IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 4, August 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

A Modified Vertical Fragmentation Strategy for Distributed Database

Gamal H. Eladl

Information Systems Department, Mansoura University, Faculty of Computers and Information Sciences,
El-Mansoura, 35516 D.K, Egypt

Abstract - In the last decade; distributed database (DDB) has been developed for solving the problem of huge amount of uploaded data over different network sites. Fragmentation is one of the most important strategies/solutions in order to distribute the organization's database according to its three type (vertical, horizontal, mixed/hybrid). Furthermore, it reduces the amount of irrelevant data, disk access, and a single source bottleneck. In this paper, a modified vertical fragmentation strategy that is divided any single relation's attributes into two or more partitions but by using the advantages of fuzzy logic concept. The modified strategy allows the distributed database designer to specify the weight (membership function) of all attributes' importance before partitioning it in order to get accurate projection. The application of this strategy will be led to accurate fragmentation strategy.

Keywords - *DDB, Vertical Fragmentation, Projection, Fuzzy logic, Weight, Membership Function*

1. Introduction

DDB is a collection of fragment/distributed data unit that logically belongs to the same system and spreads over a computer network sites. Nowadays, developing DDB systems become more feasible and urgently needed. Thus, there are major research problems associated with DDB. One of them is designing an efficient and an accurate DDB fragmentation strategy. Therefore, the primary issue of distributed database is to design the fragmentation and allocation strategy. Distributed database management system (DDBMS) is the software system that allows the management of DDB and makes the process of distribution more transparency. There are many methods to distribute the organization data such as replication, allocation and fragmentation strategies. Thus, the general goals of DDB are to grantee reliability and functionality with high performance.

Fragmentation of relations is the most common methods to distributed database. So that, most distributed database books named the fragment of a relation as the basic unit of distribution. On the same context, fragmentation is a very important task that needs critical selections and DDB designer's participation. These selections are based on analyzing quantitative and qualitative information that focus on historical information of application's query which concentrates on the types of access (read/write) and the frequency of queries, site and where the query is run.

Fragmentation strategy must be chosen carefully according to the database model. It is clearly appearing here that the model is relational database model. There are many reasons of using fuzzy logic in this paper because; fuzzy set treats with an ambiguity/vagueness, and sensitive data. The core of fuzzy set operations that work between the zero and one (gray level). It can measure the degree and membership of unit that belongs to or not by using the membership functions (weight) $\mu(0\sim1)$. All the time, fuzzy set leads to good results and accurate data rather than crisp/logic data.

The content of this paper is organized as follows: Section 2 reviews the related work. Section 3 contains a brief about the concept of fuzzy logic. Section 4 presents the different fragmentation strategies and its issues. In section 5 the modified strategy. Finally, the conclusions are drawn in section 5.

2. Related Work

In the last few years, several research studies were introduced to discuss this subject of discourse. For example, S. K. Verma [1] introduced a recent research about fragmentation techniques for DDB. His study focused on showing the difference between centralized DB architecture and distributed DB architecture with 3 examples (vertical, horizontal, and Mixed) of fragmentation techniques. The conclusion of his study summarizes that you should manage an appropriate

methodology for data fragmentation in order to utilize the available resources with an efficient and accurate fragmentation strategy. Furthermore, a review of fragmentation techniques in DDB is also presented in [2]. R. Basseda [3] presented a paper that compared between three fragmentation allocations techniques (NNA, FNA, and BGBR) in distributed database. Furthermore, it focused on an algorithm evaluation such as response time which defined as the average delay for receiving the response for a fragment request, average time spent for moving data from one node to another and fragment data migration time [3, 4]. There are several distributed design alternatives such as top-down and down-top approach. Top-down approach means defining the global framework, after then focusing on the details. But, down-top approach is first defining the details, after then focusing on the global framework.

In general, there are six steps for distributing the database that we focus on the third item in details:

- Analysis the application requirements
- Design the global schema
- Design the Fragmentation
- Design the distribution schema
- Design the local schema
- Design the local physical layers

3. Importance of Fuzzy Logic

Fuzzy logic is a precise problem-solving methodology that is able to handle multivalued numerical data and linguistic knowledge. In fuzzy logic, a statement can assume any real value between zero and one, representing the degree to which an element belongs to a given set. It makes the development of any system much simpler and easier. Fuzzy logic mechanisms can result in higher accuracy and smoother control. Finally, fuzzy logic deals with degrees of truth /crisp set (zero, ones) and degrees and membership. Membership functions characterize the fuzziness in a fuzzy set whether the elements in the set are continuous or discrete. It can represent in a graphical form for eventual use in the mathematical formalisms of fuzzy set theory such as Trapezoidal membership functions [5-6]. Table 1 shows and summarizes the six steps that are essentials to convert any crisp system into a fuzzy-based system.

