

An Enhanced Approach to Query Optimization in Homogeneous Distributed Databases

Obasi Emmanuela C.M¹, Nlerum Promise A²Research Scholar, Department of Computer Science, Federal University, Otuoke, Bayelsa State, Nigeria¹Research Scholar, Department of Computer Science, Federal University, Otuoke, Bayelsa State, Nigeria²

ABSTRACT: The complexity involved in optimizing fuzzy query processes needs to be addressed. This is because fuzzy query processing supports sorting using multiple criteria at once. This further causes the issue of latency and vagueness in the query results. In this study, an enhanced approach to query optimization in Homogenous Distributed Databases (HDD) was adopted as the proposed system. Furthermore, Structured System Analysis and Design Methodology was carried out in the study. The benefit of the study encompasses enhanced query optimization techniques such as query-documentation, query-storage and query-transformation. In addition, the study could be beneficial to Database analysts, software developers and any researcher with penchant for Query Optimization.

KEYWORDS: Fuzzy Query, Homogeneous Distributed Database, Query Optimization

I. INTRODUCTION

The need for reduction of complexities in fuzzy query processing is highly indispensable. In order to address this need, the issue of query optimization must be critically looked into. Optimization in query processing is simply finding the best possible plan for the database engine to execute a given query and further display the results in quick response time. Vagueness and ambiguity of data are among the numerous complex issues that are associated with fuzzy query processing. However, the usage of the fuzzy data in real-life emanates from the human thoughts, cognition processes and decision-making. The only way humans can value the development of information systems is when there is ease associated with information retrieval from a database. The study intends to improve fuzzy query optimization in distributed databases with special emphasis to homogeneous databases. The proposed improvement encompasses an enhanced query-storage, query-documentation and query transformation.

II. RELATED WORK

Emoghene[1] looked at information gathering methods and tools: a comparative study. The work illustrated the need to aid numerous system analysts in the use of appropriate techniques/tools for specific situations under certain conditions during Requirement Elicitation (RE). RE techniques, also known as information gathering methods/tools are methods used by analysts to determine the needs of customers and users. The result of the work showed that using the right information gathering tools such as data mining algorithms, will go a long way in improving database performance.

Rupali [2] studied distributed data mining algorithm and trends. The work spotlighted the concept of distributed data mining. It defined it as a field which deals with analyzing distributed data and proposes algorithmic solutions to perform different data analysis and mining operations in a distributed manner by considering the resource constraints. The result of the work showed in-depth application and importance of distributed data mining in spatial databases.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

Hoang et al [3] looked at Range Query Processing for Monitoring Applications over untrustworthy clouds. The study addressed the problem of privacy-preserving range query processing on clouds. Several solutions have been proposed in this line of work, however, they become inefficient or impractical for many monitoring applications, including real-time monitoring and predicting the spatial spread of seasonal epidemics. Furthermore, the work demonstrated a solution that out-performs Pined-RQ.

Ryan et al [4] looked at Neo: A learned Query Optimizer. They introduced Neo (Neural Optimizer), a novel learning-based query optimizer that relies on deep neural networks to generate query executions plans. Neo bootstraps its query optimization model from existing optimizers and continues to learn from incoming queries, building upon its successes and learning from its failures. Furthermore, Neo naturally adapts to underlying data patterns and is robust to estimation errors. Experimental results demonstrate that Neo, even when bootstrapped from a simple optimizer like PostgreSQL, can learn a model that offers similar performance to state-of-the-art commercial optimizers, and in some cases even surpass them.

Patrick et al [5] researched on towards a benefit-based optimizer for interactive data analysis. They introduced several ideas around the optimization of Interactive Data Analysis (IDA) tasks. With an eye on traditional query optimization (QO) in Relational Database Management Systems (RDBMS), they further suggested that IDA tasks should be specified through high-level statements and optimized globally, particularly by maximizing the number and significance of insights that can be automatically collected for the task. Furthermore, they also envisioned architecture for IDA and propose in the context of IDA what corresponds to statistics, cost model and execution plan pruning strategy in relational systems.

