A Unified Framework Of Distributed Spatial Database Orienting The Mass Remote Sensing Image Data

Jun Ma[a, b], Yanan Wang[a, b*], Daojun Han[a, b]

[a]Computer and Information Engineering College, Henan University, Kaifeng Henan, China
[b]Institute Of Data and Knowledge Engineering, Henan University, Kaifeng Henan, China
e-mail: wyn_gx@163.com

Abstract—During the process of applying remote sensing image data, how to save the mass remote sensing image data of different times and use them effectively is an important problem. Researches show that the traditional large commercial relational database cannot storage the mass remote sensing data effectively and also cannot support unlimited data storage. In this paper, we combine the file system, relational database SQL and embedded database SQLite to build a distributed spatial database framework for storing and processing massive remote sensing data. On the one hand, data storage in this framework can break from the traditional large commercial relational database; on the other hand, it can satisfy the wide application demand, support multiple attribute data retrieval condition, have "instant" response ability. Experimental results show our method is effective, and it could be enhance the global search efficiency.

Keywords-remote sensing data;slice;sqlite files;distributed spatial database

I. INTRODUCTION

In recent years, the remote sensing image data rely on its features that wide coverage and short accessed cycle, it has been widely used in the land resources, mapping, urban planning, agriculture, geological surveying, marine, environmental protection and many other industries [1] . Thus, how to manage mass remote sensing data is becoming an important issue. Many academic institutions have carried out the relevant research plan or project and achieved remarkable results. Jin peidong [2] and others propose with mature relational database and GIS platform, this database can improve the efficiency of image management effectively Using the Oracle Spatial and InterMedia plug in which provided by Oracle, Tao zhiyu [3] and Zhang Nan [4] realized the image database based on the commercial database. Using Web services and ArcGIS Server, Yuan deyang [5] and Xing Yong Chao [6] achieve the remote sensing data interoperability and image resources effectively sharing. Shen zhanfeng [7] and others realize remote sensing image distributed processing system it solves the demand of remote sensing image storage management and quick query and positioning. Wang ailing [8] combines the file and database to set up the multi-scale remote sensing image database and realizes the image data management and different scale image link and meet the various needs of the users. Luo JianCheng [9] and others design the distributed remote sensing image database which based on Web Service

and puts forward to use Oracle managing data and realize network transmission through the SOAP/XML. In the summary, the management of massive image data in three ways [2-9]: The traditional relational database, co memorial database-based and NAS and SAN universal mass storage-based. Advantages and disadvantages of them are described as Table 1.

TABLE I. ADVANTAGES AND WEAKNESSES OF THE MANAGEMENT SYSTEM FOR MASSIVE IMAGE DATA

	Advantages	Disadvantages
The traditional relational database	Mature DBMS technology, System development cost is low.	Image data to support limited, System structure and data management ability by DBMS, DBMS acquisition cost is higher.
Based on the commercial database	Mature DBMS technology, Directing at the remote sensing image, Rich access interface, System development cost is low.	System structure and data management ability by DBMS, Data management is too heavy, Commercial database acquisition cost is very high.
Based on the NAS and SAN universal mass storage	Data transmission speed, Storage capacity big.	Universal storage, no special for remote sensing data, expensive.
Based on the distributed file system	Not use DBMS manage data and reduce the acquisition cost, Form file storage level to optimize system performance.	Need develop Data management and data access layer interface，difficulty is greater and development cost is higher.

From the analysis of table 1, we can know that the remote sensing image database system architecture at present from centralized to gradually, from small-scale cluster to large-scale cluster and from relying on large commercial database to light small database. This paper first analyzes the characteristics of the distributed spatial database, and combined with the advantages of the file database, relational databases SQL and embedded relational database SQLite; followed, we select HJ out data for example, design a distributed spatial database for massive remote sensing images storing according to the characteristics of remote sensing image data and data optimized split ways; finally, we give the test results of the storage system. And experimental results shows that this system has the high efficient concurrent retrieval access ability, data storage environment

is scalable distributed environment and it can meet user's quick access requirements, data storage and management completely break from the traditional relational database restriction.

II. DISTRIBUTED SPATIAL DATABASE

Distributed Spatial Database (Distributed Geo - Spatial Database, DSDB) is a unified space Database which connect the space Databases that physically dispersed but the management and control centralized inordinately together [10]. In the network every database server can be placed in a place and per database server saves a complete copy of the database management system. The entire database server forms a complete database system through the computer network and provides users with global view [11]. Spatial information acquisition will be done by different organizations, so it is difficult to focus copy to the same node; the distributed system conforms to the characteristics of the spatial databas [12].

