The research on the query optimization on the distributed heterogeneous database based on the response time

Zhang Zhenyou
Collage of Information Engineering
Hebei United University
TangShan, China
youzhenadd@163.com

Luo Bin
Collage of Electrical Engineering
Hebei United University
TangShan, China
luobin11@163.com

Cao Zhi
Collage of Information Engineering
Hebei United University
TangShan, China
526818718@qq.com

Abstract—**The user's response time was an important indicator for the heterogeneous database which locate in the different network. This paper describes the structure of the distributed heterogeneous database system of query optimization system based on the Hibernate technical for the B/S structure, and analysis the calculation of the response time. Finally an example was used to demonstrate the feasibility and practicability of query optimization.**

Keywords- heterogeneous database; query optimization; the response time; Hibernatee

I. INTRODUCTION

The query optimization in the traditional relational database usually adopts query tree to show the optimization process of the query operation. In the centralized database management system, because of didn't need to consider the factors of the network, the query tree can well performance query and its optimization process. However, in the distributed database systems, network factors should be considered, and the query tree cannot fully reflect this. Network factors include the data transfer price in the distributed environment. That is to say the data query with a price. So it is must to consider the response time of the data source, communication channel is available, etc. This is the main factors in the distributed query optimization. The main idea was with the different response time to choose the database queries.

II. THE QUERY RESPONSE TIME ANALYSIS ON THE DISTRIBUTED HETEROGENEOUS DATABASE SYSTEM

A. The Definition of ResponseTtime

The calculation problem of the response time in distributed heterogeneous database query optimization system. In fact the response time in distributed heterogeneous database query optimization system was divided into two parts, one was the communication part to establish a connection with heterogeneous data source and the other was data transmission parts in heterogeneous data source. The sum of the two parts referred to as "the response time in the system query optimization. But the response time in distributed heterogeneous data source must only be calculation and the response time in established heterogeneous data connection, which ready to communication part of the response time. But it cannot calculate the data transmission time in heterogeneous data sources , Because the data in each a distributed heterogeneous data sources was different, this leads to different data transmission time in distributed heterogeneous data sources , the more the amount of data the more transmission time, suitability was short. So it only calculate the first part time in designed scheme in the query the optimization.

B. The Calculation of the Rresponse Time

The main strategy in query optimization for the distributed heterogeneous database was to use a kind of algorithm based on experience, and to query optimization for the distributed heterogeneous data source. This strategy according to the visit response time, so analysis the experience data in these algorithm, and finally get distributed heterogeneous data access priorities. Priorities were calculated according to the experience of the response algorithm. An experience value of the priority: experience value for the each of heterogeneous data sources = Σ each visit time/visits, it visited which one if the value smaller than the others at first, it visited which one if the value bigger than the others at last. So, according to the response time of the distributed heterogeneous data source before to predict the future response time of the distributed heterogeneous data source.

III. THE DESIGN OF THE QUERY OPTIMIZATION SYSTEM BASED ON THE HIBERNATE

The data integration of the query optimization system on the distributed heterogeneous database was from the bottom to the up, step by step, put the bottom data which was in the heterogeneous database released up integration, thus ensures that the transparency of the data. The system consists of three layers include the application layer, middleware layer, data source layer, as shown in figure 1. The advantage of the design of the three layers system, it was easy to coordinate the project personnel to develop each layer, and easy to system integration when each team exploited every layer. This can greatly shorten the envelopment time in the data integration. It was mainly that there was enough time to develop middleware layer, middleware provides interface for each layer, down for data, and up for the customer. Middleware was the foundation of the other two level of design. So the design of the three layers structure was very reasonable, and also to convenient for the maintenance of the late. The paper detailed the three layers.

1) the application layer

The application layer of the data integration in the distributed heterogeneous database query optimization design, the main business logic was that users can search and display the search results. It was mainly the interaction layer the user WEB browser and WEB server. System users put the request information sent to the WEB server, and then the middleware in the WEB server get the data from the distributed heterogeneous database, so data would be feedback to the user. Finally the data shows on WEB browser interface of the distributed heterogeneous database query system.

2) middleware layer

The middleware layer in the distributed heterogeneous database query system from higher level to shield the difference of data source, and distribution and heterogeneous. The system users think that all the data for the request was the local data sources, were on the same network. The system users simply won't feel that the database was the source all over the world. The middleware was responsible for search data that users need, and process these data, and feedback to the WEB server after integrated, so the WEB server submitted to the user. The middleware layer can conversion the parameters of the data query into the corresponding the query data structure, also was the corresponding middleware query language. Use the mechanism of the object relational conversion, the query language of the middleware translation into query language for the heterogeneous database data sources. When the heterogeneous data back the data, using object relational conversion mechanism, data translated into the data that the middleware can identify (object data).Finally it integrate all heterogeneous data sources, back to the application layer.

