Middleware Development for Remote Sensing Data Sharing and Image Processing on HIT-SIP System

Jianqin Wang[1], Yong Xue[1,2,*], Chaolin Wu[1], Yanguang Wang[1], Yincui Hu[1], Ying Luo[1], Yanning Guan[1], Shaobo Zhong[1], Jiakui Tang[1], and Guoyin Cai[1]

[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] Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK

Tian1.wang@163.com, y.xue@londonmet.ac.uk

Abstract. Sharing spatial data derived from remote sensing is a very significant thing. Grid computing and Web Service technology provides fundamental support for it. In this paper we mainly discuss architecture and middleware of sharing spatial data derived from remote sensing and processing. Because middleware of automatically transferring and task execution on grid is the key of the architecture, we study the middleware. It can effectively protect the owner of data and middleware's property through giving users their required result not just simply copying data and codes resource to them. Based on this sharing architecture and middleware technology, a data and middleware transferring example is showed.

1 Introduction

The Grid (http://www.globus.org/about/faq/general.html#grid) refers to an infrastructure that enables the integrated, collaborative use of high-end computers, networks, databases, and scientific instruments owned and managed by multiple organizations. Grid applications often involve large amounts of data and/or computing and often require secure resource sharing across organizational boundaries, and are thus not easily handled by today's Internet and Web infrastructures. Ian Foster *et al.* [1][2][3] offered several definitions of grid: "A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities". Grid computing is concerned with "coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations (VO). … [A VO is] a set of individuals and/or institutions defined by [some highly controlled] sharing rules."

There are several famous grid projects today. Access Grid (ww.fp.mcs.anl.gov/fl/access_grid) lunched in 1999 and mainly focused on lecture and meetings-among

* Corresponding author.

O. Gervasi et al. (Eds.): ICCSA 2005, LNCS 3482, pp. 173 – 180, 2005.

scientists at facilities around the world. European Data Grid sponsored by European union, mainly in data analysis in high-energy physics, environmental science and bioinformatics. Grid Physics Network (GriDPhyN) [4] lunched in 2000 and is sponsored by US National Science Foundation (NSF) mainly in data analysis for four physics projects: two particle detectors at CERN's Large Hadron Collider, the Laser Interferometer Gravitational Wave Observatory, and the Sloan Digital Sky Survey. Information Power Grid [5] is the NASA's computational support for aerospace development, planetary science and other NASA research. International Virtual DataGrid Laboratory (iVDGL) [6] was sponsored by NSF and counterparts in Europe, Australia, and Japan in 2002. Network for Earthquake Engineering and Simulation labs (NEESgrid) (www.neesgrid.org) intended to integrate computing environment for 20 earthquake engineering labs. TeraGrid (www.teragrid.org) is the general-purpose infrastructure for U.S. science: will link four sites at 40 gigabits per second and compute at up to 13.6 teraflops. UK National Grid is sponsored by U.K Office of Science and Technology[7]. Unicore (www.unicore.de) is a seamless interface to high-performance Education and Research computer centers at nine government, industry and academic labs. The famous Grid focused on spatial information includes SpaceGrid, EnvirGrid and EarthObsevation Grid. ESA's SpaceGrid is an ESA funded initiative (http://sci2.esa.int/spacegrid).

EnvirGrid main goals are generalization of Earth Science application infrastructure to become GRID-aware, extend GRID access to European Environmental and Earth Science application to large science communities, to value adding and commercial communities, …, and demonstrate collaborative environment for Earth Science.

Dozens of satellites constantly collecting data about our planetary system 24 hours a day and 365 days a year. Large scale of satellite data needed to be processed and stored in real time or almost real time. So far real time processing in remote sensing confronts much difficulties in one single computer, or even impossibility. Computing grid that is integrated by series of middleware provides a way to solve this problem [8].

Serials of middleware [9] of remote sensing image processing on grid platform have been developing by Telegeoprocessing and Grid study group in Institute of Remote Sensing Applications, Chinese Academy of Science. But as common users, they can have neither remote sensing data nor processing middleware except requirement of result. Automatically transferring and task execution (ATTE) middleware on grid can meet the requirement through farthest sharing remote sensing data and processing middleware.

