Dynamic Vertical Fragmentation of VLDB Using Active Rules

Lisbeth Rodriguez Mazahua and Xiaoou Li Zhang

Departamento de Computación, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, CINVESTAV
Av. Instituto Politécnico Nacional 2508, Col. San Pedro Zacatenco, México, DF.
lisbethr@computacion.cs.cinvestav.mx, lixo@cs.cinvestav.mx

Abstract. In this project an algorithm of generation of active rules to fragment Very Large Databases (VLDB) dynamically will be developed. Information related to the access of the applications to the attributes will be introduced to the algorithm and the algorithm will carry out the vertical fragmentation. The algorithm will be executed whenever the structure of the database (modify the attributes of the database: to add, to eliminate or to modify) and the users accesses frequencies undergo changes. The VLDB will be fragmented according to the scheme of vertical fragmentation, since this scheme is suitable for applications where data objects tend to be very large and many of the user queries do not require access to the complete objects [1], in this form the impact to use active rules in the process of dynamic vertical fragmentation of a VLDB will be analyzed.

1 Introduction

Modern application, such that document management, multimedia and hypermedia applications, manipulates increasing amounts of data of great size produced and stored in VLDB. The performance of these applications is degraded as increases the amount of data that is stored in the VLDB, for that reason it is necessary to carry out a distribution of the database to improve the performance of these applications.

Vertical fragmentation of databases is a technique for facilitating efficient execution of end-user applications, because it reduces the number of disk accesses needed for executing a query by minimizing the number of irrelevant attributes accessed. This is accomplished by grouping together the frequently accessed attributes as vertical class fragments. In the case of the VLDB where attributes tend to be very large and many of the user queries do not require access to the complete records, the vertical fragmentation would imply a quite substantial cost saving [1].

There exist a great number of vertical fragmentation algorithms, nevertheless one of problems of these algorithms to apply them in VLDB, is that the majority of them realises a static fragmentation [2], [3], [4], [5], [6], that is to say, settles down with the initial design of the distributed database (DDB), which remains

fixed until the administrator of the system takes part to realise changes. This reasoning is valid for those applications that their access patterns to the DDB do not change. Nevertheless, the nature of a VLDB is dynamic, with changes in the patterns and frequencies of access, nodes, costs, and resources. A VLDB with a static vertical fragmentation can be degraded severely with time in its performance by its incapacity of answer to the changes in the operation and resources of the system. The VLDB would be more efficient if it could verify dynamically if the vertical fragmentation scheme continues being good when the frequency of access to the data and the database scheme undergo changes in order to determine if a new fragmentation is necessary [7], [8].

The active databases have the capacity to monitor events of and to respond to these events automatically, this capability allows the database system to deal with problems concerned with observation objects and situation (for example, frequencies of access, state of the database) and executing some predefined actions in a timely manner when certain events are detected [9]. In this sense, the incorporation of active rules in the VLDB provides a platform adapted for the implementation with global reactions to the dynamic changes of the applications and the users. Until now works related to the use of the active rules to realise dynamic vertical fragmentation of VLDB have not been reported. Therefore, the idea to create an algorithm of generation of active rules to fragment a VLDB dynamically arises with the aim of improving the performance of end-user applications.

2 Problem statement

When one looks for to improve the performance of VLDB, it is not possible to think about realising a static vertical fragmentation because the initial fragmentation is realised on the basis of the information of the queries and of the database scheme that is obtained before realizing the fragmentation, the dynamic nature of VLDB causes that the information of the database changes constantly, therefore a suitable fragmentation scheme can be degraded little by little being in a reduction of the database performance. Due to the necessity of efficient tools or techniques for the management of VLDB in this doctoral thesis an algorithm of generation of active rules to fragment VLDB dynamically will be developed to improve the performance of the end-user applications. The problem that is tried to solve in this doctoral thesis is to study the impact that has the use of active rules in the process of dynamic fragmentation of VLDB. In order to achieve this the vertical fragmentation algorithms were analyzed to select the most suitable for VLDB since the algorithm that will be developed will be based on the selected algorithm, if changes to the database structure (to add, to eliminate, or to modify attributes) or changes in the frequencies of users accesses are done, the algorithm will have to detect them and to generate the new vertical fragmentation scheme. Multimedia databases are a special case of VLDB, for that reason, the algorithm will be implemented and evaluated in a multimedia application called SEREMAQ [10], who presents to the user the

multimedia information of equipments of the company Servicios de Reparacin de Maquinaria, such information is stored in a distributed database created with the Object Oriented DataBase Management System (OODBMS) DataBase For Objects (DB4O) [11].

3 Methodology

The process to develop the algorithm of generation of active rules to fragment VLDB dynamically consists of four stages:

Algorithm Determination In this stage it is necessary to realize a detailed study of the vertical fragmentation algorithms in order to select the most suitable for VLDB.