Compared with conventional database, distributed spatial database has some special features, such as a large quantity of data storage, widely used for the purpose, in the physical distribution and good network environment for data sharing. Those features determine the complexity of the database environment configuration during the database design and construction process. In order to supply fast response when dealing huge amounts of data, distributed spatial database must adopt certain mechanism to ensure the mass data transmit rapidly between the service nodes. Obviously, this mechanism is a key consideration during the construction of the distributed space database.

A. Remote Sensing Image Data Split

Due to mass remote sensing image data storage capacity up to TB and even to EB, which need to add a data split processing while construct the spatial data model to realize its distributed storage. A good split data organization model can improve storage speed and efficiency, retrieval speed and memory usage efficient of every storage node. Therefore, subdivision model is an important component while design the mass remote sensing image distributed storage system. Observe that the most common remote sensing image data, generally including satellite images and aerial image and which has multiple formats, HDF, H5, TIF, TIFF are most used. Therefore, Image Pyramid was selected as a popular subdivision model. In the ideal circumstances, tile pyramid should be a continuous resolution model, but in practical application, the really resolution continuous change is not exist. Deng Xueqing [12] points out that the magnification method usually used to form a plurality of resolution levels, that is, from the bottom of the pyramid to the top, the plurality of sampling points were combined for a reduced resolution between adjacent layers, all layers range indicated table. Kang Junfeng [13] proposes a method based on threshold value of the remote sensing image pyramid create algorithm, and Yang jingYu [14] puts forward the remote sensing image pyramid model tile data and tile index separation, both through the pointer associated can separate pyramid storage model. They have improved in the creation

of image pyramid and retrieval speed, but the application of the image accuracy problem was omitted. Obviously, the combination of remote sensing image data and 3D display and simulation system is a main trend for the application of the industry, display and processing of the tiles should be the greatest degree of true. To solve this problem, A Five-story fifteen remote sensing data organization structure is proposed by Wang Dong [15].Different from the traditional pyramid section way, the section way that Five-story fifteen remote sensing organization structure model adopt is as follows: In the layer ratio is 5:2. 5:1; the proportion between layer and layer is 10:1; different resolution image data will mean corresponding to different levels; each section will stand for a pair of 1000 * 1000 image. This section way can ensure that the different image segmentation in each level resolution has consistency and practicability and the tile image accuracy is higher than traditional pyramid tiles. And image structure pyramid model based on Five-story fifteen layers could be applied to some GIS applications conveniently.

B. File Naming Conventions

Remote sensing image data has its unique characteristics, one of the important features is filename, which contains some image meat information, and we can obtain the unique corresponding relation by filename analysis. We select HJ1A data as an example, Figure 1-a represent the primitive compressed files, and Figure 1-c expressed the files that belong to compressed, then, section is mainly processing documents the four band image files, such as band 1:

HJ1-IRS-454-63-20100624-L20000331138-1.tif

In figure 1-d , sliced file directory can be observed under the sliced naming rules of figure 1-b. The band number part also may be Zenith Angle and Azimuth Angle (Azimuth/Zenith), its sliced name determine according to the requirement of system.
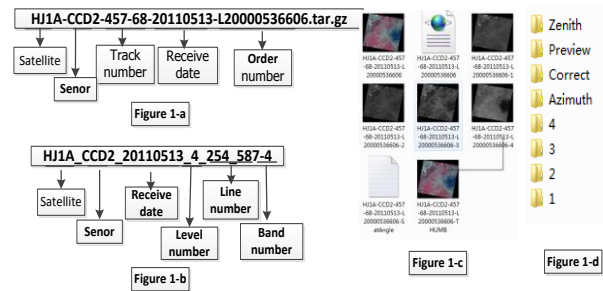

Figure 1.   File Naming Convention

III. DISTRIBUTED SPATIAL DATABASE DESIGN FOR REMOTE SENSING IMAGE DATA

Based on the analysis about the distributed spatial database, the characteristics of remote sensing image data and file naming conventions, we carried out a solution for massive remote sensing image data. At present, HJ data and CEBERS data are the main store data saved in distributed spatial database that designed for remote sensing image data.

Figure 2 indicate the network topology structure, and the whole system is divided into three parts: storage site machine parts, the database server part and control part of the server. Followed, the every function of three parts and relations of them are introduced.

A. Storage Site

Storage site machine part denotes the actual location of the data storage. In this part, the complete image data and processed slice data are stored using independence files. Obviously, load balance is an important issue. In order to ensure all sites load balancing, and make the data transmission and storage not focus on a storage site, we publish Web Service on all storage sites and get the station site hardware information to control machine. That information are rich and necessary, which include that the host machine whether damaged, the amount of disk available space, and the current network traffic. Comparing all of the site machine's states and choosing the best site machine as data storage site when there is data need to store. Under the direction of load balancing, the system can prevent the unnecessary troubles and maximize use of the network to accelerate file transmission, for example, the conditions, system is lack of disk space or some of its hardware become invalid can be avoided. The purpose of store total data is to make use those data freely in practical applications.

