Database application research in real-time data access of accelerator control system

CHEN Guanghua* CHEN Jianfeng WAN Tianmin

Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Zhangjiang Campus, Shanghai 210200, China

Abstract The control system of Shanghai Synchrotron Radiation Facility (SSRF) is a large-scale distributed real-time control system. It involves many types and large amounts of real-time data access during the operating. Database system has wide application prospects in the large-scale accelerator control system. It is the future development direction of the accelerator control system, to replace the differently dedicated data structures with the mature standardized database system. This article discusses the application feasibility of database system in accelerators based on the database interface technology, real-time data access testing, and system optimization research and to establish the foundation of the wide scale application of database system in the SSRF accelerator control system. Based on the database interface technology, real-time data access testing and system optimization research, this article will introduce the application feasibility of database system in accelerators, and lay the foundation of database system application in the SSRF accelerator control system.

Key words Accelerator, Oracle RAC, XML, Partitioning

1 Introduction

The Shanghai Synchrotron Radiation Facility is a third-generation of synchrotron radiation light source which electron energy is 3.5 Gev [1], referred to as the SSRF. The SSRF accelerator control system realizes the full operational control on all of devices including the linear accelerator, booster and storage ring, and executes control tasks in accordance with subsystems including control of Magnet Power Supply System control, Vacuum System control, Beam Instrumentation System control, Radio Frequency System control, Injection and Extraction System control, Timing System control and Safety Interlock System control.

In running period of the accelerator control system, large amounts of data generated will provide the necessary machine parameters and reference information for accelerator machine debugging, analysis, closed-loop operation and system simulation. The popular EPICS is used in the SSRF accelerator control system to establish the software architecture of control system, and EPICS Channel Archiver toolkit is also used as the data access tool. The Channel Archiver toolkit which has specific data storage format is not convenient to do scientific computing on large-scale data, and the way of its data storage is also not flexible enough. And, only the data in some specific format can be stored on the local machine. And because the toolkit supports just a few query conditions such as time stamps, it can't meet more detailed query needs of users.

Therefore, it's required to establish a database system to achieve standardized data management and operation.

Database systems have wide application prospects in the large-scale accelerator control system. Its high throughput, high reliability, standard interfaces and rich application software make the database achieve obvious advantages in the accelerator control system. It is the future development direction of the accelerator control system, to replace the

Supported by Shanghai Synchrotron Radiation Facility project
* Corresponding author. *E-mail address:* chenguanghua@sinap.ac.cn
Received date: 2011-11-16

differently dedicated data structures with the mature standardized database system.

In this article, the Oracle RAC technology is used as the starting point to analyze the RAC technology principles, characteristics and functions. On the basis of this analysis, the SSRF database test platform is designed to do the tests for fast access to real-time data, system load balance and failover testing. And through the feasibility study of Oracle RAC database system in the accelerator application, it lays the foundation of Oracle RAC database system application in the SSRF accelerator control system.

2 Oracle 10-g RAC Database

2.1 Cluster Technology

Cluster consists of more than one computer (or node) to form a whole set to provide services to users. Cluster technology has the most widely usage in Linux operating system with high extendibility, high performance and high availability. High Availability Cluster (HA) focuses on improving system availability, which is through the integrated hardware and software fault tolerance to achieve the overall service availability. If a cluster node fails, then there will be another node to replace it. Even if multiple nodes fail, the entire system environment will also ensure users can access.

2.2 Oracle 10-g RAC

Oracle Clusters RAC has a shared cache architecture, which runs on multiple machines connected through the cluster interconnect and the shared storage subsystem. Data files are stored in shared storage devices (usually disk array). It allows multiple Oracle instances to access a database storage device, and the cache and data files are shared between instances to ensure that all the data is the same at any time for all nodes[2].

Oracle RAC introduces CRS (Cluster Ready Service) to implement the unified management and distribution for the software and hardware resources on all cluster nodes. In the RAC, the CRS services can do unified coordination of all the work of the cluster nodes, automatically delete or add cluster nodes in real time based on hardware fault conditions, realize the dynamic management and maintenance of the cluster, and provide automated workload management with flexible extendibility, low cost and high availability. So, it becomes a good solution in the high availability system cluster applications[3].

3 RAC Platform Construction and Testing

3.1 Four-nodes RAC Testing Platform

The Oracle RAC platform includes the clusterware installation, configuration of shared disks, Oracle database installation, Oracle RAC software installation and database creation. In the Oracle RAC 10g, through the virtual IP addresses and service switching, failover and load balance can be completed to ensure high-availability (HA) of the database. In each RAC node, public, private and virtual IP addresses are configured.

