Research on the Integration and Query Optimization for the Distributed Heterogeneous Database

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

Zhang Jingjing
Collage of Electrical Engineering
Hebei United University
TangShan, China
tsphinix@163.com

Liu Shu
College of Information Science & Engineering
Yanshan University
Qinhuangdao , China
jishyusama@126.com

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

***Abstract*—Heterogeneous database query optimization should study not only the node database integrated in different database, but also to solve the strategy problem which the data query optimization. In this paper, it studied integration among the node database in the distributed heterogeneous database system and the query strategy for the data in heterogeneous databases. It was put forward the database of integration solutions based on Hibernate and query optimization strategy that used the tiny user inquires the response time , and for this scheme design in this concrete solution.**

Keywords- Heterogeneous database; system integration; the query optimization strategy; the tiny response time of user querye

I. INTRODUCTION

The research in system integration for the distributed heterogeneous database system was to many different types of database system, which allows users to transparent query the different qualitative database system. And lets users got the tiny query the response time in of query optimization was the main research.

First of all, the problem must be solved was many heterogeneous database system in the different quality integration to a platform. Also was the integration of heterogeneous database node. Next, it want to analysis on the main indicator of query optimization--the main factors of user response time. Finally puts forward the solution to optimize the response time of user query.

II. SYSTEM ANALYSIS FOR DISTRIBUTED HETEROGENEOUS DATABASE IN QUERY OPTIMIZATION

A. System Integration of the Heterogeneous Database

System integration was designed to provide a unified interface of heterogeneous data sources to visit, the user need not consider the heterogeneous data model. Users only need the concerns of the specified data, and don't have to concerns how to get this data. System integration platform for users to store the data source on each node of the heterogeneous databases. And so the user can access to all heterogeneous databases. As shown in figure 1 :

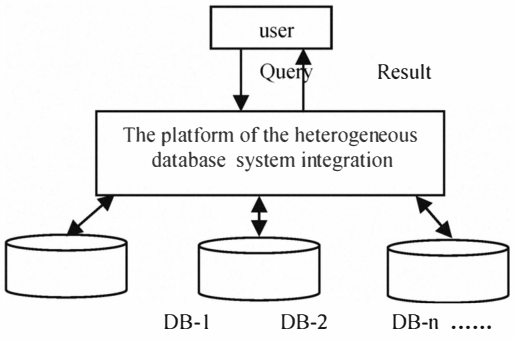

Figure 1 platform of the data integration

B. The Research in the Integration of Framework Based on Hibernate

Hibernate was object-relational mapping framework, it was the very lightweight object encapsulation to JDBC. It allows using the XML config file put the ordinary Java objects mapped to a relational database table. The framework manage the whole JDBC layer . So the application programming of the data access layer above in Hibernate, and completely scratch from basement data model.

The data integration scheme based on the Hibernate have lot's of advantages, first it could transparent access the distributed heterogeneous data sources with Hibernator's own technology through, and the HSQL query language which the Hibernate technology has can realize to the unity query for the lower of the heterogeneous data source. That was to said, a statement can be achieved on the data query for all the distributed heterogeneous data sources. This would avoid used the query module. It could be directly call the HSQL query

language of the Hibernate to query the different heterogeneous data source. The data integration scheme based on Hibernate could reduce the writing code of the query module and could be very good for distributed heterogeneous data management.

C. *The Main Factors and Strategy in Distributed Heterogeneous Database Query Optimization*

The query optimization system in the distributed heterogeneous database based on B/S model, would consider the response time of the heterogeneous data access, if the response time was too long, could lead to the WEB server resource exhaustion, lead to a distributed heterogeneous database query optimization system collapse.

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.

Because of various environmental parameters in fixed condition if the network was stable, this experience algorithm was more practical. It can easily predict the future development. If the network was not so stable, the algorithm was very poor. It cannot adapt quickly to the situation changes. In this case, it need administrator manually adjust response time parameters of the distributed heterogeneous data source to control access priority level of the distributed heterogeneous data source.