Timi and Asagba[6] looked at Query Optimization of a Distributed Database System using Fuzzy Logic. The authors design a distributed system that optimizes queries and processors, and to develop effective query execution plans for mining distributed database with the use of fuzzy logic. The system was implemented by developing a suitable framework for query processing in a distributed system that is time dependent, and a fuzzy logic approach was employed based on linguistic terminology for flexible user query. The result obtained shows fuzzy queries will always take less response time as compared to normal SQL queries.

III. DISTRIBUTED DATABASES

A distributed database can be defined as a system that stores two or more files in different sites either on the same network or on different networks. Distributed databases are usually interrelated especially in a logical form. According to Hiremath and Kishor [7], distributed databases are very important for industrial environment. They also defined a distributed database system as a system located at different machines on a network. In other words, numerous computer systems (or nodes) bear the load of the database. Also, the computers in a distributed database system are connected wirelessly or via a communication device. Figure 1.0 shows a simple illustration of a distributed database system.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

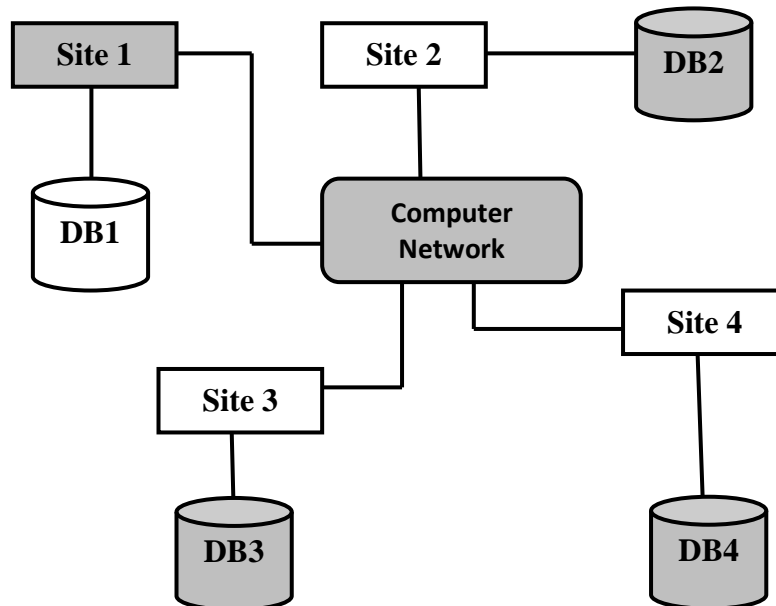


Figure 1.0: Sample Structure of a Distributed Database System
(Source: [7])

IV. RESULTS AND DISCUSSION

4.1 Methodology

The Methodology for the Proposed System Design is Structured System Analysis and Design Methodology (SSADM). Structured Systems Analysis and Design Methodology is a systems approach to the analysis and design of information systems. SSADM was produced for the UK government office concerned with the use of technology in government, from 1980 onwards. System design methods are a discipline within the software development industry which seeks to provide a framework for activity and the capture, storage, transformation and dissemination of information so as to enable the economic development of computer systems that are fit for purpose.

The Methodology application to the proposed system involved the following stages:

- **Stage One:**
Analyzing Existing ways deployed to tackle the issue:
There was in-depth analysis on current performance of the Existing System in Recommending reliable information to users.
- **Stage Two:**
Existing System Performance and Identification of Anomalies:
This phase involved studying result performance of the Existing System in order to spot any anomaly that should be corrected.
- **Stage Three: Review of any spotted anomaly from the Existing System:** This phase involved the deployment of the proposed system algorithm to any spotted anomaly from the Existing System Performance.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

- **Stage Four:**
Comparison of the Existing and Proposed Systems Performance Result: This phase involves comparing results of the Proposed and Existing Systems Performance.
- **Stage Five:**
This phase involves the adoption of the most efficient result that tackles the mentioned issue due to the outcome of the comparison between the Existing and Proposed Systems.

