A METHOD ON CONNECTION POOL SERVICE FOR DISTRIBUTED HETEROGENEOUS DATABASES IN URBAN GEOGRAPHIC INFORMAITON PUBLIC PLATFORM

Dezhu GUI[a,b], Gang LI[a], Chengcheng Zhang[a,b], Pengcheng Yin[a]

[a]School of Environment Science and Spatial Informatics,
China University of Mining and Technology, Xuzhou, Jiangsu, 221008,China
[b] Chinese Academy Surveying and Mapping, Beijing 100039,China - guidezhu0930@163.com

KEY WORDS: Distributed Database, Connection Pool, CSCW, Data Process, Public Platform

ABSTRACT:

Today with the fast advancement of spatial data acquisition ability, the update circle of data becomes shorter and shorter and the volume of spatial data increases more and more, and distributed storage and extraction becomes the primary management means of spatial data. The geographic information is collected and managed by a multitude of private and public agencies. Distributed infrastructure in urban geographic information public platform(UGIPP) has traditionally tackled the problems associated with the inter-connection of a number of computer systems in urban departments and companies. Thus we have to be up against the access difficulty of heterogeneous distributed spatial data and collaboration spatial data process. Efficiency of database connection and collaborate spatial data process becomes one of the main factors that affect WebService, and how to optimize it is very important. This paper summarizes the results of research and development towards internet-based support of collaborative spatial data process, and discussed key issues in applying CSCW technologies in distributed spatial data environment. The results indicate that connecting pool, collaborative spatial date process, groupware principles can be effectively brought together and applied to provide better solutions to implement distributed heterogeneous spatial data share and service.

1. INTRODUCTION

With the development of spatial information science, the amount of spatial data acquisition increased rapidly. Web-based geographic information share and service is a coupling of web service and geographic information system, which changes spatial data management into various data and function services in Internet. Distributed spatial databases and collaborative spatial data process are the core of urban geographic spatial data share and service. The availability and practicality of share and service are determined by quality of services and both aspects have influence on the popularity of service. A high efficiency database connection is the need of high performance of WebService. Connection pool is an efficient strategy for the database connection and collaborative process of distributed heterogeneous databases. The internet-based computer supported cooperative work(CSCW) concepts and related technologies have been increasingly integrated in many areas to support collaborative process in distributed work environment, through which collaboration among group of people located at various geographical locations can be realized. Aiming at solving this problem, based on the analysis the signification of distributed heterogeneous database management in urban geographic spatial data share and service, this paper puts forward the dynamic connection pool service to dynamically respond requests, assign database connection and collaborative spatial data process of urban geographic information public platform and focuses on the efficient querying and quality of service of this distributed infrastructure.

2. ESSENTIAL OF URBAN GEOGRAPHIC INFORMAITON PUBLIC PLATFORM

Spatial databases are prominently used in Geographic Information System (GIS) applications. However, many of the current architectures rely on a centralized data repository. The next evolution will be GIS applications that utilize and integrate a multitude of remotely accessible data sets, for example via Web services. The combination of spatial database systems and Web services promises to form the foundation of powerful new applications that dynamically integrate data from multiple, distributed repositories. The use of a Web services infrastructure allows direct programmatic access to remote data and applications have the opportunity to obtain a plethora of data without storing and managing a lot of information locally. There are clearly a number of advantages to this concept. First, data is often maintained by specific entities or organizations. By allowing remote access to their data sets, these entities continue to have administrative control over their data. Second, the correct data set for a specific calculation can be downloaded automatically without manual user intervention, reduced data acquisition time. Third, updates and changes to the data are almost instantly available to remote applications.

The geographic information is collected and managed by a multitude of private and public agencies, such as the planning bureau, the land resource bureau, etc. Given such a federation of spatial database servers, agencies that depend on distributing geospatial data over the Internet benefit it. Companies, public and government departments need geographic information for various applications in commercial trade, daily life and urban management, e.g., mapping of natural resource, transportation and urban plan. For instance, land resource management agency relys on available geographic data for producing land use maps,

from which real estate developers evaluate land values and calculate house price. A distributed geographic information management and exchange platform based on a spatial database engine to maintain these geographic data sets is essential.