Table (1). Six steps for developing Fuzzy-based system

1-	**Determining a set of fuzzy rules,**
2-	**Fuzzifying the input using the input membership functions,**
3-	**Combining the fuzzified inputs according to fuzzy rules to establish a rule strength,**
4-	**Finding the consequence of the rule by combining the rule strength and the output membership function,**
5-	**Combining the consequences to get an output distribution, and**
6-	**Defuzzifying the output distribution (if a crisp output is needed).**

4. Fragmentation Types And its Issues

Indeed, a fragment of a relation is a unit of distribution. There are three major categories of fragmentation that are depicted in Fig. 1.

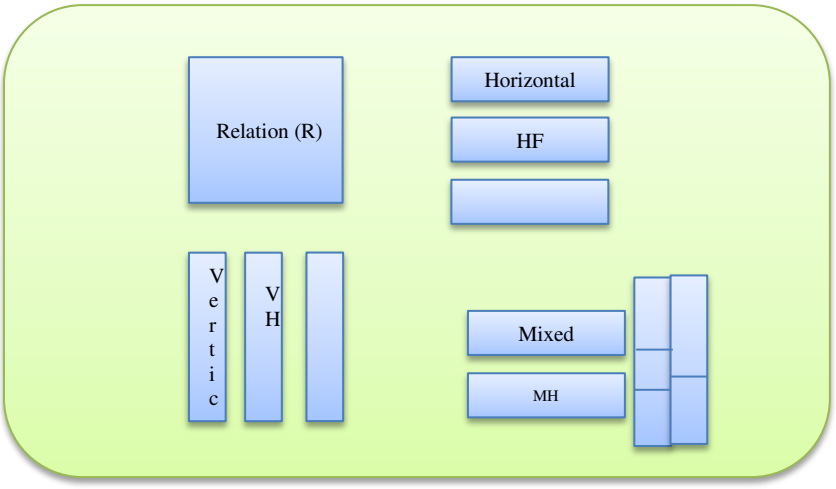


Fig.1 Three Types of Fragmentation (HF, VH, MH)

Horizontal fragmentation (HF) means the partition of a relation into a set of records/tuple. HF is defined as selection operation of the relational algebra as explained in the following equation (1).

$$HF=\boldsymbol{\sigma}\mathbf{p}(R) \quad \text{where P is tuple, R is the database relation} \quad (1)$$

Vertical fragmentation (VF) is defined as the partition of a relation into a set of smaller relations. In other words, it means the projection operation of the relational algebra as explained in the following equation (2).

$$VF=\pi\, A1, A2\ (R) \quad \text{where A is attribute, R is the database Relation} \quad (2)$$

Mixed/Hybrid Fragmentation (MF) is defined as a combination of horizontal and vertical fragmentation. MF is defined as a combination of selection and projection operations as explained in the following equation (3).

$$MF=\sigma p(\pi\, A1, A2\ (R)) \quad (3)$$

Hence, there are distribution issues for the process of fragmentation. We can represent these issues in raised questions in order to find its answer. Table 2 contains a summary of the most issues and its proposed solution.

Table 2: Fragmentation issues and its solutions

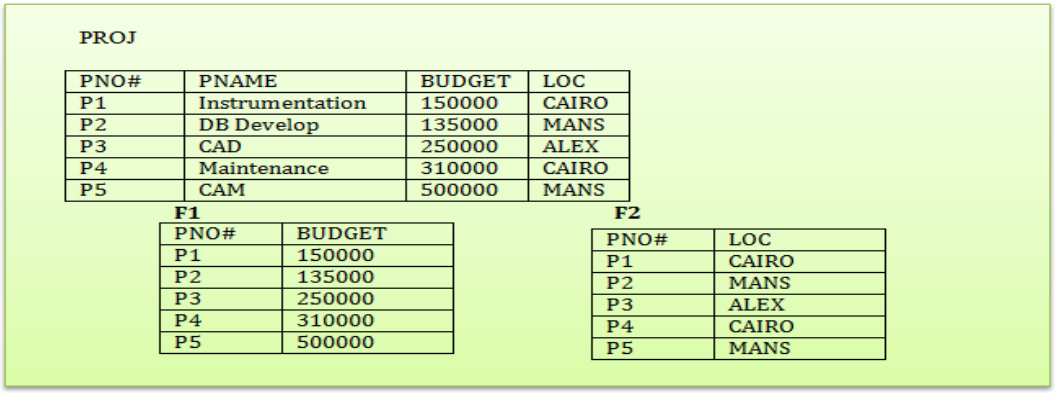
Item #	***Issues/Questions***	***Answer***
1	How to do fragmentation?	A1- Three types (HF,VF,MF)
2	How much to fragment?	A2- Data item is distributed to the sites where needed
3	How to test the correctness?	A3- Ensure (completeness, Reconstruction)
4	How to optimize the allocation?	A4- Quantitative and Qualitative Information (Query)
5	How to ensure the accuracy of attributes selection?	***A5- A modified Fragmentation strategy (Fuzzy Selection)***

