

# Efficient Query Optimization in Distributed Database Using Decision Tree Algorithm

Kamaljit Singh, Dr. Dheerendra Singh  
Shaheed Udham Singh College of Engg, Tangori, Mohali, Punjab, India  
[er.kjsingh@gmail.com](mailto:er.kjsingh@gmail.com), [professordsingh@gmail.com](mailto:professordsingh@gmail.com)

## ABSTRACT

This paper presents semantics query optimization on distributed database using Decision Tree algorithm based on that. Whenever user query from server the master sever can process the query of user from different **server's** database in Heterogeneous environment on alternative basis which will reduce the time of processing query and less time of response time of user. we will use all query hints based on that whenever user query anything from server our master sever will make SQL query hints to generate more than query using specified rules like query of 2 or more tables query of 1 tables and get required data from database and revert back to user screen in Table format.

**Keywords:** query optimization, decision tree algo, data mining.

## 1. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

This white paper provides an introduction to the basic technologies of data mining. Examples of profitable applications illustrate its relevance to today's business environment as well as a basic description of how data warehouse architectures can evolve to deliver the value of data mining to end users.

## 2. QUERY OPTIMIZATION

It is a function of many relational database management systems. The **query optimizer** attempts to determine the most efficient way to execute a given query by considering the possible query plans. Generally, the query optimizer cannot be accessed directly by users:

once queries are submitted to database server, and parsed by the parser, they are then passed to the query optimizer where

optimization occurs. However, some database engines allow guiding the query optimizer with hints. A query is a request for information from a database. It can be as simple as "finding the address of a person with SS# 123-45-6789," or more complex like "finding the average salary of all the employed married men in California between the ages 30 to 39, that earn less than their wives." Queries results are generated by accessing relevant database data and manipulating it in a way that yields the requested information. Since database structures are complex, in most cases, and especially for not-very-simple queries, the needed data for a query can be collected from a database by accessing it in different ways, through different data-structures, and in different orders. Each different way typically requires different processing time. Processing times of the same query may have large variance, from a fraction of a second to hours, depending on the way selected. The purpose of query optimization, which is an automated process, is to find the way to process a given query in minimum time. The large possible variance in time justifies performing query optimization, though finding the exact optimal way to execute a query, among all possibilities, is typically very complex, time consuming by itself, may be too costly, and often practically impossible. Thus query optimization typically tries to approximate the optimum by comparing several common-sense alternatives to provide in a reasonable time a "good enough" plan which typically does not deviate much from the best possible result.

## 3. METHODOLOGY

In this whole Research work concept we will use Network programming Tcp/ip Sockets to show interaction of master sever with another sever who process query also the communication of Client Machine to Master Master.

We will use Multithreading concept of java based on that we will implement master server that can listen the query of more than user simultaneously each Thread can handle query of each user We will use mysql database on each server having window operating system for Homogenous environment.

#### 4. OBJECTIVES

1. To process the query of users by more than one server each server handle one query of user one by one.
2. Handle traffic of more than queries of user based on that bandwidth of server can utilise efficiently.
3. Anylisis and check the performance of user query response timing in to network.
4. Reduce the user query processing time

#### 5. ACKNOWLEDGMENTS

The author is highly grateful to Dr. Manjit Singh Grewal, Director Shaheed Udham Singh College of Engineering and Technology (SUSCET), Tangori (Mohali), for providing this opportunity to carry out the present thesis work. I would like to express a deep sense of gratitude and thanks profusely to my Supervisor Dr. Dheerendra Singh, Professor and Head, Department of Computer Science & Engineering,. Without the wise counsel and able guidance, it would have been impossible to complete the thesis in this manner. I express gratitude to other faculty members of Department of CSE, for their intellectual support throughout the course of this work.

