© 2019 IJSRST | Volume 6 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology
DOI : https://doi.org/10.32628/IJSRST196219

Study of Query Optimization in Cloud

Y. Supriya, K. Gayathri
G. Pullaiah college of Engineering and Technology, Department of CSE, Kurnool, Andhra Pradesh, India

ABSTRACT

Cloud computing in very simple terms, is basically where a company uses someone else's computing services (usually over the internet) instead of having to run that software on their own computers. Today, cloud computing plays an important role in service-oriented technologies. The main purpose of cloud computing is, it allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access and it allow users to easily and efficiently calculate and save resources. The recent approach is to process data expression and search. To improve cloud performance, it is necessary to optimize the processing time. Our research provides a comprehensive overview of the different models and methods used to optimize queries to reduce execution time and improve resource utilization. We conducted various query optimization research activities for the classic SQL and Map Reduce platforms.
Keywords : Cloud Computing, Map-Reduce, Service Level Agreement, Query optimization, Conventional SQL.

I. INTRODUCTION

Cloud computing is a very successful paradigm for service-oriented computing [1]. The most popular cloud services are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Expanding this concept is a Database as a Service (DBaaS) or Storage as a Service. Cloud computing improves common computer and storage capacity for a number of database applications. "The observed number of applications that have affected multiple platforms in the clouds has greatly increased the amount of data generated and used by these applications" [2]. "Cloud is the basis for cloud computing applications and search algorithm and cloud organization and search algorithm." The new research topic was how to organize and manage this huge amount of data. Get users useful information about cloud computing "it's the core of the cloud application" [3] How to get data fast, accurate and secure Play an important role in a successful job data model in the cloud. In cloud computing, resources need to be automatically and quickly acquired and released during business hours to provide Service Level Agreements (SLAs) between the client and the cloud provider [4]. Using virtual machine clusters, cloud computing users can rent large amounts of resources for a short period of time to effectively execute large-scale complex queries [5]. Lease duration can be further reduced by using better query optimization techniques [6]. Therefore, it is necessary to investigate effective query optimization techniques to reduce query time and response time. It will also improve the use of computing resources in the cloud. "Query optimization leads to optimization of resource lease time in the cloud environment". Query optimization techniques in centralized and distributed platforms are extensively researched in conventional SQL and Map Reduce methods. An example of processing a request in the cloud is shown in Figure 1. In our work we looked at the field of this research and analysed different query optimization methods.

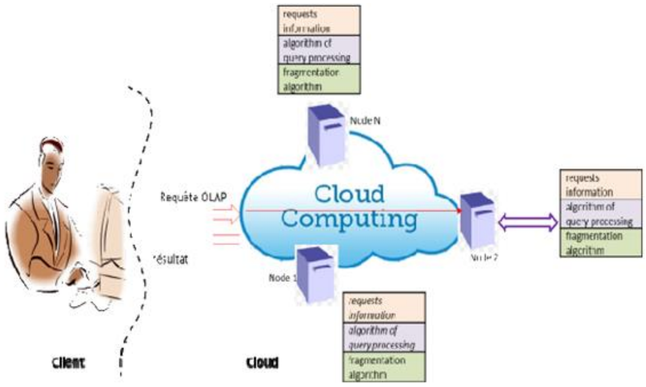

Fig 1. Cloud query processing

II. RELATED WORK

Arduino In this article, we have analysed query optimization techniques. Understand theoretical analysis of optimization methods. We have classified various methods used to optimize the query.

1.1 Spatial Query Optimization

Spatial On-Line Analytical Processing [7] performs operations in data cubes integrating both spatial and multidimensional operations. Spatial operations and multidimensional operations are performed using Spatial Data Warehouse data using multidimensional data views and pre-data counting data. Multidimensional data views are provided with multidimensional operations. Spatial operations are functions that make important elements of a model that measure position data, then analyse and provide output information. These processes are known as the Spatial On-Line Analytical Process.

1.2 Selectivity Using Histograms

1) Histogram Structure: The histogram contains the values of different cubes. In digital data, we can determine a certain range, then the appropriate group. In the case of categorical data, we need to divide the data into a ranking relative to the letters.

2) Gradual maintenance of histograms: Here we find the error of evaluation of each attribute; if the error is above a certain threshold, then we need to update the histogram; otherwise we will use the same old histogram to provide a selectivity assessment.

3) Diagram of error estimation method: Here we find error estimation using standard deviation between old data value and value of data restored in histogram groups.

1.3 Data Flow Style

In this approach [8], Data Integration System (DIS) uses the data flow query execution mode. This data flow execution model has four types of elements.

1) Query Plan: Contains a set of sub-queries for data sources and operators that indicate how the sub-query result is combined.

2) Query: The processing group is generated after the query plan and sent to the μ Engine.

3) Sender: sends the request to the μ Engine.

4) μ Engine: application process.

