International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

Optimized Query Processing In Distributed Database System Using Hybrid Model

Eze, Chinyere[1], Ojekudo, Nathaniel[2]
[1]Computer Science Department, Ignatius Ajuru University of Education, Port Harcourt, Rivers State Nigeria
[2]Computer Science Department, Ignatius Ajuru University of Education, Port Harcourt Rivers State Nigeria

Submitted: 25-07-2021 Revised: 04-08-2021 Accepted: 06-08-2021

ABSTRACT

The formation of a very high-functioning and competent distributed database system that uses the superlative aspect involving object-oriented and relational methodologies needed in the server-client machines design which may completely utilize server-client machine hardware and software in a flexible way. Major model for server-client query evaluation are data and query shipping techniques. We initially described these courses of actions (techniques) concerning the limitations they attached within operator node selections throughout query optimization in distributed environment. We will also define and estimate a more adaptable technique known as hybridized-shipping technique that combines the query and data shipping techniques, that may implement queries in the server, client, or the joining of more than one technique. Hybridized-shipping technique as described to an extent combine the optimal of two main techniques, while in most instance, works optimally than either of the query and data-shipping technique. The application of hybridized-shipping technique generates an onerous complication involving query optimization as the search space needed to evaluate any given query is comparative large than either of the two-technique used. The methodology we used in this application is Object-Oriented-Analysis-Design (OOAD). This work is developed in ASP.Net using C# programming language as the front end and Microsoft SQL Server as the back end. Experimental result shows that the use of data shipping, query shipping and hybridized technique to query company employee's data from different locations at client-server machine, number of data entry, CPU time taken, Disk space for data entry which is measured in kilobyte(kb) as parameters to check for query processing was able to optimized the distributed database.

I. INTRODUCTION

The Optimization of a query within distributed database environment involves the process whereby multiple query plans are generated to satisfy that query. The equivalent plans were studied while the best query evaluation plan is acknowledged. "The actual updating and retrieval of data is performed through various low- level operations" (Reza et al., 2001). The user (Application program or end user) communicates with the database by submitting requests for data selection (queries) or management (updates) in a distributed system. However, given the existence of exhaustive enumeration algorithms that can find the optimal solution for queries with dozens of relations very quickly (Moerkotte and Neumann, 2008; Fender and Moerkotte, 2013), there are few cases where resorting to heuristics or disabling bushy trees should be necessary.

For every query that is initialized whether in a centralized or distributed database system, it must be optimized by the optimizer to generate the best and the less expensive query evaluation plan from a set or sequence of equivalent query evaluation plan. Object-Oriented database system which gives more strong data modeling capabilities are usually found in Data shipping technique, whereby, the needed data or information is sent into a client system which will be worked on, and perhaps stored, within. "That is why relational vendors are moving towards integrating object-oriented features into their systems, e.g., the emerging SQL3 standard" (Kulkarni et al., 1994). That is, if a database system which has relational functionalities is later integrated with object-oriented functionalities, the database system will be more efficient and more beneficial.

On the other hand, "Relational models and its offspring are usually built upon the query shipment or shipping technique, whereby, most of the queries being processed are carried out in the servers machine, while, vendors that have object-oriented systems are adding more powerful query

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

functionalities" (Cattell et al., 1994). If this is achieved the database system that has an object-oriented functionalities, will be more effective as well, when query facilities is added. "Because of these differences, and complementary strengths, it has become apparent that database systems combining the best aspects of the relational and object-oriented approaches are likely to gain acceptance across a large range of applications" (Stone Braker et al., 1990). In other words, we have decided to combine the best aspect of data and query shipping (hybridized) techniques to achieve an efficient distributed system

1.2 Statement of the Problem

In the distributed database environment when data-shipping and query shipping technique are applied separately, the following problems occur;

i. The scan operator of data-shipping technique does not scan (parser) query at the server machine.
ii. It has a limitation of specifying that all operators of a query that would be optimized in distributed database system, must be implemented in the client- machine site where the queries were initialized.
iii. Query-shipping technique has a limitation of which, all queries are completely evaluated at the server machine.
iv. 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.

Since the Data-shipping technique has a very good advantage in the exploitation of the client resources, (memory, CPU, and disk) and on the other hand, the query shipping technique exploit the server resources, by having entry to huge series of horizontal rows within the relation, There is need to construct an optimizer that will combine the advantages of both techniques as a complimentary strength to achieve a hybridized shipping technique, to take advantages of the two techniques to provide a robust technique that will help to retrieve the results of queries faster and more efficiently.

