Original Article

A HYBRID APPROACH TO OPTIMIZING QUERY PROCESSING IN DISTRIBUTED DATABASE SYSTEMS

Briggs Tamunoemi Chukwuemeka and Diri Oghenekaro Blessing

Department of Computer Science, Rivers State University, Port Harcourt, Nigeria
DOI:https://doi.org/10.5281/zenodo.15431638

Abstract: The issue of optimizing query is a cost sensitive process and with respect to the number of associated tables in a query. Its number of permutations grows exponentially. Major model for server-client query evaluation are data shipping and query shipping techniques thus, are conventionally used despite some limitations like speed. Based on this, this study was intended to actualize hybrid technique using a combination of these existing techniques. This present work adopted constructive research which aided the hybrid technique for optimizing query processing in a distributed database as a designed. In this study, object oriented design for software development and hardware materials were used. The data shipping techniques was fast in query retrieval as opposed to the query shipping technique however, the had a synergic strength from the two existing techniques to actualize an efficient performance (runtime) This result was similar with previous finding. For server machine, query shipping technique is recommended while for client machine, data shipping technique. Query shipping technique is good for the sever machine while data shipping is good for the client machine however, the proposed hybrid technique is recommended for distributed database to achieve best results for the smooth running.

Keywords: Hybrid Technique, Data Shipping, Query Shipping, Optimization.

1. Introduction

There is a progressive rapid increase in the quantity of data in recent time and as a result the need for a database management system (DBMS) is paramount in order to properly store and manage data in a more efficient manner. Query language is an effective tool that creates an interface to the user to store and access that data. In time past, SQL has emerged as a standard query language as reported by some studies (Vidya Banu & Nagaveni, 2012), (Rahman, 2010), (Chaudhuri, 1998). Notably, components like query optimizer and query execution engine

(Chaudhuri, 1998) are involved in query evaluation. Query optimizer chooses query operation order with the fact that conventional relational algebra operators can be executed in a variety of Order (Badia et al., 2005).

In a distributed database when a user issues a query, it takes a lot of time to process the query, and the storage area which its activity involves storing the information or data will perform optimally slow. Data-shipping and Query-shipping are techniques used but they have certain limitations in their performances. Data-shipping has a limitation of specifying that all operators of a query that would be optimized in a distributed database system, must be implemented in the client- machine site where the queries were initialized while queryshipping technique has a limitation of which, all queries are completely evaluated at the server machine. Thus, scan operators are placed at the server where the initial replicas of tables are kept and the extra query operators (apart from the display operator) in the node of one of its producer.

The study was aimed at developing a hybrid technique for optimizing query processing in a distributed database system. The Specific objectives are evaluating queries Optimization at the Client Machine using the proposed hybrid technique, queries Optimization at the Sever Machine using the proposed hybrid Technique and a comparative analysis of the existing techniques and the proposed hybrid technique.

Significantly, the study was carried out to improve efficiency in an object-oriented and relational model in a distributed environment by hiding user intent and advancing the operation of the server and client machine. This present study explores best aspect of datashipping (object-oriented model) and query shipping (relational model) to actualize a hybrid technique that outperformed both techniques.

The study covers an overview of the processing of queries, in addition to different techniques used in distributed optimization and processing of queries. We emphasize on several degrees of data or information shown concerning the query by each user and then use our proposed hybrid technique to strike a balance between efficiency and performance respectively. However, the study was delimited to data shipping technique which does not scan queries in the server machine and the query shipping technique which also poses difficulties scanning query in the client machine whenever a query is submitted there and finally, the proposed hybrid shipping.

2. Review of Related Literature

Query optimization is the most important stage in query processing where the database optimizer has to choose a query-evaluation plan with minimized cost and maximized performance (Chaudhuri, 1998), (Badia et al., 2005), (Kabra & DeWitt, 1998), (Warshaw & Miranker, 1999), (Waluyo, 2005). Query optimization is a complex job in distributed client and server machine while data position becomes a main issue. For one to optimize queries correctly, enough information has to be obtained to ascertain what data access methods are most efficient (for instance, relation and field cardinality, group information, and catalog availability). Optimization algorithms have a significant effect on the operations of distributed query processing.

The execution of query in distributed system is seriously subjected to the competence of the optimizer to get effective query evaluation plan. It shows that query optimization is one of the most critical phases in the execution of queries in distributed environment. Query optimization is a difficult task in distributed environment because of numerous factors like data allocation, speed of communication channel, indexing, availability of memory, size of the database, storage of intermediate result, pipelining, and size of data transmission as reported by (Mullins,

2006). The execution plan which is produced by the query optimizer denotes an execution strategy as a minimum cost for any query.

