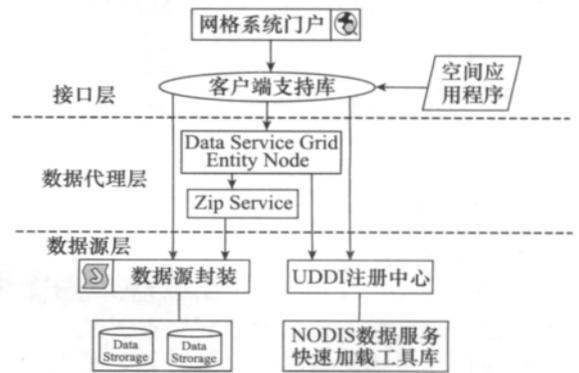


图 1 MODIS 数据服务模型组件结构

模型组件结构如图 1 所示,相关定义如下:

(1)数据源提供层包括数据源封装方法和网络资源 UDDI 注册中心。数据源层可以定义为以特定方式提供数据的查找和下载操作的逻辑服务,对于所提供数据的类型、种类、特性、组织形式等不做任何限制,能够响应空间数据源访问语言的查找或下载请求。资源注册服务方便遥感数据代理或遥感图像数据分发客户端通过注册服务找到所需要的遥感图像数据源。

(2)数据代理层可以被视为一个虚拟的数据源,它不存储任何遥感图像数据,但用户可以通过它实现对多个数据源的同时访问。代理层虽然不是管理分发系统的必须组件,客户端完全可以通过直接访问资源注册服务和遥感图像数据源来查找和获取遥感图像数据,但它的作用还是不可低估的:它从已注

册的各遥感图像数据源中搜集其提供数据的信息并将这些信息进行分析、优化和重新组织,这样数据分发客户端就可以更加简单和有效地完成对多种不同来源和不同性质的遥感图像数据的查找和访问了。从系统设计简单性的要求出发,遥感图像数据代理采用和遥感图像数据源相同的可扩展遥感图像数据源访问语言接口,以方便客户端及其支持库的设计和实现。

(3)接口层由系统客户端支持库、遥感应用、用户界面等组成。客户端支持库为遥感应用的开发者提供一组简单易用的开发接口,方便基于本系统开发各种遥感应用。通过嵌入和调用客户端支持库,遥感应用可以从本系统中大量数据源和数据代理查找和下载各种不同类型、不同格式、不同来源、不同用途的遥感图像数据。

基于以上系统组件架构,可以利用 Web Service 接口来屏蔽异构性,利用网格来实现松耦合的协作关系,利用元数据标准来达到一致性,较好地解决分布和异构带来的操作复杂性问题,表现出数据服务一致化的工作特点。

#### 4 模型可扩展性研究

##### 4.1 模型构建机制和服务扩展性支持

无论对于数据密集型网格或是计算密集型网格,数据的发现、查询、复制和代理交换等都是影响信息服务质量的重要因素<sup>[12]</sup>。通过分析,对于数据源而言,最重要的3个接口就是数据查询服务(data query)、数据获取服务(data access)和数据显示服务(data view)。服务模型以这3个方面为功能核心,同时保留扩展服务的功能接口,在对数据服务功能进行界定的基础上,构建合理高效的空数据服务模型,需要重点考虑以下几个问题:

首先,服务模型的可扩展性和灵活性。由于空数据资源的发展处于不断发展变化的进化过程中,需要数据服务模型必须具备高度可扩展性、灵活性和可保持动态更新的特点。它能够对分布异构数据源的变化进行封装和改进,并且保证这种功能的扩充不会影响空信息网格平台的整体结构<sup>[13]</sup>。同时,由于网格环境中的成员是由各种异构的硬件资源和软件资源组成的,因此数据服务模型应该具有平台独立性,和对异构平台良好的支撑性。

其次,数据服务的可交互性和协同。为了满足网格环境中不同数据资源协同工作的需求,数据服务模型应该实现异构空数据源的跨域分析和计算

处理的能力。

本文对数据资源发现和资源访问的处理流程进行分解和抽象,将相对集中的处理步骤归并为同一个处理层次,对每个功能层次进行定义,并规定了访问语言请求和响应的语法,对每个层次的输入输出参数规范为相对标准的形式,使得其可扩展性得以提高<sup>[14]</sup>。这里以数据查询为例介绍层次化功能定义,具体描述见表1:

表1 可扩展数据服务模型的层次化功能描述

Function Layer	Descriptions
Application Layer	调用数据服务的应用程序
Pre-Process Layer	对调用请求进行预处理(语义翻译和转换)
Engine Layer	将提交的请求进行资源分配和进程调度
Adapter Layer	确定待连接的数据源类型,初始化配置
Data Partner Interface Layer	不同类型数据源的具体封装和实现