4.2 Existing System Architecture of a Homogeneous Distributed Database

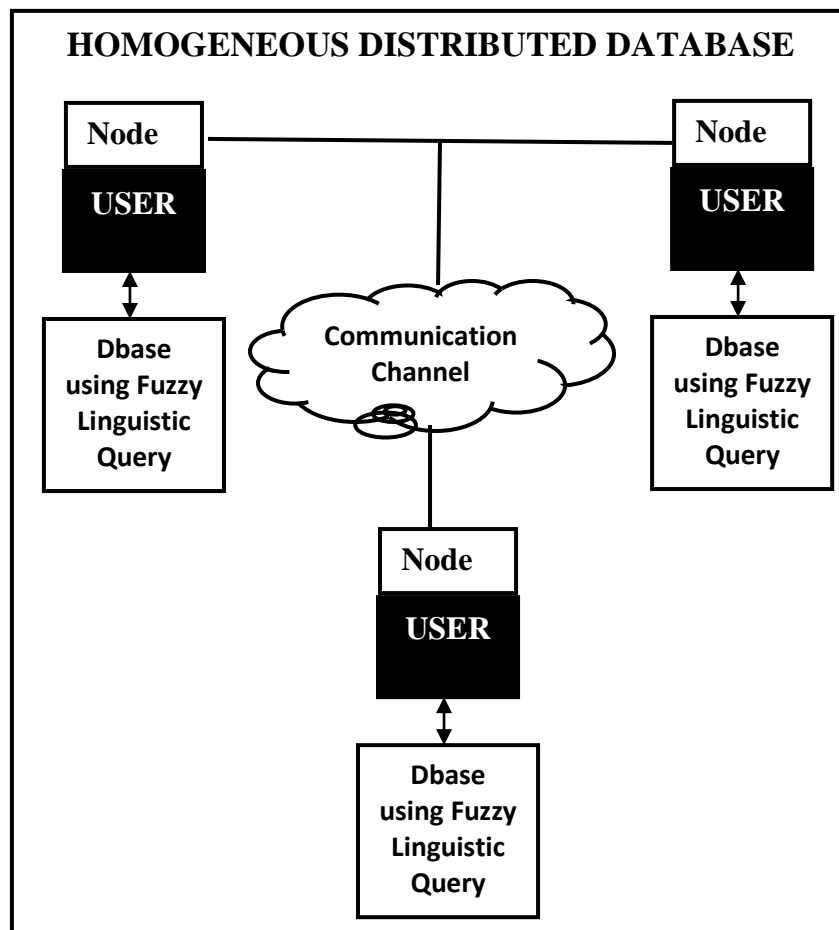


Figure 2.0: Existing Homogeneous Distributed Database Model

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

4.3 Proposed System Architecture for an Improved Homogeneous Distributed Database

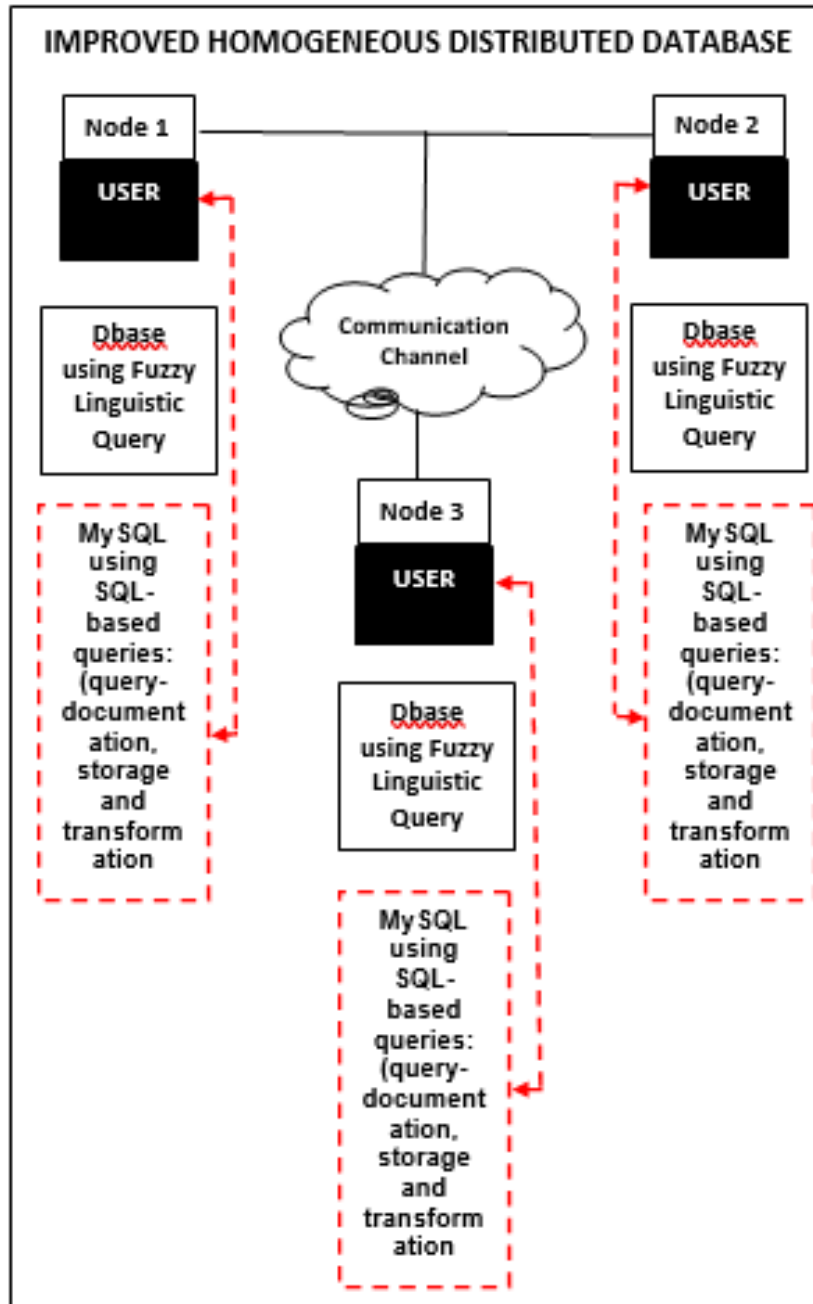


Figure 3.0: Proposed Homogeneous Distributed Database Model

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

4.4 Advantages of the Proposed Model

The following advantages of the Proposed Model are:

- Latency Reduction
- Best query result for structured and unstructured datasets
- No need for manual labeling of datasets before query processing
- The proposed system enhances query optimization through query-documentation, query-storage, and query-transformation.

V. CONCLUSION

In this study, we have developed an improved homogeneous distributed database model which further enhances query optimization and processing. Furthermore, the proposed system consists of SQL-based graphical user-interface and query techniques such as query-documentation, query-storage and query-transformation. In addition, the proposed model could be a useful tool to software developers, database analysts/administrators, and any researcher with penchant for query optimization in distributed databases.