The paper mainly describes the mechanism and process of the middleware. We describe the High throughput Computing for Spatial Information Processing System on Grid platform (HIT-SIP) developed in our research group in IRSA, CAS in Section 2.

2 Remote Sensing Data Processing Middleware on HIT-SIP Grid Platform

Remote sensing data processing is characterized by magnitude and long period of computing. Real time or almost real time processing is impossible on simple personal computer. But on the other hand, large numbers of PCs waste computing and restoring resources in certain time when they are free. So how to use the free computer. But on the other hand, large numbers of PCs waste computing and restoring resources in certain time when they are free. So how to use the free geographically distributed resources to process remote sensing data in cooperative type seems very significant and practicable. Luckily, emerging grid technology can provide large computing power in the Internet as the security type. The grid computing nodes resource distribute loosely. They may be supercomputers or common PCs with heterogeneous architecture. The grid computing nodes resource distribute loosely. They may be supercomputers or common PCs with heterogeneous architecture.

The HIT-SIP system on Grid platform has been developed in Institute of Remote Sensing Application, Chinese Academy of Science. It is an advanced High-Throughput Computing system specialized for remote sensing data analysis using Condor. Heterogeneous computing nodes including two sets of Linux computers and WIN 2000 professional computers and one set of WIN XP computer provide stable computing power. The grid pool uses java universe to screen heterogeneous characters. The configuration details of PCs in the pool are shown in Table 1. The structure of HIT-SIP system is shown in Figure 1. Figure 2 show the interface of front page of HIT-SIP system, which can be access by web service.

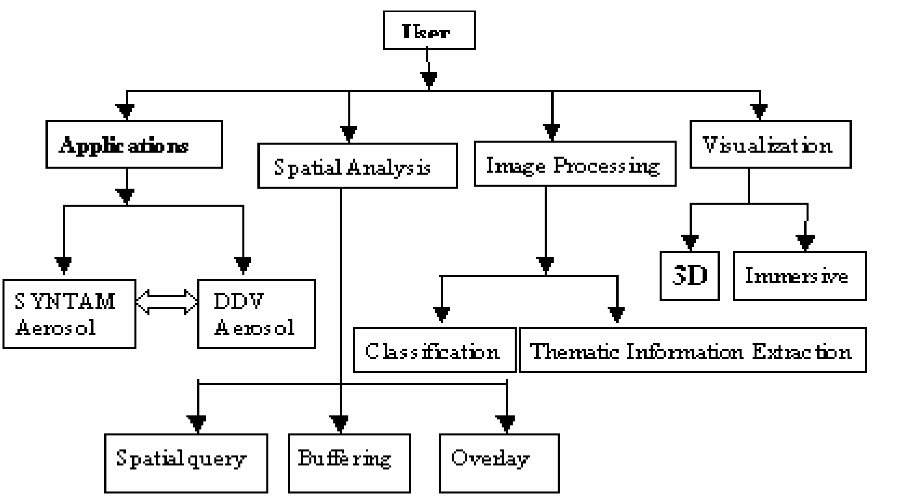

Fig. 1. The Structure of HIT-SIP system

The HIT-SIP system is a fundamental bare computing platform and can not run remote sensing image processing program. So remote sensing data processing middleware running on grid pool is inevitable. Common users can use the heterogeneous grid and share its strong computing power to process remote sensing data with middleware as if on one supercomputer. Till now, aerosol middleware, thermal inertia middleware and image classification middleware have been developed in HIT-SIP system.

