

Using Data Grid Technology to Build MODIS Data Management System in Agriculture Application

Yi Zeng¹, Guoqing Li²

¹ College of Information, Beijing Forestry University, Beijing, China

² Center of Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing, China

Zeng.bjfu@yahoo.com.cn

Abstract. MODIS data is the world's best remote sensing data of free receiving. It plays an important role in agriculture science and application. However, there is a serious problem about duplicated establishment of MODIS data receive station in China, and it has brought a tremendous waste of money. This paper presents research work on utilizing data grid technology to integrate heterogeneous distributed MODIS data sources. We studied and implemented a three-layer structure for data interoperation, and designed the extensible XML-based spatial data access language to support cross-border data computation. We developed the MODIS data catalogue service to facilitate users and improve system performance. The research helps to release the conflict between duplicated establishment of MODIS receiving station and less effectively data sharing.

Keywords: MODIS, agriculture application, data grid, data service. 1
Introduction

1 Introduction

As the world's advance earth observation system, MODIS (Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard the Terra and Aqua satellites. MODIS data has high spectral resolution, high spatial resolution and high time resolution, with a total of 44 kinds of image products [1]. Until now, it is the world's best remote sensing data of free receiving. MODIS data plays an important role in agriculture science and application, especially in crop yield estimation, drought and stalk burning monitoring, crop disease and pest monitoring [2][3].

However, there is a serious problem of duplicated establishment of MODIS data receiving station in China. China has constructed 50 MODIS station so far, and plans to establish 80 stations, which is much larger than the US and European countries. The redundant construction has brought a tremendous waste of money.

Using Data Grid technology could take advantages to integrate and share heterogeneous data. It helps to reduce the duplicated establishment cost, and eases the contradictions between data receiver and data user. Thus, it is meaningful to research

and develop the MODIS data management and distribution grid system providing support for scientific applications.

This paper introduces a three-layer structure of data management system based on data grid technology, and describes how to design a uniform interface to facilitate access heterogeneous data, without any change to the original data achieve storage. The extensible XML-based spatial data access language, which promises the data interoperability, is discussed in a great detail. We use this method in Spatial Information Grid platform for case study, and discuss the MODIS data catalogue service finally.

2 MODIS data sharing status

There is a serious problem of duplicated establishment of MODIS data receiving station in China now. The United States has 16 MODIS receiving stations, which acquiring data fully covered the whole country. Russia has 8 MODIS observation stations. Britain, Germany, France, and Italy have built one station of each, and they share data resources together. However, China has constructed 50 MODIS station so far, and plans to establish 80 stations. The quantity is far beyond the US and Europe. The expenditure of full set MODIS data receiving equipment for each station is about 1 million US dollars. China plans to import a total of 80 sets, and 30 sets have been completed at present.

Since there is not a favorable MODIS data sharing service system, large quantity of data is not take effectively use. The conflict between duplicated establishment of receiving station and less effectively data sharing becomes more and more obvious.

As a resource, the remote sensing data should be used constantly repeat. The research on integration of the existing data and building data sharing platform is a crucial approach to save expenditure and to return the huge investment.

On the other hand, with the strong needs to execute complex computation and cross-border joint data analysis in agriculture application field, it is very necessary to deploy research on geospatial data management and distribution system of scalability and interoperability. Utilizing data grid technology to build MODIS data management and distribution system is a favorable way to solve this problem.

3 The Architecture of MODIS Data Management System

The geospatial data management and distribution system play the role as a bridge to link the data providers and image users. The quality and efficiency of the system directly impact the data utilization and application performance [4].

Traditional spatial information management system mostly use web-based interface to provide services to users. Since the HTML language lacks semantics and not suitable for heterogeneous and distributed data storage, it is powerless when execute and embed distributed remote sensing applications. Open Geospatial Consortium (OGC) organization demonstrate a lot of research work on overcoming the obstacles of geospatial data distribution and heterogenous, by defining inter-

operation of spatial information and standards of system framework [5]. However, OGC does not provide resolve strategy on technical means and implementation level. The system utilizing OGC specification not has strong ability to support complex data search and discovery, and difficult to use the interfaces effectively. Thus, it is very necessary to deploy research on building geospatial data system with scalability and interoperability based on Data Grid technology.