On the condition of mass remote sensing image data is quickly stored, how to ensure the rapid retrieval of the future desired result is another major problem while construct database system. To ensure the efficient data retrieval on the basis of data fast stored, we placed in the top-level directory of each storage site storage file with a SQLite database file, which record the information of the slice file that stored on the site machine. Because of every scene image data segmentation will produce six or seven hundred small slice files, if these slice file information are stored in the database server SQL database, the amount of the data will be very big, and which will cause data retrieval is troublesome and very time-consuming. Furthermore, deploying a SQLite file on each site can trigger the slice data warehousing operation after the file upload is complete and the information are saved in the SQLite file. When retrieval file, all sites have opportunity to be retrieved concomitantly and submit the retrieval results to center control machine. The storage site is more, and the speed of retrieval is faster.

B. Central Control Server

Central control server is key component of the whole system. On the one hand, it is responsible for monitoring the state of the site machine in the entire system and getting the optimal location of the data stored. On the other hand, it is responsible for all interactions with the user's operation.

C. Database Server

The database server deploys some SQL databases. This part mainly provides and manages the information of the site machines and supply the encoding of the data tables and the relationship between them. The table structure is shown in Figure 2, and those tables are used to save the information of the site machine and source data. In Figure 3, the process steps about uploading, downloading and retrieval of the file are described.
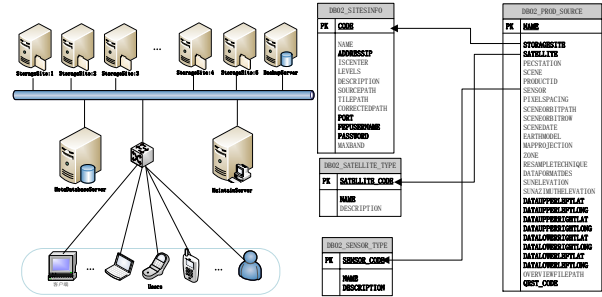

Figure 2.   network topology mechanism and database table structure
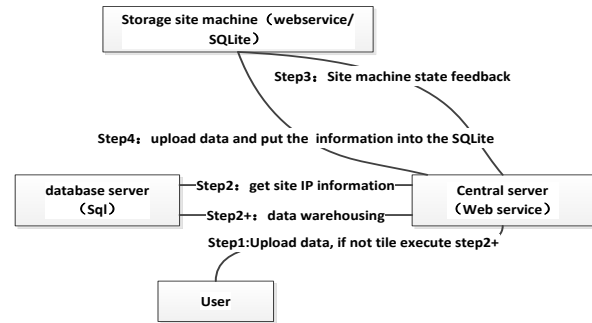

Figure 3.   simple upload processes

The system begin from a library content of database to provide database engine for the basic operation of the database system, and then to specific functional package of services of various database systems and finally integrates and release service through the database service platform. The database system utilize Web services to announce its services including data directory service, data retrieval service, data download service, data upload service, data entry service as well as other WMS, SDE type of access service.

IV. CASE STUDY

Based on the above studies, the performance of the distributed file storage in the LAN cluster environment is implemented. Laboratory environment configuration LAN speed of 100Mbps and 1Gbps, all machines have the same configuration of hardware and software, and configuration details are listed as Table2.

TABLE II.   THE CONFIGURATION OF HARDWARE AND SOFTWARE

Database software	CPU	physical memory	rigid disk	operating system
SQL Server, SQLite 3.8.7	Core 2 Quad8300@2.5GHz	4GB	500G 7200revolutions per minute	Window XP Pro 32-bit

In order to verify the efficiency of the Framework. We built a small distributed file system. The test result in single node distributed environment is show in table 3, the result in Multi-Node distributed environment is show in table 4, the

result in normal circumstances is show in table 5 and contrast result is shown in table 6.