1.3 Aim and Objectives of the Study

The aim is to develop a randomized optimizer that will take advantage of the two techniques (Data-shipping and query-shipping techniques) that will generate an efficient privacy-preserving distributed processing technique. Specific objectives are to:

(i) Design a technique that can preserve the intension (privacy) of a query in a distributed environment.
(ii) To provide and develop query parameter – number of data entry, CPU time taken, and disk space for data entry which is used for relational processing of query in the server as well as client machine.
(iii) To test and evaluate developed system efficiency.
(iv) This system would be implemented using Microsoft SQL Server 2008 as the back-end and ASP.Net using C# programing language as the front-end.

II. RELATED WORKS

A. Distributed Database

When we have one rational database which can be extended materially over client machine in many areas that are linked together through a computer network, such situation can be said to be a distributed database (DDB). "A distributed database is a database in which storage devices are not all attached to a common processing unit such as the C.P.U." "It may be stored in multiple computers located in the same physical location or may be dispersed over a network of interconnected computers" (Gupta, S., et. al., 2011).

In distributed databases, there is software known as Database Management System (DBMS) which helps in the control of data entry to different sites, if such is achieved, then we call the system a distributed database management system. "Processors communicate with one another through a communication network" (Silberschatz, A., et al., 1997; 1998).

For an effective sending and receiving of data in distributed database, there is need to have database management system with the aid of the processor to manage the retrieval of data within several nodes.

Figure 2.1 depicts a well-known design in support for client machine that has Distributed Database Management System competence. Where every node owns at least one neighboring database management system which controls data in the database that are kept in their respective nodes. It can also be noticed that, every node contains one replica of distributed database management system in addition to a correlated distributed data directory. Distributed data directory has several positions of every data within the network computer, in addition to the data descriptions, and needs for data from end users that are made use of initially through the distributed database management system.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

B. **Data Shipping Technique**

Data shipping technique does not process query at the server machine. Rather, all query processing is optimized and executed at the client machine. "A key decision for data shipping technique is the unit of transfer of data" (David J., Philippe F., et. al., 1990).

Data-shipping technique (DST) requires every operators of user query shall execute in client device where the user initialized the query. Data Shipping execution plans, for that reason, support that, the node comment of all scan and of display operator are done at the client device, while comment of every additional operator are "consumer" (note that, "display operator" in the root of tree is carried in client) some of the merits of the data shipping techniques areas follows.

. i) It exploits the resources (CPU, memory, and disk) of powerful client machines,

ii) And reducing communication in the presence of locality or large query results

iii) It permits light-weighted communication among user application and the database software, that is required to help the navigational data retrieval.

iv) It utilizes the resources (such as; Central Processing Unit, disk, and memory) of the robust client device during Optimization on distributed query processing.

v) Reduces transmission within the existence of district or huge query outcomes.

C. Query Shipping Technique

This technique ships SQL queries from the client machines to the server. The server performs all of the query-processing effort, and the answer is returned in the form of a stream of tuples" (Ivan T., et. al., 2000). Query shipping technique (QST) puts scan operators on server machines in which the basic replicas of tables is located, while every more operators (apart from the "display") in the node of one of their "producers". E.g., Join operator could be done both in the "producer" of the "inner table" or the "producer" of the "outer table". Because of that, evaluation plans which encourage query shipping techniques at no time have consumer comments or scans which are done in the client system. Some of the merits of the techniques are as follows.

i) the reduction of communication costs for high selectivity queries

ii) the ability to exploit server resources when they are plentiful

iii) and the ability to tolerate resource-poor (i.e., low cost) client machines

It uses the resources in the server effectively.

iv) It reduces transmission overheads for huge demanded queries.

v) It has capability of using server resources if they are abundant.

vi) It has capacity to stand resource-bad (that is, little overhead) client device.

Move diagram on page 54 here. LET IT BE FIGURE 1.

III. METHODOLOGY

Methodology is the documented collection of policies, processes and procedure used by a developer. The methodology we used in this application is Object-Oriented-Analysis-Design (OOAD). We adopted this methodology because it improves the quality of the system due to program reuse and maintenance.

OOP methods make code more maintainable. Identifying the source of errors becomes easier because objects are self-contained (encapsulation). In this OOAD, there is no separation between the analysis and the design phases, which improves communication with the users throughout the project development. Object Oriented Analysis Design is an industry-proven methodology for building high-valued object-oriented machines. This succeeding application development methodology, includes three features namely: Object Oriented Analysis, that works on the design needs as well as the total architecture of the client machine, which also focuses on explaining what the client machine can do as regard to the key objects within the problem domain; Object Oriented Design, which transform a client machine architecture to programming constructs (that is; classes, method descriptions, and interfaces); and Object Oriented Programming that executes these programming constructs.

