RESEARCH ON SPATIAL DATABASE MODEL IN GRID ENVIRONMENT

Liu Xiaosheng, Huang Xiaobin, Zhao Zhiyong

School of Architectural and Surveying Engineering, Jiangxi University of Science and Technology
86 Hongqi Road, Ganzhou, Jiangxi, China - lxs9103@163.com, hxb2929@yahoo.com.cn

Commission IV, SS-3

KEY WORDS: Grid, Spatial Database, Model, Work Flow, Simulation Experiment

ABSTRACT:

The paper firstly introduces grid technology and spatial database. Secondly it has designed the spatial database model in grid environment, which contains Application layer, Collective Layer and Fabric Layer; and then it has researched the working flow. At last, the simulation experiment of the spatial database model in grid environment has been carried out compared with traditional spatial database, and the result has shown the spatial database model in grid environment has more superior performance then the traditional.

1. INTRODUCTION

In the current spatial database systems, single spatial database can not obtain the data in other database in time, and the ability of spatial analysis and decision-making support is limited, so it is hard to realize spatial database sharing in heterogeneous environment such across regions, across platforms and cooperative working among departments, which can not solve a large number of application problems. The spatial database model in grid environment can regard geographic feature as database basic unit, which accords with the mode to recognize and analyze the realistic world. It can describe and express the real world completely, operate database like in camera obscura, and solve the data connection problem of multiple sources data to let us obtain spatial information in any computing platform and storage node, which will realize spatial information resources sharing and cooperative working maximally. Therefore, the author has researched the spatial database model in grid environment.

2. INTRODUCTION OF GRID TECHNOLOGY AND SPATIAL DATABSE

2.1 Introduction of grid technology

Grid technology, based on internet technology, web technology and high performance calculating technology, represents a kind of advanced technology and infrastructure, which can fully get various calculation resources and transform them into a kind of acquirable everywhere, reliable, standard and economical calculating ability to realize network resources sharing and cooperative working.

Grid refers access directly computers, software, data and other resources changed dynamically with time, therefore it has the ability of distribution and scheduling of resource and mission, safety transmission and communication guarantee, interaction between user and system and interaction among users. Classical grid system structure is five layers hourglass structure, which is a protocol-centered abstract layer structure, and emphasizes protocol in grid resources sharing and interoperation. A mechanism can be realized by protocol to make resources using negotiation and establish sharing relationship between virtual user and resources, and manage and develop sharing relationship further. The standardized open structure is good for expansibility, interoperability, consistency and cord sharing. Five layers hourglass structure is shown in Figure 1.

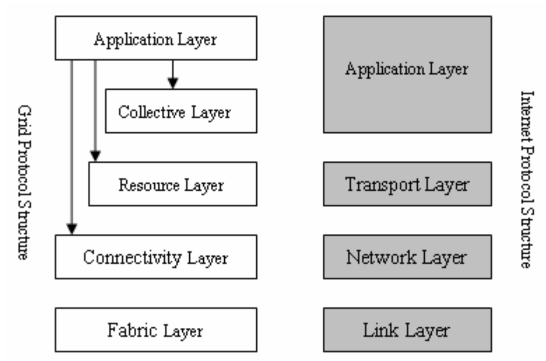

Figure 1. Five layers hourglass structure of grid

In the five layers structure, Fabric Layer in grid is the layer where resources can be shared; Connectivity Layer and Resource Layer compose commonly the bottleneck of the structure; Collective Layer coordinates various resources sharing, and Application Layer exits in virtual organization environment. The grid system constructed by five layers hourglass structure promotes single resource sharing powerfully and enhances interoperability among resources, which will make distributed computers can solve problem cooperatively. It is the successful introduction of five layers hourglass structure that make some research fields, which are hard to process for supercomputer such as weather forecast, high energy physics and aircrafts, have obtained better effect.

