GEOSPATIAL WORLD FORUM
Theme
Dimensions and Directions of Geospatial Industry
18 - 21 January 2011, Hyderabad, India

The Spatial Data Sharing Mechanisms of Geological Survey Information Grid in P2P Mixed Network Systems Network Architecture Model

Chaoling Li[1,*], Miaomiao Song[2], Xia Lv[1],Xiangang luo[2], Jianqian Li[1]
1. Development Research Center,China Geological Survey, Beijing, China
2. Faculty of Information Engineering, China University of Geosciences, Wuhan, China
* Corresponding author: LCHAOLING@MAIL.CGS.GOV.CN

***Abstract*—The spatial data sharing mechanisms is important in the technical infrastructure of cloud computing or GRID computing, Many researchers have focused on the centralized resource pool and global-local two-layer model but little work considers the availability and practicability of universal discovery description and integration and mechanisms of supporting long transaction management. In this paper, we propose an effective ways of resource organization and mechanisms of management in P2P mixed network systems network architecture model, Called RA-P2P. The variety of spatial data that is of large storage capacity and of different data format and spatial services which are various and powerful are collected and put together, therefore it needs an effective mechanism to store, manage, retrieve and discover these geo-resources. Based on the platform of GSI-Grid RA-P2P adopts a means of metadata modeling to establish resource catalogue, provides many kinds of catalogue services conforming to OpenGIS Catalogue Services Specification and constitutes Catalogue Services Aggregator (CSA) which is a unified storage container for catalogue services to realize registration, issuance and discovery of services. Heterogeneous distributed grid computing environment gives an effective solution to manage geo-resources metadata, supports aggregation and integration of these resources and meets the needs for distributive storage and integrative discovery of multi-level, multi-source, multi-type spatial information. Based on Composition and technical infrastructure of the RA-P2P, the frame of resource organization based on geological fields Ontology, share, synchronous and teamwork mechanism of global-local two-layer resource, temporary aggregate-ware of data resource for supporting long transaction mechanism, the portal configurator of virtual node resource for node self-government mechanisms have been discussed in the paper.**

***Keywords*—GSIGRID, aggregate-ware of virtual node resourc, meta-services database, temporary aggregate-ware of data resource, metadata, distributed catalogue**

I. INTRODUCTION

A. *Backgroud*

Geological Survey Information Grid (GSI-GRID) is the platform for resources share, including geological data and information, software and computer hardware and running on 3 levels (over 16 virtual nodes or provinces) network system environment all over China. According to WWW (World Wide Web) service

framework, and under the support of criterion protocol, like SOAP, WSDL, WRSF, GML, OGSA [1] and etc, GSI-Grid has implemented the share and application service of spatial information based on GRID and SOA [2].

A peer-to-peer (P2P) network is a distributed system in which peers employ distributed resources to perform a critical function in a decentralized model. Nodes in a P2P network normally play equal roles. Therefore, these nodes are also called peers. P2P networks can be classified based on the control over data location and network topology. There are three categories: unstructured, loosely structured, and highly structured. In an unstructured P2P network such as Gnutella [3]. And a loosely structured network such as Freenet [4], both the overlay structure and the data location are not precisely determined. In a highly structured P2P network such as CAN [5] and Chord [6], both the network architecture and the data placement are precisely specified. The P2P node manager of GSI-Grid platform belongs to a P2p mixed network systems network architecture model.

The RA-P2P is the most important technical infrastructure in GSI-Grid platform. It includes standards and protocols of geological survey spatial data for sharing information and resources and establishing application services. Many researchers have focused on the centralized resource pool and global-local two-layer model [7] but little work considers the availability and practicability of universal discovery description and integration and mechanisms of supporting long transaction management in recent years.

GSI-GRID Platform has been developed since 2003. A lot of problems had been found in early stage as follows. (1) Most of geological metadata can not be directly used in the platform. It must change or register to new database. Because reusability of database resource is no good, it's difficult to avoid unconformity of describing data. (2)The evaluation of mineral resources is a process of long transaction. When user requests services or computes reserves of mineral resources, besides interactive processes and feedback process, reusability of temporary retention of data is considered as an important part. If all temporary retention of data is registered in the data directory database (as you know, unregistered data can not be found), a lot of garbage data would be produced. As a result, the efficiency and performance of the platform will be Reduce. (3) If there is only one global portal of the platform, it is impossible to implement the node self-government mechanism. So, the framework and the mechanism of P2P Node manager particularly develop.