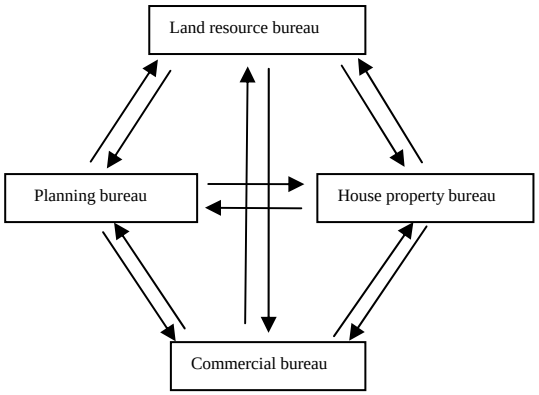

Figure 1. Traditional many to many data exchange model

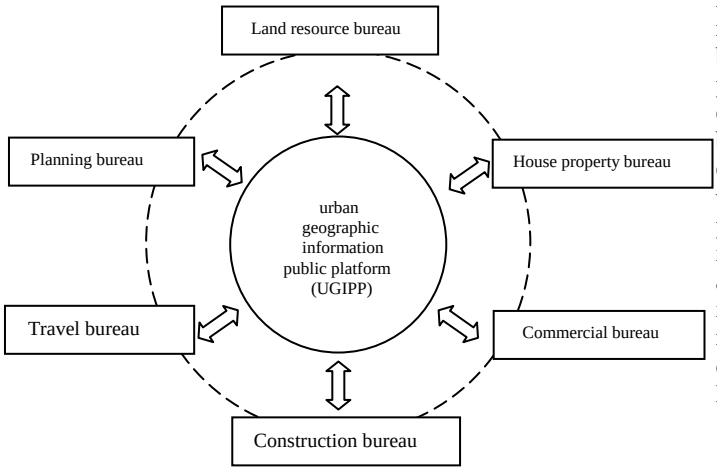

Figure2. One to many data exchange model based on UGIPP

3. EXIST PROBLEM

Distributed infrastructure in urban geographic information public platform(UGIPP) has traditionally tackled the problems associated with the inter-connection of a number of computer systems in urban departments and companies. The emergence of Computer Supported Cooperative Working (CSCW) has challenged this technological perspective by focusing on the group-working of a number of users. The internet-based computer supported cooperative work(CSCW) concepts and related technologies have been increasingly integrated in many areas to support collaborative operation in distributed work environments, through which collaboration among group of people located at various geographical locations can be realized. With every user trying to log in or log out, our geographic information public platform will issue calls to these validate user methods. Every call to validate user method will result in establishing a separate connection with the database. This is definitely not the best approach. A few problems with this approach:

1. There are many types of database in urban geographic information public platform(UGIPP), so heterogeneous databases accessing is high frequency.
2. Different users maybe access and operate the same database simultaneously.
3. Every user trying to log in 'waits' for a connection to be established with the database.
4. There is an overhead with opening and closing the connection with every request.
5. We are not checking conditions such as 'if the no. of allowed connections is over'. Such a condition will result into our validate user method not getting a valid connection.

These types of problems are easily solved with the use of a connection pool.

4. DATABASE CONNECTION POOL

4.1 Database connection pool

Connecting to a database is a time consuming activity since the database must allocate communication and memory resources as well as authenticate the user and set up the corresponding security context. A connection pool is a container of open and reusable connections. A connection pool is released from the memory when the last connection to the database is closed. The basic advantage of using connection pool is an improvement of performance and scalability while the main disadvantage is that one or more database connections, even if they are currently not being used, are kept open. Connection pool gives you an idle, open, reusable connection instead of opening a new one every time a connection request to the database is made. Multiple pools can exist in the same application domain at the same point in time, but connection pools cannot be shared across application domains. When the connection is closed or disposed, it is returned to the pool and remains idle until a request for a new connection comes in. If we use connection pool efficiently, opening and closing of connections to the database becomes less resource expensive.

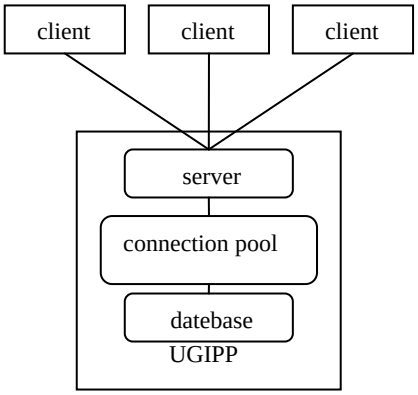

Figure3. Connecting to distributed heterogeneous databases

4.2 Basic idea of database connection pool

Database connections are expensive in terms of performance to make & re-make, so rather than making and breaking connections as required, a "pool" of database connections is maintained by the system on the platform server. The basic idea of a connection pool is follow.

4.2.1 Open and shutdown connection pool

The connection pool object when instantiated will open a predefined no. of connections (say 20) with the database. It will have methods like getConnection that will return a reference to one of these 'already established' connections. Similarly a returnConnection method will return the connection back to the pool. A servlet of UGIPP will instantiate this connection pool in its 'init' method. Note that the init method is executed only once when the servlet of UGIPP is loaded. We should always open the connections late and release them early for improving connection pool performance; in other words, immediately after we are done using it. Connections should be opened only at the time when they are actually required. The connections should not be acquired prior to its usage as it would decrease the number of available connections in the connection pool and, hence, have detrimental effects to the operation of the connection pool and the application's performance. The connections should be explicitly released immediately when we are done using it. This would facilitate better connection pool as the connection would be returned to the pool and be available for reuse.