2.2 Introduction of spatial database

Spatial database organizes geographic data according to hierarchical data object, which includes object class, element class and element dataset. Object class is non-spatial data set; element class is the element set which has the same geometrical attribute, and element dataset is the element class set which shares space reference system. Spatial database mainly manages spatial data, and describes spatial features, attribute features and time features of geographic entity concretely. Spatial feature refers spatial position and relationship of geographic entity. Attribute feature refers nomenclature, type and quantity of geographic entity. Time feature refers correlative changes of geographic entity changing with time. In order to realize integration of spatial database, it is needed to consider the data object characteristics of spatial database, not only integrate attribute data, but also graphic data. There are always two ways to integrate. The one is that combining spatial data and attribute data before data integration and then integrating. The other one is that integrating spatial and attribute data respectively and then combining them. Considering not weakening semantic of spatial database and simple realization, generally, we can combine spatial data and attribute data before data integration and realize integration by spatial database engine (SDE).

3. DESIGN AND APPLICATION OF SPATIAL DATABASE MODEL IN GRID ENVIRONMENTT

3.1 Design of grid spatial database model

We combined the spatial database model and five layers hourglass structure of the Globus grid computing. Taking use of UML, the author has established a new data model with three layers structure, and the model structure is shown in Figure 2.

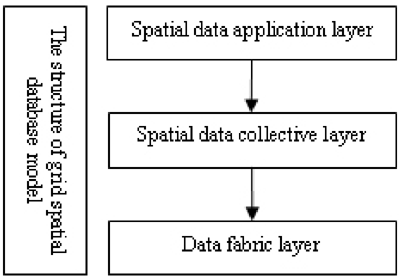

Figure 2. The structure of grid spatial database model

Grid spatial data model is designed for several layers, which has several advantages as following:

Firstly, since all components are subjunctive in grid, we provided a group of relatively uniform kernel interfaces which is the foundation of any data services. So, hierarchical structure and higher levels services can easily be constructed. These services can cross different abstract layers in a unified way.

Secondly, it's possible that virtualization images a number of examples of logic resources to the same physical resources. We do not have to consider the idiographic realization in composition of services, basing on composing of bottom resources, administrating resource in virtual organization. We can seamlessly image common services and behaviour to the foundation establishment of local platform by virtualization of grid services.

The structure of grid spatial database model is divided into three layers, which are application layer, collective layer and fabric layer.

Spatial data application layer is also the application layer in five layers hourglass structure. It's mainly developed for different users for adapting different databases and application systems, mainly taking use of data interface to satisfy users. In this layer, it provides users functions such as input, query, index, access and analyze of spatial data. It provides direct intercourse between spatial data and users.

Collective layer is also the middle layer which mainly describes common of the system resources without taking into account differences in an idiographic resource. It consists of computing model such as various data interfaces, network transmission protocol. It includes collective layer, resource layer, and connectivity layer in five layers hourglass structure. At the same time, it also includes a lot of spatial domain information. Its task is to find and record spatial data resources, filtrate and re-distribution the repeated resources. According to the data that application layer provided to process tasks, collective layer looks for spatial data of an existent directory, dealing with the tasks through distributed computing and dynamic optimization algorithm, returning the final results.

Data fabric layer is consisted of large number of spatial database, which includes not only construction layer in the bottom of five layers hourglass structure, but also variety data sources, such as computing resources, storage systems, directories, network resources and sensors. It provides the interface to access the partial resources which were controlled at the same time, and realizes the basic functions that query mechanisms to find structure and state information of resources, and controls the quality of service in resource management. Fabric layer provides the more rich functions, the more senior share operation it supports, (For example, if the resources layer supports function of obligate, it's easy to achieve the synergy scheduling services in the high layer; else services will have more additional spending in high layer). However, fabric layer provides fewer functions, so the structure is relative simply and it's relatively easy to realize it.

3.2 The work flow of spatial database in grid environment

Grid spatial database model provides unified accessing interface for grid users, presenting a single view of logical database, and the specific work flow is shown in Figure 3.