B. The Main Technical Features of GSIGRID Platform Architecture

Based on grid technologies and thought, GSI-GRID platform is a geological survey information service architecture that can implement sharing, integration and cooperation of distributed and heterogeneous geological resources. The platform is a service-oriented. The hierarchy system of GSI-GRID platform can be divided into the data layer, the structure layer, the software resource and connection layer, the discovery and integration layer and the application and representation layer. The standards or criterions with each layer is the key technologies in the architecture of GSIGRID platform, especially in discovery and integration layer. The criterion of resource aggregation description, Geological domain ontology [8][9]classification and code and the geological data model standards based on the third generation of geodatabase model have been developed for supporting architecture of GSIG-RID platform.

The GSI-GRID platform has nine parts or module. Because the distributed data resource service is the most important service in GSI-GRID platform, the organization rule and model of data has been offered in the software architecture of GSI-GRID platform. The different data model would have been organized under the constraint of different compulsory standards in order to integrate resources and avoid information island.

This research is supported in part by the National High Technology Research and Development Program of China (863 program) by grants No.2007AA120504

II. THE FRAME OF RA-P2P

RA-P2P (see Fig.1) is composed of aggregate-ware of virtual node resource, meta-services database, the portal configurator of virtual node resource and temporary aggregate-ware of data resource.

Aggregate-ware of virtual node resource is composed of a catalog table of system metadata, an application metadata table, a catalog table of counter-part metadata, a system meta-services table and an in-sync manager of resource which can update information of resource among grid nodes synchronously.

Meta-Services database is composed of meta-services and meta-computing database (scheduling meta-services, basic meta-services of grid, profession meta-services of grid, meta-services of node domain, work flow service), the interpreter of work flow, scheduler of global resource and intelligent service engine.

The portal configurator of virtual node resource is composed of in-sync manager for global users, services migration-ware of node portal, login and role manager, configuration table of role and privilege, configuration table of logical domain, binding table of resource and privilege.

Temporary aggregate-ware of data resource is composed of application metadata table of node, catalog table of counter-part metadata, resource catalog manager of temporary data.

GSI-GRID node portal is a dispersed portal or P2P portal. It executes the geographic information publication of node, the registration of node, the text ,spatial data and metadata information publication of node, invoking synchronizer of global information among all nodes. Not only resources of one node but also ones of the other nodes (including global information of others node) can be discovered in the node portal. The self control of node is the most notable characteristics of node portal. The global information such as metadata and others resource information in each node can be synchronized to update with the synchronizer trigger of global information between nodes. GSI-GRID Node portal is the entrance of platform of GSI-GRID. After entrance of platform of GSI-GRID, the resource of computation, data and software can be shared.

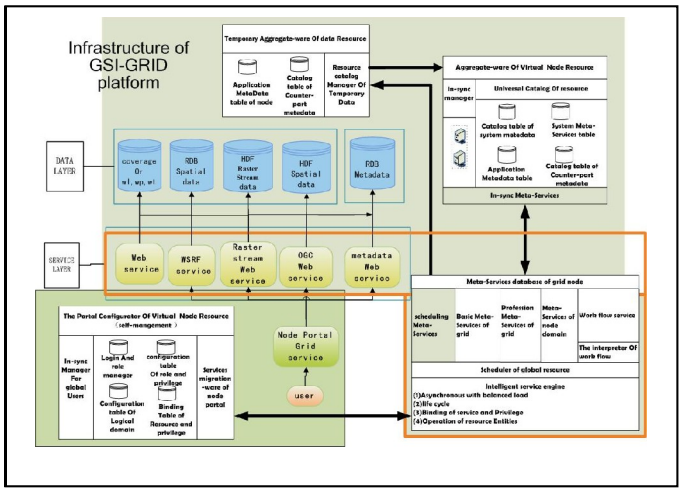

Figure 1. The frame of RA-P2P

III. ARCHITECTURE OF CATALOG SERVICES SYSTEM

A. *The Architecture of Spatial Information Catalogue Service System*

The architecture of Spatial Information Catalogue Service System contains the layer of spatial information resources, the layer of resources container and the layer of catalogue service, from the bottom to the top. The architecture is illustrated in figure 2.

