ISSN No. 0976-5697

Volume 3, No. 2, March-April 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Conceptual approach of Query Optimization for Distributed Databases applications

Deepak Sukheja *
Priyatam Institute of Technology and Management,
Indore,India
d_sukheja@rediffmail.com

Umesh Kumar Singh
Institute of Computer Science, Vikram University,
Ujjain, India
umeshsingh@rediffmail.com

Abstract: This paper presents an innovative query processing concept against the higher repetition rate in distributed databases. The new concept proposed in this work also completes the data dictionary, the overall model design and then improves the query flow. This paper shows the concept to improve the efficiency of the query, reduce the amount of intermediate data, and reduce the total cost of network communication.

I. INTRODUCTION

Distributed database system is physically distributed and logically centralized database system, is the product generated from the mutual penetration and organic integration of computer network technology and database technology. Physical distribution refers to that the data composing. The distributed database is distributed to the different computers in the network, and each site in the network has the ability to deal with implement local applications. Logically distribution is refers to that each site is a logical whole which is managed by a distributed database management system, and each site implements the global application through the network communication subsystem. In distributed database systems, data independence refers to data logical independence, physical independence and distributed independence which is also known as distribution transparency. In [1] Distribution transparency refers to that the users do not have to care about the logical sub-chip of the data, the details of data physical location distribution, the consistency of the duplicate copies, as well as do not have to care about what kind of data model the database in local field support. The distribution of transparency makes the system easier and more effective to be used. Distributed database architecture is shown in Figure 1.

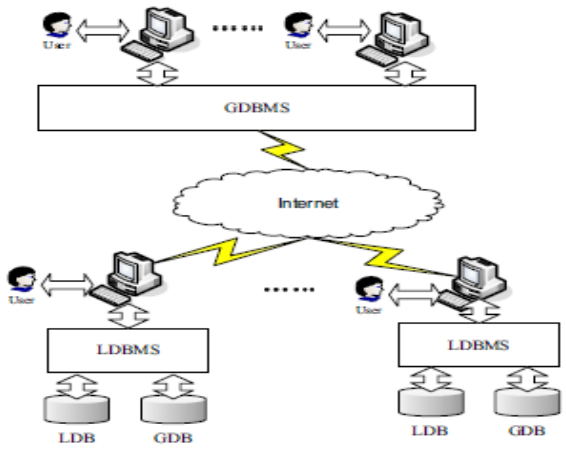


Figure-1 : Distributed database architecture

In short a *distributed database* is a collection of databases that can be stored at different computer network sites. Each database may involve different database management systems and different architectures that distribute the execution of transactions [2]. According to Ovitz Barretto [3], the objective of a distributed database management system (DDBMS) is to control the management of a distributed database (DDB) in such a way that it appears to the user as a centralized database. In recent years, with the extensive application of the distributed systems, distributed database has become an important part of information processing. Distributed data query processing is the interface of users and distributed database, and then query optimization has become a core problem of the relational database area.

II. OBJECTIVES

The major objective of this paper is to design a conceptual query optimizing model for a distributed database management system (DDBMS) to manage the complexity, minimize the response time and to make the best use of system's resources by minimizing network traffic, disk I/O, and CPU time.

III. DISTRIBUTED DATABASE QUERY OPTIMIZATION

Query processing and optimization is very important in the relational database system, and also is one of the main research questions of the distributed database.

A. Distributed Database System Mode Structure:

In a distributed database system, in order to enable users to conceptually and logically process of the data without having to care about the specific expression way of the data in computer. The distributed system uses a three-tier mode structure: inside mode, conceptual model and outside mode. And the distributed database is a logical aggregate of the centralized database based on computer network connection. Therefore, the mode structure of the distributed database system adds some new content on the basis of the centralized database system. Generally, in distributed database, the mode structure can be divided into global outside mode, global conception mode, piecewise mode, distribution mode, local conception mode, and local inside mode. In [4, 5] G. Graefe and Blakeley presented the conversion between the various modes, that can be achieved through multiple images provided by the database management system, the specific structure is shown in figure 2.

B. Query Processing and Optimization Process in Distributed Database:

In distributed database, distributed query processors automatically translate a high-level query on a distributed database, which is seen as a single logical database by users, into an efficient low level execution plan expressed on the local databases. Distributed query processor is mainly to translate the global query request into the local query request in the various interrelated nodes and return the query results of the various nodes, which includes distributed query processing and distributed query optimization. According to hoffer, Prescott and McFadden, in [6] the distributed database, the data distribution to users is generally transparent, when users request for query, they do not have to care about execution location of query, and it can be the same as a centralized database. In [7] it is defined that three steps are necessary to process global queries which are issued against the global schema: First, a global query is decomposed into global subqueries such that the data needed by each subquery are available from one local database.

