Quantitative vs Qualitative information Regarding fragmentation

Bachelor of Technology
Computer Science and Engineering

Submitted By

NAME – ABHIRUP BAG
ROLL NUMBER – 13000122082
DEPARTMENT – CSE(B)
SEMESTER – 6

PAPER – DISTRBUTED DATABASE MANAGEMENT SYSTEM (PEC-IT 601B)

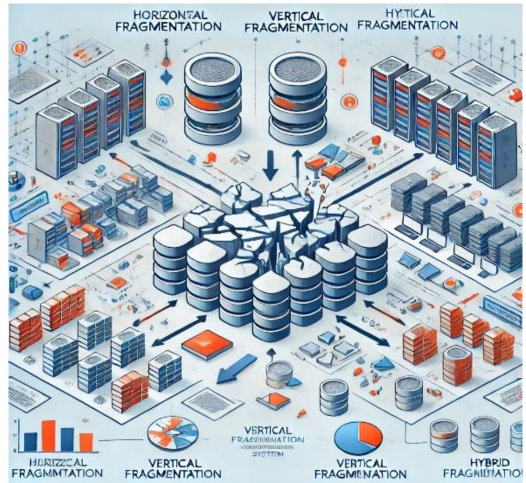
March 2025

Techno Main
EM-4/1, Sector-V, Salt Lake
Kolkata- 700091
West Bengal

Table of Contents

1. Introduction

Distributed Database Management Systems (DDBMS) are designed to manage data across multiple locations, ensuring efficient data access, storage, and retrieval. Fragmentation is a critical aspect of DDBMS, where data is divided into smaller subsets (fragments) to optimize performance, availability, and scalability. The process of fragmentation involves both quantitative and qualitative information to ensure that the fragmentation strategy is effective and efficient. This report explores the role of quantitative and qualitative information in fragmentation, focusing on minterm selectivity and access frequency as quantitative measures, and completeness and minimality as qualitative parameters.

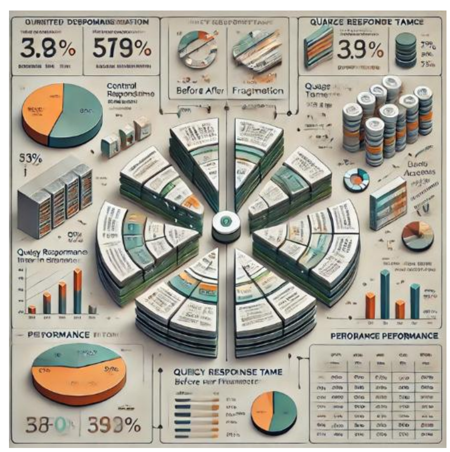


2. Quantitative Information in Fragmentation

Quantitative information in fragmentation refers to measurable, numerical data that helps in making decisions about how to divide the data. Two key quantitative measures in fragmentation are minterm selectivity and access frequency.

Quantitative analysis of fragmentation in a **Distributed DBMS** helps evaluate efficiency and performance. Key metrics include:

- **Fragmentation Granularity (FG):** Measures fragment size in terms of rows or columns.
- **Data Transmission Cost (DTC):** Cost of transferring fragments across network nodes.
- **Access Frequency (AF):** Number of times a fragment is accessed, influencing placement decisions.
- **Response Time (RT):** Time taken to execute queries involving fragmented data.
- **Redundancy Ratio (RR):** Measures data duplication in replicated fragments.
- **Fragmentation Completeness (FC):** Ensures all necessary data is included in fragments.
- **Fragmentation Disjointness (FD):** Ensures minimal data duplication (except for replication).



Efficient fragmentation minimizes **data transfer costs, improves query performance, and enhances system scalability**.

A. Minterm Selectivity

i) **Definition**: Minterm selectivity refers to the proportion of tuples in a relation that satisfy a given minterm predicate. A minterm predicate is a conjunction of simple predicates used to define fragments.

ii) **Role in Fragmentation:**

  (a) Minterm selectivity helps determine the size of each fragment. High selectivity means fewer tuples satisfy the predicate, resulting in smaller fragments, while low selectivity means larger fragments.

  (b) It is used to balance the load across distributed nodes by ensuring that no single fragment becomes too large or too small.

iii) **Measurement:** Represented as a fraction (0 to 1) or a percentage (0% to 100%).

iv) **Usage:** Helps in optimizing horizontal fragmentation by ensuring balanced partitioning based on frequently used predicates.

v) **Example:** If a minterm predicate is (Department = "Sales"), and 20% of the tuples satisfy this condition, the selectivity is 0.2. This information is used to estimate the size of the "Sales" fragment.

B. Access Frequency

i) **Definition:** Access frequency refers to how often a particular fragment is accessed by queries or applications.

ii) **Role in Fragmentation:**

  (a) Fragments with high access frequency are often placed on nodes with higher performance or closer to the users to reduce latency.