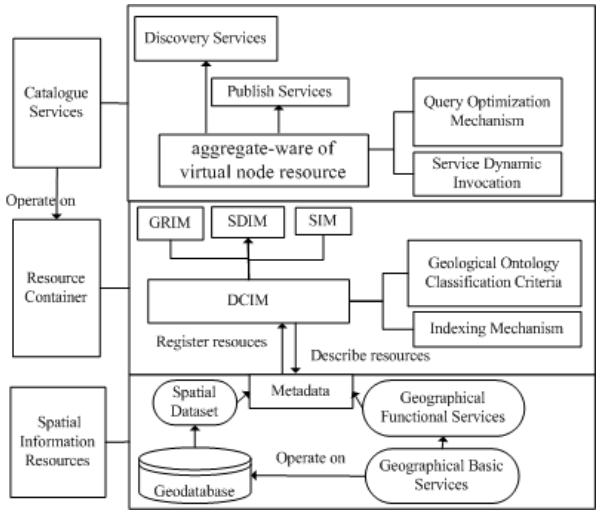

Figure 2. The technical architecture of GSIGRID

In the layer of spatial information resources, distributed geographic database is used to organize and store spatial data, which is consisted of basal geological data, hydrographic data, geochemistry data, mine data, and so on. There are also geographical functional and basic services in this layer.

In the Resource Container, data resource is extracted to build metadata which can be aggregated into Distributed Catalogue Information Model (DCIM). According to metadata content, the model is divided to three parts, Global Router Information Model (GRIM), Spatial Dataset Information Model (SDIM) and Services Information Model (SIM). In the model, an ontology classification criterion is used to encode catalogue information and index is also established.

In the layer of Catalogue Services, Aggregate-ware of virtual node resource is constructed to provide runtime environment for various services in Spatial Information Catalogue Service System. There are two types of catalogue services, discovering services and publishing services. The former provides functions of simple searching, searching by types and advanced searching; the latter can be used to register catalogue information in Catalogue Information Model. These services are elements of Catalogue Service Aggregator. In addition, in Catalogue Service Aggregator, a mechanism of ebXML/ebRIM [10] registering services is

designed, a mechanism of non-centered broadcast based on messages is used to publish information, and all services are managed in Services Pools which use a search strategy which integrate keyword search, space range search and directory search to achieve efficient discovery and dynamic selection of catalogue services.

As a catalogue container, DCIM owns three components: GRIM that stores distributed router information of every grid node; SDIM built on Geological Information Metadata Standard; SIM complying to OpenGIS Catalogue Services Specification which describes parameters of services, operations of services, status of services, source of service, Quality of Service (QoS) [11], and soon on. Additionally, the model sorts catalogue information based on a geological ontology classification criteria and creates indexing mechanism of directories.

B. *Chinese Geological Ontology Classification and Coding Rules*

In DCIM, all geo-resources should be classified, coded and indexed to implement the discovery, retrieval and management in an efficient way. Based on the study of Chinese geological ontology classification and coding rules, this article proposes cataloguing policies and coding rules of these resources.

According to the geological ontology classification criteria, every geo-resource has an identifier. The identifier of each geo-resource is well defined and is consisted of 32 characters. The first character stands for national code and "G" is generally used; the second, third and fourth character stands for basic categories code[6] which is defined in Chinese geological domain ontology; the fifth and sixth character expresses spatial identification code, for example, "0A" stands for the international framing; the next ten characters is spatial area code, like map numbers; it is year code from the seventeenth character to the twentieth one; the last twelve characters stands for geological ontology vice code which is also defined in Chinese geological domain ontology. Indexing of geo-resources is bounded to the above cataloguing policies. The cataloguing policies and coding rules is illustrated in Figure 3.

