Information Registry of Remotely Sensed Meta-module in Grid Environment

Yong Xue^{1,3}, Jianqin Wang^{2,*}, Chaolin Wu¹, Yincui Hu¹, Jianping Guo¹, Lei Zheng¹,
Wei Wan¹, Guoyin Cai¹, Ying Luo¹, and Shaobo Zhong¹

¹ State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Institute of Remote Sensing Applications, Chinese Academy of Sciences, P.O. Box 9718, Beijing 100101, China

² College of Information and Electrical Engineering, China Agricultural University, P.O.Box 142, Beijing, 10083, China

³ Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
Tian1.wang@163.com, y.xue@londonmet.ac.uk

Abstract. The term "the Grid" was coined in the mid 1990s to denote a proposed distributed computing infrastructure for advanced science and engineering. Grid computing technology is a new way for remotely sensed data processing. Special remotely sensed data processing Grid node provides convenient, consistent spatial information processing service for user. In this paper we mainly discusses information registry mechanism and method for remotely sensed data processing module and algorithm in Grid environment and gives the remotely sensed meta-module conception. Then based on the characteristic of remotely sensed data processing module and Grid technology, we describe remotely sensed data processing meta-module information registry method and give a concrete example.

1 Introduction

Grid is an integrated computing and resources environment, or a location for computing resources. Grid can absorb all kinds of resources for computing, and transfer them into available, reliable, standard and economic computing ability. Besides computers, computing resources also include web communication abilities, data, instruments and equipments, and even people. Foster *et al.* [4] defined "Grid problem", which is defined as flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources –what they referred to as virtual organizations (VO) [3].

Dozens of satellites constantly collecting data about our planetary system 24 hours a day and 365 days a year [7]. Remote Sensing data is characterized by largeness and instantaneousness. The analysis and sharing of these huge amounts of data is a big

* Corresponding author.

challenge for the remote sensing community [6]. Grid computing technology provides a good way to process remotely sensed data and sharing the processed spatial information through combining virtual super computing power storage resource on the Internet.

Today there have several famous Grid projects in spatial information field. Within the DataGrid project funded by the European Union, Work Package (WP) 9 of the Data aims to demonstrate the use of Grid technology for remote sensing applications and earth observation [5]. Earth Observation Grid (ESG) is a research project sponsored by the U.S. DOE Office of Science under the Scientific Discovery through Advanced Computing program (SciDAC). The aim of this project is to analyze climate variety all over the earth surface [1]. Information Power Grid (<http://www.ipg.nasa.gov/>) is the NASA's computational support for aerospace development, planetary science and other NASA research.

The spatial information can be collected, shared, organized, and processed with Spatial Information Grid (SIG), which has been sponsored by Ministry of Science and Technology of the People's Republic of China. It is the infrastructure that manages and processes spatial data according to users' demand (<http://www.spatial.org>). Through demonstration applications in some key fields, develop Spatial Information Grid to promote the industrialization of spatial information resources. ([http:// www.863.org.cn/english/annual_report/annual_repor_2001/200210090014.html](http://www.863.org.cn/english/annual_report/annual_repor_2001/200210090014.html))

Remotely sensed data is one of the most important spatial information sources, so the research on architectures and technical supports of Grid-based remotely sensed data processing node is the significant part of the research on SIG. Today, remotely sensed data processing modules and algorithms are the valuable resources, which can be used as a service for users who are not familiar with remotely sensed data processing methods. It can also avoid rewriting the same algorithms over and over again for different people through sharing the remote sense algorithms and module as a resource service. The key of using this valuable resource is how to let users conveniently find them and how to get use of them in Grid environment in order to share resources. To solve the above problem, we must let the remotely sensed data processing module be indexed in the Grid environment. The service information of the remotely sensed data processing module should clearly describe category of the algorithms and how to operate in Grid environment with these resources. Grid service request will be submitted to special remotely sensed data processing node with virtual super computing power after user searching the service information. Service Data technology of OGSA in Globus provides a effective method for information registry. We will discuss the metadata and method and mechanism of remotely sensed data processing module service registry in Grid environment based on Globus toolkit3.0.2 as follows.

2 Grid-Based Remotely Sensed Meta-data Service Data

Service Data technology of Globus (GT3.x) provides the fundamental method to describe service content in Grid environment.

2.1 Service Data of OGSA (Open Grid Service Architecture)

Service Data is the aggregation structure that is related to Grid service [2]. It must be conveniently indexed and classified through Service Data Element (SDE). A service data declaration is a mechanism of publicly expressing the available state information of a service through a known schema. There are two types of service data elements (<http://www-128.ibm.com/developerworks/grid/library>):

- Static:
Declared as part of the service's interface definition (GWSDL portType definition)
- Dynamic:
Added to a service instance dynamically. This behaviour is implementation-specific. The client may know the semantics (type and meaning) of the service data, or you can acquire that information from somewhere. For example, in order to process the dynamic SDE values you may need to get the schema for the SDE from a remote location.