  (b) It helps in optimizing query performance by ensuring that frequently accessed data is readily available.

iii) **Measurement**: Expressed as requests per second (RPS), queries per minute (QPM), or other time-based metrics.

iv) **Usage:** Determines optimal fragment placement to minimize network latency and data transfer costs.

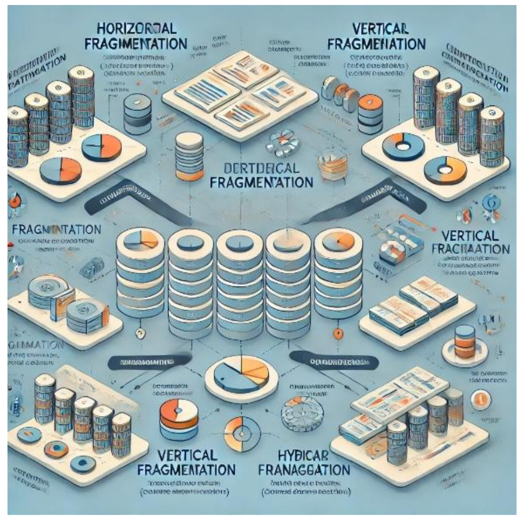
v) **Example:** If the "Sales" fragment is accessed 70% of the time, it may be placed on a high-performance server or replicated across multiple nodes to ensure availability and reduce response time.

3. Qualitative Information in Fragmentation

Qualitative information in fragmentation refers to non-numerical, descriptive attributes that ensure the fragmentation strategy adheres to certain logical and structural principles. Two key qualitative parameters are completeness and minimality.

Qualitative aspects of fragmentation in a Distributed DBMS focus on the logical and structural effectiveness of data distribution. Key considerations include:

- **Data Locality:** Ensures frequently accessed data is placed close to users to improve performance.
- **Scalability:** Supports system growth by enabling efficient distribution of data across multiple sites.
- **Reliability & Availability:** Proper fragmentation enhances fault tolerance and ensures data availability.
- **Consistency:** Maintains data integrity by minimizing redundancy and synchronization issues.
- **Query Optimization:** Well-designed fragmentation reduces query processing time by eliminating irrelevant data retrieval.



A well-planned fragmentation strategy enhances system efficiency, reliability, and user experience in distributed databases.

A. Completeness Parameter

i) **Definition:** Completeness ensures that all data in the original relation is preserved in the fragments. No data should be lost during fragmentation.

ii) **Role in Fragmentation:**

(a) Ensures that the union of all fragments reconstructs the original relation without any loss of information.

(b) Critical for maintaining data integrity and consistency in a distributed environment.

iii) **Qualitative value:** A design principle that ensures no data loss during fragmentation.

iv) **Usage:**

(a) Guarantees that fragmentation does not exclude any tuples.

(b) Affects query correctness and data integrity.

v) **Example:** If a relation Employee is fragmented into Employee_Sales and Employee_HR, the union of these two fragments must contain all tuples from the original Employee relation.

B. Minimality Parameter

i) **Definition:** Minimality ensures that the fragmentation process does not create unnecessary fragments. Each fragment should be distinct and non-redundant.

ii) **Role in Fragmentation:**

(a) Prevents over-fragmentation, which can lead to increased complexity in query processing and management.

(b) Ensures that each fragment serves a unique purpose and does not overlap with others.

iii) **Qualitative value:** A design principle that avoids over-fragmentation or redundancy.

iv) **Usage:**

(a) Prevents unnecessary data duplication across sites.

(b) Helps in storage efficiency and reducing synchronization costs.

v) **Example:** If a relation Employee is fragmented based on department and location,

4. Comparison of Quantitative and Qualitative Information

Aspect	Quantitative Information	Qualitative Information
Nature	Numerical, measurable	Descriptive, logical
Examples	Minterm selectivity, access frequency	Completeness, minimality
Purpose	Optimize performance, balance load	Ensure data integrity, avoid redundancy
Measurement	Can be calculated using formulas or statistics	Evaluated based on logical rules and principles
Impact on Fragmentation	Determines size, placement, and replication of fragments	Ensures correctness and efficiency of fragmentation strategy

5. Comparison with Parameters

Parameter	Quantitative Information	Qualitative Information
Minterm Selectivity	Proportion of tuples satisfying a predicate (e.g., 30%).	N/A (Quantitative only).
Access Frequency	Number of accesses per fragment (e.g., 1,000/day).	N/A (Quantitative only).
Completeness	N/A (Qualitative only).	Ensures no data loss during fragmentation.
Minimality	N/A (Qualitative only).	Avoids unnecessary complexity or redundancy.