3) the data source layer

The composed of database in the distributed heterogeneous database query system, made of various database design manufacturers, can be to SQL Server, ACESS, Sysbase, DB2, Oracle, MySql, etc. The database may be having different logical structure and physical structure. Each database can be put in different network node, distributed in all over the world.

Each data source has its own management system, so that the data source intricate. Then it is needs the middleware to manage the data source layer.

IV. GET THE RESPONSE TIME OF THE DISTRIBUTED HETEROGENEOUS DATABASE DATA SOURCES

The middleware which designed by the technology of the Hibernate, use integrate the distributed heterogeneous database. When the middleware to access heterogeneous data sources in low-level, it got the data form the bottom of the heterogeneous data sources and stored the response time in a database, and updated the response time and the number of visits on the heterogeneous data. If the middleware didn't find the distributed heterogeneous data sources, that is to say also with the string for the database connection to set up a connection with the distributed heterogeneous data source was fail, at this time it is to quit, and delete the data records of this distributed heterogeneous source from the database.

The mode of connection, user name and password for the every distributed heterogeneous database saved in a text file, when the middleware access the distributed heterogeneous database, it need to read the text file from the connection string, and then establish a connection with heterogeneous database. Getting the response time the distributed heterogeneous database data source was an important data collection of query optimization. Through it analysis these data with the experience algorithm, so implement the query optimal of the distributed heterogeneous data sources. The data flow diagram of the response time for the distributed heterogeneous database data sources was as follows figure 1.
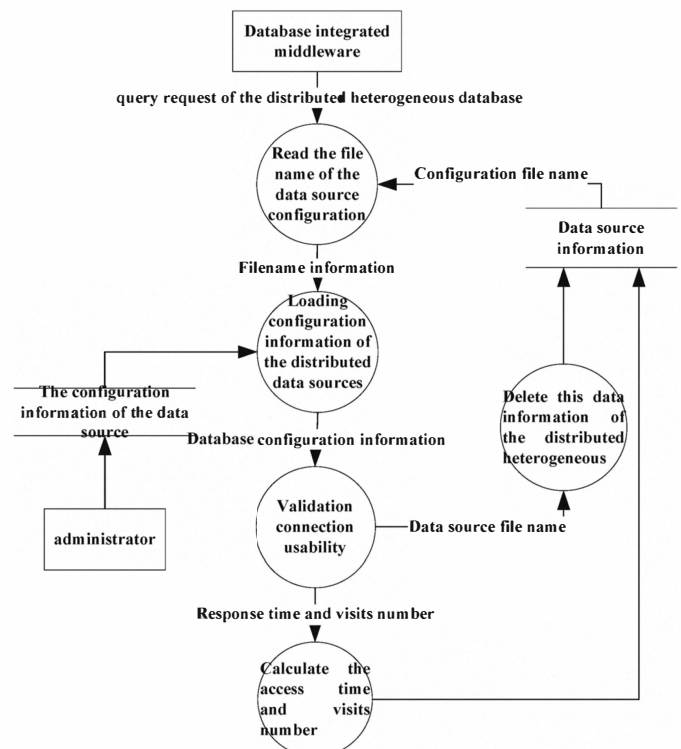

Figure 1 Get response time of data flow diagram

V. The Realize of the Query Optimization System for the Distributed Heterogeneous Database

The specific coding idea of the single node table of query optimization of the distributed heterogeneous data source:

The middleware loading the address of the configuration files of the distributed heterogeneous data source, and then to read the configuration files of the distributed heterogeneous data source according to experience algorithm, and then opens these data sources and reading of the data. It writes the response time of this operation to the database. It would support the query optimization next time. The last, it would send the query result which get from the each of the distributed heterogeneous database to the users.

Public List getAllStu () {//search for all students

//the path of the database connection file.

String filepath = this. GetClass (.) getClassLoader (.) getResource ("Hibernatecfg"). GetPath ();

//get all of the configuration file name from database, and sort the experience value for the heterogeneous database =Σeach of the response time /visits number.

List filelist = null;

Try {

Filelist = databaseDao. GetDataInfoList (

"Select * from dbsinfo order by accesstime asc", null);

} catch (SQLException e1) {

E1. PrintStackTrace ();

}

//initialization the structure sets of the students information.

List "StudentDO > List = new ArrayList < StudentDO > ();

//cycles loading each of the configurations file for every heterogeneous data sources.

For I = 0 (int; I < filelist. Size (); i++) {

File readfile = new File (filepath + "\\" + filelist. Get (I));

//if data sources not found, it deleted this data record of response time from the database which saved the response time, next time, no longer traverse this data sources.

If (! Readfile. Exists ()) {

DatabaseDao. DeleteDataInfo (readfile getName. ());

Continue;

}

//loading the configuration information of the database in heterogeneous data sources.