On the platform, HP 580G3 servers are used as the database servers, SAN storage network system used as the data storage device. And two SAN fiber cards are used for each server to connect with optical switches. The platform architecture is shown in the Fig.1.

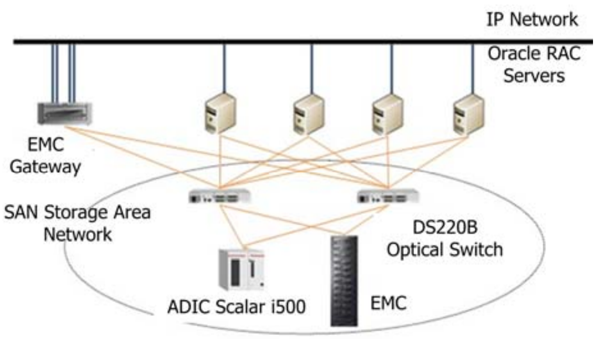

Fig.1 SSRF Database Architecture.

For the platform, two fiber channel switches are used as connection parts, EMC storage devices used as back-end. All of hosts connected to the SAN network are installed dual-host access card (HBA). In the SAN storage network, the servers with high performance and fast I/O response time are connected with EMC storage devices through multiple storage fibers and optical switches, and they can access the EMS storage directly to meet the needs of SAN applications. All the connections from hosts to back-end storage devices have several routes, which ensure the high reliability of data access.that is there is no single point of failure, in the redundant architecture from the host to the storage. All the connections from hosts to back-end storage devices have several routes, which ensure the high reliability of data access through the redundant architecture from the hosts to storage. And the duplicate routes also ensure there is no single point of failure. When the storage is needed to be extended, it's easy to extend the online storage space, just by adding the appropriate hard disks through the management software. The whole operation is online for storage capacity expansion without interrupting data access, so you can ensure business continuity. When a new server is added to the current SAN storage network by the connection to the SAN optical switches, the new server will access the EMC storage easily.

The platform is also equipped with an ADIC Scalar i500 tape library, two LTO3 tape drives with fiber interfaces and 36 tape slots, which can provide enough backup spaces for all the data in disk arrays. With the backup software modules, enabled data backup LAN-Free will help users achieve high capacity, high-speed data backup environment to maximize data protection capability.

The operation systems of four database servers are CentOS 4.5. The database software is Oracle 10g RAC. Windows database application development platform is Borland C++ Builder 2007 and API interface is OTL (Oracle, ODBC and DB2-CLI Template Library).

3.2 Table Design

3.2.1 XML Technology

Extensible Markup Language XML is a W3C specification, which is used to define the framework. It provides a format for describing structured data, so that it's easy to exchange data between the various applications based on Web. XML with simplicity, openness, scalability, interoperability and a series of characteristics brings the power and flexibility to Web-based applications.

From the beginning of the second edition of Oracle 9i database, Oracle XML DB has been seamlessly integrated with the Oracle database[4]. In order to support XML, Oracle XML DB introduces the XMLType object type, which is a native server data types. This type can be understood by the database as a column or table containing the XML. The tables or columns defined by XMLType can be used to store XML data.

3.2.2 Partitioning Technology

In running period of the SSRF accelerator control system, more than 5GB data per day and about 2TB data per year are produced, and real-time data recorded exceeds ten thousand. Partitioning technology is used to deal with these data and manage the database system with high response. Partitioning is a technology to deal with very large tables and indexes, which partitions large tables and indexes into manageable pieces. Thus, it avoids the management of each table as a large, single object, and provides a huge data scalable performance. Through the above partitions, it reduces the time for managing operations and improves the performance through enhanced parallel processing.

For the management of large data tables, Oracle database system provides range partitioning, list partitioning, hash partitioning and composition partitioning[5]. According to the need of improving real-time data access speed, range partitioning is used to partition the range of some value in a table.