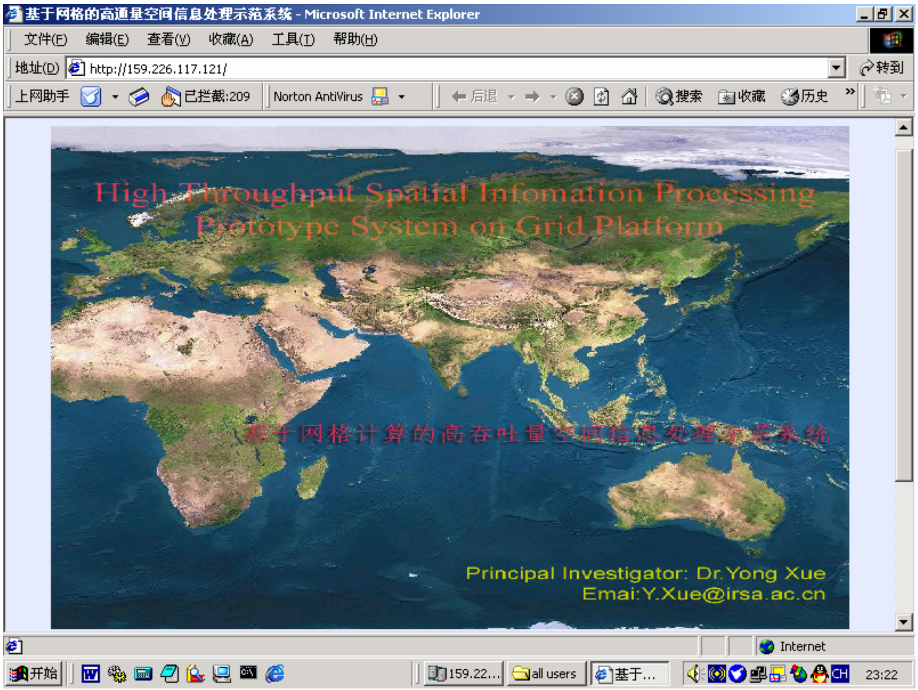

Fig. 2. The interface of front page of HIT-SIP system

Table 1. The configuration of computers used by HIP-SIP in the Grid pool

IP	Name	Arch-OS	Role
192.168.0.5	Manager.linux	Itel/Linux	Manager/client
192.168.0.3	Wjq.linux	Intel/WIN50	Client
192.168.0.102	Tele1.linux	Intel/WIN50	Client
192.168.0.111	Tele2.linux	Intel/WIN51	Client
192.168.0.6	Client2.linux	Intel/Linux	Client

3 Share of Remote Sensing Data and Image Processing Middleware

3.1 Sharing Requirement

Remote sensing data sharing became more and more significant with the digital information era coming. Grid technology also provides powerful tool for remote sensing data sharing .In the mean time, sharing remote sensing data processing middleware on grid naturally become very important. There still has a sharp contradiction between data and middleware owners and common users: as data and middleware owners, they don't just want to simply copy their production to other people because of property, but as common users, they also don't want to take long time to develop middleware. We have developed Middleware prototype of automatically transferring and task execution on grid to solve this contradiction. It can both meet users' sharing requirement and protect owner's property.

3.2 Sharing Architecture of Remote Sensing Data and Image Processing

We bring forward sharing architecture of remote sensing data and image processing middleware codes based on grid technology and Web Service.

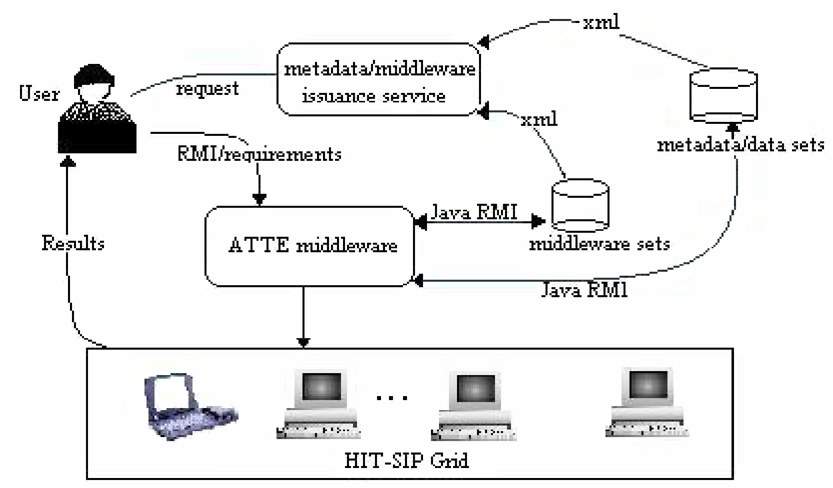

Fig. 3. Architecture of Remote Sensing Data and Image Processing middleware sharing