##### 4.2 可扩展数据源访问语言 XSDAL

MODIS 数据服务模型的构建是试图支持大量不同类型、不同来源、不同格式、不同用途的图像数据,通过数据网格技术以统一的接口加以分发。为了实现这个目标,异构分布式的空数据源访问接口必须具有足够的灵活性和平台无关性。由于 web 服务技术是空信息网格的技术基础之一,所以模型访问接口选择支持 Web Service 规范的 SOAP 传输协议,以及基于 XML 定义的可扩展空数据源访问语言 XSDAL(eXtensible Spatial Data Accessing Language)。

针对空数据服务模型定义的基本功能,可扩展空数据源访问语言 XSDAL 以 XML Schema 的形式规定了相应的请求和响应的语法,并规定了这些请求和响应的使用规程。用户只要访问数据源所提供的 web 服务,将请求传递给数据源并分析数据源传回的响应,就可以完成相应的功能,执行包括模型基本功能在内的数据查询、获取和浏览操作。

根据 XSDAL 语言的规定,相关元素定义和功能是:

(1)请求以 <query>、<access>、<view>、<getCapability>、<getStatus>或<getResult>作为其根元素,

分别表示查找图像、获取数据、浏览数据和获得数据源特性这几类不同的功能以及获取操作状态、获取操作结果的请求。

(2) 响应被分为 3 类, 以 `<response>`、`<status>` 和 `<result>` 为根元素, 分别在启动操作、获取状态和获得结果使用。对数据源的一次访问(查找图像、获取数据或获得数据源特性)需要经过发送请求启动操作, 查询操作状态和获取操作结果 3 个步骤。

具体语法规则与定义为:

(1) 客户端生成一个以 `<query>`、`<access>`、`<view>` 或 `<getCapability>` 为根元素的操作请求, 并通过 SOAP 接口作为一个字符串参数发送给数据源;

(2) 数据源接到请求后即对其进行分析, 一方面开始执行该请求所包含的操作(查找图像、整理和打包要下载的数据或准备数据源特性参数), 另一方面为该请求分配一个唯一的处理编号并形成以 `<response>` 为根元素的响应返回给客户端, 供客户端以后获取操作状态和操作结果时使用。

(3) 客户端使用该处理编号通过以 `<getStatus>` 为根元素的请求查询数据源的处理状态, 直到处理结束。最后, 客户端通过以 `<getResult>` 为根元素的请求获取数据源的处理结果, 并完成全部操作过程。

对于数据处理的每个具体环节, XML Schema 可以对参数和格式进行约束与校验, 有利用在代码层面, 实现请求条件和响应结果的标准化描述及形式化提交。

图 2 所示是查找一个特定日期的 MODIS 图像的过程示例。

#### ① 客户端发送查找图像请求

```
<query>
  <conditions relation="AND">
    <condition op="EQ">
      <param>satellite</param>
      <value>MODIS</value>
    </condition>
    <condition op="EQ">
      <param>date</param>
      <value>2008-01-01</value>
    </condition>
  </conditions>
  <orders>
    <sortBy order="ASC">ID</sortBy>
  </orders>
</query>
```

#### ② 数据源启动查找操作, 返回一个操作处理编号

```
<response>
```

```
<operationID>001235864631</operationID>
</response>
③ 客户端获取操作处理状态
<getStatus>
  <operationID>001235864631</operationID>
</getStatus>
④ 数据源返回操作处理状态
<status>
  <operationID>001235864631</operationID>
  <currentStatus>processing</currentStatus>
</status><!-- 处理中 -->
<status>
  <operationID>001235864631</operationID>
  <currentStatus>finished</currentStatus>
</status><!-- 已完成 -->
⑤ 客户端获取操作处理结果
<getResult>
  <operationID>001235864631</operationID>
</getResult>
⑥ 数据源返回查找操作处理结果
<result>
  <operationID>001235864631</operationID>
  <resultSet>
    <item>
      <id>0013256782</id>
      <spacecraft>MODIS</spacecraft>
      <sensor>TERRA</sensor>
      .....
    </item>
    .....
    <item>
      <id>0020165784</id>
      .....
    </item>
  </resultSet>
</result>
```