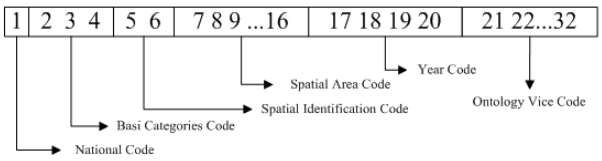

Figure 3. Encoding rules of catalogue information

IV. WORKING MECHANISM OF RA-P2P

In the aggregate-ware of virtual node resource, catalogue service interfaces are provided to implement the publish, discovery and harvest of services and geo-data metadata. According to "OpenGIS Catalogue Services Specification", the aggregate-ware of virtual node resource, implements five operations of catalogue services which are "DescribeRecord","GetRecords", and "GetRecordById" [12]. There is another operation named "HarvestRecords" providing the ability to harvest metadata records from other grid node. In the aggregate-ware of virtual node resource, registering services is implemented by the operations named "RegisterRecords" complying with "CSW-ebRIM Registry Service" [13].

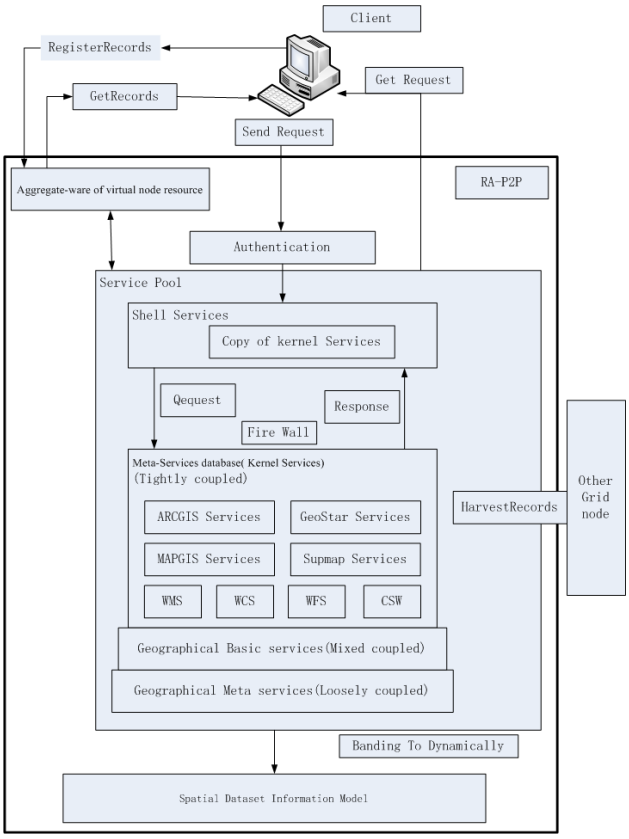

Figure 4. The working process of the RA-P2P

In the aggregate-ware of virtual node resource, services are divided into Web Map Service (WMS), Web Feature Service (WFS), WCS (Web Coverage Service), Catalog Service of Web (CSW) and other services defined by GIS Renders, by types; There are meta services, basic services and functional services [14]. In practice, the working process of the RA-P2P includes the following actions:

1) Publish Geo-Resources Metadata

Owners of geo-resources metadata publish information into a catalog table of system metadata, application metadata table, and catalog table of counter-part metadata by the interface “RegisterRecords”. Publishing action includes the register, log-on and alteration of metadata. The publisher knows the URL of the catalogue service has knowledge about the “RegisterRecords” interface and has the right to access the catalogue service. The metadata record is either successfully published to the catalogue service or publishing fails due to a non-valid metadata description [15].

2) Discover Geo-Resources

The requestor finds geographic resource metadata (including data and service metadata) from the SIM and SDIM by the “GetRecords”, and “GetRecordById” interface, then calls the service, binds of spatial data, and finally obtains maps or features.

3) Harvest Geo-Resources

Following the above example, there is another situation. That is, the metadata of "M5235" comes from some other destination grid nodes, instead of the current node. In this situation, the requestor needs to harvest results by invoking service of the destination grid node. The process is called Harvest Geo-Resources CSA provides interfaces of services to publish, discover and harvest geo-resources, offers support for resource virtualization, and improve the integration and sharing of distributed geographic information resources.