Several different blends of sub queries can be used to assess a query. Nonetheless the blends and cost of assessment are dissimilar but then each combination is assessed to the equivalent result. These blends otherwise known as combinations are called access plans or query execution plans (QEP) (Matysiak, 1995). The query optimizer usually selects the optimal cost in essence the minimum cost query execution plan amongst them; which is commonly seen as a problem of the query optimization as reported by (Matysiak, 1995). Furthermore, query optimizer generates many alternative query execution plans for selecting the optimal query plan and estimates the execution cost of each of them to choose the QEP having lowest cost. Optimal query plan selected by query optimizer is forwarded to query execution engine which is responsible for execution of query. Also, query optimizer sends access plan to query execution engine, this process of optimization is the most critical step in query evaluation because it determines both the execution time and the space complexity of query. However, query optimization on its own is very complex and expensive as a result its computational complexity is determined by the number of alternatives for access plans which must be evaluated before deciding the best query execution plan (Matysiak, 1995). The alternative planes grow exponentially with the increase in number of relations involved in a query. Some studies have reported methods to address this issue although it is a long time (Jarke & Koch, 1984), (Swami & Gupta, 1988), (Horng et al., 1994), (Steinbrunn et al., 1997).

Existing Technique(s)

Several techniques used in querying a distributed database in order to retrieve, update, delete and insert new data into the database. Traditionally, Data shipping technique, Query shipping technique, Wigan technique, and mutant query plans are commonly used to achieve task in a distributed database. The retrieval process is slow and not cost effective using the query technique even though it has the benefit of been rapid. But the opposite effect is seen in shipping technique on the client machine on the other hand, a similar result is achieved when query shipping technique is applied on server machine- the process is fast, cost is reduced and the data transfer time is small thereby making this technique reliable and good in the server machine. Therefore, these variations noticed when each technique is applied on either the client machine or the server machine is to be corrected or eliminated by our proposed system since in our proposed research, we are expected to use a combination of both data and query shipping techniques which constitute what is known as hybrid technique for the optimization of a user query.

Hybrid Technique

Hybrid technique is the combination of two or more techniques. The basic idea behind hybrid technique is that, no one technique can fully satisfy all the constraints of an optimization problem for this reason, more than one techniques is needed to satisfy all the constraints and complement each other's drawbacks. There are several hybrid techniques used for query optimization problem. A new hybrid approach according to (Kadkhodaei & Mahmoudi, 2011) uses genetic algorithm and ant colony algorithms. Nevertheless, this present study proposed a newly designed technique which is the combination of two techniques; datashipping (object-oriented model) and query shipping (relational model) to objectify a hybrid technique which is fast, reliable, flexible, has the ability

Original Article

to preserve the (privacy) and efficient algorithm to manage the query processing in a distributed database management system of a query in a distributed environment which will be implemented using Microsoft SQL Server 2008 as the back-end and ASP.Net using C# programing language as the frontend.

n India based study on Query Optimization: An Intelligent Hybrid Approach using Cuckoo and Tabu Search reported that query optimization is a vital part in designing database management systems, basically to invent an optimal query execution plan within the short possible according to (Joshi, 2013) in their study proposed a heuristic based algorithm as a solution of MJQO problem since MJQCO is an fundamental part of query optimization. The result of simulation showed some exciting results in favor of their proposed algorithm and they concluded that proposed algorithm can solve MJQO problem in less amount of time than the existing methods. Their study was however limited to an algorithm which a combination of two basic search algorithms was namely; cuckoo and tabu search (Joshi & Srivastava, 2013). Furthermore, the proposed system ensures that the right technique is applied at both the client and server machines respectively unlike the existing system that utilizes only one kind of technique no matter the machine that it is sent from. Hybrid-shipping technique permits each operator to be annotated in any pattern that is permitted through either data or query shipping technique. Between these three techniques hence, hybrid shipping technique contains best elasticity when generating query plans. It ensures the mandating of operators to nodes should be established during execution time. Query and data shipping technique combined, can give superior privacy than any of the aforementioned techniques used alone on the existing system even to the detriment of keeping extra neighboring state.

3. Methodology and Design

Research methodology is the study of how a specific research project is been carried out using some laid down techniques or approaches. It can also be seen as the scientific study of how a research problem is solved.