图 2 基于 XSDAL 的 MODIS 数据查询过程示例

查找图像时以 `<query>` 为根元素的请求中主要描述的是被查找图像所需要符合的条件, 包括卫星类型、传感器类型、图像获取时间、图像覆盖范围等参数; 同时还可以把对返回结果的排序要求、返回结果的最大数量要求等作为参数传递给遥感图像数据源。对于查找图像操作, 最后一步从数据源中返回的将是一个查找结果图像的列表, 包括图像 ID 和包括卫星类型、传感器类型、覆盖地域、图像获取时间等一些主要信息, 以便访问该数据源的遥感应用的进一步筛选。

出于模型可扩展性的考虑, 遥感图像数据源访

问语言被设计成为一种基于 XML 的可扩展语言, 无论作为请求的 <query>、<access>、<getCapability>, 还是作为响应的 <result> 都有非常大的扩展余地, 以此实现对于不同种类、不同来源和不同用途的遥感图像数据源和遥感应应用。

### 4.3 数据目录服务

在得到实现、部署并注册到空间信息网络的资源注册服务中之后, 数据源就可以被访问了, 不断加入的数据源可以为遥感图像数据分发系统提供强有力的数据支持。除了网格应用和用户界面会通过客户端访问数据源之外, 各种图像数据代理也会通过客户端访问数据源。图像数据代理是 MODIS 服务模式研究的另一个重点, 多种多样的图像数据代理能够极大地丰富系统的功能, 提高系统的性能, 方便系统的使用。

目前, 数据服务模型已经实现了在一定范围内进行 MODIS 数据信息收集、整理和索引的图像数据代理, 定义为遥感图像数据目录服务。该服务通过收集多个数据源不同种类和用途的遥感图像数据信息, 并将这些数据相关信息进行整合存储和建立索引, 为用户提供跨数据源的遥感图像数据查找功能。用户可以通过遥感图像数据目录服务从多个数据源中查找到符合某特定条件的图像数据, 再向相应的数据源发起数据获取请求, 下载和使用图像数据。

遥感图像数据目录服务主要由两部分组成: 用于收集、整合和索引来自不同数据源的遥感图像数据信息, 并将其存入数据库中的后端; 用于响应用户请求并在数据库中执行图像数据的查找, 置于数据库的前端。这两部分之间没有直接的调用关系, 它们通过共享同一个数据库来实现图像数据信息的传递, 其结构如图 3 所示。

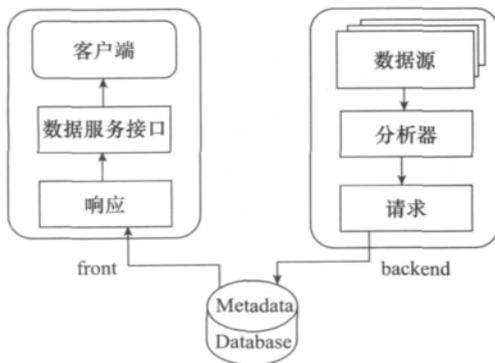


图 3 数据目录服务组件

遥感图像数据目录服务采用 SOAP 协议和可扩展遥感图像数据源访问语言作为其访问接口, 主要由负责访问数据源并收集图像数据信息的数据源客户端、分析和整理这些信息的分析器、用于存储这些信息的数据库、在这些信息中进行查找和产生查找结果的服务逻辑、实现服务的服务接口等部分组成。其中, 服务接口调用服务逻辑从数据库中读取信息, 分析器周期性地调用数据源客户端从数据源获得信息并分析和存储结果到数据库中, 数据库作为桥梁将前端和后端联系在一起, 组成一个完整的遥感图像数据目录服务。

## 5 实验与小结

目前, 在 SIG 运行平台上, 已集成了 3 个数据提供单位的 MODIS 数据源。这 3 个运行性 MODIS 数据源有着互异的存档格式和组织方式, 通过统一定义的元数据标准进行转换。

为了实现数据的准实时发布, 该项工作的重点主要集中在两点: 首先, 必须建立一套自动化的服务部署模板和流程; 其次, 由于 MODIS 数据量较大, 需要设计合理的数据传输机制。针对前者, 在对 MODIS 数据文件组织形式进行详细分析的基础上, 结合行业用户的需求, 我们分别对两个数据源设计服务快速部署模板。

针对第一种数据源类型, 我们采用自动扫描 FTP 目录的方法, 建立数据属性索引。MODIS 数据文件采用 23 位编码, 定义分别如表 2:

表 2 MODIS 数据文件分析

码位	名称	说明
第 1 位码	卫星名称代码	A 为 TERRA 卫星, B 为 AQUA
第 2-4 位码	传感器名称代码	MOD 是 MODIS 的缩写
第 5-6 位码	数据级别定义代码	01 为 L1A 数据, 02 为 L1B 数据, 03 为 GEOLOCATION 数据; 二级以上数据编码另行规定
第 7-9 位码	数据分辨率代码	1KM 为 1 公里分辨率数据, HKM 为 500 米分辨率数据, QKM 为 250 米分辨率数据, 以及 OBC 数据
第 10-23 位码	时间代码	数据采集年月日时分秒