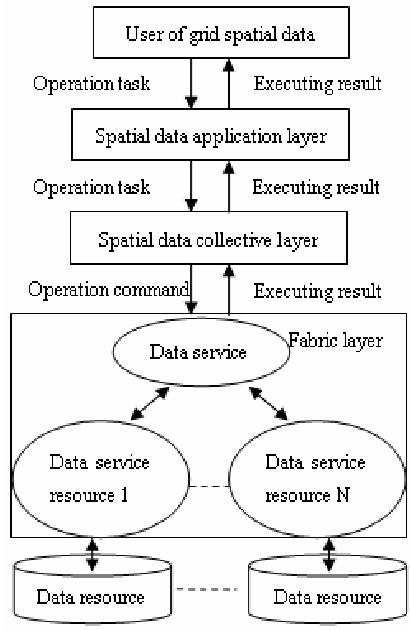

Figure 3. The work flow of grid spatial database model

The concrete steps of work flow are as follows:

1. The users of grid space send data request;

2. Spatial data application layer parses the user's requests according to registration information of data resources, and then builds an XML executing document. The document could be identified and executed by relevant module in system model;

3. Spatial data collective layer parses the executing document which receives from spatial data application layer, positioning the operation to grid spatial data resources, interpreting executing document and operation request, and sends the information to data fabric layer.

4. Data fabric layer executes the command from spatial data convergence layer, positioning and confirm the resources again. The results will be returned to spatial data convergence layer, in the end, spatial data convergence layer returns the organized process and results information to application layer in documents.

3.3 The simulation experiment

According to the designed grid spatial database model, the experiment of system simulation has been carried on. So we selected the spatial database resources at seven different areas as the grid nodes, it is shown in table 1:

Stand	Computing unit (Number)	Memory cell (GB)
Nanchang	0	30
Wuyuan	1	3.3
Sanqingshan	1	6.6
Lushan	1	5.0
Ganzhou	1	8.0
Jinggangshan	1	3.3
Longhushan	1	5.0

Table 1. The resource table of grid spatial database

The network topology which is in these resources stands as well as the network bandwidth are shown in Figure 4, in which these network resources provide the link for the transmission of spatial data in the various stands.

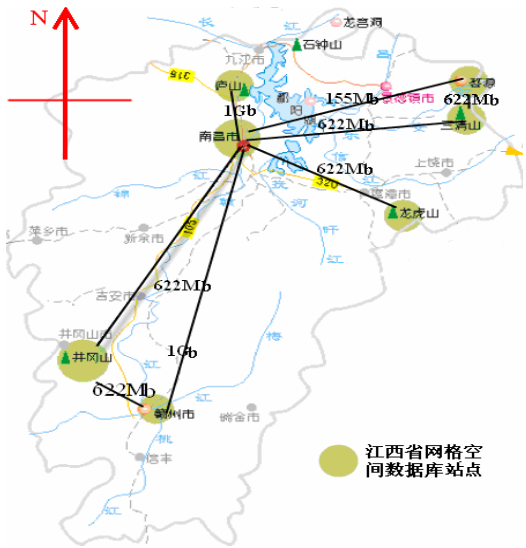

Figure 4. The network topology of grid spatial data stands

The disposition situation of spatial data file is shown in table 2, in which the general data file is used by each kinds of spatial data work, the spatial data file provides attribute data and layer data for the special spatial work, the file size and the processing time carry on certain proportion to reduce based on the actual content of usual process, and the index is used in the optimization algorithm for prediction.

The disposition table of spatial work is shown in table 3, because each stand all may process these works, the work selected probability comes from the statistical information of usual work-handle times.