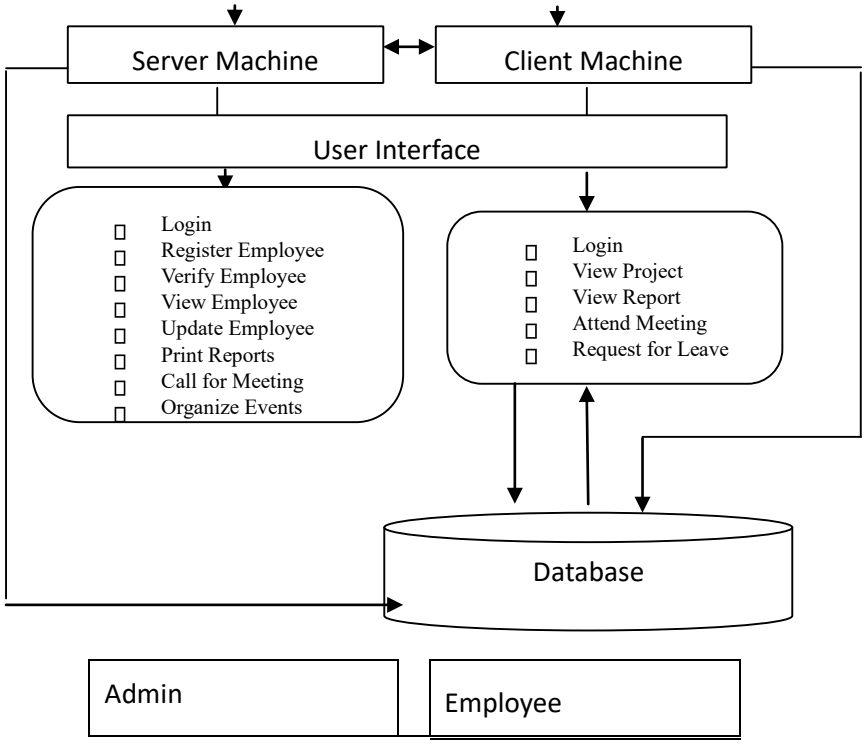
a. Research Methodology

This research adopted a constructive research according to (Booch, 1994). The constructive research approach used here aided the hybrid technique for optimizing query processing in a distributed databases and it is mostly used in software engineering and computer science.

b. Design Methodology

This work adopted this method because the steps involved in this methodology are best suitable in achieving the aim of this research. Object-oriented design methodology can help in the design of the system by first, analyzing the system's requirements; which are the user requirements, functional requirements, performance requirements and the design requirements.

Original Article



Architecture and Data flow of the Proposed Distributed Database System

Architecture of the Proposed Distributed Database System is shown on figure 1 demonstrating the two machines namely the server and the client for admin and employee respectively with a user interface. Data flow here are of three levels from the Distributed Database for Admin (Server) & Employee (Client). See figures 1 & 2 for details.

Fig. 1: Architecture of the Proposed Distributed Database System

4. Experimental Results

The experimental analysis of this research was performed using Java, MYSQL database and NetBeans IDE text editor were the application packages used.

S/N	Number Of Samples
1	1000
2	3000
3	5000
4	7000
5	10000
6	15000

Original Article

Table 1 shows, the number of entries in a relation that will be processed using data-shipping, query-shipping technique and hybrid technique

Comparison of Existing Technique and the Proposed Technique at Client Machine The Comparison of Queries Optimized using Data-shipping Technique against Queryshipping Technique (Existing Technique) and Hybrid Technique (Proposed Technique) reported changes in the different variables as seen on table 2. However, the hybrid technique had a lesser runtime.

Table 2: Comparison of Existing Technique and the Proposed Technique at Client Machine

Number Of Entry	**Query percent**	**Disk space (kb)**	**Time (s)**	**Disk space (kb)**	**Time (s)**	**Disk space (kb)**	**Time (s)**
		Data Shipping Technique		**Query Shipping Technique**		**Hybrid Technique**	
1000	5	84	0.7	1200	6.4	125	0.4
3000	7	125	0.8	1300	6.9	150	0.7
5000	9	150	0.9	1500	8.0	121	0.8
7000	12	171	1.2	1900	10.1	171	1.1
10000	15	221	1.9	2500	13.3	364	1.7
15000	17	450	2.4	3000	16.0	450	2.2

Table 2: Comparison of Existing Technique and the Proposed Technique at Server Machine

Comparison of Existing Technique and the Proposed Technique at Server Machine

Comparison of Queries Optimized using query-shipping Technique against Data Shipping Technique (Existing Technique) and Hybrid Technique (Proposed Technique), the result showed variations in the disk space and time at different entry numbers as shown on table 2.

