An intelligent approach to improve the performance of a data warehouse cache based on association rules

Walid Moudani*
Mohammad Hussein
Mirna Moukhtar
Lebanese University
Doctorate School of the Sciences and Technologies
Tripoli
Lebanon

Félix Mora-Camino*
Air Transportation Department
ENAC/DGAC, 7 Avenue Edouard Belin
31055 Toulouse
France

Abstract

In the world of business applications, it is significantly important to reflect patterns and trends of customers, all this to make tactical and strategic decisions. The data warehouse holds information management and turns it into meaningful management information, from which, very interesting patterns can be discovered by applying knowledge discovery process. The use of Online analytical processing with other related technologies such as data mining, can meet the needs related to business management analysis of an organization. Most analytical activities are completed remotely, and because of the huge data size of the Data Warehouse systems. So we need tools that strengthen applications to access the requested information quickly. As the update of the Data Warehouse is not too frequent, it is possible to improve query performance while storing the data retrieved by them in a cache. However, the most powerful systems have a small capacity to store the entire database in memory cache. The caching chunks technique is designed to keep in cache the query results in the form of chunks of values, instead of storing them in large tables. In this paper, we propose a new technique for caching multidimensional queries based on association rules. Using this technique will allow all users to enjoy the benefits of Data Warehousing in the best manner, and also to improve performance and also increase the use of the system while reducing the response time. The technique is build using an architecture comprising a data

*E-mail: wmoudani@ul.edu.lb

Journal of Information & Optimization Sciences
Vol. 33 (2012), No. 6, pp. 601–621
© Taru Publications
warehouse, a memory cache on the server and a one on each user’s machine, in which the association rules and query results are stored. These results are kept in the form of chunks to enjoy all the advantages of the technique of fragmentation into chunks. This approach has been implemented and tested over a real huge data followed by displaying the results and analyzes.

**Keywords:** Data Warehouse, Data Mining, OLAP, Association Rules, Cache based on chunks.

1. **Introduction**

In the modern business environment, Business Intelligent System should provide data access for managers, analyzers and decision makers. Business Intelligent System has become a key contributor to mark a significant competitive advantage for companies. It includes a Data Warehouse (DW) environment [9], a system for On Line Analytical Processing (OLAP), tools for Data Mining (DM) [2, 12, 14, 15, 17], etc. Many companies have adopted DWs from operational databases. Thus, the DW associated to analysis tools OLAP are an effective solution for business intelligence. These systems are based on the multidimensional paradigm, which, based on the concepts of dimension, fact, measurement and OLAP operators, enable a multidimensional analysis of large amounts of data. There are two types of DW: ROLAP in which data is represented in a relational database and MOLAP in which data is represented in multidimensional data cubes [1]. In this work, we focus on the second one. The large size of most DWs leads to a significant cost of queries processing and leads us to find a cache efficient technique for reducing delay in multidimensional processing. The DW stores information that is collected from multiple heterogeneous information sources to answer complex queries and analyze information [5, 6, 8, 10, 16]. It collects copies of data from remote data sources and integrates this information into a repository for reporting and monitoring of strategic decisions. An OLAP system provides tools to explore and navigate through data cubes to extract valuable information [6]. Systems or applications of decision support running both queries in multidimensional data using complex techniques, and also use DM techniques to extract valuable information and useful in the considered application domain. DM techniques provide powerful extraction and knowledge discovery such as Association Rules (ARs), classification, and segmentation, etc. A DM algorithm is a well-defined procedure that takes input data and