The choice of partitioning key is a critical step in the partition scheme. When the distribution of partition key values gets more average, the partition spaces will become more equal, and the load of disk I/O will become more balanced. The SSRF accelerator system includes the linear accelerator, booster and storage ring, and the storage ring is divided into 20 units. The accelerator, in its running process, will produce pv_id values with unique variable definition. The distribution of pv_id values is divided into 12 areas according to different systems and units, so as to ensure that the number of variables in each range of areas is balanced. So, pv_id is selected as the partitioning key, and the values of pv_id are divided into 12 ranges. The table structure is defined as followed:

```
create table pvdata
(  svr_time      timestamp,
   pv_id         number,
   xml_data      xmltype,
```

```
 CONSTRAINT PK_PVDATA PRIMARY KEY
(pv_id,svr_time) USING INDEX LOCAL )
TABLESPACE CHENGH
  partition by range(pv_id)
  ( partition PDATA_SR0102 values less than
(2000001),
      partition PDATA_SR0304 values less than
(4000001),
      partition PDATA_SR0506 values less than
(6000001),
      partition PDATA_SR0708 values less than
(8000001),
      partition PDATA_SR0910 values less than
(10000001),
      partition PDATA_SR1112 values less than
(12000001),
      partition PDATA_SR1314 values less than
(14000001),
      partition PDATA_SR1516 values less than
(16000001),
      partition PDATA_SR1718 values less than
(18000001),
      partition PDATA_SR1920 values less than
(20000001),
      partition PDATA_SR21LA values less than
(22000001),
      partition PDATA_BS0102 values less than
(24000001)).
```

3.3 Performance Testing

3.3.1 Data Access Speed Testing

The first step is to prepare 12 copies of the same text (RTest01~RTest12), each size of text 10.6MB with two million records. The range of pv_id values is divided according to partition value, that is Rtest01, the pv_id ∈ [1,2000000]),…, Rtest12, the pv_id ∈ [1,24000000]).

The programming sentences are inputted through Borland C++ as followed.

```
char      svr_time[20];
    int         pv_id;
    char        xml_data[80];
insert into pvdata
values(to_date(:f1<char[20]>,'yyyy-mm-dd
hh24:mi:ss'),:f2<int>,xmltype(:f3<char[80]>));
```

There are 12 processes for writing data, each process writing data records in the related text file to table pvdata (RTest01−RTest12). The relationship between the number of processes and the speed of writing data is shown in Fig.2.

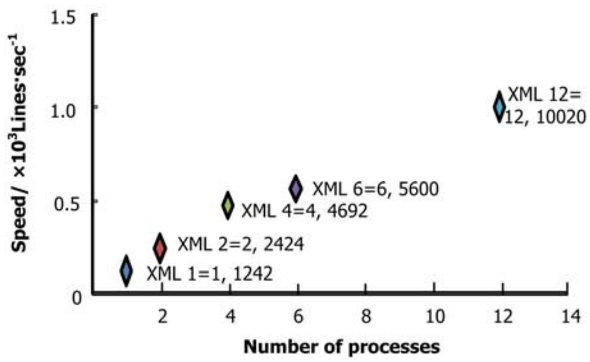

Fig.2 XML Data Writing Performance Testing.

The results are shown that the appropriate and reasonable partitioning for tables can enable data-written in parallel, and double the write speed for real-time data. Moreover, while the twelve processes of writing data were running, all of CPU usages of the four servers were very low.

3.3.2 Load Balance Testing

Through two Shell scripts, the session connection and query connection are established respectively and circularly. Test results are shown that the session numbers of four nodes were average distribution, and the new added session was automatically connected to the relative free node. It is shown that the load balance function was distributed to all nodes automatically.

3.3.3 Failover Testing

The first step is to link the database through SQLPlus. The second one is to execute a query for large data (tens of thousands), while querying gv$session view to identify the cluster node related with current connection. The third one is to use the srvctl command to close the current connection. The results are shown that the data query was interruptd within a few seconds, then query was continued, and finally the number of shown records was the same as the actual number of data in the table. Through querying GV$SESSION view, it is shown that the connection was transferred to another node, and in the current active node, through the NETSTAT command it was discovered that the virtual IP of closed node had been successfully transferred.

4 Conclusion

Oracle RAC, as a highly available cluster, has prominent features in failover, load balancing and concurrent processing. This paper describes the analysis and construction of four-node cluster testing platform. Based on the running state of SSRF accelerator control system and the feature analysis of Oracle RAC, it introduces the table partitioning and XML technology. The performance tests are shown that the application achieved the high availability of database, and load balancing and failover capabilities, and improved real-time data access speed.

References

1 SSRF Instoduction. http://ssrf.sinap.ac.cn/engilish/1/Instruction.htm.

2 Gu WeiTao, Fang Yu.Computer Knowledge and Technology, 2010, **6:** 5137−5139.

3 Matthew Hart, Scott Jesse.Oracle database 10g high availability with RAC, Flashback & Data Guard. California (USA): Brandon A. Nordin, 2004, 200−210.

4 Lan Xiaoji, Ren Jintong. Journal of Geodesy and Geodynamics, 2009, **6:** 139−143.

5 Qi WeiWei. Computer Knowledge and Technology, 2010, **6:** 4602−4604.