4.2.2 Monitor connection and manage spatial data collaborative process

It will keep track of returned and 'in use' connections. Status servlets allow the state of the connection pool to be viewed at any time, and every connection has a description that identifies it's current and last use, as well as the last time it was accessed. Connections can also be "queued" - if there is a specified maximum number of connections to the database (due to limitations on either the database itself or for performance reasons), then objects requesting a connection can be "put on hold" for a few moments until another object releases its connection. The "waiting time" can also be configured to avoid unacceptable delays. In addition, background and low-priority jobs can be suspended or queued when interactive requests are waiting for connections. It is necessary to check the connection pool

4.2.3 Manage spatial data collaborative process

Conflict problem often appears in the spatial data processing. The system deepfreezes and even entire paralysis because of violating predetermined constraints and regulations. According to the characteristics of spatial data, cause of conflict is the dependence or independence relationship of different users of the UGIPP. The resource, capability and strategy of problem processing of each user is different, so, the dependence and independence are always contradictory.

Spatial data collaborative process between the inter-connection of a number of computer systems is very important. When a user start to process, he(or she) transfers the operation command to another user through internet. The core of the process mode is the control of the platform connecting centre.
The method of process includes these parts:

1. Other users can browser, but can not amend the object that under processing and uncompleted
2. Sequencely operation and lock process during data access.
3. Keep data integrity and consistency when data handling exception.

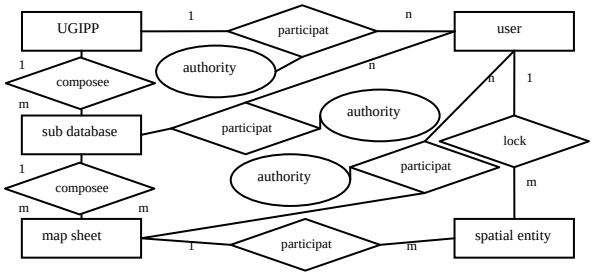

Figure4. Conceptual model of collaborative spatial process

4.2.4 Allocate and release connection

Recycling and reusing already existing connections to a database is more efficient than opening a new connection. Once created, connection pools are not destroyed until the active process ends. Maintenance of inactive or empty pools involves minimal system overhead. When the maximum pool size is reached, the request is queued. The object pool satisfies these requests by reallocating connections as they are released back into the pool. If the time-out period (determined by the Connect Timeout connection string property) elapses before a connection object can be obtained, an error occurs. The object pool will remove a connection from the pool if the connection lifetime has expired, or if the pool detects that the connection with the server has been severed. Note that this can be detected only after attempting to communicate with the server. If a connection is found that is no longer connected to the server, it is marked as invalid. The object pool periodically scans connection pools looking for objects that have been released to the pool and are marked as invalid. These connections are then permanently removed.

5. CONCLUSION AND PROSPECT

Each time a resource attempts to access a database to process spatial data, it must connect to that database. A database connection incurs overhead, it requires resources to create the connection, maintain it, and then release it when it is no longer required. The total database overhead for an application is particularly high for Web-based applications because Web users connect and disconnect more frequently. In addition, user interactions are typically shorter, because of the nature of the Internet. Often, more effort is spent connecting and disconnecting than is spent during the interactions themselves. However, we should use connection pool appropriately as using it inappropriately might have negative effects to the overall performance of our applications. MSDN says, "Connection pooling is a powerful functionality that can improve your applications' performance. But if you aren't a good lifeguard, your connection pools can become a detriment instead of a benefit." This article has discussed connection pooling and how we can use it efficiently for improving the performance of and scalability of spatial data collaborative process. Although web-based geographical information services provides proprietary ways to allow users to quickly access, display and query spatial data over the web, the heterogeneity of existing spatial database and the file-level data sharing systems over the web causes problems for time-critical spatial information management that need real-time data access to the most up-to-date information. Database connection pool will solve a lot of our performance problems but there are still a few more things to be considered: the ability to balance load with multiple servers of urban

geographic information public platform(UGIPP); fault tolerance of spatial data collaborative process and so on.

REFERENCES

Jizhou, Wang., 2003. Research on connection pool service for distributed heterogeneous databases in WebGIS. *Geography and Geo-information Science*, 19(3), pp. 18-21.

Ruyan, Cheng., 2007. Database connection pool technology based on JavaBean in WebGIS. *Geospatial Information*, 5(4), pp. 76-78

Cunming, Zhao., 2005. QoS and key issues in WebGIS. *China University of Geosciences*.Wuhan, China

Jiping, Liu., 2004. A research on spatial decision-making for e-government. *PLA University of Information Engineering*. Zhenzhou, China

Yaser, B., 2000. Internet based large distributed geospatial databases.*International Archives of Photogrammetry and Remote Sensing*, 33(4), pp. 126 -132.

Stasik, M., 1999. Design considerations for space and time distributed collaborative spatial decision, *Journal of Geographic information and decision analysis*,1(1), pp. 1-8

Haiyang, Hang., 2000. B/ S based, distributed, heterogeneous spatial database intergration.*Journal of Remote Sense*, 4(1), pp. 77-81.

Jianshou, Kong., 2002. An implementing technology for web-based synchronization CSCD system, *Journal of Nanjing University of Science and Technology*,26(3), pp. 251-254

Yingwei, Luo., 2002. The model of distributed GIS oriented multiagent system. *Acta Scientiarum Naturalium Universitatis Pekinensis*,38(3), pp. 375-376

Yanyou, Qiao., 2001. Techniques in distributed spatia l database management. *Journal of Image and Graphics*, 6(9), pp. 874-876