The purpose of building MODIS data management system is to provide a data service platform, and collaborate data sources in diverse types and stored across multiple sites. The infrastructure is three-layer architecture, including data source layer, data agent layer and interface layer.

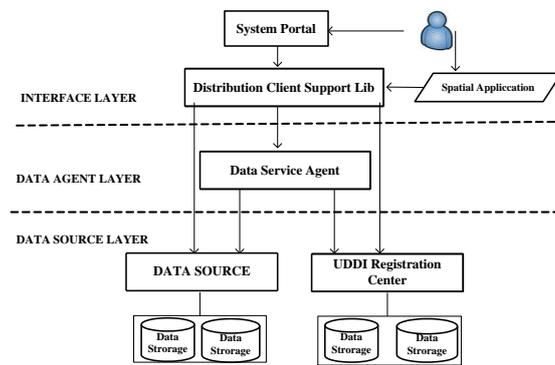


Fig. 1. Three-layer architecture of data distribution system.

Data source layer includes data providers and grid resources UDDI[6] registration center. By invoking data service, this layer is defined to provide functions and logic operations as data discover and data access. It not restrict to the specific data providers, or forms of organizations. This layer is able to respond the extensible XML-based spatial data access language, which will give detail description in the following chapter, and to find request or download data. grid resources registration center facilitate data service agent or distribution client to search data image.

Remote sensing data agent can be regarded as a virtual data source. It does not store any remote sensing image, but users can achieve it on multiple data sources at the same time visit. Agent layer is not the necessary component of data distribution system. However, its role can not be underestimated. The information can be analysis, optimize and re-organize from the original data collections. Thus the distribution client can complete data query and access from variety data sources through a more simple and effective approach.

Interface layer include client support library, geospatial applications, user interface and other components. Client support library provide a set of interface easy to use for the development of geospatial applications. Through embedded and invoke client support library, users can discover and download data from different sources and types, and also facilitate the development of various MODIS data applications.

4 Scalability of Data Service Model

4.1 Layer design of data service

The mission of data service is to provide logic support for all the distributed sources, and designed to act as middleware between data providers and users. As a data intensive grid, the capability of data service is essential for improving data discover and access [7].

By analyzing the data flow and control flow, the service processing procedure is decomposed and then composed into some function layers. For each layer, the request and response conditions are loosely defined. We design relatively standard language for data definition as well as parameter specifications, and all I/O parameters are based on XML. In addition, both query and access conditions are loosely defined, which decrease the dependence of specific spatial data storage organization. Therefore, the flexible mechanism and standard interface simplify the task of building applications to manage geospatial data sources, and the extensible capability of data service is ensured [8].

Table 1. Function layer descriptions of data service

Layer Name	Function Description
Application Layer	Grid applications call the data service with specific interface to query and access data information.
ServiceEngineLayer	In charge of allocation grid sources to support and schedule the processing and computing services.
Adapter Layer	Initialize configurations of related connections and environment. Determine the type of geo-data resource.
Geo-dataPartner Interface Layer	Provide wrapper methods for heterogeneous and distributed data sources.

4.2 Extensible XML-based Spatial Data Access Language

The heterogeneous structure and distributed location are the main two characters of spatial data source which bring much obstacle for data sharing, and take much difficulty for MODIS data integration. Thus, the interface design of MODIS data management and distributed system must have sufficient flexibility and platform independent. The interface is implemented on SOAP and extensible XML based spatial data access language [9].

