

Selection and Research for Online Registration System's Database System

Yabing Chen¹, Haiyong You²

¹Department of mechanical and electrical engineering, Yantai Engineering & Technology College, Yantai, China; ²Electrified education centre, Yantai Engineering & Technology College, Yantai, China.

Received 2013

ABSTRACT

Based on the Online Registration System (ORS) characteristics and key technology analysis, this paper points out that that a good performance and high stability of the ORS lies in the choice of the system database. Database clustering technology which has merits such as concurrent processing, easy expansion, and high security is proposed to achieve database subsystem of ORS, and the design of the database cluster system framework is available in this paper. Finally, we also explore the database load balancing of the cluster system, heterogeneous database replication technology.

Keywords: Online Registration System; Database System; Database Cluster System; Load Balancing

1. Introduction

With the rapid development of modern information technology, the internet characterized with convenient, fast, as well as rich information becomes ubiquitous in most cities and even countryside. Furthermore it is also run widely in various educational resources including the common entrance examination, examination of the adult, independent recruitment examination institutions at all levels, even distance education as well. As the most focused interface of the network education, the online registration technology has been developed rapidly.

Project of Online Registration System (ORS) research and development was audited by the Shandong Provincial Planning Leading Group of the Education & Science in May 2011. The choice of database system, one of topics in the project is fully researched in this paper.

2. The Characteristics and Key Technologies of ORS

In my studies the main framework adopts three B/S layers which are browsers and negligible size of maintenance base installed by clients, the Web server kept in the middle layer to probe the client's requests, as well the underlying database which is visited by the web server.

Unique features of online registration technology mainly focus on below points.

1) The data access period converged relatively. In usual, simple work just to do the routine maintenance because of seldom accesses by user. Nevertheless during the examination stage, amount of database access will

surge abruptly, not only by administrators to frequently maintenance but by users to access.

2) The lethal factor of data security. Once the information leakage occurs, it is not only the personal information leakage problems, but even affect the entire test process. The risk of large-scale examination security is increased proportionally.

3) Data types which are mainly text data and a few of date or figures type that supports candidates photos are relatively simple. However, the surged candidates take vast volume data for their private information storage.

In the ORS technology, data processing is consisted with several complicated working processes. One of them, the data response speed is the key technical point. Therefore, capability of soon data response depends on the high reliability and fast response ability of database in the whole project.

In light of the above analysis, the ideal state for the ORS technology should meet the flexible requirement that the usage of the database system should occupy less hardware resources as possible in usual while be able to facilitate expansion, integrated the necessary hardware and software resources to meet the spurt of the great access demand. Ideally, the structure of ORS is shown in **Figure 1**.

3. Technical Requirements Towards the Database by ORS

To meet the requirements of surge accesses and tight security towards users signed in information, the database of the ORS should adopt the following strategies:

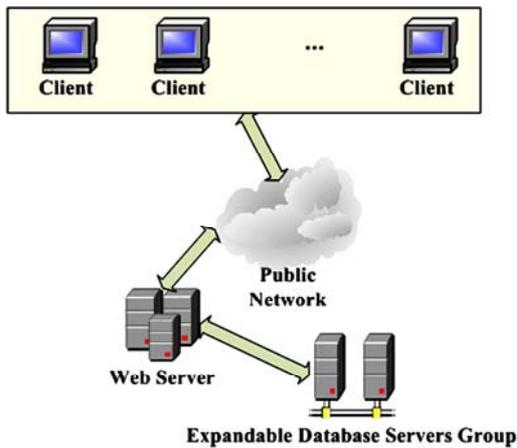


Figure 1. The favorable structure of ORS.

- 1) The huge data storage, processing, processing capacity
- 2) The ideal response time for user requests
- 3) The stability and robustness towards physical database running
- 4) The facility of the database itself
- 5) The security of the data confidentiality.

For most stylish database system, desirable services are available towards above stability and robustness. And most of databases have not addressed the security audit data analysis tools, in order to improve and perfect the current databases security audit mechanism, we can use data mining technology, data mining technology to the database security audit, the characteristics of database records for analysis.