The main proposal after this object-oriented language is object decomposition that separate difficult system application into its several objects, that have data and the functions which work on the data towards one unit.

MOVE DIAGRAM ON PAGE 56 LET IT BE FIGURE 2.

Figure 3.7: Use Case Diagram of the Proposed System

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

IV. RESULTS AND DISCUSSION

The time and disk space taken for the processing and optimization of each data entry had their S. I. unit in second (s) and kilobyte (kb) respectively. WE examine our application in order to discover the advantage of using hybridized technique over the other techniques. When a query is sent to the database, the program checks if it is to the client machine or server machine and if it is a client machine, the application uses data shipping technique in order to achieve better result at the lowest cost and time.

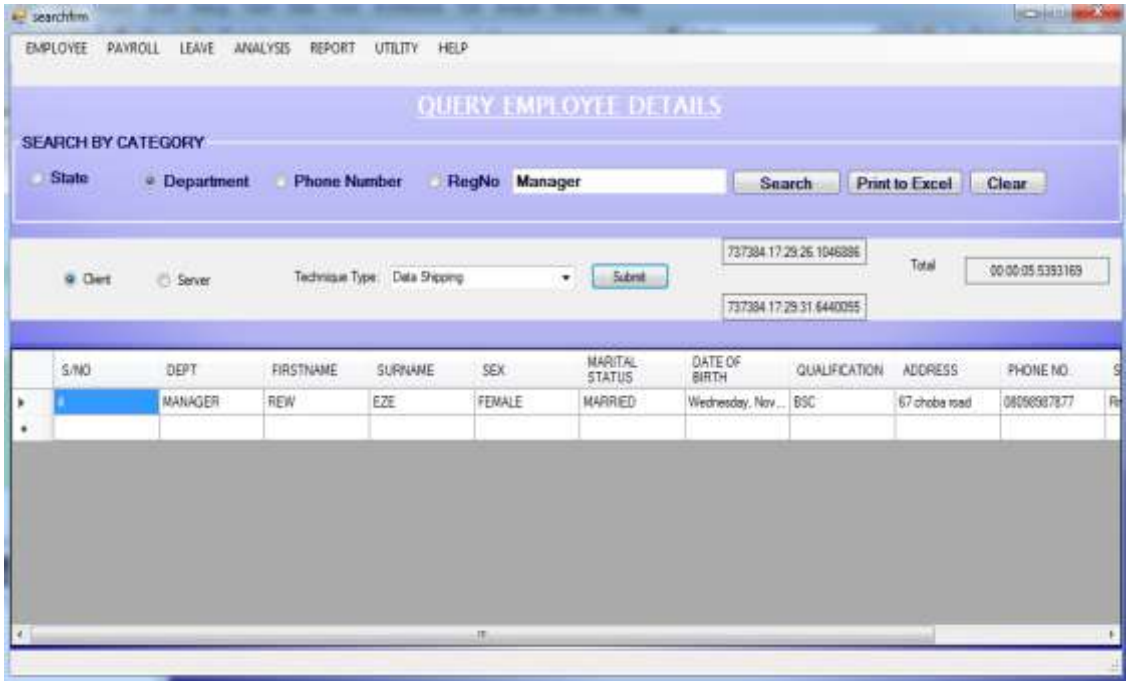

Figure 3:**Data Shipping Execution at client machine**

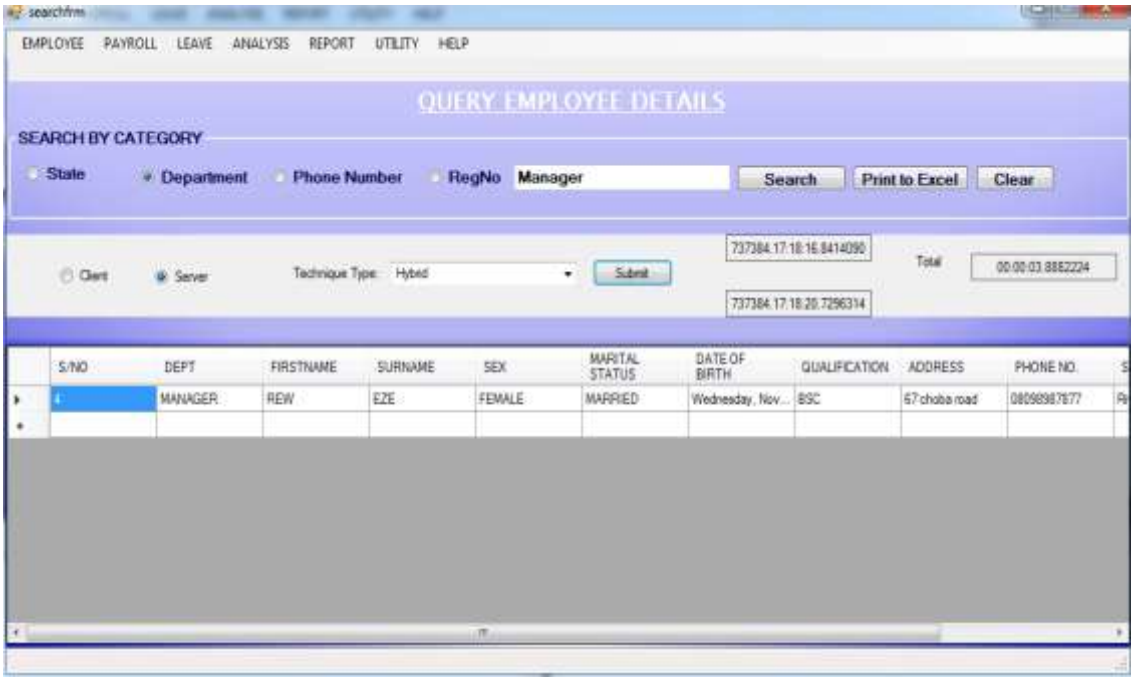

Figure 4: **Hybrid Execution at Server machine**

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

Table 1 contains standard values for the experimental setup results in the table below. Every data entry is accounted for during execution. The time and disk space taken for the processing and optimization of each data entry had their S. I. unit in second (s) and kilobyte (kb) respectively.

Table 1.0 Queries Optimized at the Client Machine

Number of entries	**Percentage due to query %**	**Data-shipping Technique**		**Query-shipping Technique**	
		Disk space (kb)	Time (s)	Disk space (kb)	Time (s)
300	15	450	2.4	3000	16
280	13	364	1.9	2800	14.9
250	5	125	0.66≈0.7	2500	13.3
190	9	171	0.91≈0.9	1900	10.1
150	10	150	0.8	1500	8.0
130	17	221	1.17≈1.2	1300	6.9
120	7	84	0.4	1200	6.4

Table 4.4 illustrated how, the number of entries in a relation can be worked on by the use of data-shipping and query-shipping technique. The percentage due to query, is gotten from a particular structuredenglish query language (sequel), stating some conditions that involves the simplest expression (mapping) in a relation, it is only possible when data-shipping technique is used in the client machine and query-shipping technique in the server machine. We observed that that disk space and time spent during query processing and optimization is considerably smaller using data-shipping technique in the system than query-shipping technique.

In Figure 5 we indicate Y as the number of entries in a relation, X1 is data-shipping technique which performances optimally at the client environment and finally, X2 is the query-shipping technique that uses more disk space and time.