III. ANALYSIS ANALYSIS OF THE QUERY OPTIMIZATION SYSTEM IN DISTRIBUTED HETEROGENEOUS DATABASE

A. *Analysis of the System Integration in Heterogeneous Database*

Using Hibernate to the heterogeneous database system integration based on B/S structure, the key was to design of the query middleware. The middleware got user query request when user query data, and then search connection whether to have established a connection in database connection pool. If there was no established in database connection pool , then read connection configuration file of the heterogeneous data, so load object mapping file and object of the relationship with database conversion table, and then establish a connection and keep the connection. Thus, the Hibernate generate query in dynamic, the Hibernate realize the searching operation to data source, and so to realize the transparent data access. Its data flow diagram was shown in figure 2.

B. *Scheme Analysis the Response Time in Querys*

Query optimization in traditional relational database usually adopts query tree to show the optimization process of the query operation. However, in distributed database systems, network factors should be considered, and query tree cannot fully reflect this. Network factors was data transfer price in distributed environment, also was the data query with a price. So it must consider the data source response time, and communication channel was available, etc, that was this main factors in the distributed query optimization.

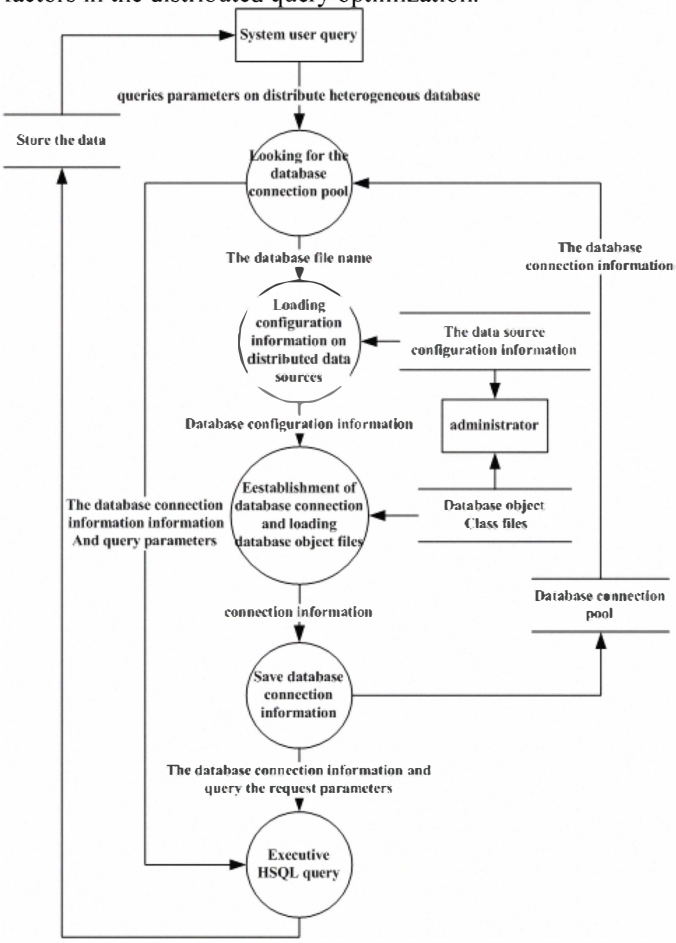

Figure 2 data flow diagram of the system integration

It research using different response time to query the database. How to calculation the response time in distributed heterogeneous database query optimization system and experience algorithm in detailed.

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.

Experience algorithm used experience value for calculation of the response time in distributed heterogeneous data source, and use this experience value to sort the access order the distributed heterogeneous database, and then, according to the order to visited the distributed heterogeneous database .

C. The Realization of the Scheme Analysis on user data Query

1) Data Query Optimization of Single Node Table in Distributed Heterogeneous Data Sources.

