

---

*CSCI 5408*

*Data Management, Warehousing, And  
Analytics*

---

---

*Assignment 2 - Problem 1*

*Performing a systematic literature review and providing summary.*

---

**Prepared By**

Bhavisha Oza (B00935827)

## **Problem-1: Perform a systematic literature review and provide summary.**

### **“A comparative Analysis of Data Fragmentation in Distributed Database” [1].**

#### **1. Summary:**

*A comparative Analysis of Data Fragmentation in Distributed Database* [1] paper proposes a comparative analysis of data fragmentation in distributed database systems. It begins by highlighting the significance of distributed systems in modern computing technology and explains that distributed databases consist of data partitions or fragments that are distributed or replicated across multiple physical locations. The paper emphasizes the importance of the distributed database design process, which includes initial design, redesign, and materialization phases.

The introduction of the paper emphasizes the significance of distributed systems in modern computing technology, particularly in handling large amounts of data. Distributed databases are logical databases that are physically distributed across multiple locations connected by a network. The distribution of data involves fragmentation, replication, and allocation processes. The authors highlight the advancements in communication technology, software, and hardware, making distributed database systems more feasible and efficient.

The initial design, redesign, and materialisation of the redesign phases of the distributed database design process are covered. Algorithms for allocation and fragmentation are used in the early design to reduce the expense of transaction processing. New fragmentation and allocation methods are created throughout the redesign process based on modifications to the distributed database environment. The new fragmentation and allocation mechanism is put into place during the materialisation phase.

Related works in the field of distributed database systems are presented, including studies on fragmentation schemes and allocation schemes. Various researchers have proposed different approaches for vertical and horizontal fragmentation in distributed databases, aiming to improve query performance and minimize communication costs. They emphasise the benefits and drawbacks of fragmentation and discuss several strategies put forth by earlier researchers.

The paper then introduces three types of fragmentations: horizontal fragmentation (HF), vertical fragmentation (VF), and mixed fragmentation (MF). A relation or class is divided into disjoint tuples or instances using horizontal fragmentation, with each fragment being stored at a separate node. Vertical fragmentation divides a relation into sets of columns or characteristics, each set containing one or more of the table's main key properties. Combining horizontal and vertical fragmentations, known as mixed fragmentation, enables the creation of more complex fragmentation schemes.

Data fragmentation is explained as the process of breaking a single object into two or more fragments or segments. The three types of fragmentations are illustrated using the example of a Human Resources table, showing how the table can be horizontally fragmented into multiple fragments, vertically fragmented into separate sets of columns, or mixed fragmented with both horizontal and vertical partitions.

Finally, the research discusses fragmentation accuracy rules such as completeness, reconstruction, and disjointness. Each data item in the original relation must be discovered in at least one fragment to be complete. The existence of a relational operator capable of reconstructing the original relation from its fragments are required for reconstruction. Disjointness assures that the fragments do not overlap or have redundancy.

In conclusion, the paper provides an overview of data fragmentation in distributed database systems. It compares horizontal, vertical, and mixed fragmentations and discusses the advantages and disadvantages of each. In the end, the authors suggest that the choice of data fragmentation for distributed database depends on the specific requirements and objectives of the system. Understanding the different types of fragmentation and applying the correctness rules can help in designing efficient distributed databases.

## 2. Scope of Improvements:

There are few areas in the paper where some improvements can be made. As the paper focuses on the fragmentation types and its advantages, disadvantages, it should shed some lights on the methodologies as well. The paper does not mention the methodology used for the comparative analysis. It is important to describe the criteria and metrics used to evaluate the different types of fragmentation and how the analysis was conducted. Providing a clear methodology will enhance the credibility of the study and allow readers to understand the basis of the comparisons made.

It would be beneficial to provide insights into future research directions or potential areas for improvement in data fragmentation techniques in distributed databases. The paper does not address potential challenges or limitations of data fragmentation in distributed databases just like “*When applied to more complex applications such as CAD/CAM, software design, office information systems, and expert systems, the relational data model exhibits limitations in terms of complex object support, type system, and rule management*” [2]. It would be valuable to discuss issues and the issues which still requires a more work such as Sensitivity, Complexity, Transaction, Replication, Interface, Design, Interconnection [2].

## References:

- [1] A. Al-Sanhani, A. Hamdan, A. Al-Dahoud and A. Al-Dahoud "A comparative Analysis of Data Fragmentation in Distributed Database," *2017 8th International Conference on Information Technology (ICIT)*, Amman (11733), Jordan, 2017, pp. 724-729. DOI: 10.1109/ICITECH.2017.8079934.
- [2] M. T. Ozsü and P. Valduriez, "Distributed database systems: where are we now?," *in Computer*, vol. 24, no. 8, pp. 72-78, Aug. 1991. DOI: 10.1109/2.84879.