Coupling Materialized View Selection to Multi Query Optimization: Hyper Graph Approach

Ahcene Boukorca, LIAS/ISAE-ENSMA, Poitiers University, Poitiers, France
Ladjel Bellatreche, LIAS/ISAE-ENSMA, Poitiers University, Poitiers, France
Sid-Ahmed Benali Senouci, Mentors Graphics, Montbonnot-Saint-Martin, France
Zoé Faget, LIAS/ISAE-ENSMA, Poitiers University, Poitiers, France

ABSTRACT

Materialized views are queries whose results are stored and maintained in order to facilitate access to data in their underlying base tables of extremely large databases. Selecting the best materialized views for a given query workload is a hard problem. Studies on view selection have considered sharing common sub expressions and other multi-query optimization techniques. Multi-Query Optimization is a well-studied domain in traditional and advanced databases. It aims at optimizing a workload of queries by finding and reusing common sub-expression between queries. Finding the best shared expression is known as a NP-hard problem. The shared expressions usually identified by graph structure have been used to be candidate for materialized views. This shows the strong interdependency between the problems of materialized view selection (PVS) and multi query optimization (PMQO), since the PVS uses the graph structure of the PMQO. Exploring the existing works on PVS considering the interaction between PVS and PMQO figures two main categories of studies: (i) those considering the PMQO as a black box where the output is the graph and (ii) those preparing the graph to guide the materialized view selection process. In this category, the graph generation is based on individual query plans, an approach that does not scale, especially with the explosion of Big Data applications requiring large number of complex queries with high interaction. To ensure a scalable solution, this work proposes a new technique to generate a global processing plan without using individual plans by borrowing techniques used in the electronic design automation (EDA) domain. This paper first presents a rich state of art regarding the PVS and a classification of the most important existing work. Secondly, an analogy between the MQO problem and the EDA domain, in which large circuits are manipulated, is established. Thirdly, it proposes to model the problem with hypergraphs which are massively used to design and test integrated circuits. Fourthly, it proposes a deterministic algorithm to select materialized views using the global processing plan. Finally, experiments are conducted to show the scalability of our approach.

Keywords: Hypergraph, Materialized Views, Multi-Query Optimization, Query Interaction, VLSI

DOI: 10.4018/ijdwm.2015040104
1. INTRODUCTION

Success or failure of enterprises depends on their DBMSs being able or not to analyze data from very large data warehouses (DW) and give timely suitable decision information by using complex and numerous queries (Fomkin & Risch, 2007). The necessity to have efficient methods to access and use data becomes more important and urgent than ever. The main technique to increase the query performance in DW is to reuse intermediate results from queries which reduce expensive access to base tables and globally optimize a set of queries. Stored intermediate results in DW are called materialized views (MV). They constitute one of the most important redundant optimization structures that improve considerably query processing time. But they need to be refreshed when changes occur to the base tables; therefore it is not possible to materialize all views due to the maintenance time constraint and to the storage space limitation. Finding and exploiting the best intermediate results is known as multi-query optimization (MQO).

The most important problem in DW design is how to identify, store and maintain the most appropriate set of MV such that they optimize both the query processing cost and the maintenance cost. i.e., the challenge in the materialized views selection problem is to find a trade-off between performance and maintenance cost under space constraint (Yang & Karlapalem & Li, 1997). The materialized views problem involves two main tasks: (1) views selection problem to select the best views (Yang & Karlapalem & Li, 1997), and (2) views maintenance problem to update views affected by changes in the data of base relations (Gupta, 1999; Wang & Gruenwald & Zhu, 2004). Furthermore, a less studied problem is the view adapting problem to adjust the views schema when the schema of based relations changes (Gupta & Mumick & Ross, 1995).

Several academic and industrial efforts are conducted on the well-studied materialized views selection problem (VSP) (Elghandour & Aboulnaga & Daniel & Zilio, 2013). Most works propose an algorithmic vision, i.e., focus on developing views selection algorithms without considering the MQO problem. Existing algorithms can be classified in three main categories: (1) deterministic algorithms (2) randomized algorithms and (3) hybrid algorithms. These approaches focus on enhancing the performance of the algorithms that select the views from a large search space. To reduce the search space of views, a couple of works have been proposed which combine MQO and materialized views problems (Mistry & Roy & Sudarshan & Ramamirtham, 2001; Yang & Karlapalem & Li, 1997). In these works, an algorithm uses MQO to identify the intermediate results and produce a global processing plan, which is handed to another algorithm that will use this global plan to select the candidate views to be materialized. To the best of our knowledge, Yang et al (Yang & Karlapalem & Li, 1997), is one of the sole work that uses MQO to select the views. It deals with two interdependent problems: (a) constructing an optimal global plan of queries using MQO and (b) use this plan to select views to be materialized. The plan is the result of merging all query logical plans, called Multiple Views Processing Plan (MVPP).

Since MQO algorithms are based on individual query plans, whose generation is a hard task, the complexity of MQO is doubly exponential. With the advent of extremely large databases, queries became voluminous and each query may be represented by an enormous number of possible logical plans. To overcome this scalability obstacle, we propose an approach which can generate a global processing plan without using individual query plans, and groups the queries in several disjoint components where each component contains queries that highly share common sub-expressions between them, and where shared sub-expressions between components are minimal. Each component will be used independently to generate a multiple view processing plan (MVPP) by using global common sub-expressions, and all components will be merged in one global plan which will be used to select materialized views.
Related Content

Classification of Failures in Photovoltaic Systems using Data Mining Techniques
[www.igi-global.com/chapter/classification-of-failures-in-photovoltaic-systems-using-data-mining-techniques/150220?camid=4v1a](www.igi-global.com/chapter/classification-of-failures-in-photovoltaic-systems-using-data-mining-techniques/150220?camid=4v1a)

A Novel Multi-Secret Sharing Approach for Secure Data Warehousing and On-Line Analysis Processing in the Cloud
XML Tree Classification on Evolving Data Streams
Albert Bifet and Ricard Gavaldà (2012). XML Data Mining: Models, Methods, and Applications (pp. 199-218).
www.igi-global.com/chapter/xml-tree-classification-evolving-data/60910?camid=4v1a

An Efficient Method for Discretizing Continuous Attributes
Kelley M. Engle and Aryya Gangopadhyay (2010). International Journal of Data Warehousing and Mining (pp. 1-21).
www.igi-global.com/article/efficient-method-discretizing-continuous-attributes/42149?camid=4v1a