

Research and Application of Synchronization Update of Distributed Heterogeneous Database

Xin Yingxiu

(Qingdao Technical College, QingDao ,ShanDong 266555 China)

Abstract: According to the different equipment, different system and heterogeneous database have be information "isolated island" problem, and the data of equipments can be updated in real time on the business node. The paper proposes a program of data synchronization platform based on J2EE (JMS) and XML, and detailed analysis and description of the workflow system, its frame structure and the key technology. Practice shows that this scheme has the advantages of convenient and real-time etc..

Keywords: data synchronization; XML; JMS; database; real-time updates

0 Introductions

With the widespread application and the rapid development of the computer and network technology, there are all kinds of internal information systems in schools, factories and research institutes. For example, oscilloscope of internal use, signal generator, data acquisition and other devices in research Institute that has function of wireless WIFI or local area network to access internet. These systems can improve the efficiency of research, through the Internet of things technology, researchers at home can real-time monitoring instrument data, and improve the efficiency of their study. But at the same time it also becomes an "isolated information islands". That is to say, a lot of equipment has the function of the Internet, but they cannot be timely sharing of data between each other, or update data, that result management difficulties in school. For the distributed network equipment, how to maintain the high efficiency and safety between heterogeneous database synchronization, input information update, is one of the important problems we need to consider. Heterogeneous network operating system is different, or different database system, and between the same internal different forms of heterogeneous database. Through the component, maintain synchronization between the databases, and can improve the reliability and efficiency of the system and data transmission. At the same time, bring convenience for maintenance system.

At present, synchronous update of distributed heterogeneous database have many methods and many researchers in the field in domestic. But their field research emphases are different, and researchers can study and rarely achieve results.

Gong Tao [1] put forward "return decomposition of the SQL statement" in the "Research on real time synchronization of heterogeneous database", the author put forward new ideas, and the article mainly studies on data conflicts, especially the synchronization technology. At the same time, the author also explains the important role of synchronization in the presence of table structure, but it does not explain how the table structure conflict, field conflict etc., or the author cannot analyze on the synchronous data to XML mapping and XML mapping table structure to the system, so, in general the system have been certain restrictions in cross platform, scalability aspects.

XiongXian[2] is gives a synchronization scheme between Oracle, SQL Server, DB2 and other mainstream database, and

finally presented the solutions in synchronous systems: the trigger and AOP. AOP technology mainly is aimed as support database without trigger. The author puts forward "publish / subscribe" mode is message oriented middleware, interpretation of communication can realize the time, space, process. But the article not related to the table structure conflict, composite mode synchronization, file mapping synchronization problem. So this method is also subject to certain restrictions.

Yang Peng [3] introduce combination of the trigger and the log table to capture in detail, and through the log table changes combine with time stamp synchronization strategy. The strategy of using the same trigger to capture the net change data, it improves the synchronization system performance; and the author uses middleware: XML and SyncML, it can go to the transmission of synchronous data according to the protocol, completely solve the heterogeneous database or the data synchronization between different devices interconnect problems, and the system is easy to expand and synchronization upgrade, so as to adapt to the more data sources. But the system also has some problems such as: the article not to XML mapping is given the concrete design and implementation of data synchronization code.

Analysis of the above three schemes, which mainly represent the common current implementation of heterogeneous database in distributed computing, the advantages and disadvantages are mainly manifested in:

(1) Each database is established user interface: all kinds of heterogeneous databases distributed in each node is established user interface, but not integrated, and each subsection management and remote login access. This scheme is simple, but the user and access rules, when the increase of heterogeneous database new, that must increase the user interface, maintenance and use of complex method; (2) the establishment of a global model between heterogeneous databases. The system centralized management mode to physically dispersed distribution and logically related parts of centralized management, operation on any database is like in local for user; they do not have to care about the physical location of data model and other details. Because the database integration, when user face this heterogeneous database, they set up global schema with many difficulties. (3) The decentralized and centralized management model. Centralized management of different database autonomy and members of the database,

the data transfer technology to solve the consistency problem of distributed data processing and multi node cluster. (4) The general acquisition mode ignores system. (5) Effects of mapping ignore synchronous data to the XML on the performance of the synchronization.

In this paper, in order to effectively solve the packet flow, reduce the flow (such as the wireless sensor network), data synchronization, fundamentally solve the "information isolated island" problem of distributed heterogeneous database, which can integrate various business terminal data, realize the real - time data synchronization update. The frame synchronization platform JMS and based on XML Technology

1 System framework and implementation

Figure 1 gives system frame structure and the main module. Among them ①front end equipment, including the equipment data acquisition, data storage, database, database interface API; ② synchronous data platform, it is the core part of the system, it mainly includes the message server Admin API, JNDI API, JMS API; ③management center. It mainly includes Admin Tools, LDAP/ANDI Provider, JMS Server, and Database.

Among them, the message server is the core of the system, it follow JMS (Java Message Server) standard JMS Server service. Firstly, the message queuing on the JMS server, use normally "publish / subscribe" model of message queue. In this model, there is no limit to the number of target can be sent to message end, and each message can be sent to any number of subscriptions to the receiving end

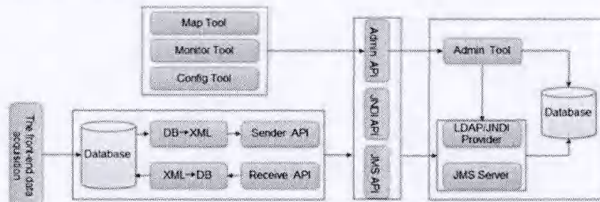


Fig.1 The diagram of system architecture

From the functional perspective, the block diagram of Figure 1 can be seen that the function module of the system, they are the source of the data register module, form mapping XML module, message monitoring module, data is mapped to the XML module, data reading module and conflict processing module. As shown in Figure 2, the task of each function module is:

1) The source data module: this part mainly read the database login information, such as: ID, Password, IP, Address etc.. The module can login, delete, registered user information.