The Service Data Element (SDE) model provides standard mechanisms for querying, updating, adding and removing data associated with each Grid service instance

2.2 Remote Sensing Meta-module Service Data in Grid Environment

2.2.1 Conception of Grid-Based Remote Sensing Meta-module

We give the conception of Grid-based remote sensing meta-module as following: The metadata that describe serials of remotely sensed data processing modules and algorithms in order to publish and manage these resources in Grid environment. The description of Grid-based remote sensing meta-module must include function express of the algorithm, input and output description, control parameter description etc. As Grid-based remote sensing meta-module, they must match these conditions:

- They must be described as the form of the Service Data Element (SDE) of Grid Service.
- Expression schema must match with the rule of the Grid Web Service Description Language (GWSDL)
- They can clearly describe the name, function, input and output parameter of the remotely sensed data processing module.
- Service Data used to describe remote sensing module must be static, which will not change until the module and algorithms become update.

2.2.2 Meta-module Service Data Structure in OGSA

The structure of the remote sensing meta-module seems as dendriform organization (Figure 1). RSService Instance will be produced by remotely sensed data processing service factory. We can give above figure embodiment: RSService Factory can be supposed as an service factory to produce remotely sensed image classification service instance. The same classification algorithm has different Service Data Set, which is made up of several Service Data Elements. Also we can suppose that Service Data Element A describes that classification algorithm service can provide classified

image after the task, and Data Element B describes that classification service can provide description document or log file about this task.

If users not only want to get the classified image but log files about this task, they can choice RSService Instance A that is described by Service Data Set A. If users only want to get the classified image, they can just choice RSService Instance B that is described by Service Data Set B.

Fig. 1. Remote sensing Service Data structure

2.2.3 Grid-Based Remotely Sensed Meta-module Service Data Item

In order to clearly describe remotely sensed data processing module service information, Grid-based remote sensing meta-module Service Data must include description of the module and algorithm, input data, threshold value, output information, running time. Service Data items are shown in the following table (Table1). Keyword item mainly contains description content of Grid-based meta-module. Model_Description item mainly express the name of the algorithm, function, developed date of the module. Input_Data item seems complex because of diversity of the remotely sensed data format. Output_Data item has the flexibility according to user’s demand.Remote sensing information service can provide image or document result. Data Type item is the XML element type description in Grid Web Services Description Language (GWSDL).

Table 1. Item of Grid-based remote sensing meta-module Service Data

	Keyword		Data Type	Note
Description	Model_Description		Description	Whole description
Input	Input_Data		GridRS_Meta	Input data format
Threshold	Threshold_Con		xs:float	Threshold and control parameter
Output	Output_Data	Image, Document	xs:string	Output format
Running time	RunTime_MByte		xs:duration	Seconds for per Mbyte

RunTime_Mbyte means how many seconds this module process per mega-byte remotely sensed data using a certain algorithm. Though this service data item ,user can easily estimate the time by simple multiplication because most of the algorithm analyze pixel one by one. But the result of RunTime_Mbyte must be tested before a meta-module service data being published. There are many remote sensing data types which is described as complex type in Grid Web Service Description Language(GWSDL). Input_Data Service Data is the description of remote sensing data needed by the algorithm and module in the following table (Table 2).

Table 2. Item of Input_Data Service Data

Data Format	XML ComplexType	Data Type	Note
GridRS_Meta	Data_Quality	xs: int	Full cloud covered is the max value(100) 0-100 scale
	Data_Format	xs:string	Data format name
	Imaging_Time	xs:dateTime	Collect data time

2.2.4 Service Data Description of Grid-Based Remotely Sensed Meta-module

Service Data Description (SDD) can be used to express remote sensed meta-module Service Data. Remotely Sensed meta-module SDD XML Schema is made up of six parts: domain name, version copyright, module information, module algorithm and running time, input information, output information.