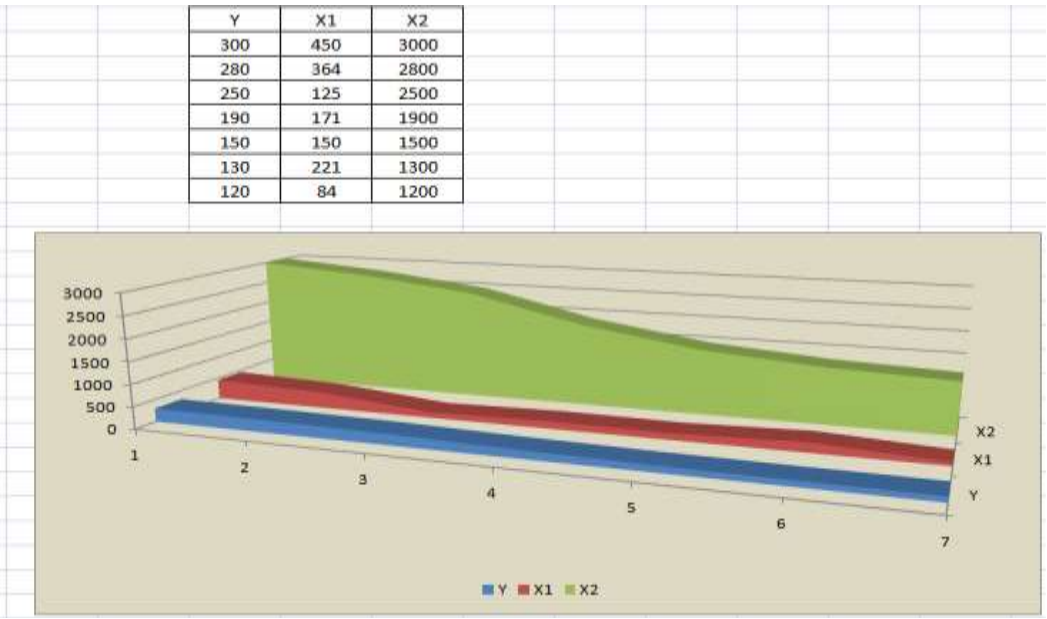

Figure 5: Queries optimized at client machine using data and query shipping technique

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

Table 2 Queries Optimized at the Sever Machine

Number of entries	Percentage due to query %	Query-shipping Technique		Data-shipping technique	
		Disk space (kb)	Time (s)	Disk space (kb)	Time (s)
300	15	450	2.4	3000	16
280	13	364	1.9	2800	14.9
250	5	125	0.66≈0.7	2500	13.3
190	9	171	0.91≈0.9	1900	10.1
150	10	150	0.8	150	8.0
130	17	221	1.17≈1.2	1300	6.9
120	7	84	0.4	1200	6.4

Table 2.0 above shows queries that are processed and optimized at the server machine, here query-shipping technique performed optimally when compared to data-shipping technique. Lesser disk space and time were spent on data.

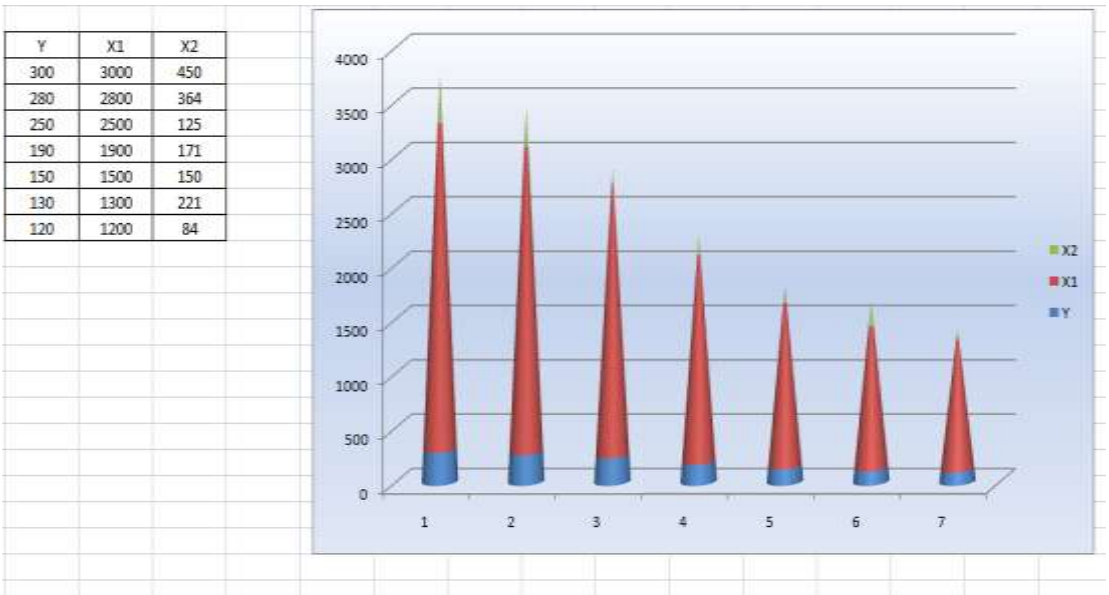

Figure 6.0 Number of data queries optimized at server machine

Figure 6 above indicates that query shipping technique uses fewer disk space than data-shipping technique at the server environment