REFERENCES

- [1] Emoghene O., Information gathering methods and tools: A Comparative Study, Research gate Journals, <https://www.researchgate.net/publication/326688869>, 7(9), 1-6, 2018.
- [2] Rupali C., Study of Distributed Data Mining Algorithm and Trends, IOSR Journal of Computer Engineering 4(4), 41 – 47, 2016.
- [3] T.Hoang, N. Zeng, U. Kong, and B. Jin, Range Query Processing for Monitoring Applications Over Untrustworthy Clouds, Over untrustworthy Clouds, Open Proceedings, 10.5441/002/edbt.2019.82, 444-447, 2019.
- [4] S. Ryan, B. Neslpo, and D. Pisco, An Improved Neo Query Optimizer, International Journal of Computer Applications (IJCA), 7(8), 114 – 121, 2017.
- [5] D. Patrick, O. Ninioma, and S. Sunju, Query Processing Based Compressed Intermediates, Journal of Software Development Research (JSDR), 10(7), 28-34, 2019.
- [6] Timi O.T. and Asagba P.O., Query Optimization of a distributed database system using fuzzy logic, International Journal of Engineering Science Invention (IJESI), 7(1), 67 – 74, 2018.
- [7] Hiremath D.S and S.B Kishor, Distributed Database Problem Areas and Approaches, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278 – 0661, P-ISSN: 2278-8727, 15 – 18, 2016.