V. THE MECHANISM OF RESOURCE SYNCHRONIZATION MANAGER

The centralized management for grid data resources is in the most resource manager now, which makes sure eliminating data redundancy and data consistency. Once the centralized database has a physical breakdown, the data resource can not be shared. If we have the copies of centralized management in each node the preceding question can be avoided. The above mechanism has been set up in the resource synchronization manager of GSI-GRID platform.

The message mechanism with the operations such as upload, download, append, update and deleting between different nodes has been applied to the resource synchronization manager based on the penetration firewall speciality of WEB service (see figure 5).

The penetration of message queue table in GSI-GRID is confined to local node without sending the message. The local node calls the other node web service to implement the synchronization operations of the data. It is different from traditional message queue table. Thought each node has the same structure of message queue table each node is complete independence.

The principle of synchronization mechanism is to obtain unprocessed information through reading message queue table. After the analysis processing of the message, the resource synchronization manager in the local node gets the key information, and then makes the synchronization operations of the data. The advantage in message queue table lies in physical structure, the conception of the centralized node has been reduced. It implements the architecture of p2p and the communication between different nodes.

The resource synchronization operation can be described 8 steps as figure 5.

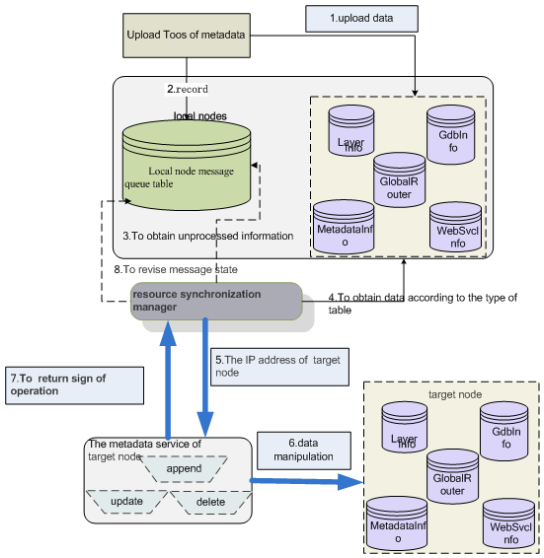

Figure 5. The flow-process of resource synchronization manager

VI. CONCLUSION

To sum up the above arguments, Through the P2P node manager, the distributed computer resources can become a virtual computer with network connections under the support of GSI-GRID platform.

Based on composition and technical infrastructure of the P2P node manager, the frame of resource organization based on geological fields ontology, share , synchrone and teamwork mechanism of global-local two-layer resource, temporary aggregate-ware of data resource for supporting long transaction mechanisms, the portal configurator of virtual node resource for node self-government mechanisms, the achievement has been got in GSI-GRID platform as follows:(1)The resource(including data, hardware, webpage and software) of the node portal has the features of self-government ,share, synchrone and cooperation between nodes portal.(2) The data format can be integrated GSI-GRID database in ARCGIS, MAPGIS, GEOSTAR, SUPMAP with the RDB, HDF(raster stream data), HDF(spatial data),COVERAGE and files without any changing of the data format or data model.(3) The resources of distributed data can be found based on the ways of words and map, geographic range, administrative divisions, metallogenic belt, international map subdivision and metallogenic province. (4) Through the temporary aggregate-ware of data resource for supporting long transaction mechanisms, the interim deliverables and final data can be published and reused, and the environment off-line breakpoint can be protected for long transaction computation, and the mechanisms of users interactions and process of feedback of a long transaction for grid computation on-line has been founded.

There are over fifteen demonstration nodes in Beijing, Tianjin, Xi'an, Chengdu, Hebei, Shengyan, Zhengzhou, Jinan, Lasa and Kunming now. Base on the standards of universal description, organization, discovery, and integration, all distributed database are organized in GSI-GRID. The data service can be provided online now from over ten GSI-GRID nodes. The data volume are 1 TB in GSI-GRID including 1:5,000,000,1:2,500,000 1:200,000 and 1:500,000 geological map database, 1:200,000 natural heavy mineral database of 27 provinces, 1:200,000 hydro-geological map database, isotopic dating database, national mineral deposit database, regional geochemical database, geological exploration levels of 1900-2010, mineral resource planning map.