Algorithm design For design the algorithm it is necessary to specify the set of rules of vertical fragmentation, this consists of three stages: rule extraction, rule analysis and rule update [12]. It is tried to use the technique of data mining in this stage, because it is a non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [13]. The objective of the rule extraction in this project is to find the semantics of the vertical fragmentation and to represent them using the ECA rule paradigm. For this the discovered rules will be organized and represented based on general-specific (GS) patterns [14] it organizes the discovered rules in an intuitive hierarchical fashion, which enables user to focus his/her attention on the interesting aspects of the rules and apply the discovered rules appropriately. The analysis of rules checks if rule behaviour is correct and the rule update is indispensable when the vertical fragmentation semantics change or mistakes have been made during the rule design. To analyze the set of rules will be used the Conditional Colored Petri Net (CCPN) [15] since integrates rule representation and processing in only one model.

Algorithm implementation The algorithm will be implemented in the multimedia application SEREMAQ.

Algorithm evaluation The algorithm implementation will be evaluated on application SEREMAQ to know the impact of using active rules in the process of dynamic vertical fragmentation of a VLDB.

4 Preliminary results

The state-of-the-art with regard to the techniques and vertical fragmentation algorithms was analyzed, tests of execution with the different algorithms [2], [3], [4], [5], [6], [16] in the multimedia application SEREMAQ were realised and the

one that obtained a better performance was selected, also the importance of the vertical fragmentation scheme used was demonstrated. According to the study of the state-of-the-art and the realised comparison we determined that the algorithm that will be used in the thesis is [16] because its cost model considers the size of the data, which is very important in VLDB because data have very varied sizes, in addition it realises the fragmentation and allocation simultaneously with low computer cost, is easier to implement and takes into account to the complex data, therefore can be implemented in VLDB.

References

1. Fung, C., Karlapalem, K., Li, Q.: An Evaluation of Vertical Class Partitioning for Query Processing in Object-Oriented Databases. IEEE Transactions on Knowledge and Data Engineering, 14(5), 1095–1118 (2002)
2. Navathe, S., Ceri, S., Wiederhold, G., Dou J.: Vertical Partitioning Algorithms for Database Design. ACM Transactions on Database Systems, 4, 680–710 (1984)
3. Marir, F., Najjar, Y., Alfaress, M., Abdalla, H. I.: An Enhanced Grouping Algorithm for Vertical Partitioning Problem in DDBS. 22nd International Symposium on Computer and Information Sciences, 39–44 (2007)
4. Abuelyaman, E.S.: An Optimized Scheme for Vertical Partitioning of a Distributed Database. IJCSNS International Journal of Computer Science and Network Security, 8(1), 310–316 (2008)
5. Son, J. H., Kim, M. H.: An Adaptable Vertical Partitioning Method in Distributed Systems. Journal of Systems and Software, 73(3), Pages 551–561. (2004)
6. Hui, M., Schewe, K. D., Kirchberg, M.: A heuristic approach to vertical fragmentation incorporating query information. Seventh International Baltic Conference on Databases and Information Systems (IEEE Cat. No. 06EX1364), 69–76 (2006)
7. Zhenjie, L.: Adaptive Reorganization of Database Structures trough Dynamic Vertical Partitioning of Relational Tables. Master Thesis,University of Wollongong, (2007)
8. Sleit, A., AlMobaideen, W., Al-Areqi, S., Yahya, A.: A dynamic object fragmentation and replication algorithm in distributed database systems. American Journal of Applied Sciences, 613–618 (2007)
9. Vanapipat, K., Pissinou, N., Raghavan, V.: A dynamic framework to actively support interoperability in multidatabase systems. Proceedings. RIDE-DOM 95. Fifth International Workshop on Research Issues in Data Engineering, 148–153 (1995)
10. Rodriguez, L.: Desarrollo de una Aplicación de Negocios que Implemente una Base de Datos Distribuida Multimedia Utilizando un Sistema de Gestión de Bases de Datos Orientado a Objetos. Tesis de maestría, Instituto Tecnológico de Orizaba, División de Estudios de Postgrado e Investigación, (2007)
11. DataBase For Objects, http://www.db4o.com/
12. Dai, M., Huang, Y.: Data Mining Used in Rule Design for Active Database Systems. Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 4, 588–592 (2007)
13. Fayyad, U.: Advances in Knowledge Discovery and Data Mining. AAAI Press/MIT Press, (1996)
14. Dai, M., Huang, Y.: Organizing the Discovered Association Rules Based on General-Specific (GS) Hierarchical Patterns. The Fourth International Conference on Machine Learning and Cybernetics, 2206–2212 (2005)

15. Li, X., Medina. J. M., Chapa, S. V.: Applying Petri Nets in Active Database Systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 37(4), 482–493 (2007)
16. Hui, M.: Distribution Design for Complex Values Databases. PhD Thesis, Massey University, (2007)