



## An Idea of Extraction of Information Using Query Optimization and Rank Query

Vivek Shrivastava<sup>1</sup>

Computer Science Department(SRIT,Jabalpur)

viveksri\_tit@yahoo.com

Brajesh Patel<sup>2</sup>

Computer Science Department(SRIT,Jabalpur)

patel.brajesh@indiatimes.com

**Abstract**— Information retrieving and storing are the basic tasks of the relational database. Query optimization is used for accessing database in an efficient manner, analysing and choosing an optimized plan is a complex process. Today’s end-users are more interested in specific results rather than getting response in short time. Ranking queries (or top-k queries) are support for accessing specific results based on user’s demand. Efficient execution of top-k queries is becoming a major challenge for relational database. Ranking queries produce results that are ordered on some ranking attribute. In this paper some light has been thrown on query optimization and its challenges, and rank query has been discussed.

**Keywords-** Query Optimization, Top-k Query, Rank Function, Information Retrieval.

### 1. Introduction

Relational query languages provide a high-level declarative interface to access data stored in relational databases [7]. Structured Query Language is a standard for storing and retrieving data from relational database. Query optimization is the way to reduce execution time of the declared query. Query optimizer is a main part in optimizing process which handles a large input space of complex query. The query optimization process itself is complex task. By reducing execution time end users get response in short time, but apart from getting response today’s end users are very much interested in specific results based on their inputs. Ranking query or top-k query plays an essential role in retrieving specific information. Top-k queries intend to provide only the top-k results of a query, according to a user-specified ranking function. The growing significance of top-k queries has caught the attention of the researches. A top-k query only returns the top k results according to a user-specified preference, which generally consists of two components: a selection condition and a ranking function [6].

### 2. Query Optimization by Ranking Queries on Relational Database

#### 2.1. Relational Database Management System (RDBMS)

A relational database management system (RDBMS) is the most well known database, now in use, is based on the relational database model [1]. An imperative characteristic of relational systems is that a database can be divided in several related tables and data can be accessed without restructure the database tables.

#### 2.2. Information Retrieval and SQL

Information retrieval is a process of accessing data from relational database, in other words it is used to make queries into database. On the other hand Structured Query Language (SQL) is a programming language designed for organization or manipulation or retrieval of data in/from relational database management systems (RDBMS). A Query Execution Plan (QEP) is then founded which represented as a query tree includes information about the access method available for each relation as well as the algorithms used in computing the relational operations in the tree. The next step is to generate the code for the selected QEP; this code is then executed in either compiled or interpreted mode to produce the query result [8, 9]. Figure1: shows the different steps of Query Processing.

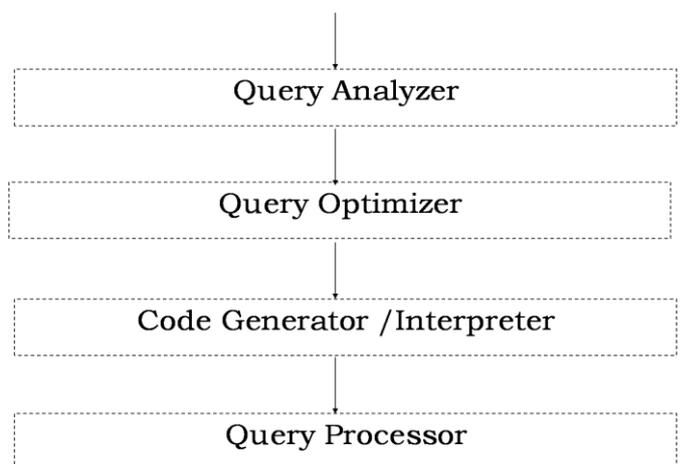


Figure 1: Query Traversal

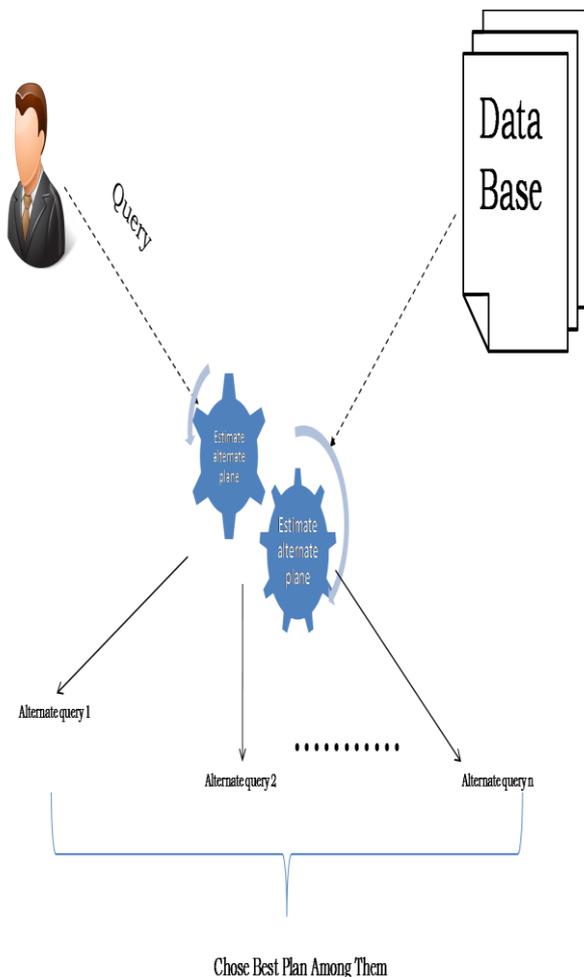
### 3. Query Optimization

Query optimization plays a crucial role in retrieving information from the relational database, a query can be executed in multiple fashions, wherein several query arrangements are checked up and a fine query plan is identified for further process. There are two factors one is the amount of time spent in finding out the best plan and the amount in executing the query plan. Different database have different methods of balancing these two factors [1].

Cost based query optimizers evaluate the resources of various query plans. In general the resources (resource cost) are CPU time, amount of memory space, storing and fetching time. Possible access ways and diverse relational tables join techniques are examined for efficient query plan. The searching space may be rather large depending on the complexity of the SQL query.

#### 3.1 Advantages of Query optimization

The main intention of query optimization is to choose an efficient execution plan. The query optimization process is shown in fig.



- Faster processing of query
- Lesser cost per query

- High performance of the system
- Lesser stress on the database
- Efficient usage of database engine

#### 3.2 Main Complicatedness of Query Optimization

The main objective of query optimization is to choose effective execution plans for defined query. In this process an optimal execution plan is selected from many alternatives, depending on the optimization parameters, an optimized plan may be based on response time or amount of memory because time and space are the most important parameters in case of data retrieving and data storing. The other parameters of the query optimization process and their limitations are as follows:

- I. **Complex Query and Several Execution Plan:** For every entered query, the query optimizer thinks about a large number of execution strategies, and for every plan there is need to analyze and check the validity. A complex query contains a lot of SQL clauses and filters due to these a large number of alternative execution plans are possible, after a particular limit it is not possible to analyze every possible execution plan. So, this is the main difficult task to select or optimal plan for execution.
- II. **Optimization time:** “JOIN” is a keyword in SQL, and it plays an important role in making complex query and optimizing query. If optimizer considers all possible execution plan it may takes more time to find out the optimized plan than the time taking in retrieving data from the data base. The problem of finding the optimal join order in query optimization is NP-hard [1]. Thus, in many cases the query optimizer has to select a plan that is nearly optimized.
- III. **Tuple assessment:** An optimal query execution plan is always depend on number of tuples used in query, it means Query optimizer primarily rely on statistical information to make tuple assessment, and query optimizer is always depend on the accurateness of the assessment of the tuples. Increase the qualities of the selection process of an optimal execution plan rely on additional CPU cost and increased memory consumption.
- IV. **Cost estimation:** Cost estimation models are mathematical algorithms or parametric equations used to estimate the costs of a query execution in terms of time or memory consumption [4].

### 4. Rank Query or Top-k Query

Top-k queries are leading in many applications such as web databases, multimedia databases, and data mining. Rank query gives ability to database system to efficiently retrieval of information based on user’s ranked attribute. Top-k queries work in applications where users have relatively flexible preferences or specifications for certain attributes. The working of top-K query is simply an assignment of target

values to the particular attributes of a relation. To answer a top-K query, a database system identifies the objects that best match the user specification, using a given ranking function.

**For Example:** A user is interested in finding a hotel; consider a relation that is hotel in the Mumbai City. There are many tuples in this relation like address, nearest landmark, rent etc. There are three conditions, first is hotel must be near railway station, it should be three stars, and must not be expensive, these three parameters will be considered for ranking. The result to this query is a list of the 10 hotels that match the user's specification closest.

## 5. Conclusion

Query optimization has an exceptionally impact on the performance of a DBMS and it constantly grows with new optimization strategies. The main objective of query optimization is to choose effective execution plans from many alternatives. Query optimization process is a complex process. Top-k queries are leading in many applications such as web databases, multimedia databases, and data mining. Rank query gives ability to database system to efficiently retrieval of information based on user's ranked attribute. Top-k queries work in applications where users have relatively flexible preferences or specifications for certain attributes. The working of top-K query is simply an assignment of target values to the particular attributes of a relation.

## References

- [1] Leo Giakoumakis, and Cesar Galindo-Legaria, "Testing SQL Servers Query Optimizer: Challenges, Techniques and Experiences", IEEE, 2008.
- [2] M.A. Kashem, Abu Sayed Chowdhury, Rupam Deb, and Moslema Jahan, "Query Optimization on Relational Database for Supporting Top-k Query Processing Techniques", JCIT, 2010.
- [3] Neha Singh, P.K. Pandey and Anil Kumar Tiwari, "A Study on Optimization of Top-k Queries in Relational Databases", IJDE.
- [4] Dong Xin, Jiawei Han, Hong Cheng and Xiaolei Li, "Answering Top-K Queries with Multi-Dimensional Selections: The Ranking Cube Approach", ACM, 2006
- [5] Surajit Chaudhuri, "An Overview of Query Optimization in Relational Systems".
- [6] Alaa Aljanaby<sup>1</sup>, Emad Abuelrub<sup>1</sup>, and Mohammed Odeh<sup>2</sup>, "A Survey of Distributed Query Optimization" The International Arab Journal of Information Technology, Vol. 2, No. 1, January 2005
- I. S.K. SINGH, "DATABASE SYSTEM CONCEPTS, DESIGN AND APPLICATIONS", PEARSON EDUCATION, FIRST IMPRESSION, 2006 ISBN 81-7758-567-3