ACKNOWLEDGMENT

The authors would like to thank Luo Xiangang, Liang Wu for the help to design related experiments, and Gou Deke for the detail revisions and supplement.

REFERENCES

[1] Ian Foster, Carl Kesselman, Jeffrey M. Nick4 et al. The Physiology of the Grid-An Open Grid Services Architecture for Distributed Systems Integration. Draft paper, 2002

[2] Mao xinping, " The practice or methods or principles of SOA", Publishing House Of Electronics Industry, Beijing, 2007

[3] Yatin Chawathe , Sylvia Ratnasamy , Lee Breslau , Nick Lanham , Scott Shenker, "Making gnutella-like P2P systems scalable", Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, August 25-29, 2003, Karlsruhe, Germany

[4] I. Clarke, O. Sandberg, B. Wiley, T. W. Hong. "Freenet: A distributed anonymous information storage and retrieval system", In Proc. ICSI Workshop on Design Issues in Anonymity and Unobservability, 2000

[5] S. Ratnasamy, P. Francis, M. Handley, R.M. Karp. "A scalable content-addressable network," In Proc. ACM SIGCOMM, 2001

[6] I. Stoica, R. Morris, D. Karger et al. "Chord: A scalable peer-to-peer lookup service for Internet applications", In Proc. ACM SIGCOMM, 2001

[7] Foster and C. Kesselman, " The Grid2: Blueprint for a New Computing Infrastructure," Morgan Kaufmann Publishers, 2004.

[8] W.N.Borst, Construction of Engineering Ontologies for Knowledge Sharing and Reuse, PhD thesis, University of Twente, Enschede, 1997.

[9] A.Gomez-Perez, R.Benjamins, "Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods", In Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5), Morgan Kaufmann, 1999.

[10] LIU Shu-Lei, LIU Yun-Xiang, ZHANG Fan, TANG Gui-Fen, and JING Ning "A Dynamic Web Services Selection lgorithm with QoS Global Optimal in Web Services Composition," Journal of Software,Vol.18,No.3,March 2007,pp.646-656

[11] China Geological Survey Bureau, Geological Information Metadata Standard,pp.2. December 2006, NO. DD2006-05

[12] Open Geospatial Consortium Inc.,OpenGIS® Catalogue Services Specification. OGC 07-006r1, 2007-02-23.

[13] Open Geospatial Consortium Inc., CSW-ebRIM Registry Service - Part 1 ebRIM profile of CSW. OGC 07-110r2, 2008-02-29

[14] Open Geospatial Consortium Inc., OpenGIS® Catalogue Services Specification 2.0.2-ISO Metadata Application Profile, p.31. 07-045, 2007-07-19

[15] Zhong Xie, Miao-Miao Song, Xian-Gang Luo, Resource and Information Sharing Mechanism Based on Spatial Information Grid, 2008 Seventh International Conference on Grid and Cooperative Computing,Oct.24-26 ,2008, pp.221-225

Author's Information

Paper Reference No.: PN-75
Title of the paper : The spatial data sharing__ mechanisms of geological survey information grid in p2p mixed network__ systems network architecture mode
Name of the Presenter : Chaoling Li
Author (s) Affiliation : Department of Information Engineering, Development Research Center of China Geological Survey
Mailing Address : 45 Fuwai Street,
Xicheng District Beijing, P.R. China
Email Address :lchaoling@mail.cgs.gov.cn
Telephone number (s) : 86-10 58584309
Fax number (s) : 86-10 58584307
Author(s) Photograph

Brief Biography:
Lichaoling, Ph. D. and Prof., the director of department of information engineering, the development and research center of China Geological Survey. As a team leader my duties include: geological data model and information standards for CGS ,technology of Grid and GIS,digital geological survey technology(iMapping and iExploration software). now in charge of the many projects in the informatization of geological survey of CGS and one item in "863 Program" now .4 new invention in iMapping have been patented, and the 2 first prize and 1 second prize of M.L.R have been achieved in recent years.