<?The first part: domain definition>

```
...
<wsdl:definitions name=" ModelData" xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/">
...
<wsdl:types>
```

<?The Second part: copyright statement>

```
...
```

<?The third part: Information description of module>

```
<complexType name=" ModelDataType ">
<sequence>
<element name="Model_Description" type="ModelDescription"/>
<element name="Input" type="InputInfo"/>
<element name="Output" type="OutputInfo"/>
```

```
...
```

<?The fourth part: algorithm and time>

```
<complexType name=" ModelDescription ">
<sequence>
<element name="ArithmeticName" type="string"/>
<element name="Function " type="string"/>
<element name="ModelDate " type="date"/>
</sequence>
</complexType>
```

```
<?The fifth part: input information>
  <complexType name=" InputInfo ">
    ...
  <complexType>
    <element name=" format " type=" string "/>
    <element name=" imagingDate " type=" dateTime "/>
    <element name=" imageQuality " type="ImageQuality "/>
  <element name="ImageQuality" >
  <sequence>
  <element name="cloud" type="int" minOccurs="0" maxOccurs="1"/>
  ...
<?The sixth part: output information>
  <complexType name=" OutputInfo" type="OutputData">
    ...
  <element name="ImageFormat" type="string"/>
  ...
</wsdl: definitions>
```

3 Grid-Based Remotely Sensed Meta-module Service Registry

Combing Globus Open Grid Service Infrastructure (OGSI) rules and remotely sensed meta-module above, we provide the remotely sensed Grid service registry method (Figure 2). Serials of remotely sensed data processing module service data is registered in the registry container of the special remotely sensed data processing Grid node through Service Data provider, such as NDVI Service Data Provider, aerosol computing Service Data Provider, image classification Service Data Provider, etc. For user, they can search the proper service from the service data through special portal made up of remote sensing service set. There are two ways to find service through Grid service, one is based on demand line, and the other is Grid browser (Figure 3). User can use the browser to find the proper service description while inputting the service Uniform Resource Locators (URL) of remotely sensed data processing service.

Fig. 2. Grid-based remotely sensed meta-module registry process

Fig. 3. Grid-based remotely sensed processing service searching through browser

4 Conclusion

Grid technology is a very effective method for remotely sensed data processing service. Through it the existing remotely sensed data processing modules and algorithms resource can be shared seems like the common service in Grid environment. It much enhances the resource utility rate. In order to publish, describe, and manage Grid service data, the paper puts forward a Grid-based remotely sensed meta-module conception and corresponding structure. Combing the rule of this conception and OGSi technology, we define the Grid-based remotely sensed meta-module service data and Service Data Description language schema. Through remotely sensed Service Data Provider, the remote sensing service can be register in the Registry Container of the special remotely sensed data processing Grid node. Users can also search and choice the proper service by demand line or Grid browser.

After that, he can send a service request to Grid computing node according to the service data description.

Acknowledgement

This publication is an output from the research projects "Grid platform based aerosol fast monitoring modeling using MODIS data and middlewares development" (40471091) funded by National Natural Science Foundation of China (NSFC), China, "Dynamic Monitoring of Beijing Olympic Environment Using Remote Sensing" (2002BA904B07-2) and "863 Program - Remote Sensing Information Processing and Service Node" (2003AA11135110) funded by the MOST, China, "Digital Earth" (KZCX2-312) funded by Chinese Academy of Sciences, China, and "Research Fund for Talent Program" funded by China Agricultural University.

References

- [1] Allcock, B. I. Foster, V. Nefedova, A. Chervenak, E. Deelman, C. Kesselman, J. Leigh, A. Sim, A. Shoshani, B. Drach, D. Williams. SC 2001.High-Performance Remote Access to Climate Simulation Data: A Challenge Problem for Data Grid Technologies., November 2001.
- [2] Chen, D., Demichev, A., Foster, D., Kalyaev, V., Kryukov, A. et al. OGSA Globus Toolkit3 evaluation activity at CERN Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 534, Issues 1-2, 21 November 2004, Pages 80-84.
- [3] Foster, I. and Kesselman, C. (eds.). The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999.
- [4] Foster, I. C. Kesselman, S. Tuecke The Anatomy of the Grid, *Intl J. Supercomputer Applications*, 2001.
- [5] Giovanni, N. A., Luigi, F. B., and Linford, J., 2003, Grid technology for the storage and processing of remote sensing data: description of an application. In *Proceedings of the society of photo-optical instrumentation engineers (SPIE)*, Vol. 4881, 677-685, 2003.
- [6] Hu Yincui, Xue Yong, Tang Jiakui, Zhong Shaobo, Cai Guoyin, 2005, Data-parallel Georeference of MODIS Level 1B Data Using Grid Computing. *Lecture Notes in Computer Science*, Vol. 3516, pp883-886.
- [7] Wang, J. Q., Xue, Y., and Guo, H. D., 2003, A Spatial Information Grid Supported Prototype Telegeoprocessing System. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'2003) held in Toulouse, France on 21-25 July 2003*.
- [8] Xue Yong and Wang Jianqin, Wang, Y. G., Wu, C. L. and Hu, Y. C., 2005, Preliminary study of Grid computing for remotely sense information. *International Journal of Remote Sensing*, Vol.26, No.16, pp3613-3630.