Typically, the main and most frequently used operations of remote sensing data are data query, access and image view. Based on the data service model we discussed in the above section, the XML Schema of the extensible data access language describes the grammar of data request and respond, as well as the application rules. By invoking

the corresponding web service, the request will be transmitted to data sources, and then the analysis of the returning response is also achieved. Figure2 shows an example of MODIS data query.

```
1)To send image request
<query>
  <conditions relation="AND">
    <condition op="EQ">
      <param>satellite</param>
      <value>MODIS</value>
    </condition>
    <condition op="EQ">
      <param>date</param>
      <value>2009-07-01</value>
    </condition>
  </conditions>
  <orders>
    <sortBy order="ASC">ID</sortBy>
  </orders>
</query>
2)To launch data discover operation and return result ID
<response>
  <operationID>001235864631</operationID>
</response>
3)Client access the processing status
<getStatus>
  <operationID>001235864631</operationID>
</getStatus>
4)Data source return the processing status
<status>
  <operationID>001235864631</operationID>
  <currentStatus>processing</currentStatus>
</status> <!--be in processing -->
<status>
  <operationID>001235864631</operationID>
  <currentStatus>finished</currentStatus>
</status> <!-- be finished -->
5)Client get the discover result
<getResult> <operationID>001235864631</operationID>
</getResult>
6)Data source send back result info list
<result>
  <operationID>001235864631</operationID>
  <resultSet>
    <item>
      <id>0013256782</id>
      <spacecraft>MODIS</spacecraft>
      <sender>TERRA</sender>
      .....
    </item>
    .....
    <item>
      <id>0020165784</id>
      .....
    </item>
  </resultSet>
</result>
```

Fig. 2. Example of extensible XML based spatial data access language.

According to the XML schema of the extensible data access language, data query function use the tag of <query>, <access>, <getCapability>, <getStatus> or <getResult> as the root element, which define the function of data search, download, metadata access and operational status access. The respond message use <response>, <status> and <result> as the root element, and describe the operation commands, status and result respectively. The workflow of data source operation is a three-step process, including sending request, check operational status and access result.

Firstly, the client program generate a request used <query>, <access> or <getCapability> as the root element, and send the parameters to data source through SOAP interface. After analysis and execute allocation, a unique code referring the operation will return to the client. Then, the client handles the request status through <getStatus> root element by using the unique code until the end of treatment. Finally, the client processes the results and completes a full course of operation by using <getResult>.

Data access process is similar to data query. The first step is sending the request through the <access> root element, and obtain the ID list and operation details. This processing workflow can be describe as a processing chain, which including the steps of image filter, image format transform and data packaging, etc. In the final step, one or a group of <dataURL> will return the information to client as the element of <result>.

The data source layer is the foundation of the spatial data management and distribution system. Because of the heterogeneous and distributed characters, all the data storage is organized by plug-in profile mechanism, provided by JAVA class. The profile designated the data source and call the related plug-in and parameters.

4.3 Data Catalogue Service

Besides grid application program and client, data agents also query and access the data source. Data agent is another implementation focus in MODIS data management and distribution system. It can greatly enrich the functions of the system, improve performance and facilitate system users. We develop a MODIS data image agent to collection information and set up image directory. This service is defined as spatial Data Catalogue Service.

By collecting various information of a number of different types of data storage, the agents federate the data sources and build the index according to a specific time interval. The catalogue service facilitates users to find, download and use a certain MODIS image in data federation from a multiple sources.

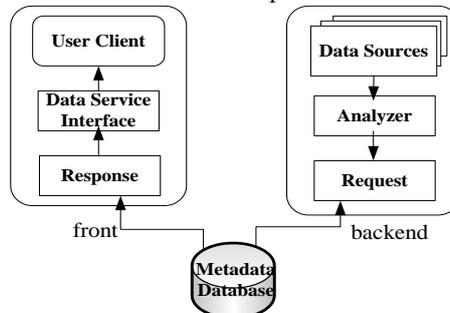


Fig. 3. MODIS Data Catalogue Service.

The MODIS data catalogue service also based on SOAP and the extensible XML based spatial data access language. It includes two components. One is in charge of collecting and indexing data information from databases in the backend. And the other is respond to different user request and executes database operation in the front. They exchange information by sharing the same database.