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

Y	X1	X2
300	16	2.4
280	14.9	1.9
250	13.1	0.7
190	10.1	0.9
150	8	0.8
130	6.9	1.2
120	6.4	0.4

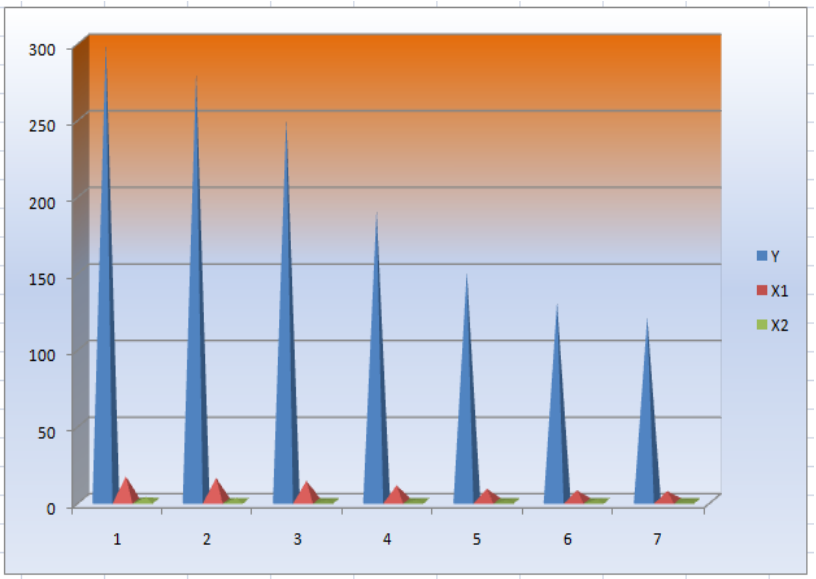

Figure 7.0Time taken for data entries for data and query shipping technique at server machine

Figure 7 depicts time spent on data entries when we made use of both techniques, here query-shipping technique uses fewer than data-shipping technique.

Table 3 Queries Optimized at the Client Machine using Hybridized Technique

Number of entries	**Percentage due to query %**	**Data-shipping technique**	
		Disk space (kb)	Time (s)
300	15	450	2.4
280	13	364	1.9
250	5	125	0.66≈0.7
190	9	171	0.91≈0.9
150	10	150	0.8
130	17	221	1.17≈1.2
120	7	84	0.4

Table 3 shows how the proposed technique (hybridized-shipping technique) uses data-shipping technique when query is submitted in the client machine, since it has data-shipping technique functionalities imbedded in it. X indicate the hybridized-shipping technique.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

Y	X
300	450
280	364
250	125
190	171
150	150
130	221
120	84

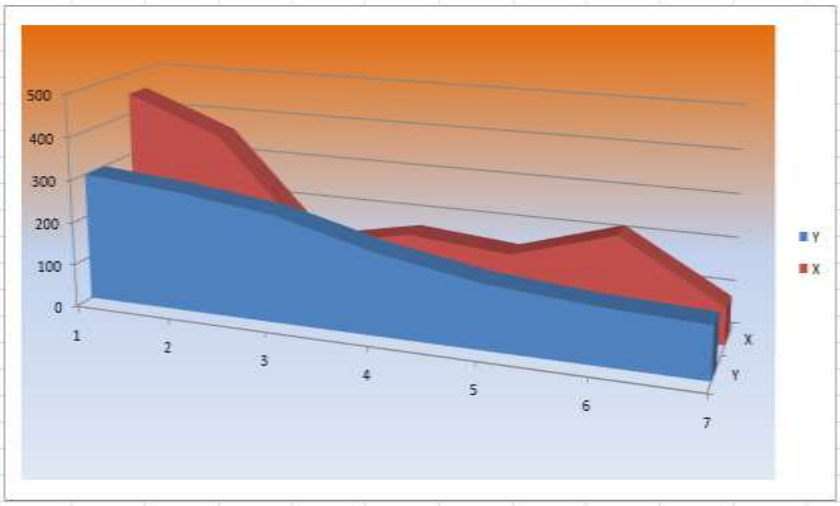

Figure 8 The volume of disk space used for hybridized technique at the client machine

Figure 8 depicts the volume of disk space spent when the proposed technique is used to optimized data entries in the client system.