InputStream is = this. GetClass (.) getClassLoader ()

. GetResourceAsStream ("hibernatecfg/"+ readfile getName. ());

Properties Properties = new Properties ();

Try {

Properties. Load (is);

} catch (IOException e) {

E.p rintStackTrace ();

}

//hibernate loading object-relational mapping file *. HBM. XML

Configuration CFG = new Configuration ();

CFG. AddProperties (properties);

CFG. AddClass (StudentDO. Class);

//response time timing

Long begin = System. CurrentTimeMillis ();

//the establishment of database connection, and reading of the data

SessionFactory SessionFactory = CFG. BuildSessionFactory ();

Session Session = sessionFactory. OpenSession ();

//end response time

Long end = System. CurrentTimeMillis ();

Query q = session. CreateQuery (" from StudentDO ");

List list_temp = q.l ist ();

List. AddAll (list_temp);

Session. Close ();

SessionFactory. Close ();

//The response time saved to the database which saved the response time.

DatabaseDao. UpdateDataInfo (filelist. Get (I). ToString (), end-begin);

}

// return the traversed result set of the heterogeneous data sources.

Return list;

}

}

VI. The Examples of the Query Optimization of the Distributed Heterogeneous Database

It used a simple example to explain the principle of the distributed heterogeneous database query optimization scheme. If there were three different distribution heterogeneous data sources that locate in the network node in the distributed heterogeneous database query optimization system, they were MYSQL01, SQLSERVER01 and ORACLE01, and each database contains a student database which contains the student information table and class information table. And each of the

students information table and class information table has the same structure, because of it was the premise of the research with the same table structure. Figure 2 was the network topology structure of the distributed heterogeneous database.
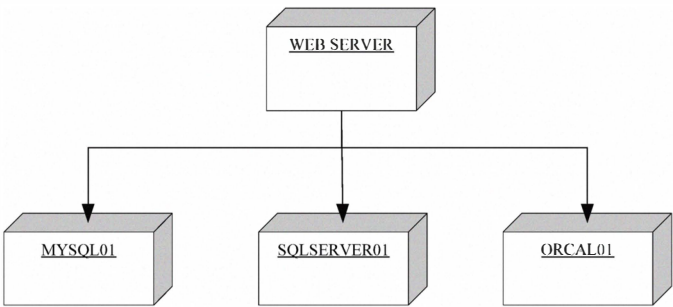

Figure.2 Distributed heterogeneous database network topology

If the WEB SERVER has the different response time and the different access number to the three distributed heterogeneous data source. Table 1 was the specific data below.

Table 1 Distributed heterogeneous data sources response time

database	response time for the server to the data source	visited time
MYSQL01	1000ms	
MYSQL01	2000ms	2
SQLSERVER01	3000ms	
SQLSERVER01	1000ms	3
SQLSERVER01	2000ms	
ORACLE01	1000ms	
ORACLE01	2000ms	
ORACLE01	3000ms	4
ORACLE01	4000ms	

It calculates the experience value with the experience algorithm for the distributed heterogeneous data source according to the data in the table.

The experience values according to the experience algorithm for the MYSQL01, SQLSERVER01 and ORACLE01 distributed heterogeneous data source was to the following table 3. Through the table 2 was very easy to get distributed heterogeneous database of traverse sequence, first is MYSQL01, followed by SQLSERVER01, and finally ORACLE01.

Table 2 Experience the value of distributed heterogeneous data sources

database	response time for the server to the data source	experience value
MYSQL01	1000ms	
MYSQL01	2000ms	1500
SQLSERVER01	3000ms	
SQLSERVER01	1000ms	2000
SQLSERVER01	2000ms	
ORACLE01	1000ms	
ORACLE01	2000ms	
ORACLE01	3000ms	2500
ORACLE01	4000ms	

VII. SUMMARY

The research on query optimization of the heterogeneous relational database was the emphasis field in the heterogeneous database. In the study, the Hibernate technology makes system can dynamically linked many heterogeneous database system, and queries the different user information table of the different database. The tiny query response time make the heterogeneous database of database query optimization more practical.

REFERENCES

[1] Nick Roussopoulos. MOCHA: A Extensible Database Middleware System for Distributed Data Sources [C]. SIGMOD Conference, 2009, 8 (12) : 2130-2240.

[2] XuHongJi, ZhangAnJiang, LiJianYe. The design and implementation on the A cross-platform database access layer model [J]. Application research of computer, 2007, 13 (19) : 144-146.

[3] ChenYingHui, YuJun. The reaearch on the query optimization method for the distributed heterogeneous database [J] Microcomputer Information 2009, 25 (12) : 22 to 28.

[4] YangFang, TaoShiQun. The integrated design for the heterogeneous database system based on hibernate [J]. Computer engineering and design, 2007, (7) 19:-110

[5] Wang Ning, WangNengBin.The realization of the query decomposition and optimization on heterogeneous data source integration system [J]. Journal of software, 2008, (2) : 222-228..