But on the view of strategies huge data processing and response time for user requests, efforts to improve possibility of efficient database system usage by users are never suspended as well. In fact, the strategies huge data processing and response time for user requests require not only a database management system (DBMS) single effort but perplexing factors including a database system hardware configuration, Web connected and so on. Consistent with the needs of huge data processing and response time for user requests, it is conventional argued to improve the capability of database server and storage volume, adjust the DBMS parameters and other means while database system selection and preparation. However the factual effect is controversial because of the wasted cost of mainframe and limited improvement of overall performance of the database system. Paralleled or distributed database systems are evolutionary methods. Paralleled database systems with high availability and scalability focus on parallel data processing capability [1] and emphasize one single database system that each of its node does not have autonomy so much so flexibility is constrained [2]. Distributed database systems emphasize the centralized control for the system. It is characterized

with the distribution of data and database coordination, as well it aims to realize the autonomous site and data sharing globally and transparently without requiring the usage of every node in the network to improve system performance [3]. Therefore, we prefer to use a database cluster system to meet the online application system for database system requirements.

4. The Choice of Database Cluster System

As far as the single database system selection, it is key to choose the proper DBMS which is selective from rich and varied candidates on current IT market. Mainstream DBMS products include DB2, Oracle, Microsoft SQL Server, Sybase SQLServer, Informix and MySQL which have their respective features and advantages. However, to the database designers, it seems obsolete DBMS products on project schedule until the operation funds and other preparations are available sometimes no matter how latest he chooses. Theoretically speaking, it is valid to use database cluster system to effectively alleviate the suffering mentioned and deal with more and more complex demanding for database system nowadays. Database cluster system refers a group of database server as one whole to manage through the network so that they can work coordinately to supply services for users, which can be heterogeneous. The cluster system has few limitation for the server operating system and all database servers are mutually independent to perform a given user request by itself, at the same time it is available to replicate data for backup, recover data and other functions in the group. When there are concurrent requests arrive, it can be invisibly for user to choose one certain database server node to process requests [4-5]. On the basis of the characteristics of ORS, one framework is proposed to use a database cluster system as shown in Figure 2 which can

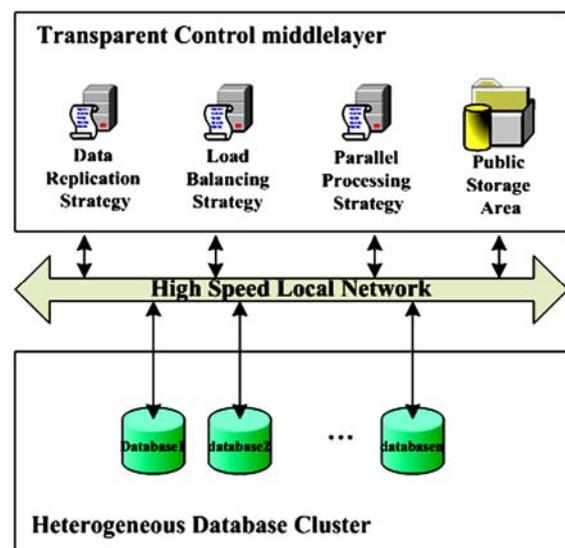


Figure 2. database cluster system framework.

conceive the volume of access in ORS to adjust on database cluster. It is negligible effect on other databases when a single database is adjusted, as well the storage area can be used as public data center.