Y	X
300	2.4
280	1.9
250	0.7
190	0.9
150	0.8
130	1.2
120	0.4

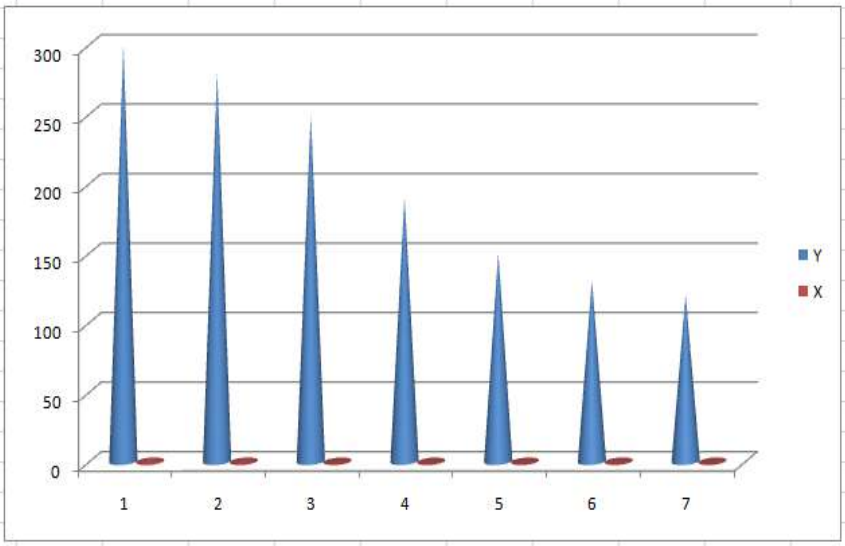

Figure 9 Time taken for data entries for hybridized technique at the client environment

Figure 9 shows the time spent on data entries after the application of the hybridized-shipping technique during the execution of query at the client environment.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

Table 4. Queries Optimized at the Server Machine using Hybridized Technique

Number of entries	Percentage due to query %	Hybridized-shipping technique	
		Disk space (kb)	Time (s)
300	15	450	2.4
280	13	364	1.9
250	5	125	0.66≈0.7
190	9	171	0.91≈0.9
150	10	150	0.8
130	17	221	1.17≈1.2
120	7	84	0.4

Table 4. shows the value of disk space and time spent when processing and optimization of query is done at the server machine using our proposed technique. The technique optimized query optimally since it has query-shipping technique functionality.

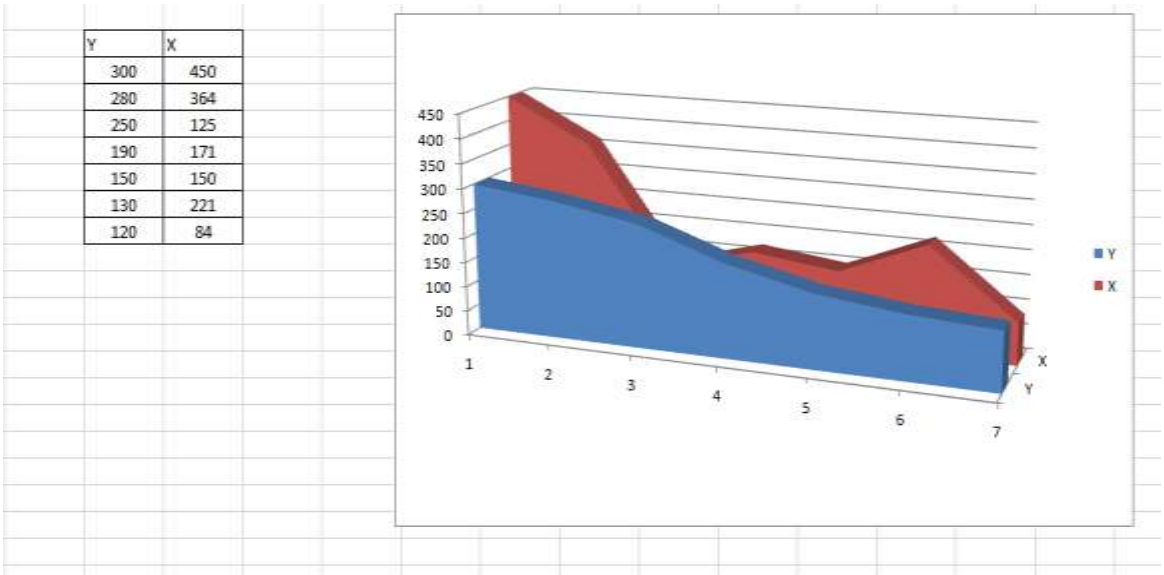

Figure 9 The volume of disk space used for hybridized technique at the server system

Figure 9 depicts the amount of disk space used for any given number of data entries, at using the hybridized-shipping technique at the server system.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4111-4121 www.ijaem.net ISSN: 2395-5252

