

A Survey of Various Query Optimization Techniques

Rini John

Department of Computer Engineering of PIIT
Mumbai University, New Panvel, India

Nikita Palaskar

University of Memphis

ABSTRACT

Data storage and retrieving the data on a specific time frame is critical for any application today. So an efficiently designed query lets the user get the results in the desired time and creates the credibility for the corresponding application. Here in this paper various techniques which are currently used or proposed in recent years and to get a better perspective in this field of query optimization are explored. Automatic external SQL-query optimization method is where the principle of building queries regardless of applied database management system and its settings is explored. It would be interesting to explore the work of authors where they have proposed the energy-efficient query processing and optimization based on a database accelerator.

Keywords

Query Optimization, Tensilica RISC processor, Hadoop, OLTP, OLAP, Multi Query Optimization (MQO).

1. INTRODUCTION

It's an age of information overload where the user can get the required query through any medium. In this paper, the various techniques used in query optimization field and would be useful to know about these methods and how they can be incorporated or combined for the future applications are discussed. In any of the Database systems today the query optimizer is the most anticipated component on which it depends how the results will be queried in the fastest possible way. Through this component how the SQL statements can be efficiently improved. Based on these statement the query optimizer gets the proficient execution plan, this also depends on the query structure, statistical data of the objects etc.

2. QUERY OPTIMIZATION

Query optimization is a promising area where there is lot of scope of development as today data has huge importance, be it for creating your user base, research and market analysis etc. Tons of data is generated each day due to the huge success of the web world. It's important for any organization to have their system in peak of performance. Query optimization is a field which will help them to achieve these targets presently and in future.

3. LITERATURE SURVEY

In this section the significant past literature that use different query optimization techniques are cited. Most of the researchers have combined techniques of the fields to get the most effective results. These techniques are used in variety of various systems like Image Retrieval, Hadoop, JIT Compilation-based Unified.

The authors Nicoleta Angelescu, Henri George Coanda, Ion Caciula*, Catalin Dragoi and Felix Albu have proposed optimization of SQL query in Content based image retrieval systems [1]. The proposed approach have been verified by these results on infantile hemangioma images and the

efficiency is proved based on the implementation and the details about the textual features.

The database which they have used for image retrieval contains two table for storage, one is for image and the other for sub-image features. Web technologies for the server they used is Apache for the scripts running in PHP and MySQL database system for the SQL queries. For the faster retrieval of the queries, internally the indexing have been done with the GLCM matrix. The results they have shown are promising i.e. loading the system with 200 images of hemangioma cases. The average running time of the query is 2-4 milliseconds which is better as compared to the existing systems.

The author Varun Garg have proposed a framework for the Optimization of Multiple Queries for Big Data with Apache Hadoop/Hive[2]. Data warehouse organization is a critical factor to get the correct results for data analysis and for various other purpose like data mining. Hive is one of the tool used for this to process the data in Hadoop. The proposed framework transforms a set of Hive queries into more efficient queries resulting in less execution time. The results of the proposed system have shown encouraging results by 20-50% reduction in getting the results.

They have proposed the modified version of Hadoop Hive where the Multiple Query Optimizer is added which executes the group of correlated queries and produces a single HiveQL command. Also the other components like the Distributed Hive Layer is used to improve the execution plan and shown the results by which there is 50% improvement in the execution plan.

The author Sebastian Haas and co-authors have proposed A Database Accelerator for Energy-Efficient Query Processing and Optimization [3]. As everyone is aware in this age of information overload where the data is continuously growing exponentially. Here they have focused more on how to integrate the specialized hardware with database operators. And presented the Tensilica RISC processor and used the SLP CMOS technology which have shown to improve the energy improvement by 1000x on database operators. This processor is the newly developed instruction set which based on the conventional database operator.

The authors Myungcheol Lee, Miyoung Lee, ChangSoo Kim have proposed A JIT Compilation-based Unified SQL Query Optimization System [4] where they have concentrated on the Online Analytical processing and transactional queries using the JIT compilation for query optimization along with the procedure workloads to improve the efficiency of the underlying hardware.

The proposed system few modules which are as follows:

Query Workload Analyzer which analyze the compile of JIT and based on which it uses aggregate functions so as to improve the performance of SQL queries.

IR Generator uses the execution plan of SQL to generate the IR.

Stored Procedure manager helps managing the input and output of the user queries using the procedures.

SQL Semantics Optimization Rules are based on the SQL queries semantic to transform the internal indexing of the queries which is changed on the basis of these rules.