#### 5. REFERENCES

- [1] Preeti Aggarwal, **M. M. Chaturvedi** "Application of Data Mining Techniques for Information Security in a Cloud: A Survey", "International Journal of Computer Applications (0975 – 8887) Volume 80 – No 13, October **2013**".
- [2] K. Srinivasa Rao\* & B. Srinivasa Rao, "The Standard International Journals (The SIJ)", ISSN: 2321 – 2381", Vol. 1, No. 3, July-August 2013".
- [3] **S.Hameetha Begum** "Data Mining Tools and Trends – An Overview", "International Journal of Emerging Research in Management & Technology ISSN: 2278-9359", February 2013".
- [4] Manjunath T. N., Ravindra S. Hegadi, Umesh I. M., and **Ravikumar G. K.** "Realistic Analysis of Data Warehousing and Data Mining **Application in Education Domain**", "International Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012"
- [5] Periakaruppan Sudhakaran1, Shanmugasundaram **Hariharan2 and Joan Lu3**, "Research Directions, Challenges and Issues in Opinion Mining", "International Journal of Advanced Science and Technology Vol.60, (2013), pp.1-8 <http://dx.doi.org/10.14257/ijast.2013.60.01>".
- [6] **N. Anitha, B. Anitha, S. Pradeepa**, "Sentiment Classification Approaches – A Review", "International Journal of Innovations in Engineering and Technology (IJJET)" ISSN: **2319 – 1058**", Vol. 3 Issue 1 October **2013**"
- [7] Padmapani P. Tribhuvan , S.G. Bhirud , Amrapali P. Tribhuvan., "A Peer Review of Feature Based Opinion Mining and Summarization", Padmapani P. Tribhuvan etal, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 247-250"
- [8] Vibha Soni, **Meenakshi R Patel**, "Unsupervised Opinion Mining From Text Reviews Using **SentiWordNet**", "International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 5 –May 2014"
- [9] **Jayashri Khairnar\***, **Mayura Kinikar**, "Machine Learning

- Algorithms for Opinion Mining and Sentiment **Classification**", "International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 ISSN 2250-3153"
- [10] Erik Cambria, National University of Singapore, **Bjrn Schuller**, Technical University of Munich, Yunqing Xia, Tsinghua University, Catherine Havasi, **Massachusetts Institute of Technology**, "New Avenues in Opinion Mining and Sentiment Analysis", March/April 2013 1541-1672/13/\$31.00 © 2013 IEEE 15 Published by the **IEEE Computer Society**"
  - [11] Grigori Sidorov1, Sabino Miranda-Jiménez1, Francisco Viveros-Jiménez1, Alexander Gelbukh1, Noé Castro-Sánchez1, Francisco Velásquez1, Ismael Díaz-Rangel1, Sergio Suárez- **Guerra1, Alejandro Treviño2, and Juan Gordon**, "Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets",
  - [12] Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2, 1–135 (2008)
  - [13] Liu, B.: Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing. Ed(s) Indurkha, N. and Damerau, F.J., 2nd ed. (2010)
  - [14] Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the ACL-2004 (2004)
  - [15] Pang, B., Lee, L., and Vaithyanathan S.: Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL, pp. 79–86. Association for Computational Linguistics (2002) [16] Polanya, L., Zaenen, A.: Contextual Valence Shifters. Computing Attitude and Affect in Text: Theory and Applications In Computing Attitude and Affect in Text: Theory and Applications, Vol. 20 (2006)
  - [17] Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: Opinion Finder: a system for subjectivity analysis, Proceedings of HLT/EMNLP on Interactive Demonstrations, pp. 34–35, Vancouver, British Columbia, Canada (2005)
  - [18] Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford University, Stanford, CA. (2009)
  - [19] Martínez Cámara, E., Martín Valdivia, M.T., Perea Ortega, J.M., Ureña López, L.A.: Técnicas de Clasificación de Opiniones Aplicadas a un Corpus en Español. Procesamiento del Lenguaje Natural, Revista n° 47, pp. 163–170 (2011)
  - [20] Aiala, R., Wonsever, D., Jean-Luc, M.: Opinion Identification in Spanish Texts. Proceedings of the NAACL HLT (2010)
  - [21] Padró, L., Collado, M., Reese, S., Lloberes, M., Castellón, I.: FreeLing 2.1: Five Years of Open-Source Language Processing Tools. Proceedings of 7th Language Resources and Evaluation Conference, La Valletta, Malta (2010)
  - [22] Das, S., Chen, M.: Yahoo! For Amazon: Extracting market sentiment from stock message boards. In Proceedings of the 8th Asia Pacific Finance Association Annual Conference (2001)
  - [23] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1 (2009)
  - [24] EAGLES. Recommendations for the morphosyntactic annotation of corpora, Eag-tcwgmac/r, ILC-CNR, Pisa (1996)
  - [25] Esuli, A., Sebastiani, F.: SentiWN: A Publicly Available

Lexical Resource for Opinion Mining. In: Fifth international conference on Language Resources and Evaluation (LREC2006), pp. 417–422 (2006) [26]Díaz-Rangel, I., Sidorov, G., Suárez-Guerra, S.: Weighted Spanish Emotion Lexicon.(submitted) (2012)

[27]Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, pp. 37–46 (1960).

[28] Jalaj S. Modha,Prof & Head Gayatri S. Pandi Sandip**J. Modha** ,” Automatic Sentiment Analysis for Unstructured Data”,” Volume 3, Issue 12, December 2013 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering”