IJCSC International Journal of Computer Science & Communication Vol. 1, No. 1, January-June 2010, pp. 235-237

Query Optimizer Model for Performance Enhancement of Data Mining Based Query

Deepak Sukheja[1] & Umesh K Singh[2]

[1]Privatam Institute of Technology and Management Indore
[2]Institute of Computer Science, Vikram University, Ujjain
[1]sukheja@rediffmail.com, [2]umeshsingh@rediffmail.com

ABSTRACT

In present scenario almost applications are built upon data mining & OLAP tools and allow Users to answer information requests based on a data warehouse, that is managed by a powerful RDBMS. This paper is focused on query optimization technique which generates sequences of SQL statements in order to produce the requested information. The analysis for this paper is exposed that many sequences of queries generated by commercial tools are not very efficient. Semantic query optimizer architecture is suggested for these applications. The main component is a SQO optimizer that accepts previously generated sequences of queries and rewrites them according to a set of optimization strategies, before they are executed by the underlying database system. The advantages of this proposed architecture are discussed and this is an appropriate approach to optimize query sequences for data warehousing & Data mining based applications.

1. INTRODUCTION

Query optimization is of great importance for the performance of a relational database, especially for the execution of complex SQL statements. A query optimizer determines the best strategy for performing each query. The query optimizer chooses, for example, whether or not to use indexes for a given query, and which join techniques to use when joining multiple tables. These decisions have a tremendous effect on SQL performance, and query optimization is a key technology for every application, from operational Systems to data warehouse and analytical systems to content-management systems.

In present scenario data warehouses & mining turned out to be the common basis for the integration and analysis of data in modern enterprises. Data mining based applications are used to analyze data on the operational level as well as on the strategic level. This includes techniques like online analytical processing (OLAP) and data mining. Additional tools are used for the preprocessing and integration of data from different sources. A lot of work has been done on data warehousing & data mining and their optimization.

Query processing and optimization process work together to execute any kind of queries. Query processing is concerned with execution of query or refers to the activities involved in extracting data from a data warehouse. Query processing determines what data is to be fetched but does not define the method by which the data manager searches the database. On the other hand, optimization process deals with the efficiency of the query. It defines the execution plans, the strategy of execution of query and chooses best execution plan. In other words Query optimization is the process of selecting the most efficient query evaluation plan from among the many strategies. The goal of query optimization is to minimize the response time of query and to make the best use of the server's resources by minimizing network traffic, disk I/O and CPU time. This goal can only be achieved by understanding the logical and physical structure of the data warehouse.

2. BACKGROUND

The purpose of query optimization is faster response to queries. The semantic optimizer knows more about its data rather than user. Therefore it can replace the user's query with a different query which will generate the same result set efficiently in less time. The new query is faster because it will do less work when extracting the selected result tuples from the database

3. LITERATURE REVIEW

A large number of research papers are published in the area of Data warehousing & data mining and Query Optimization Techniques but Previous research does not make clear under which conditions, what types of query optimizer is likely to be carrying great weight.

In the paper [2, 3 and 4] query driven rule discovery, rules are inferred from the restriction clauses of queries arriving at the database and the results they produce. In other words, the approaches in which two syntactically

different queries produce the same answer and the cost of each query is different.

In the paper [5] describe the principle of semantic query optimization which is to use semantic rules, such as to reformulate a query into a less expensive but equivalent query, so as to re-duce the query evaluation cost.

In the paper [6] propose an architecture for processing complex decision support queries involving multiple, heterogeneous data sources, introduce the notion of transient-views and also develop a cost-based algorithm that takes a query plan as input and generates an optimal "covering plan", by minimizing redundancies in the original plan.

4. OBJECTIVE

In this paper, we focus on a model of Optimizer Architecture for data Warehousing based (decision support systems) application. According to this model, the application generates a sequence of SQL statements, which is processed by the (OLAP Server) DWDBS.

5. COMMON QUERY OPTIMIZATION MODEL

Stander query optimization model is described in figure 1.1 and figure 1.2. Basically Query optimization process consists of two phases:

First Phase: Query Rewrite

Second Phase: Cost Based Optimization

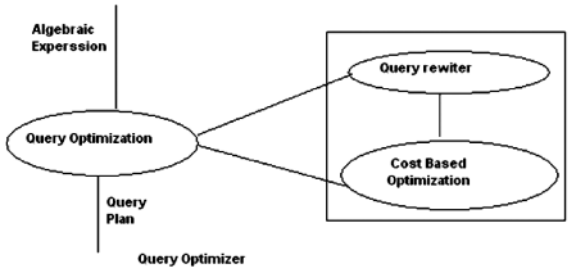

Fig: 1.1

Query Rewriter Phase is responsible to write same query in different way which should be more efficient and cost effective and second phase is responsible for calculate the cost of all possible query those are written in Phase one and original query

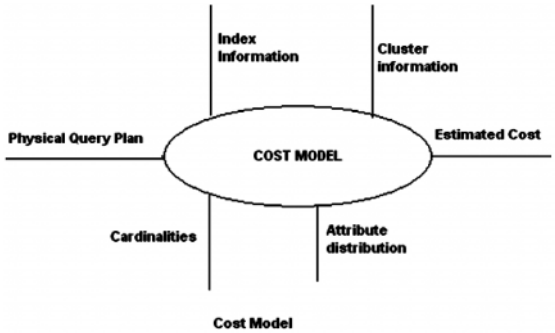

Fig: 1.2

6. PROPOSED QUERY OPTIMIZER MODEL

SQL is used as query language because most data warehouses are based on relational or extended relational database system. As the information requests of the users are likely to be very complex. In order to reduce the complexity of the query generation process and in order to preserve portability to other database systems proposed semantic query optimization architecture is very useful.

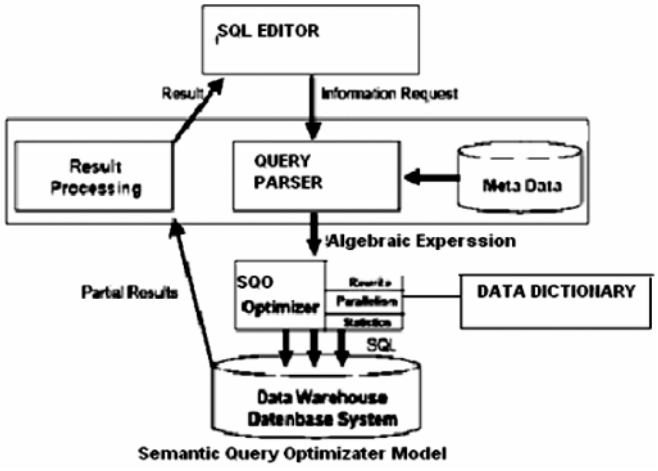

Fig: 2

In the context of RDBMS, previous research does not make clear the unique role that database integrity constraints can play in SQO. In addition, the close relationship between SQO and the enforcement of database integrity constraints has not been established at the time of query execution.

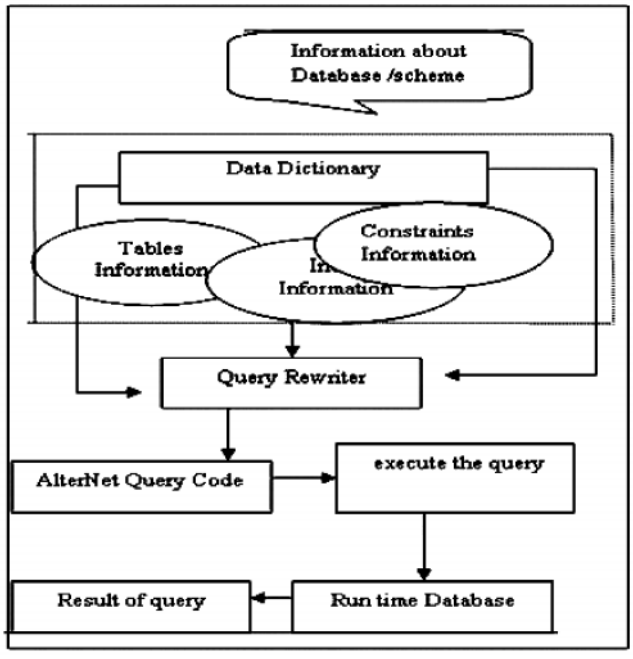

Figure 3: Internal Architecture of Query rewriter for Semantic query optimizer

In general, an abstraction of the query optimization process in a DBMS. Given a database and a query on it, several execution plans exist that can be employed to answer the query. In principle, all the alternatives need to be considered so that the one with the best estimated performance is chosen. An abstraction of the process of generating and testing these alternatives which is shown in Figure 3, which is essentially a modular architecture of a semantic query optimizer.

7. CONCLUSIONS

It is a general perception of the users that information can always be retrieved efficiently, assuming that the database is designed properly. However, there are limitations of query processing and query optimization. Queries formulated using SQL query language provides little predictive information useful for estimating query performance. Internal knowledge of the database structure, data distribution, and semantic query optimizing strategy are necessary to develop effective query execution plan. This is possible if and only if when the indexing, referential integrity and naming conventions are properly defined at the database designing level so that the semantic query optimizer uses indexing, referential integrity to rewrite a new query.

REFERENCE

[1] H. Schwarz, R. Wagner, B. Mi tschang. Improving the Processing of Decision Support Queries Strategies for a DSS Optimizer in Technical Report TR-2001-02 .

[2] C.T. Yu and W. Sun. Automatic knowledge acquisition and maintenance for semantic query optimization. IEEE Transactions on Knowledge and Data Engineering.

[3] M. Siegel, E. Sciore, and S. Salveter. A method for automatic rule derivation to support semantic query optimisation. ACM Transactions on Database Systems.

[4] S. Shekhar, B. Hamidzadeh, A. Kohli, and M. Coyle. Learning transformation rules for semantic query optimization: A data-driven approach. In Special Issue on Learning and Discovery in Knowledge-Based Databases,

[5] Ullman, J. D. (1988). Principles of Database and Knowledge-base Systems, **I,II.** Palo Alto, CA: Computer Science Press

[6] Cost-based optimization of decision support queries using transient-views by Subbu N. Subramanian, Shivakumar Venkataraman. ACM SIGMOD international conference on Management of data

[7] S.sudarshan, Korth: Database System Concepts , second edition.

[8] Date, C. An Introduction to Database Systems, Third edition. [9].http://www.oracle.com/technology products /bi / db / 10g / pdf twp _ general _query _ optimization_10gr2_0605.pdf.

[10] http://www.dbaoracle.com/art_otn_cbo.htm