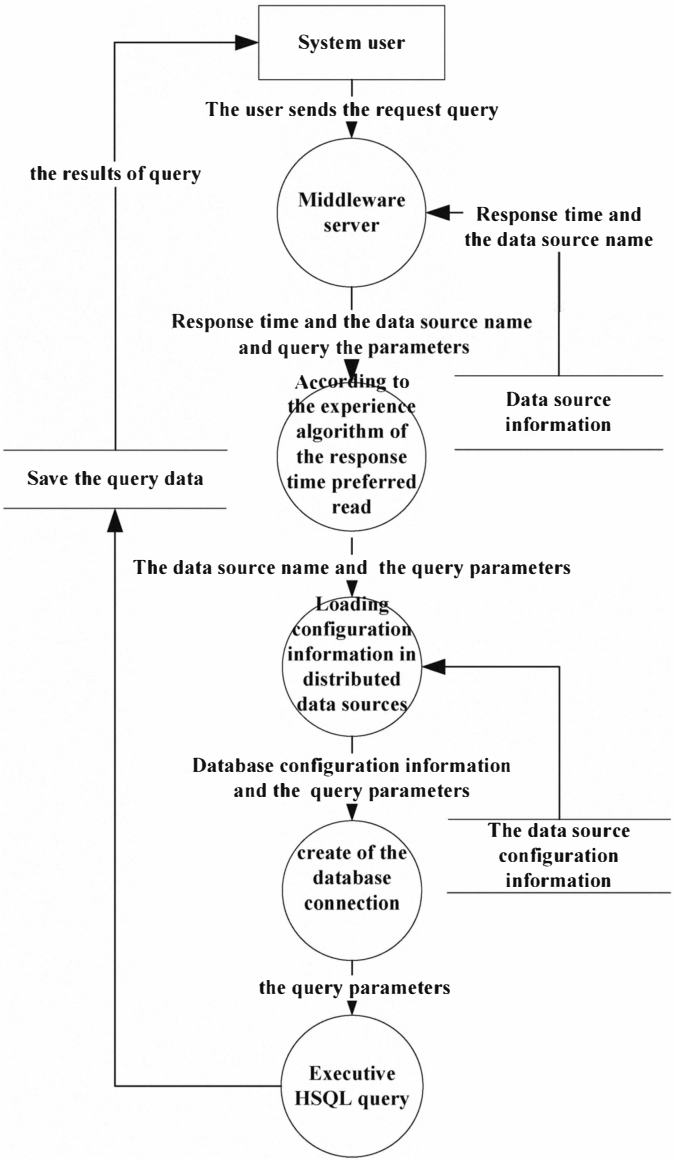

Figure 3The data flow diagram of single node table query

The table query to the same node of the heterogeneous data sources was the data query from a single database in the distributed heterogeneous databases, and the query didn't involved the cascade inquires in the data table in the different nodes, only involved the cascade inquires in the data table its own database or just a cascade inquires form a data table, then all of the data would be integration in the distributed heterogeneous database. This was the important problem in this research. The table query to the same node of the heterogeneous data sources which according to the experience algorithm with the response time can be fully realized the query optimization of the distributed heterogeneous data sources.

First of all, users send a query request, and then middleware server read the response time and data source name, so calculate the experience value of the response time according to experience algorithm in each distributed heterogeneous data source, and then to traverse sequence in order in this experience value on distributed heterogeneous source, then successive to traverse all distributed heterogeneous data source and data integration of data information to return to the user. design specific data flow graph as shown in figure 3 in data query optimization of single node table in distributed heterogeneous data sources.

2) Data Query Optimization of Multi- table in Distributed Heterogeneous Data Source .

The multi-table query in the different node table of the heterogeneous data source was the cascade data query in each of the distributed heterogeneous database. The data query involve different node table. Also was to use a data table in the A data source to find the data in the B data source.

For example: a class table in the A database, a student table in the B database. To use class information in the A database to find the student information in the B database. This was the multi-table query in the different node table of the heterogeneous data source. If it wants to query the student information corresponded the method has two kinds:

The first, it got a record from the class table which in A database, and then cascade query each of the distributed heterogeneous data source, then saved the results of the query. And then got anther record from the class table which in A database to cascade query each of the distributed heterogeneous data source again and so on, until take all the class records, the result of query to integrate to feedback to the user.

Another method, it sent the class information in the A database sent to a distributed heterogeneous data source (called B database), and then put the local cascade query in B database, so the query result would be stored. And then sent the class information in the A database sent to anther distributed heterogeneous data source, and so on, until the entire distributed heterogeneous data source traverse completed. Then integrate the data would send to the users. The order of traverse was accordance with the response time to experience algorithm for the distributed heterogeneous

database.