Therefore, all sub-queries can get their results and overall overhead costs can be reduced. The results of the study show that when the DIS executes a set of parameter queries, this new calculation can reduce the normal request processing time from 39 to 58% if the DIS executes an unwanted query group, the time for normal consultation can be reduced from 28 to 38%.

III. NAVIGATION OF QUERY OPTIMIZATION METHODS

To improve the high performance of large-scale data systems, the use of centralized or distributed platforms with traditional SQL or MapReduce technology query optimization is required. Query optimization is done in two phases in the first phase, creating a search space, and selecting the optimal search space scheme in the second phase [9]. Researchers discover many approaches that complement one or more query optimization techniques, such as removing surplus estimation, continuous or repetitive processing, queries or intermediate capture results, materialization, and pipeline. Some of these techniques reduce the cost of communications; others focus on reducing query execution time and others on the proper use of system resources in the cloud.

a) Continuous or Iterative Processing

The continuous query optimization method differs from the traditional method that focuses on compiling and disseminating data statistics before executing a query. Bruno N et al. [10] proposed a method to continuously monitor query execution, collect real-time execution statistics and adjust execution plans when needed during execution of the request. Query Optimization is enabled when new runtime statistics are available. If a better plan is found, the proposed technology intelligently adjusts the current execution plan with minimal change. Optimization methods are based on accurate data statistics to select the best execution plan.

b) Pipelining

In the cloud environment used by MapReduce, the total cost of data processing for heavy load increases simultaneously. Anyanwu K et al. [3] introduced a Nested Triple Group Data Model and Algebra (NTGA), which reduces total processing costs by reducing the number of MapReduce cycles. Based on MapReduce, enters the cost of storing intermediate results. The task is divided into different tasks. The task reads the result of the previous task to continue processing. The AQUA (Automatic Query Analyzer) for MapReduce reduces the amount of intermediate results generated to reduce storage and network costs. Adopt a two-step optimizer: Step 1 creates a group of association operators to reduce the total number of MapReduce jobs to assess the query. Phase 2 is added to group interim results to get final query results.

IV. CONCLUSION

In the submitted survey, we conducted different approaches on both the centralized platform and distributed query optimization based on the usual SQL and MapReduce techniques. Cloud computing platforms must have a standalone solution to extract and release resources during work to provide a secure service to the customer. However, traditional query optimization strategies do not determine the future availability and release of resources; therefore its performance may be poor compared to the MapReduce strategy.

V. REFERENCES

[1]. D. J. Abadi, “Data Management in the Cloud: Limitations and Opportunities”, IEEE Data Eng. Bull., vol. 32, no. 1, (2009), pp. 3-12.

[2]. D. Agrawal, S. Das and A. E. Abbadi, “Big data and cloud computing: current state and future opportunities”, In Proceedings of the 14th International Conference on Extending Database Technology, ACM, (2011), pp. 530-533.

[3]. L. Zhou, K. He, X. Sheng and B. Wang, “A survey of data management system for cloud computing: models and searching methods”, Research Journal of Applied Sciences, Engineering and Technology, vol. 6, no. 2, pp. 244-248

[4]. H. Hacigümüs, J. Tatemura, W. P. Hsiung, H. J. Moon, O. Po, A. Sawires and H. Jafarpour, “CloudDB: One size fits all revived”, In Services

(SERVICES-1), 2010 6th World Congress on IEEE, (2010), pp. 148-149

[5]. T. Dokeroglu, S. A. Sert and M. S. Cinar, “Evolutionary multi-objective query workload optimization of Cloud data warehouses”, The Scientific World Journal, (2014).

[6]. P. Doshi and V. Raisinghani, “Review of dynamic query optimization strategies in distributed database”, In Electronics Computer Technology (ICECT), 2011 3rd International Conference on IEEE, vol. 6, (2011), pp. 145-149

[7]. A. Tripathy, L. Mishra, and P. K. Patra, “A query optimization strategy for implementing multidimensional model in spatial database system,” in Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, vol. 2, pp. 64–68, IEEE, 2010.

[8]. G. Chen, Y. Wu, J. Liu, G. Yang, and W. Zheng, “Optimization of sub-query processing in distributed data integration systems,” Journal of Network and Computer Applications, vol. 34, no. 4, pp. 1035–1042, 2011.

[9]. Y. N. Silva, P.-A. Larson, and J. Zhou, “Exploiting common sub expressions for cloud query processing,” in Data Engineering (ICDE), 2012 IEEE 28th International Conference on, pp. 1337–1348, IEEE, 2012

[10]. H. Zhao, S. Yang, Z. Chen, S. Jin, H. Yin, and L. Li, “Mapreduce model-based optimization of range queries,” in Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on, pp. 2487– 2492, IEEE, 2012.

Cite this article as :

Y. Supriya, K. Gayathri, "Study of Query Optimization in Cloud ", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 6 Issue 2, pp. 91-94, March-April 2019. Available at doi : https://doi.org/10.32628/IJSRST196219

Journal URL : http://ijsrst.com/IJSRST196219