2) Center DB → XML Adapter module: After source database registration, it will extract all the user table structure, such as: registered table name, field names, field type, and turn it into a Table_Info.xml file in the database, the file is called Sender API, is also based mapping files.

3) Monitoring module (Monitor Tools): according to the news information, monitoring module real-time capture and refresh the heterogeneous network information, it is the basic requirement for

synchronization. The module is: when the source database to form data change, through certain method to extract synchronous data.

4) Synchronous data to the main task of the module of XML: is carries on the design to the synchronous data acquisition module extraction, and convert it to Data_xml.xml file. Each node of the Data_xml contains the source type of database, the source database name, the source table name.

5) Conflict processing module: analysis and record maintenance fault system in operation.

6) Data loading module: This module read data and sent to the destination table, read after removal of the corresponding control information in the C_TabChange table; if a read error, message rollback, and retain the corresponding control information in table C_TabChange, exception handling, and convenient system.

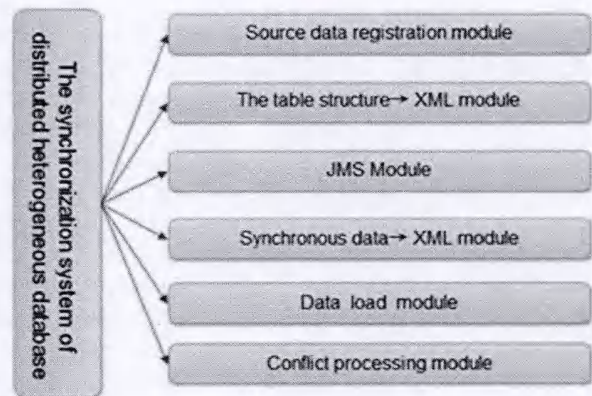


Fig.2 The system function block

2 System process and working principle

2.1 The flow of the system realization

1) For the relationship between the source and target databases, firstly the system read the source database and target database login information and all the users' table structure information;

2) Create a list of records in the source database (c_change), and create trigger data exists in the source database that share all the table, when the operation of the DML on the source table, the trigger is awakened, extraction of synchronization control information in the source table write into the record in the table;

3) read record of records table in row, according to the recorded content extracted from the database data and generate the Data_xml.xml file, according to the table in the value of UPDATETYPE to generate the XML file format design;

4) Based on the source and target tables metadata, it establishes the mapping file (MappingFile.xml), the mapping file needs to have the hierarchy and regularity, and as far as possible to sort out the conflict situation exists;

5) Using Dorn parse XML documents, according to the Data_xml in the source table find node information in the mapping file for the target table information, and determine whether there is a

variety of complex mapping mode. We combined and restore SQL statement according to DataXml.xml and Mapping "before the synchronization ile.xml files, the synchronous data loaded into the target table.

6) After the success of data synchronization, delete the source records the corresponding information records in the table, if the disruption caused by the network or other unknown reasons, stop all operations, the current change records the transaction is rolled back, waiting for the next cycle of synchronous operation, and save the synchronous operation record.

2.2 The key code of the system realization

Data synchronization between heterogeneous devices: The caches S-Table will convert to MAP which export from XML and convert XML to MAP format of cache table D-Table.

The export MAP Format:

```
<? Xml version="1.0" encoding="GBK" ? >
<TABLE2XML>
  <DataSources>
    <DataSource Name="Source-A">
      <DbType> </DbType>
      <DbType> </DbType>
      <User Name> </User Name>
      <Password> </Password>
      ...describe source data
    </DataSource>
  </DataSources>
  <Tables>
    <table Name="cache-TableA" DataSource Name="
Source-A" USql="DELETE cache-TableA" WhereCon-
dition="">
      <Field Name="ADB-SEQUENCE" DataType="">
```

Import the MAP format:

```
<? xml version="1.0" encoding="GBK" ? >
<XML2TABLE>
  <DataSources>
    <DataSource Name="Destination-B">
      <DbType> </DbType>
      <DbType> </DbType>
      <User Name> </User Name>
      <Password> </Password>
      ...described the DataSource
    </DataSource>
  </DataSources>
  <Tables>
    <table Name="cache-TableB" XMLNAME="cache-
TableA" DataSource Name="Destination-B" Action="
INSERT">
      <Field Name="ADB-SEQUENCE" XName="
```

3 Conclusions

With the development of information technology, large internet equipment due to differences in environmental factors and type of database selection, that caused a lot of Information Island. However, a lot of data is the association for this relatively independent information service, management system and the database system. This requires that we must ensure each data consistency in the database for node. This paper proposed the establishment of a safe, generic platform to solve the data between distributed heterogeneous database synchronization problems. Synchronization method presents a "mapping files and JMS message" to capture the synchronous data in this paper, at the same time, using XML as the intermediate data representation, easy to use, which can effectively avoid the unify problem between heterogeneous database standard, to ensure the professional data and shared data, database and spatial database consistency has important significance.

Reference

- [1] Gong Tao. Research on real time synchronization technology of heterogeneous database [D]. Huazhong University of Science and Technology. 2010
- [2] Xiong xian. Design and implementation of heterogeneous database synchronization system of JAVA based on XML [D]. Shanghai Jiao Tong University. 2007
- [3] Yang Peng. Heterogeneous database change capture and synchronization strategy [J]. Computer engineering, 2008,34 (16): 53-55
- [4] Shen Min, Xu Hua-hu, et al. Implementation of data synchronization for distributed heterogeneous database [J]. Computer Engineering and Application, 2005, 41(5):184-186.