The other modules used are IR Syntax Optimization Rules, IR Manager, Macro Optimizer, Micro Optimizer and the

Machine Code Generator. They have proposed a system which generates efficient machine code for OLTP, OLAP for various heterogeneous environments.

The author M. Zagirnyak, P. Kostenko have presented an external optimization of SQL-query under conditions of databases structural uncertainty[5] they have used a LMCP algorithm which is a local model of controlled process which allow the rationality and the correctness of the process thus getting the query optimized by developing the mathematical model of IS.

Table 1. Summary of Query Optimization Techniques

| Sr. No. | Query Optimization Techniques | Description | Application |
|---------|---|---|--|
| 1 | SQL Query Optimization in Content Based Image Retrieval Systems | The proposed approach for the particular tumor images which is based on the concept of details of the image and textual features and hence they have proved the results based on these concept. | This application can be incorporated into various medical planning system particularly in infantile hemangioma images. |
| 2 | Optimization of Multiple Queries for Big Data with Apache Hadoop/Hive | The method used here is Mutli-Query optimization to enhance the overall performance of the Hadoop Hive tool. Also they have shown based on the experimental results how they have reduced the execution time. | Can be applied on the Hadoop systems to improve the Query optimization. |
| 3 | A Database Accelerator for Energy-Efficient Query Processing and Optimization | The paper proposed a database accelerator based on a Tensilica RISC core with a newly developed instruction set extension for fundamental database operators. | The experimental results have shown the energy efficiency of the database operators to most efficient level and hence be incorporated into many processors to increase the speed of the queries. |
| 4 | A JIT Compilation-based Unified SQL Query Optimization System | In this paper proposed the JIT compilation for query optimization. | This system can be used with OLAP, OLTP, and Stored Procedures for generating a queries for faster results. |
| 5 | External optimization of SQL-query under conditions of databases structural uncertainty | They have used the LMCP algorithm which is a local model of controlled process to optimize the query. | Data mining applications. |

4. CONCLUSION

This paper summarizes the techniques which will be helpful in optimizing the query for getting the results in desired time duration. It is interesting to know about how a Tensilica RISC processor extended with an instruction set to accelerate basic database operators which are used in the modern database systems. They have shown how there chip measurements show a 1000x energy improvement on selective database operators. This would be good way to explore plus expand the possibility of this technique to improve the performance in other areas of query optimization. Multi Query Optimization (MQO) is a technique to enhance the overall performance of Hadoop (Hive which is a data warehouse infrastructure tool to process structured data in Hadoop) where in a unified SQL

query optimization system using JIT compilation of OLTP, OLAP and stored procedures are used to enhance the performance on current hardware. They have backed up their achievements with results showing 20-50% improved performance as compared to the conventional systems. Query optimization using the clustering and distributed database systems would be a highly anticipated area to be explored as many of the organizations specially hospitals, banks and research companies are heavily relying on such systems. The experiment results show that the proposed Genetic algorithm is effective for query processing of multi-join, and plays important role in improving the performance of distributed database.

5. REFERENCES

- [1] Nicoleta Angelescu, Henri George Coanda, Ion Caciula*, Catalin Dragoi and Felix Albu., “SQL Query Optimization in Content Based Image Retrieval Systems”, IEEE 2016 International Conference on Communications (COMM)
- [2] Varun Garg, “Optimization of Multiple Queries for Big Data with Apache Hadoop/Hive.”, IEEE 2015 International Conference on Computational Intelligence and Communication Networks (CICN).
- [3] Sebastian Haas; Oliver Arnold; Stefan Scholze; Sebastian Höppner; Georg Ellguth; Andreas Dixius; Annett Ungethüm; Eric Mier; Benedikt Nöthen; Emil Matúš; Stefan Schiefer; Love Cederstroem; Fabian Pilz; Christian Mayr; René Schüffny; Wolfgang Lehner; Gerhard P. Fettweis, “A Database Accelerator for Energy-Efficient Query Processing and Optimization”, 2016 IEEE Nordic Circuits and Systems Conference (NORCAS) in Proc. Int. Conf.
- [4] Myungcheol Lee; Miyoung Lee; ChangSoo Kim, “A JIT Compilation-based Unified SQL Query Optimization System”, 2016 IEEE 6th International Conference on IT Convergence and Security (ICITCS).
- [5] M. Zagirnyak; P. Kostenko, “External optimization of SQL-query under conditions of databases structural uncertainty”, 2016 IEEE 17th International Conference Computational Problems of Electrical Engineering (CPEE)