Geographically distributed remote sensing data sets and middleware codes data sets can be involved in this architecture. XML (extensible Markup Language) acting as a new standard for data interchanging on Internet provides a powerful tool for Web Service. Metadata of remote sensing and middleware codes will be registered at clearinghouse in certain suited type. So common users can find required remote sensing data and middleware through metadata matching module. Nowadays, there exists several metadata standards about geo-spatial data derived from remote sensing, of which "Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata" is the most famous one made by Federal Geographic Data Committee (FGDC). Metadata on the grid must be different with the FGDC's standards but with reference of it. Metadata of remote sensing data processing middleware must including task description files and middleware description files that must match the metadata format of remote sensing data.

Through module of metadata describing middleware and remote sensing matching service, user can easily find matched data location. Java Remote Method Invocation (RMI) is an advanced java technology used for distributed application program. Middleware of automatically transferring and task execution (ATTE) on grid can

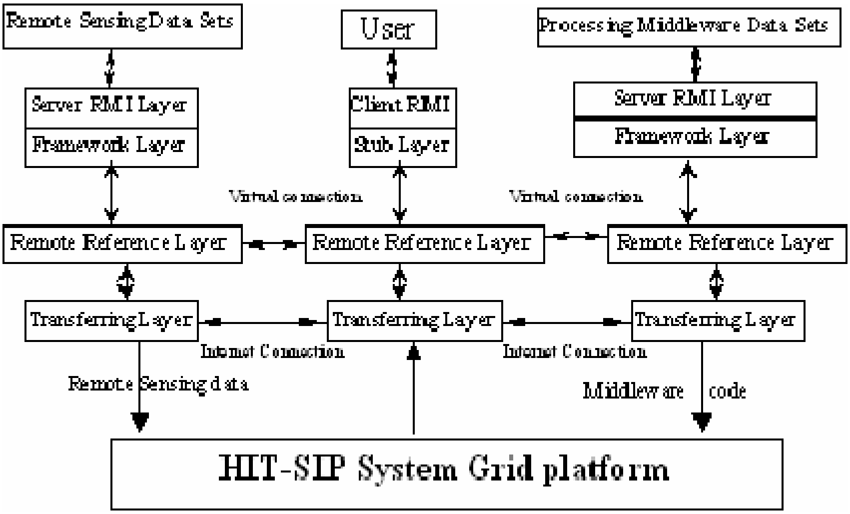

Fig. 4. Architecture of ATTE Middleware

long-distance transfer geospatial data derived from remote sensing satellite and remote sensing data processing middleware codes. When user gets the URL (including directory) of remote sensing data and processing middleware codes, he can easily starts ATTE. Matched data and middleware codes can be executed on grid platform and computed result be automatically transferred back to user.

3.3 Middleware of Automatically Transferring and Task Execution on Grid

Recently Java RMI has been popular. It is easier to use than CORBA because applications may only be in one language. Object can be easily serialized and transferred long-distance but not needed to execute on local computer.

The remote sensing data sets and processing middleware data sets nodes will be acted as server in the system architecture. Accordingly, user node will be acted as client. Application program on user node includes interface with server transferring module and task execution module on grid platform. RMIServer object will be serialized into Sub; so executing client sub is the same as executing server.

3.4 A Practical Example Using ATTE Middleware and Processing Middleware on Grid

On HIT-SIP grid platform, we test this sharing mechanism. Unsupervised Classification of Remotely Sensed Images middleware has been chosen as processing middleware codes resource. Overall resource and Internet configuration as following:
The Table 2 shows remote sensing data sets resource located at Tele2 (IP: 192.168.0.110) and unsupervised classification image middleware codes at TGP (IP:192.168.0.110). Client (IP:192.168.0.3) can be the node in the grid pool or not. In this test we choose it as the grid node and share its computing resource.