Original Article

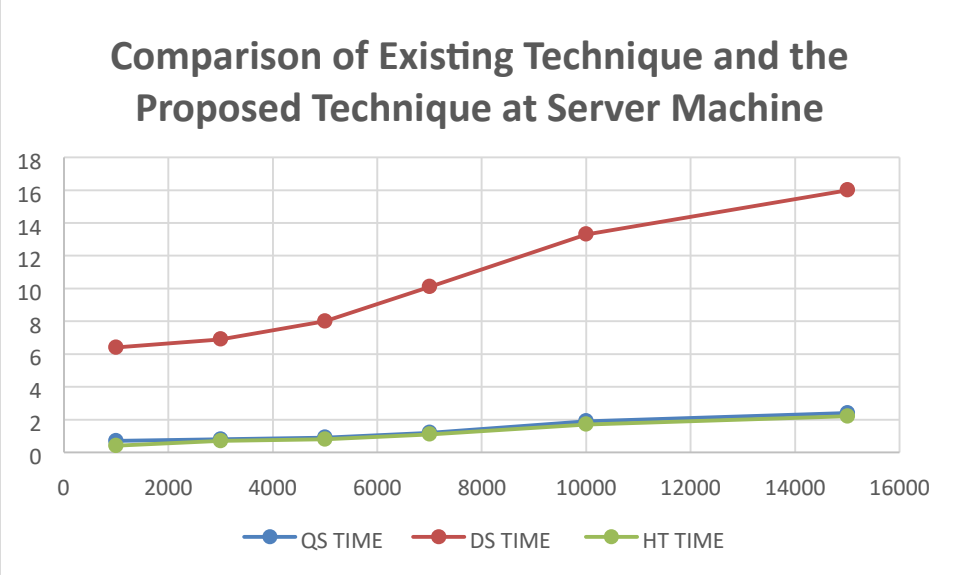


Fig. 2 graphical representation of Existing Technique and the hybrid Technique at Server Machine with QS as Query shipping time, DS, data shipping Technique time and HT, read as hybrid technique time

Table 3: Queries Optimized at both Client and Server Machine Using Hybrid Technique

Number of Entry	Query percent	Disk space (kb)	Time (s)	Disk space (kb)	Time (s)
		Queries Optimized at the Client Machine		Queries Optimized at the Server Machine	
1000	5	125	0.4	125	0.4
3000	7	150	0.7	150	0.7
5000	9	121	0.8	121	0.8
7000	12	171	1.1	171	1.1
10000	15	364	1.7	364	1.7
15000	17	450	2.2	450	2.2

Table 3: Comparison of Existing Technique and the Proposed Technique at Server Machine

Graphical Representation of the result of the comparison of the techniques. See table 2 & 3.

Original Article

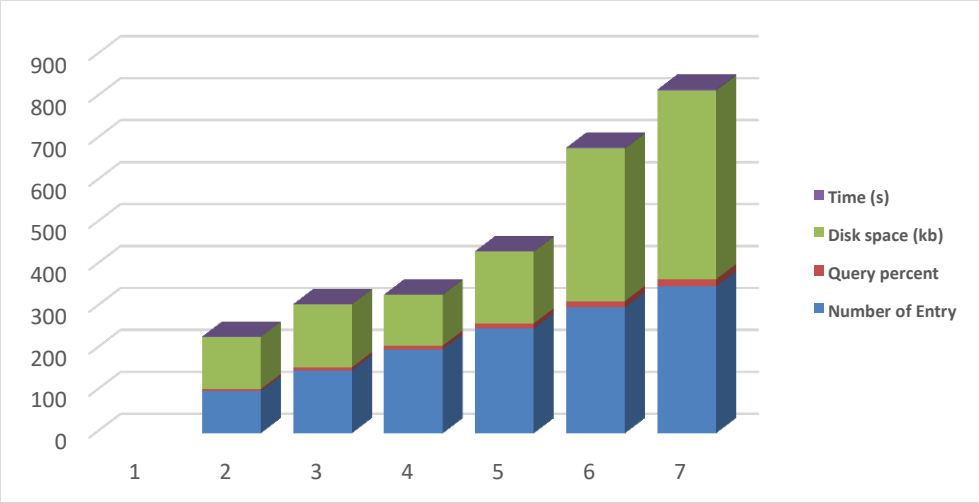


Fig. 3 Comparison of Existing Technique and the Proposed Technique at Server Machine

5. Discussion of Results

This experiment was done using a laptop running Windows 7 operating system with 2.3GHz core i3 processor and 2GB of RAM memory. The analysis was performed using one set of data to test the various techniques data shipping, query shipping and hybrid techniques. Result from the experiment showed six ranges of data used to test the efficiency of the system beginning with 1000 and incremented to 3000, 5000, 7000, 10000 and 15000 records as revealed in the tables from the previous chapter. By the end of the first experiment as revealed in table 3 data shipping technique shows to be the best execution plan used by the hybrid technique to optimize the run time of the queries.