结合地学应用和计算需求,抽取使用最频繁和重要的元数据属性,设计文件索引项如表3所示。根据这些信息可以设计MODIS数据查询标准,并建立自动化的数据服务加载模板工具库。

表3 MODIS文件索引属性

属性名称	数据类型	具体含义
ID	Int	数据文件标识
FileName	Varchar	MODIS数据存档名称
Location	Varchar	数据存放位置相对路径
Imaging_Data	Date	图像数据成像时间
QuickView	Varchar	快视文件名称
ZipFileName	Varchar	压缩数据文件名称
CraftType	Varchar	飞行器名称 TERRA\AQUA
Resolution	Varchar	成像分辨率
Sensor	Varchar	传感器名

针对第二种和第三种数据源类型,在SIG网格管理中心注册后,通过权限验证,数据服务通过Web服务接口直接调用图像元数据索引,实现数据的互操作。

系统在图像数据分发客户端的基础上用JSP

实现了一套基于Web的用户界面,用户可以通过这套界面按照图像获取时间、卫星和传感器种类、图像覆盖范围等条件对系统中已经存在的几个数据源进行数据查找,并对查找的结果执行诸如下载全部数据、查看快视图、获取原始元数据、下载指定波段的数据等操作。

通过在SIG试验床上部署该系统,用户可以同时从异构的MODIS数据源中查找需要的数据,并对这些数据进行下载数据、查看快视图等操作。该实验表明,基于数据网格技术的MODIS数据管理分发模型能够将多种不同种类、不同来源、不同用途的遥感图像数据通过一个单一界面加以发布,提供各种集成分布式地学计算和行业应用。由于数据服务良好的可扩展性和平台无关性,通过测试,我们加载一个新的数据节点只需修改配置文件中相关信息,半天的时间即可完成一个新的数据源加载。通过开展基于数据网格技术的空间数据服务模型的研究,可以大大减少MODIS数据接收站的建设成本和开支,并且为地球科学领域的用户提供方便高效的数据访问平台。

致谢:感谢中国科学院对地观测与数字地球科学中心数据技术部对文本研究工作的大力支持。

### 参考文献

- [1] Zhan X, De Fries R, Townsend J R G. The 250 m global land cover change product from the moderate resolution imaging spectroradiometer of NASA's earth observing system[J]. International Journal of Remote Sensing, 2000; 1433-1460.
- [2] King M D, Greenstone R. 1999 EOS Reference Handbook: A Guide to NASA's Earth Science Enterprise and the Earth Observing System[M]. Greenbelt: NASA/GSFC, 1999; 1-361.
- [3] 徐冠华. 关于建设创新型国家的几个重要问题[J]. 科技日报, 2006.
- [4] 方建中, 邹红. 实现科技资源共享: 科技创新体系建设的基本任务[J]. 科技与经济, 2005, 18(2).
- [5] 任建民. 大型科研设备: 利用率缘何不到三成[J]. 人民日报, 2003.
- [6] 孙九林, 李爽. 地球科学数据共享与数据网格技术[J]. 中国地质大学学报, 2002, 27(5): 539-543.
- [7] A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke. The data grid: Towards an architecture for the distributed management and analysis of largescientific datasets[J]. Journal of Network and Computer Applications, 2001(23): 187-200.
- [8] 周成虎, 欧阳, 李增元. 我国遥感数据的集成与共享研究[J]. 中国工程科学, 2008, 10(6).
- [9] K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman. Grid information services for distributed resource sharing[C]//Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing, IEEE Press, 2001.
- [10] OGC Consortium[EB/OL]. <http://www.opengeospatial.org/standards>.
- [11] B Allcock, J Bester, J Bresnahan, AL Chervenak. Data management and transfer in high-performance computational grid environments[J]. Parallel Computing, 2002, 28(5): 749-771.
- [12] John Graybeal, S. W. , Anthony Isenor. Sensor metadata interoperability workshop report[R]. 2007.
- [13] Zhenchun Huang, Guoqing Li. Building data grid for spatial information applications by meta-data adapters[C]//International Conference on Computer Science and Software Engineering, 2008: 18-21
- [14] Liu Dingsheng, Zeng Yi. General model of data service in spatial information grid[C]//IEEE International Geosciences and Remote Sensing Symposium Proceedings, 2007.