TABLE III. SINGLE NODE DISTRIBUTED ENVIRONMENT

file size (GB)	MB card takes(s)	MB card transmission speed(MB/s)	gigabit time-consuming(s)	gigabit transmission speed(MB/s)
2.36	235.6	10.15	24.85	97.48
224.9	21987.3	10.23	2309.6	97.76
742.1	68020.1	10.91	7500.3	96.64
1470.2	135377.6	10.86	14841.2	99.11

TABLE IV. MULTI-NODE DISTRIBUTED ENVIRONMENT

file size(GB)	MB card takes(s)	MB card transmission speed(MB/s)	gigabit time-consuming(s)	gigabit transmission speed (MB/s)
2.42	200.7	12.18	24.58	102.62
524.3	41743.2	12.56	5170.2	101.41
1534.6	120313.7	12.75	14761.7	103.96
2762.6	221527.2	12.77	26924.8	102.65

TABLE V. NORMAL CIRCUMSTANCES

file size(GB)	MB card takes(s)	MB card transmission speed(MB/s)	gigabit time-consuming(s)	gigabit transmission speed (MB/s)
2.36	297.8	8.13	36.8	67.52
224.9	27394.6	8.21	3297.2	68.21
742.1	87203.3	8.51	11136.2	66.64
1470.2	180171.6	8.16	21907.3	67.11

TABLE VI. CORRELATION DATA

operating environment	storage space	MB card transmission speed(MB/s)	gigabit transmission speed (MB/s)
normal circumstances	limited	8.31	67.24
Single node distributed environment	limited	10.54	97.58
Multi-Node distributed environment	theoretically infinite	12.62	101.43

From experimental result we can see that the transmission speed in distributed concurrent system is much higher than it in centralized network system. The distributed transmission consuming is significantly lower than normal conditions and its transmission is 12.5MB/s or so, it is higher than the general environment 8.7MB/s. Under the distributed circumstances, the transmission efficiency is relatively stable. System will consume a certain amount of time and efficiency because of the need for data warehousing and automatic load balancing while the file transferring. The efficiency is very higher in distributed multi-node environment than it in the single-node; the test proves that with the increase of the number of nodes, automatic load balancing of distributed systems gradually improved and stabilized and will eventually achieve the best efficiency. Theoretically, in the case of nodes suitable, the bottlenecks restricts data

transmission will depend on the card and the network environment itself.

V. CONCLUSIONS

This paper associative file system, relational database SQL and embedded database SQLite design distributed spatial database for remote sensing image data to solve the problems of traditional database efficiency and accuracy in dealing with massive remote sensing image file, and achieve the desired result that managing effectively and using massive remote sensing image data. However, with the increase of distributed nodes, node failure will be the norm, and node failure causes the data unavailable, and it will reduce the reliability of the system itself, so the future work will focus on file redundancy mechanisms in the distributed storage system and improving system performance and reliability.

REFERENCES

[1] Liu Wei, Liu Lu. Mass remote sensing image data storage technology research [J].Computer engineering .2009:05.

[2] Du Xiao, Jin PeiDong et al. Researchs on the rapid importing of remote sensing images and metadata acquiring system [J]. Remote sensing for land & resources.2006,18(3).

[3] Tao ZhiYu, Ma DongYang et al. Design of Multi-Resolution Remote Sensing Image Data Based on Oracle [J].Journal of institute of surveying and mapping. 2005,22(1).

[4] Zhang Nan. Research on Key Techniques of Content-Based Optical Remote Sensing Image Retrieval [D]. National University of Defense Technology.2008.

[5] Yuan DeYang, Nie Juan et al. Design and implementation of metadata-based multi-source remote sensing image database integration technology [J]. Science of Surveying and Mapping.2012,37(3).

[6] Xing YongChao, Ma RongHua et al. Remote sensing image retrieval system design and implementation Based on ArcGIS Server[J]. Microcomputer Information.2012(02)

[7] Shen ZhanFeng, LuoJianCheng et al. Distributed remote sensing image database design and implementation based on WebService. 2006(02)

[8] Wang AiLing, LiJing et al. Multi-scale remote sensing image database management design and implementation [J]. Computer Engineering and Application. 2007(03).

[9] Luo JianCheng, ZhouChengHu et al.The Design of Middleware-Based Grid-GIS. Istic Pku. 2002,4(3).

[10] DiBenedetto M.G, De NardisL, and Junk M et al..(UWB) 2: Uncoordinated,Wireless,Base-born ,medium access control for UWB communication networks . Mobile Networks and Applications, Mobile Networks and Applications Special Issue on WLAN Optimization at the MAC and Network Levels, 2005,10(5):663-674.

[11] Yan Bin. Based on distributed space database engine design and practice [D].University of Electronic Science and Technology of China.2011.

[12] Deng XueQing. Grid spatial data service system structure and algorithms [D].The Chinese people's liberation army information engineering university.2003.

[13] Yang Jingyu, Zhang Yongsheng. Based on wavelet transform multi-resolution analysis characteristics of the remote sensing image pyramid model of dynamic reconfiguration[J]. Science of Surveying and Mapping.2007,32(5):50-52.

[14] Wang Dong. Ecological products parallel production subsystem and related technology research[D].Henan University.2011