

OLAP query optimization : A Framework for Combining Rule-Based and Cost-Based Approaches

Ladjel Bellatreche*, Arnaud Giacometti **, Dominique Laurent ***,
Patrick Marcel **, Hassina Mouloudi**

* LISI/ENSMA

Téléport 2, 1, ave. Clément Ader 86960 Futuroscope - France
bellatre@ensma.fr

** Université Francois-Rabelais de Tours, LI

Master SIR, Antenne Universitaire, 3 place Jean Jaurès 41000 Blois
(arnaud.giacometti, patrick.marcel, hassina.mouloudi)@univ-tours.fr

** Université de Cergy-Pontoise

2, ave. A. Chauvin - BP 222 - 95302 Cergy-Pontoise Cedex - France
dominique.laurent@dept-info.u-cergy.fr

Abstract. To optimize queries in relational databases, two categories of optimization techniques have been proposed : the Rule-Based Approach (RBA), and the Cost-Based Approach (CBA). In the RBA, the optimizer uses rule transformations using the relational algebra. In the CBA, the optimizer uses a cost model to estimate the potential cost of each operation using statistics about the database and the tables involved in the query. Usually both categories are implemented by commercial DBMSs and are often intermixed. In multidimensional databases however, most of query optimization techniques follow only the CBA to select optimization structures such as : materialized views, advanced indexing schemes and data partitioning. No approach has been proposed yet to rewrite OLAP queries using a multidimensional algebra. In this paper, we show that the RBA can be applied to multidimensional databases by rewriting each OLAP query to obtain an efficient rewritten query that can be executed using a CBA. In particular, we show that the RBA can be used to take into account one of the specificities of OLAP which is the visualization of the OLAP query result. We propose a multidimensional algebra that represents the core of our RBA optimization, and we show how rewritten queries can be processed using the CBA proposed for multidimensional databases.

1 Introduction

A data warehouse (DW) integrates massive amounts of data from multiple sources. In a DW, users access very large databases to carry out strategic analysis for maintaining business competitiveness by executing complex OLAP queries [Karloff et Mihail, 1999]. This complexity is due to the presence of join and aggregation operations. Therefore, an efficient query processing becomes a critical issue. To optimize these complex queries, several techniques were proposed that we can divide