

Recent Database Challenges in China on Data Consolidation and Integration

Yun Wang

IBM

CSDL, 8/F, Building A, Power Creative,
No. 1, East Road, Shang Di, Beijing 100085 P.R. China
(86-10) 82782244 x 3489
wang@cn.ibm.com

ABSTRACT

During the first phase of IT exploitation in China, a lot of isolated systems were developed and deployed over all the sites where the business took place. Each system typically dealt with one simple local business application. Since the mid 90's, data consolidation and data integration became critical to address the business need for cheaper cost, shorter time to the market as well as better business decision and control. Without consolidated and integrated data, there is no way to support new and complex applications for rapid and large scale business growth in China across lots of industries such as financial, telecom, government, manufacture and distribution.

When database is used for companies and government agencies in China, the large Chinese population makes the database volume huge. A nation-wide bank or an insurance company has hundreds of millions accounts easily. A province-level telecom service provider has tens of millions subscribers. In certain domains, such as banking and telecom, many applications are executed as on-line transactions which ask for large transaction throughput with good response time. A province-level telecom service provider expects to handle up to 100 billion calls in a month. In this case, most of the phone calls will be pre-paid type and these calls need to be handled in near-real time.

In order to satisfy the performance requirement, some applications implement application cache on application servers. This approach improves the application performance in getting data but it also introduces the complexity and longer delay when a failure in application occurs. Database management systems can provide client cache for powerful application servers in these days. As more and more application servers are added to respond large number of concurrent users and high volume transaction throughput, database management system must be highly scalable. Both logical and physical contentions should be avoided or greatly reduced. Continuous availability and high reliability are more critical for a consolidated large database because many users and applications rely on it. Both planned outage and unplanned outage should be avoided or minimized through appropriate duplication and redundancy. Smooth transition through cooperative versions of data helps reduce planned outages.

Swift switch over to hot stand-by database system helps minimize the unplanned outages. Database management systems are expected to be autonomic and remain healthy in resource utilization with minimal administrative involvement. As a summary, data consolidation in China asks for low-cost, high performance database systems with high scalability, availability and reliability on very large database volume.

Data integration in China often deals with both a horizontal layer of multiple lines of business and a vertical organizational structure. For government agencies, there are at least four levels across nation, province, district and county. There are a few industry specific data models and solutions to address the requirements of data integration. However, a common data integration platform has been constantly requested by many IT projects of large scale. In many cases, business control processes need to integrate business operation processes to support near-real time decision and control. There is a high demand for a practical and well-proven methodology with tool sets which can support data integration life cycle of architect, design, development, deployment and change management.

Rapid economic and business growth in combination of the join to WTO created big demand in IT evolution across China business, industry and government. There are many demands and opportunities of data consolidation and integration with large data volume and system complexity. The demand of low-cost solutions makes it even more challenging.

Categories and Subject Descriptors

H.2.4 [Database Management]: Systems – *Transaction processing, Relational databases, Parallel databases, Distributed databases, Currency.*

General Terms

Performance, Design, Reliability.

Keywords

Data Center Consolidation, Information Integration