File type	File number	File size (GB)	File accession No	work-handle time (second)
Universal	ty1,ty2	1.0,1.0	0,1	5
spatial data 1	ts11,ts12,...,ts18	0.5,0.5,…0.5	10,11,12,…,17	2.5
spatial data 2	ts21,ts22,...,ts28	0.3,0.3,…,0.3	20,21,22,…,27	2.5

Table 2. The disposition table of spatial data

Work type	Required file number	Computing unit which can process work	Selected probability
Spatial work 1	ty1,ty2,ts11,ts12,...,ts18	All stands which have computing unit	0.68
Spatial work 2	ty1,ty2,ts21,ts22,...,ts28	All stands which have computing unit	0.32

Table 3. The disposition table of spatial work

On the basis of above configuration, different work amount has been selected as a comprehensive comparison to measure the performance of grid spatial database. According to grid spatial database, the experiment tests have been operated for spatial data work of 500, 1000, 1500 work amount. To compare the differences of data sharing and handling capability in the grid environment and non-grid environment, tests have been carried out separately under two environments, among them, table 4 is the work result under the traditional spatial database environment and table 5 is the work result under the grid spatial environment.

Work amount	Utilization efficiency of memory cell	Average operation time （second ）	The harvesting integrity of Isomerism data	Network utilization
500	64.70%	80.582	95.70%	63.10%
1000	64.70%	79.905	96.50%	59.46%
5000	64.70%	79.169	98.89%	58.14%

Table 4. The simulation results of processing spatial data in grid environment

According to the experiment data, the results are as follows:
Based on non-grid mode, the average operation time becomes very unstable, the network utilization is very high, the network resources are consumed high, the utilization efficiency of memory cell is low, and the harvesting rate for the isomerism spatial data is not high.

In grid environment, with the increasing work amount, all the memory cell can be used enough; the average operation time decreases; the network utilization shows a downward trend as a whole, which has great improvement for the harvesting rate of the isomerism spatial data.

4. CONCLUSIONS

Based on the analysis for the advantages of grid computation, the deficiency of traditional spatial database and the requirement in special domain, this paper has built the spatial database model in grid environment. In comparison with simulation experiments, the result has shown that the performance of the spatial database in grid environment is better than traditional spatial database, with the former model, the client is more transparent, the spatial database has better hierarchical structure, and the operation efficiency is higher, which is beneficial to improve the isomerism spatial data sharing.

REFERENCES

Jin Qiangyong, Li Guanyu, 2002. Development and Present Situation of Heterogeneous Data Integration Technology. *Computer Engineering and Applications*, 38(11), pp.112-114.

Li Deren and Guan Zequn, 2001. *Integration and Realization of Spatial Information System*, Wuhan Survey Scientific and Technical University Press, Wuhan, pp. 152-163.

W.H.Inmon, 2000. *Data Warehouse*. China Machine Press, Beijing, pp. 46-52.

Wang Shan, Zhang Kunlong, 2004. Database System on the Grid. Computer Applications, 24(10), pp.1-3,23.

Wang Yijie，Xiao Nong, Ren Hao, 2002. Research on Key Technology in Data Grid. *Journal of Computer Research and Development*, 39(8), pp.943-947.

Written by Microsoft Corporation, translated by YunZhou Workroom, 2002. *Microsoft SQL Server 2000 Resource Kit*, China Machine Press, Beijing, pp.112-124.

Xiao Xiaocong, He Xiaojian, Zhu Jianjun, 2006. A Preliminary Study on Grid Services Based on Ontology. *Journal of Chongqing Institute of Technology*, 20(8), pp.113-115.

Xu Zhiwei, Li Wei, 2002. Research on Vega Grid Architecture. *Journal of Computer Research and Development*, 39(8), pp.923-929.

Yu Zhihui, Chen Yu and Liu Peng, 2002. *Grid Computing*, Tsinghua University Press, Beijing, pp.67-84.

Zeng Yi, Zhang Wanjun, Yang Jin, Li Guoqing, 2005. An encapsulation method for the database access in grid application based on OGSA. *Journal of Huazhong University of Science and Technology. Nature Science*, 33(1), pp.32-33, 37.

Zhang Bo, 2001. *Research on Distributed Software Architecture Based on XML*, Institute of software, Chinese Academy of Sciences, Beijing, pp. 53-60.