Table 2. Remote sensing aata and processing arrangement in the Grid pool

IP	Name	Resource	Role	Grid node (yes/no)
192.168.0.6	Condordlient2	Computing		yes
192.168.0.3	WJQ	Computing	Client	yes
192.168.0.110	Tele2	RS data & Computing	Server	yes
192.168.0.1	TGP	Processing middleware	Server	No

Following is the remote sensing data and middleware description: In this test, we use IKONOS satellite dada as data resource, which IKONOS satellite images were acquired on 26 April 2002 with approximately 1 m and 4 m spatial resolution in the visible (panchromatic) and relatively broad bands in the blue, green, red, and near-infrared portions of the spectrum, respectively. We use bands of blue, green, red, and near infrared for about 4m spatial resolutions. Test ground surrounds Olympic park in Beijing. The middleware [9] is the no-alternative unsupervised calcification algorithm. We image that each job partition data should be computed individually and be submitted only once. After computing they can be fused and displayed. Through deep analysis we find that we can try to solve this problem for several reasons. Object can be distinguished and classified whatever in one whole image or in divided parts according to brightness of every pixel. The most important issue we care about is which type every object can be classified and the difference among different objects .We can do not care about the brightness of the whole core center. We set the classified number of each divided part. The problem may be that each divided part must be classified into given class number but it may be less the number in just one divided part. Through amalgamating dynamically class number through critical threshold we can adjust the class number in each divided part. Then classifying effect in each part can be almost the same as in one whole image after properly matching colors.

4 Conclusions

Fig. 5. Classification Result on Client Node

Architecture of sharing remote sensing data and image processing middleware codes provides a integrated reference framework for resource sharing. Grid computing and Web Service technology is really a very effective method for sharing remote sensing data and processing middleware codes. Metadata of Remote sensing and middleware will be issued at cleaning house. So user can find the correct URL of matched data resource. Through ATTE middleware that we have developed, the owner of data and middleware need not just simply copy the data to client because of property. User just gets the required result through ATTE middleware on grid instead of getting raw data. It is very convenient not only for user but also for owner.

Acknowledgement

This publication is an output from the research projects "CAS Hundred Talents Program" and “Monitoring of Beijing Olympic Environment” (2002BA904B07-2) and “Remote Sensing Information Processing and Service Node” funded by the MOST, China and "Aerosol fast monitoring modeling using MODIS data and middlewares development" (40471091) funded by NSFC, China.

References

[1] I. Foster, “What is the Grid: A Three-Point Checklist”, *Grid Today*, 1 (6), 2002.
[2] I. Foster, C. Kesselman (Eds), *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publishers, 1998.
[3] I. Foster, C. Kesselman, S. Tuecke, “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”, *Inter. Journal of Super-computer Applications*, 15(3), 2001, 200-222.
[4] Grid Physics Network.. http://www.griphyn.org/index.php.
[5] W. E. Johnston, IPG Chief Architect. Information Power Grid. http://www.ipg.nasa.gov/
[6] IVDGL Grid . http://www.ivdgl.org/
[7] UK e-Science Grid Support Center. http://www.grid-support.ac.uk/
[8] Jianqin Wang, Yong Xue, and Huadong Guo, 2003, A Spatial Information Grid Supported Prototype Telegeoprocessing System. In Proceedings of 2003 IEEE/IGARSS’2003 held in Toulouse, France on 21-25 July 2003, v 1, 345-347.
[9] Jianqin Wang, Xiaosong Sun, Yong Xue, Yanguang Wang, Ying Luo, Guoyin Cai, Shaobo Zhong, and Jiakui Tang, 2004, Preliminary Study on Unsupervised Classification of Remotely Sensed Images on the Grid. *Lecture Notes in Computer Science,* Vol. 3039, pp.995-1002.