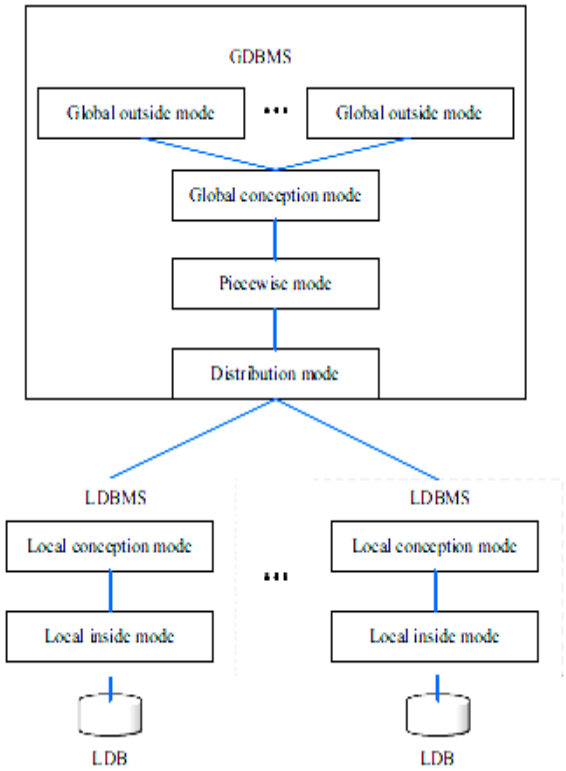


Figure 2: Distributed database model

In second, each global subquery is translated to the queries of the corresponding local database system and sent to execution and finally the results returned by the global subqueries are combined into the final answer (process shown in figure-3). Before produce the final answer, query optimization process also involve during query processing steps. According to XUE Lin In [8], to produce an efficient optimization concept for the global query in distributed database systems, two different objectives are often used to consider the query optimization. One objective uses the standard of minimum total cost, and in addition to considering CPU and I/O cost like the centralized database system, the total cost also includes the data network transmission cost. Which is due to the fact that the data distribution and redundancy make the communication cost needed by transmitting data among the sites in the query processing should be considered, which will cause the total cost increased. Another objective uses the standard of minimum response time of each query, which has a great signification in the distributed database system. The distributed database system is a system composed by multi computers, in which data distribution and redundancy has also increased the possibility of parallel processing also available, thus the response time of query processing can be reduced and the query processing speed will be speeded up.