For the database cluster system, features of reliability, high availability and high extensibility are quite suitable for ORS. Meanwhile there are some issues such as the load balancing, heterogeneous database replication, transparent access and database cluster technology which should be fixed in the practical application. The weighted least connection method is adopted in the processing of load balance which administrator assigns corresponding weight value for each node based on the database server nodes (hereinafter referred to as the node) performance because the value is representative of its processing performance and default is 1. So when a connection request arrives, the higher priority will assign to the nodes according to their weight value. That is to say, higher value node will get more links. After a while, it is possible that the link number of nodes assigned to becomes proportional to nodes weight value [6-7]. Based on this thought, the database administrator can consider total quantity value node according to the results by formula (1).

$$SN=W \times B \quad (1)$$

In the formula, SN means the maximum number of links that a node will receive. W means weight values, and B means average minimum number of links by each node. The W is valued by the hardware and software performance of database server node, while B is composed of number of database cluster head nodes and instant total visit volume. When the SN is beyond or close to the node DBMS tolerance value, it means cluster node number is too small and need to increase the database server nodes, whereas the SN is too low, means database node surplus and it is better to reduce the node properly.

On the side of heterogeneous database replication, most major database vendors provide exuberant schemes, such as Oracle converter, Sybase LTM solution, IBM CCD solutions, as well as the Microsoft Publisher / subscriber scheme [8]. One unified data interface is proposed in paper [9] for heterogeneous databases. It is also one kind of anticipated scheme to solve the data replication problem through the JDBC and XML.

It works on middleware form to fix the issue of access the database cluster transparently [10-11]. Middleware is the best option for the communication protocol of database cluster standardization and modularization. Middleware also can provide mechanism to solve heterogeneous problems among database cluster system.

5. Summary

In this paper, the analysis contents focus on the ORS basic structure and work characteristics, and then some

research on the database system proceeded. According to the study, it is lethal factor of the whole application system under the circumstance of surge access during the register period. Because it is hard to meet needs of the performance of the database with conventional methods including purely raising the performance of database hardware, modifying the DBMS parameters and using connection pool technology, this paper puts forward one database cluster system with favorable concurrency, flexible expansion and high-security based on the framework of database system. What's more, researches also proceed in designing of the database cluster system framework, and issues of load balancing, data replication.

REFERENCES

- [1] R.Buyya, T.Cortes, H.Jin. Single System Image (SSI). *Journal of High Performance Computing Applications*, 2001,15(2):124~135
- [2] Gong Weihua. Study on the key issues of database cluster system[D]. Huazhong University of Science and Technology, 2006.
- [3] Li Chongxin. Design and implementation snapshot for distributed database HBase[D]. Zhejiang University, 2011.
- [4] Murthy, V.K. High performance cluster computing using component-oriented distributed systems[C].1st International Conference on e-Science and Grid Computing, Melbourne, Australia: First International Conference on e-Science and Grid Computing, e-Science 2005, 2005:522~529.
- [5] Z Shao, H.Jin.HARTS. High availability cluster architecture with redundant TCP stacks[C].IEEE International Performance, Computing and Communications Conference, Phoenix, AZ, United states: Institute of Electrical and Electronics Engineers Inc., 2003:28~29.
- [6] Keong Loh Peter Kok, Hsu Wen Jing, Wentong Cai. How network topology affects dynamic load balancing[J]. *IEEE Parallel and Distributed Technology*, 1996, 4(3):25~35.
- [7] Zhu Xiongfeng. Research and analysis on database cluster middleware MySQL proxy[D]. Wuhan University, 2011.
- [8] Gai J Y, Zhang Z N, Xiao H. Data replication technique analysis and application in distributed database system [J]. *Computer Applications and Software*, 2005, 22(7) : 36-38, 41(in Chinese) .
- [9] Zhang D P, Chen C, Xu Z. Research and implementation of heterogeneous database replication technology [J]. *Journal of Graduate University of Chinese Academy of Sciences*, 2012, 29(1) : 101-108.
- [10] C.L.Pape, S.Gancarski, P.Valduriez. Refresco: Improving query performance through freshness control in a database cluster. in: *Proceeding of On the Move to Meaningful Internet Systems 2004:CoopIS, DOA and ODBASE*, Agia Napa Cyprus.Berlin: Springer-Verlag, 2004, (1):

- 174~193
- [11] S.Narayanan, U.Catalyurek, T.Kurc et al. Applying database support for large scale data driven science in distributed environments. in: Proceedings of the Fourth International Workshop on Grid Computing.Los Alamitos:IEEE Computer Society,2003.141~148