Distributed Query Optimization: Importance and Challenges

Mittal Nirbhik Desai
BCA Department, Parul Institute of Computer Application, Parul Group of Institutes, Limda, Waghodia, Vadodara, Gujarat, India , bhatt2008@yahoo.com

ABSTRACT: Now a days Distributed Database Management System (DDBMS) is widely used in real world, so the task of Distributed Query Optimization is one of the most important aspect of study. Here we are discussing the important criteria of query optimization on the basis of search technique.

Keywords: Distributed Database Management System (DDBMS), Distributed Query Processing and optimization.

1. INTRODUCTION

Now days the database becomes the large component for any organization those who wants their database is going to accessed from network. A Distributed database is a database that is under the control of central Database Management System (DBMS) in which storage devices are not attached to common CPU, may be stored at multiple computers or dispersed over a network.[1].

2. RELATED WORK

2.1 Distributed Query Processing

It is the process of converting relational calculus / SQL query into more efficient relational algebraic query. These processes comprise of three steps:

1. Translate the query into an equivalent relational algebra equation after checking the syntax and verifying relations.
2. Generating the plan for optimizing the cost of query.
3. In this step we need to execute the query on optimization plan and return the answer to the query.

2.2 Importance of Distributed Query Processing and Optimization

Distributed query processing is highly important for distributed computing like Grid and Cloud. These Distributed System implements Decentralized RDBMS and the main function of RDBMS is the processing query and generating data according to client request. In the case of Decentralized RDBMS the data is divided into the numbers of different locations, and the query should be processed from that multiple different locations. This type of distributed environment needs high communication for responding the query result.

- Objective of query optimization is to **minimize** the following cost function:
  I/O cost + CPU cost + communication cost (2)
- The query optimization can be done with respect to many aspect like search technique, optimization timing etc.
- In case of Cloud Computing and Grid Computing the data is distributed over different locations, so to answer any query the cost of gathering data from that distributed locations become very important aspect.

3. DISTRIBUTED QUERY OPTIMIZATION CHALLENGES

We mainly focused on optimization techniques and the optimization timing issues aspect of optimization of query.

Optimization Technique: The optimization technique mainly categorized into two search techniques

4. HEURISTIC QUERY OPTIMIZATION

From the higher level language such as SQL the query is first transformed into the Relational Algebra from the Relational Calculus. An internal representation of query is then created as Query Tree or Query Graph. Heuristic optimization transforms the query-tree by using a set of rules that typically (but not in all cases) improve execution performance:

- Perform selection early (reduces the number of tuples)
- Perform projection early (reduces the number of attributes)