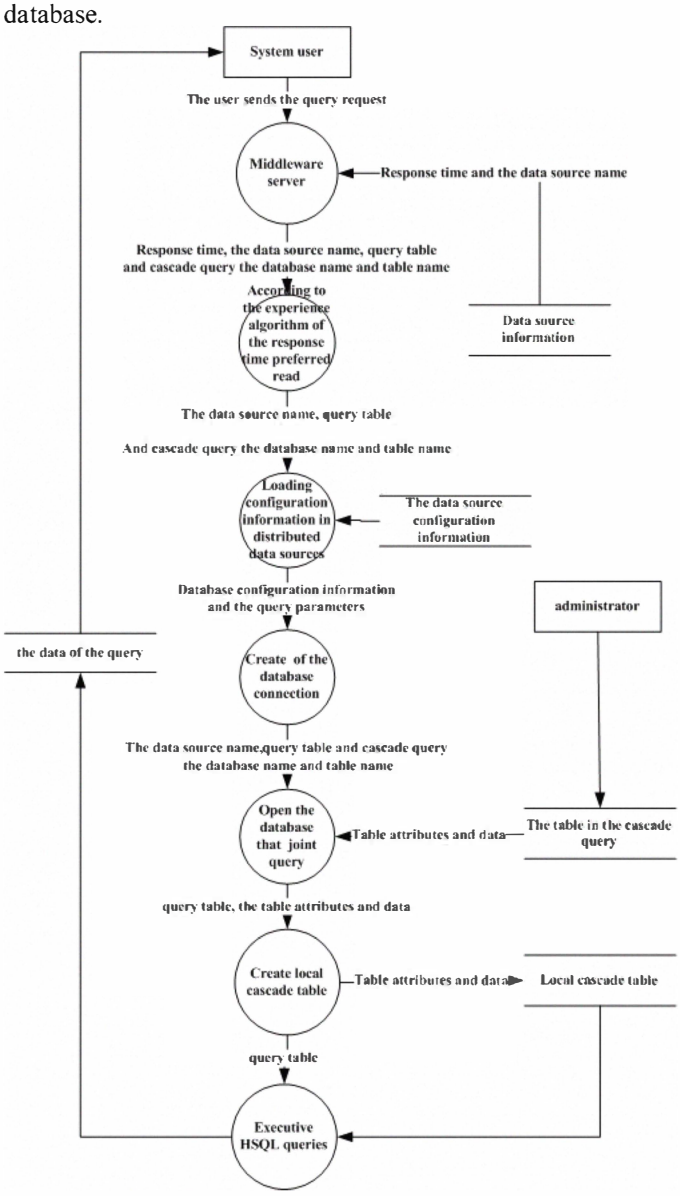

Figure 4 Different nodes of distributed heterogeneous data sources multi-table

And it similar to the single node table query optimization design, that according to experience algorithm to traverse distributed heterogeneous data sources in a distributed heterogeneous database. And so it detailed description on the multi-table of query optimization design. First user sending the query request, and then middleware read the information of the distributed heterogeneous data source, and so according to experience algorithm to calculate the experience value in each distributed heterogeneous data source, through this sorted experience value to traverse sequence distributed heterogeneous source. Then read the database cascade table in the query database, and so table structure and the data sent to distributed heterogeneous data sources, and data source which accept table structure create that table, and would be kept in this table data, and then the local multi-table cascade query taken place. In this way , all distributed heterogeneous data cascade query would take place , and then cascade inquires data would returned to the user in heterogeneous data sources . design specific data flow graph as shown in figure 4 in data query optimization of multi-table in distributed heterogeneous data sources.

IV. SUMMARY

The research on query optimization in heterogeneous database was the field of emphasis in heterogeneous database study. In the study, it can dynamically link multi-heterogeneous database system using the Hibernate technology, query the different user information table in different database. The tiny respond time make heterogeneous database query optimization more practical.

REFERENCES

[1] YangFang, TaoShiQun. Integrated and design the heterogeneous database system based on hibernate [J]. Computer engineering and design, 2007, (7) :-110.

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

[3] XiaoXianJian. The reaearch on the query optimization mechanism of the database middleware [D]. Nanjing: hohai university, 2005,.

[4] Li G Y, Zhang J, Xie Y W.T he Research of Common Data Model for Heterogeneous Data Integration [C]. ICMSE/Russian, 2008, (7) : 126-130.

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