6. Interplay Between Quantitative and Qualitative Information

A. Balancing Act: While quantitative information focuses on performance optimization, qualitative information ensures that the fragmentation strategy is logically sound and adheres to database principles.

B. Example Scenario:

i) A high-access-frequency fragment (quantitative) may need to be replicated across multiple nodes for performance, but the minimality parameter (qualitative) ensures that the replication does not lead to unnecessary redundancy.

ii) Minterm selectivity (quantitative) determines the size of fragments, while completeness (qualitative) ensures that no data is lost during fragmentation.

7. Challenges and Considerations

A. Trade-offs: Balancing quantitative and qualitative parameters can be challenging. For example, optimizing for access frequency may conflict with minimality if it leads to excessive replication.

B. Dynamic Environments: In distributed systems, access patterns and data distribution may change over time, requiring periodic reevaluation of fragmentation strategies.

C. Complexity: Managing both quantitative and qualitative information increases the complexity of the fragmentation process, requiring sophisticated algorithms and tools.

8. Practical Application

Quantitative Analysis	Qualitative Analysis
Use minterm selectivity to identify frequently accessed data subsets and optimize query performance.	Use completeness to ensure data integrity and compliance with business or regulatory requirements.
Use access frequency to allocate resources (e.g., placing high-traffic fragments on high-performance nodes).	Use minimality to design a fragmentation strategy that is efficient and scalable.

9. Example scenario

Consider a distributed DBMS for a organisation with database schema:

- **PAY (TITLE, SAL)**
- **EMP (ENO, ENAME, TITLE)**

A. <u>PAY (TITLE, SAL) TABLE:</u>

TITLE	SAL	
MANAGER	65K	PAY-1
ELECTRICAL ENGINEER	52K	
PREGRAMMER	45K	PAY-2
ACCOUNTANT	42K	

B. <u>Simple Predicate</u>

P_1	TITLE = MANAGER
P_2	TITLE = ELECTRICAL ENGINEER
P_3	TITLE = PREGRAMMER
P_4	TITLE = ACCOUNTANT
P_5	SAL $\leq$ 50K (PAY - 2)
P_6	SAL $>$ 50K (PAY - 1)

C. Quantitative Information

i) Some Minterm Predicates And Minterm Selectivity [Sel(M_i)]

M_i	Minterm Predicate	Sel(M_i)
M_1	$P_1 \wedge P_2$	0
M_2	$P_1 \wedge P_6$	1
M_3	$P_3 \wedge P_6$	1
M_4	$P_4 \wedge P_6$	0

ii) Access frequency:

The Sal > 50K is accessed 50 times per day.

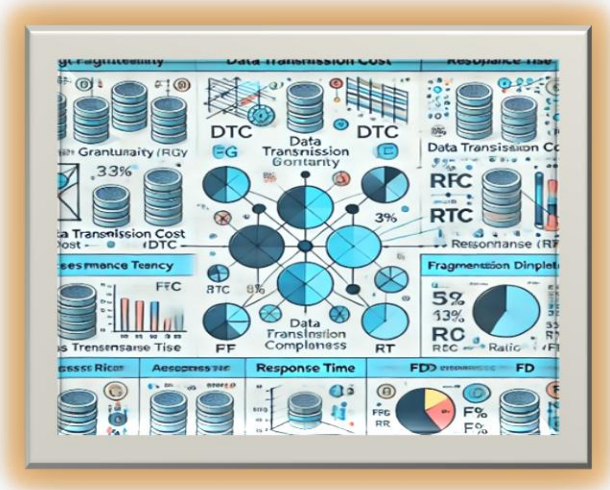
D. Access frequency:

i) **Completeness:** All order records are preserved across fragments, ensuring no data loss.

ii) **Minimality:** The fragmentation strategy avoids creating too many small fragments, reducing management overhead.

10. Conclusion

Quantitative and qualitative information play complementary roles in the fragmentation process in distributed DBMS. Quantitative measures like minterm selectivity and access frequency provide the numerical basis for optimizing performance and resource utilization. On the other hand, qualitative parameters like completeness and minimality ensure that the fragmentation strategy adheres to logical and structural principles, maintaining data integrity and efficiency. A successful fragmentation strategy must strike a balance between these two types of information to achieve optimal performance, scalability, and reliability in a distributed database environment.



11. References

- **"Distributed Databases: Principles & Systems"** – Stefano Ceri, Giuseppe Pelagatti
- **"Principles of Distributed Database Systems"** – M. Tamer Özsu, Patrick Valduriez
- **"Fundamentals of Database Systems"** – Ramez Elmasri, Shamkant B. Navathe
- **"Distributed Database Management Systems: A Practical Approach"** – Saeed K. Rahimi, Frank S. Haug
- **"Database System Concepts"** – Abraham Silberschatz, Henry F. Korth, S. Sudarshan