The result obtained using the proposed technique showed no observed disparity between the Server and Client Machine in terms of query processing in a distributed database as shown in table 2

Comparatively Data Shipping Technique (Existing Technique) when used to optimize query at the server machine showed a high result in terms of runtime and disk space on the contrary, when used to optimize query at the client machines showed no observed difference as regards to the proposed hybridized technique. Similarly, same was observed when Query Shipping Technique was applied at the client machine.

From the observed empirical evidence, the above reported limitations (difficulties in processing queries) particularly when any of the two existing technique is used separately, these limitations is been bridged by the synergic effect of the proposed hybridized technique. To this end, the experiment and given result proved the better performance of the proposed technique (hybridized) rather than other existing techniques given in the literature. The main reason for the efficiency of this technique is the ability of all operator of a query to be able to optimize a query in a distributed database system i.e. to work on both the server machine and the client machine with at an optimal runtime and disk utilization which the existing techniques could not achieved.

6. Future Works

Future research should consider other algorithms and techniques to ascertain more efficient ways to optimize query in a distributed database on the datasets. Also, further research should be carried out to increase the data size to improve efficiency and accuracy.

Original Article

References

Badia, A. Doorn, V. Ferraggine, L. & Rivero, J. (2005). Advanced query optimization. doi:10.4018/978-1-59140-560-3.ch003.

Booch, W. (1994). Method: A Rational Approach. Benjamin-Cummings Pub Co.

Chaudhuri, S. (1998). An overview of query optimization in relational systems. In Proceedings of the ACM Symposium Principles Database Systems (Pods '98), Seattle, Washington (pp. 34–43).

Horng, J. T. Kao, C. Y. & Liu, B. J. (1994). A genetic algorithm for database query optimization. In Proceedings of the First IEEE Conference on Evolutionary Computation, (Vol. 1, pp. 350-355). .

Jarke, M. & Koch, J (1984). Query optimization in database systems. ACM Computing Surveys, 16(2), 111–152. doi:10.1145/356924.356928.

Joshi, M & Srivastava, P. R. (2013). Query Optimization: An Intelligent Hybrid Approach using Cuckoo and Tabu Search. International Journal of Intelligent Information Technologies (IJIIT), 9(1), (pp. 40-55.)

Kabra N. & DeWitt, D. J. (1998). Efficient mid-query re-optimization of suboptimal query execution plans. In: Proceedings of the ACM SIGMOD Conference, Seattle, WA.

Kadkhodaei, H. & Mahmoudi,F. (2011). A combination method for join ordering problem in relational databases using genetic algorithm and ant colony. In Proceedings of the IEEE International Conference on Granular Computing (GrC), 3, 12-317.

Matysiak, M. (1995). Efficient optimization of large join queries using Tabu Search.

Information Sciences, 83(1-4), 77–88. Doi: 10.1016/0020-0255 (94)00094-R.

Mullins, C. S. (2006). Distributed Query Optimization, Technical Support.

Rahman, N. (2010). Incremental load in data ware-housing environment. International Journal of Intelligent Information Technologies, 6(3), 1–16. doi:10.4018/jiit.2010070101.

Steinbrunn, M. Moerkotte, G. &Kemper. A (1997). Heuristic and randomized optimization for the join ordering problem. The Very Large Data Bases Journal, 6(3), 191–208. Doi: 10.1007/s007780050040.

Swami, A. & Gupta, A. (1988). Optimization of large join queries. In Proceedings of the Acm Sigmod International Conference on Management Data (Sigmod '88) (pp. 8–17).

Original Article

VidyaBanu, R. & Nagaveni, N. (2012). Low dimensional data privacy preservation using multi-layer artificial neural network. International Journal of Intelligent Information Technologies, 8(3), 17–31. doi:10.4018/jiit.2012070102

Waluyo, A. B. Srinivasan, B. & D. Taniar. Research in mobile database query optimization and processing, Mobile Information Systems, vol. 1(4), IOS Press.

Warshaw, L. B. & Miranker, D. P. (1999). Rule-based query optimization revisited. In: Proceedings of the Eighth ACM International Conference Information and Knowledge Management, Kansas City, MO.