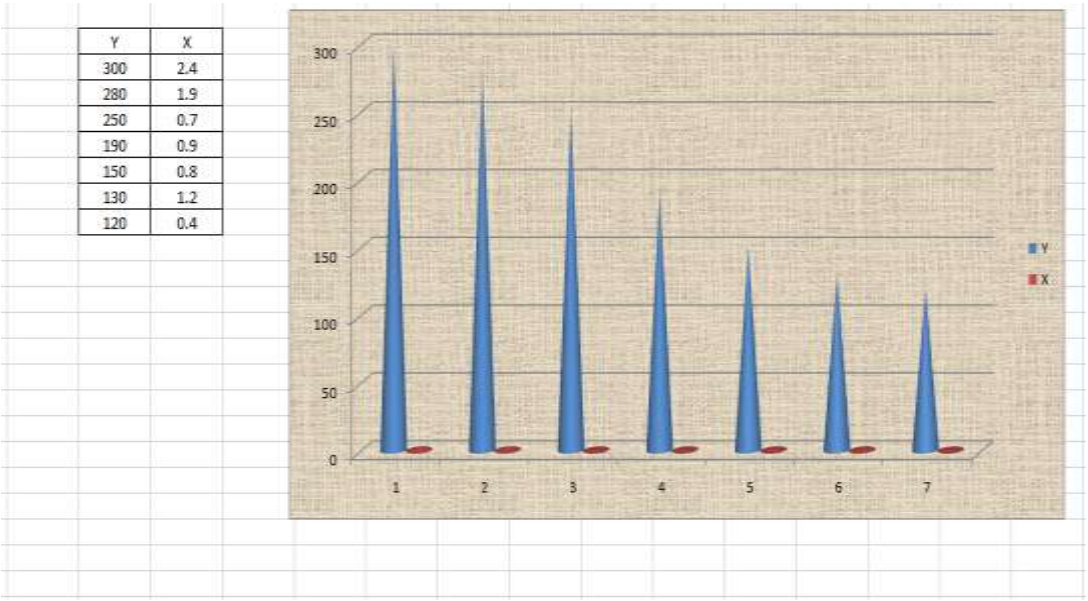

Figure 10 Time taken for data entries for hybridized technique at the server environment

Figure 10 illustrates the total time spent on any given data entry after execution, by our proposed technique.

V. CONCLUSION

In this research work, We designed an application that when a user either at the client machine or server machine logs in, the ID and username of the person is displayed to show us the particular user. Both query and data chipping techniques were applied to attain the requirement of this research paper. The major models for server-client query evaluation are query and data shipping techniques. Initially, we describe the course of actions relation to the limits they put on operator node selection throughout the optimization of a query in distributed environment

This paper "A Hybridized Technique for Optimizing Query Processing in a Distributed Database System" is a computerized solution for storing the details of employees in an organization and task assigned to an employee by an organization. Therefore, we can conclude that the application has been developed using the recommended techniques in order to reduce cost and harness both human and natural resources.

VI. FUTURE WORK

Any student who want to research in future on this topic, should use web and if there is any additional package or technique to be applied on the hybridized technique, he should apply it in order to have better results as the case may be. Improvements should be done on this topic so as to have the required result.To the user of the server machine, it is good to use query shipping technique to query the server because it is cost effective and fast in retrieving, updating and inserting data.

REFERENCES

[1]. Aggarwal M., Bawa P., Ganesan H., Garcia M., Kenthapadi K., Motwani R., Srivastava U., Thomas D., and Xu Y. (2005). Twocan keep a secret: Adistributed architecture for secure database services. In Conference on Innovative Data Systems Research.

[2]. Farnan. Nicholas L., Adam J., Lee., and Ting Y. (2010). "Investigating privacy-aware distributed query evaluation". Proceedings of the 9 th annual AC workshop on privacy in the electronic society-WPES 10.

[3]. Reza, S., Carlo, Z., Amir, Z., andJafar, A. (2001). Optimization of SequenceQueries in Database Systems. In Proceedings of the twentieth ACMSIGMOD -SIGACT-SIGART symposium on Principles of database systems.

[4]. Ivan T., Bowman and G. N. Paulley G. N (2000). Join enumeration in a memory constrained environment. In Proc. IEEE Conf. on Data Engineering.

[5]. Jim W., Geoff W., Ann W., and Sam K. (1994). A note on distributed computing.

[6]. Manuel R., and Nick R. (2000). Automatic deployment of application-specific metadata and code in MOCHA. In Proc. 7th EDBT Conference, Konstanz,Germany, March.

[7]. Manuel R., and Nick R. (2000). MOCHA: A self-extensible database middleware system for distributed data sources. In Proc. ACM SIGMOD Conference,Dallas, Texas, USA, May.