5. A Modified Vertical Fragmentation Strategy

The selection of the data elements in a table is based on the attribute identifier of the data. In this case, there are different projections that compose the content of the fragment. In this modified strategy, we add the features of fuzzy set's membership function that can help the DDB's designer to specify the importance ratio for each attribute according to analyzing the application and its queries. Because membership function gives imprecisely measurement results. So that, each attribute is checked against every application usage by giving its membership function. For example, if an attribute A1 belongs to Fragment F1 with membership function $\mu 1(0.7)$ and also belongs to Fragment F2 with membership function $\mu 2(0.3)$. Operations of fuzzy logic will be better than crisp logic such as MAX/union, MIN/intersection, and complement. Table 3 shows fuzzy relations matrices e.g. attribute-fragment relation R(Ai, Fj, μ) Where i= number of attributes, j= number of fragment. Figure 2 shows an example of fuzzy vertical fragmentation strategy after applying the result of fuzzy relation matrices. Table 4 shows a comparison between fuzzy set and crisp logic set of fragmentation.

Table 3: Fuzzy relations matrices (Attribute-Fragment relation)

$R(A_i, F_j, \mu)$	F1(info. about project budget)	F2(info. about project location)
PNO	1 (primary key)	1 (primary key)
PNAME	0.5	0.5
Budget	0.9	0.1
LOC	0.3	0.7
MAX /Union Operation	***PNO, Budget***	***PNO,LOC***
MIN operation	PNAME,LOC	PNAME, Budget



PROJ

PNO#	PNAME	BUDGET	LOC
P1	Instrumentation	150000	CAIRO
P2	DB Develop	135000	MANS
P3	CAD	250000	ALEX
P4	Maintenance	310000	CAIRO
P5	CAM	500000	MANS

F1

PNO#	BUDGET
P1	150000
P2	135000
P3	250000
P4	310000
P5	500000

F2

PNO#	LOC
P1	CAIRO
P2	MANS
P3	ALEX
P4	CAIRO
P5	MANS

Fig 2 an Example of Fuzzy Vertical Fragmentation Strategy

Table 4: A Comparison between Crisp Fragmentation and Fuzzy Fragmentation

Elements	Crisp	Fuzzy
Membership Degree	Belongs 100% (1) or Zero(0)	Belongs with membership function $\mu(0\sim1)$
Operation s Support	Binary operation	Fuzzy operation MAX/ MIN/ Complement
Attribute Selection	Semi-accurate Selection	Accurate Selection

6. Conclusion

Distributed database is an important research area because it needs more effort to utilize the available resources with the correct placement of data/fragment and the program/application across a computer network sites. The fragmentation represents the unit of distribution. The success of any fragmentation strategy depends on the accurate selection/choice of appropriate partitions on the suitable site according to the organization's need. We explore the vertical fragmentation technique using the concept of fuzzy logic. As mention above, all the time fuzzy is better than crisp logic. It deals with numerous ranges of data (numeric or linguistics). A guideline steps are introduced to convert any system into a fuzzy-based system. DDB's designer is responsible for determining and calculating the importance of any attribute (weight) by using the application's query log to assign the accurate value of its membership function. This modified vertical fragmentation strategy can be applied to any optimized technique to find the best fragmentation allocation .

References

[1] S.K. Verma," Fragmentation techniques for distribution database: A Review", IJICSE, Vol. 3, Issue 2; April 2016.

[2] A. Patel , et. al " A review on Fragmentation Techniques in Distributed Database", IJMTER, Vol. 01, Issue 03, September 2014.

[3] R. Basseda, "Fragment Allocation in Distributed Database Systems", online available at https://pdfs.semanticscholar.org/00dd/a697ceef89bea2d4c9fddf738aa54b48bd06.pdf last accessed on May 2017

[4] A.B. Deshmukh, et. al. "Horizontal Fragmentation Technique in Distributed Database", IJSRP, Vol.2, Issue 5, May 2015

[5] Suksawat, Bandit. "Diameter Error Prediction Using Fuzzy Logic for Cast Nylon 6 Turning Operation", IERI Procedia, 2014.

[6] Huang, Shouren, Yuji Yamakawa, Taku Senoo, and Masatoshi Ishikawa. "A Pre-compensation Fuzzy Logic Algorithm Designed for the Dynamic Compensation Robotic System", International Journal of Advanced Robotic Systems, 2015.