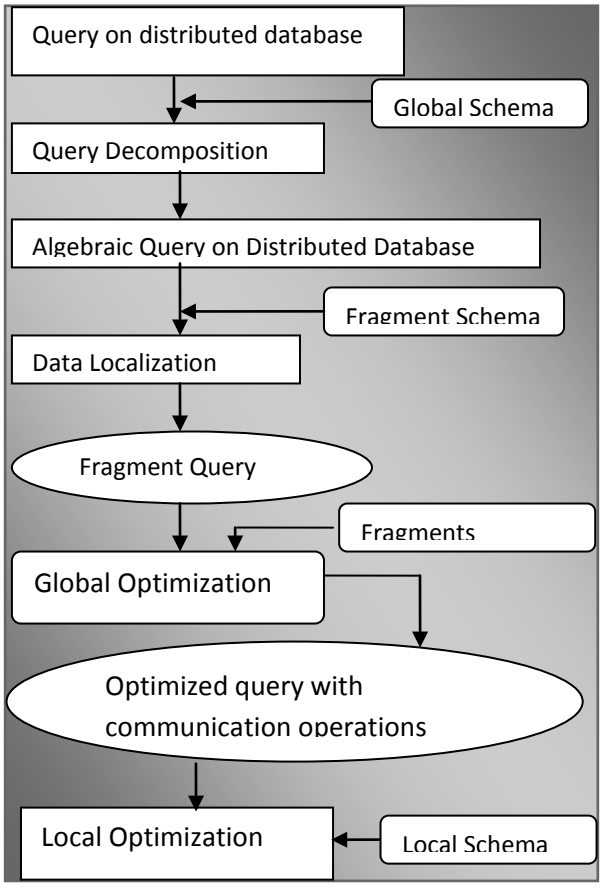


Figure-3.Query processing in DDBMS using optimizer

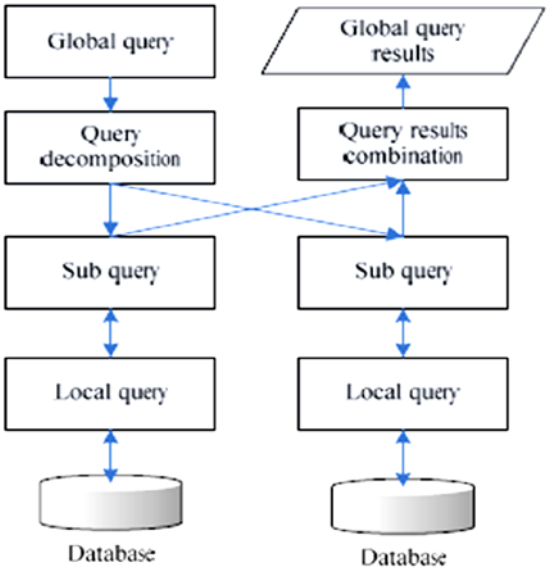
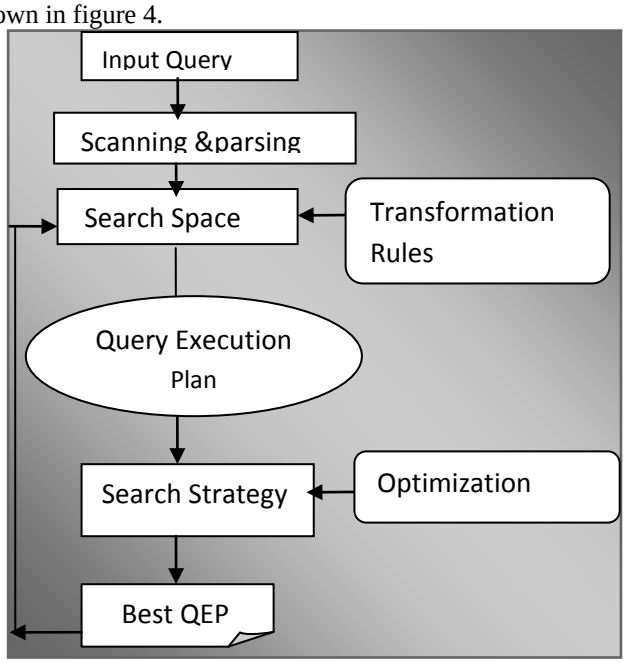


Figure: 4

IV. CONCEPTUAL IMPLEMENTATION OF DISTRIBUTED DATABASE QUERY OPTIMIZATION.

Based on the above analysis a distributed database query processing and optimization process flow is presented, as shown in figure 4.



Figure–4.query processing and optimization process

Distributed data user module analyzes the user query request according to the query optimization criteria the user selected and thus the corresponding query processing method will be confirmed. Scanning & parsing module analyzes the query sentence the user sent and generates the corresponding query tree. Query tree transformed module will convert the query tree received from the anterior stage into the inside expression which is another semantically equivalent and more conducive to optimizing module optimization. Optimization module will process the local query tree according to query optimization strategy to make the total cost smallest. Order processing module will distribute the mission to the corresponding server and return the server processing results to the user. If the operation related to global or local data dictionary, there will be corresponding operations on the data dictionary, in which information statistical operation will update the query sentence table in local data dictionary in accordance with the query frequency; information update operation will update the global data dictionary in accordance with the local data dictionary and the latest changes of the local server performance, and inform other sites through information broadcast operations.

V. CONCLUSION

In this paper, on the basis of the research on the existing distributed query optimization, against the higher repetition rate, a new query processing program was presented, which completed the data dictionary, the overall model design, and then improved the query flow. Query sentence table was added in the data dictionary, which made the local query results frequently used be stored, thus the transmission of large amounts of data in query can be avoided. Experiments showed that the strategies improved the efficiency of the query, reduced the amount of middle data, and effectively reduced the total cost of network communication.

VI. REFERENCE

[1]. Sanjib Kumar Das, Sayantan Das gupta,"Middle Layer Java Software for the Implementation of a Homogeneous Distributed Database System", DBMS LAB – 2007.

[2]. Rick Noel, "Scale Up in Distributed Databases: A Key Design Goal for Distributed System", http://www.cs.rpi.edu/~noel/distr_scaleup/distriuted.html, May 17, 2004.

[3]. Ovita Barretto, "Security Issues with Distributed Databases", http://www.students.depaul.edu/~obarratt/Security.htm, May 20, 2004.

[4]. G. Graefe, "The Cascades Framework for Query Optimization", Bulletin of the TC on Data Engineering, Vol 18 No. 3, Pages:19-29, September 1995.

[5]. Blakeley, J.A.; Cunningham, C.; Ellis, N.; Balaji Rathakrishnan; Wu, M.-C. "Distributed / heterogeneous query processing in Microsoft SQL server".Data Engineering", Proceedings. 21st International Conferenceon Volume 5, Issue 8, April 2005 Page(s): 1001 – 1012, ICDE 2005.

[6]. Jeffery A. hoffer, marry B. Prescott and fred R. McFadden, " Mordern Database Management, ," 5th Ed., PHI 2007.

[7]. Query Optimization in Oracle database10g Available from www.oracle.com/technology/products/bi/db/10g/pdf/twp_general_query_optimization_10gr2_0605.pdf.

[8]. XUE Lin, "Query Optimization Strategies and Implementation Based on Distributed Database", ICCSIT, IEEE Xplore 2009, page 480 – 484.