5 Experiment and conclusion

We present an effective architecture of seamless integration of distributed MODIS data while keeping the original data source robust. In order to facilitate user's interaction on MODIS data management and distributed system, we develop a web portal for integration and scheduling in Java, and apply this method in Spatial Information Grid for case study.

Users can discover MODIS data according to different query conditions, such as sensor type, data provider, and acquisition time or coverage range, etc. Furthermore, the result image can be accessed by different operations as check quick view, download archive metadata and designate target area. The figures below show some test examples of data operation through this portal deployed on Apache Tomcat.

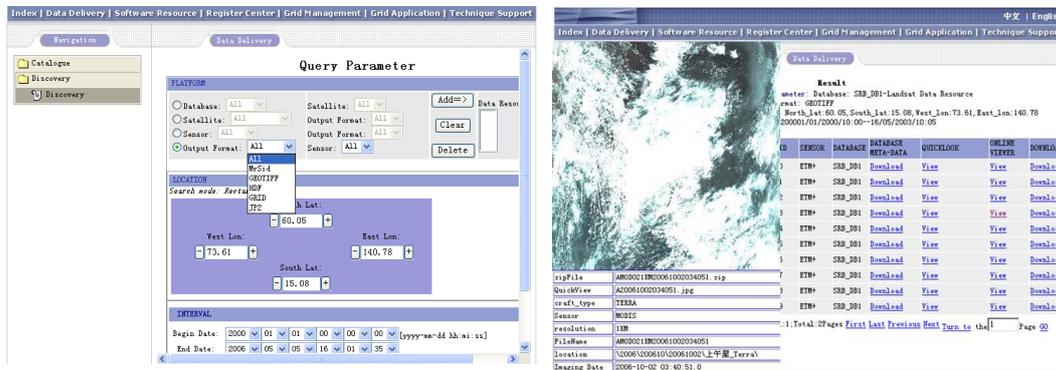


Fig. 4. The page of query condition input and result of quick view and data download.

The experiment shows that using data grid technology to establish MODIS data management and distribution system could enhance the efficiency of MODIS data utilization. Users can search and process data by different functional requirements through a uniform interface. It preserves the advantages of heterogeneity and distribution of data sources, and keeps the original data archive providers operational and robust.

Since the data service infrastructure is extensible and platform independent, it is only to spend a half day to modify configuration files and made sources grid enable when add and load a new spatial data source node.

This research can greatly reduce the expenditure on duplicated building MODIS data receiving stations, provide a user-friendly approach to discover data and access abundant MODIS image products, expand user's choices, and federate the information integration to realize rapid processing for agriculture science and applications.

Acknowledgments

Many of experiments are conducted in data technology division, Center of Earth Observation and Digital Earth, Chinese Academy of Sciences. The authors express great appreciation to the contributions and supports for all the colleagues.

References

1. MODIS data introduction. MODIS website. <http://modis.gsfc.nasa.gov/data/>
2. Dorigo, W. A., R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Intl. J. Applied Earth Observation and Geoinformation* 9(2) 2007, pp.165-193.
3. Paul C. Doraiswamy, Thomas R. Sinclairb, Steven Hollingerc, Bakhyt Akhmedovd, Alan Sterna and John Pruegere. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*. Volume 97, Issue 2, 30 July 2005, pp. 192-202
4. Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke. "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets". *Journal of Network and Computer Applications*, vol. 23, pp187-200, 2001
5. OpenGIS® Standards and Specifications. Open Geospatial Consortium (OGC). <http://www.opengeospatial.org/standards>
6. UDDI Technical White Paper. www.uddi.org/pubs/Iru_UDDI_Technical_White_Paper.pdf
7. K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman. Grid Information Services for Distributed Resource Sharing. *Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing*, IEEE Press, August 2001.
8. D. Liu, Y. Zeng, G. Li, F. Huang, "A General Model of Data Service in Spatial Information Grid". *IEEE International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, July 2007
9. G. Li, Z. Huang, "SIG-DDS: a Grid-based Remote Sensing Data Distribution System". *Proceedings of the Second International Conference on Semantics, Knowledge, and Grid*, Guangxi, China, Nov 2006.