

The Application of Fragmentation Techniques on Distributed Database: An Overview

Sharifah Hafizah Sy Ahmad Ubaidillah^{1,a}, A Noraziah^{1,2,a}, Mohd Helmy Abdul Wahab³,
Ainul Azila Che Fauzi¹, Roslina Mohd Sidek¹

¹Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, 26300 Kuantan, Pahang.

²IBM Centre of Excellence, Universiti Malaysia Pahang, 26300 Kuantan, Pahang, Malaysia

³Department of Computer Engineering, Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia

Corresponding's author: sh.hafizah1989@gmail.com, noraziah@ump.edu.my

Abstract. Distributed processing is an effective way to improve performance of database systems. Hence, fragmentation and proper allocation of fragments across various sites of the network is considered as a key research area in distributed database environment. Data can be stored in different computers by fragmenting the whole database into several pieces called fragments. Each piece is stored at a different site. Fragmentation techniques had been proved by many research in improving the performance of the database system, reduce the transfer cost and access time. This paper presents an overview of various fragmentation techniques on distributed database.

Keywords: Fragmentation, Distributed Database, Partitioning, Transfer Cost, Computational Intelligence

1. INTRODUCTION

One of the biggest problems that Data Grid users have to face today is data redundancy. Many organizations deploy the integrate database application systems in order to manage their business operation¹. One way to overcome this problem is by applying fragmentation process. The “data fragmentation” concept is similar with “data partitioning”. Data fragmentation is a fundamental characteristic of database distributed systems². Fragmentation in distributed database is very useful in terms of usage because usually, applications study with only some of relations rather than entire of it³. Data fragmentation allows you to break a single object into two or more segments or fragments. The object might be a user’s database, a system database, or a table. Each fragment can be stored at any site over a computer network. Information about data fragmentation is stored in the distributed data catalogue (DDC), from which it is accessed by the TP to process user requests⁴.

Fragmentation is done to reduce network transfer cost and communication costs. In data distribution, it is better to study with subsets of relations as the unit of distribution. The other benefit from fragmentation is the efficiency. Data is stored close to where it is most frequently used and for data that is not needed, it is not stored. By using fragmentation, a transaction can be divided into several sub queries that operate on fragments. So, it will increase the degrees of parallelism. Besides, it also good for security as data not required for local applications is not stored. So, it will not available to unauthorized users³.