

Implementation of ARM in Distributed DB

Prathima V.R¹, Dr. Fayaz.K²

Research Scholar, Department of CSE, Rayalaseema University, Kurnool, Andhra Pradesh, India¹

Professor, Department of CSE, S.K.University, Ananthpur, Andhra Pradesh, India²

ABSTRACT: Data mining and machine learning has been an important research topic in recent years. ARM is an hot data mining research area and most ARM algorithms works for centralized environment. However, changing centralized data mining to analyse and discover meaningful patterns in distributed database isn't always feasible due to the fact of merging data sets from different sites produces huge network communication costs. Therefore, our research is to develop a distributed algorithm for universally distributed data sets that reduces communication costs. Our proposed system uses EDMA algorithm where the databases are distributed over different sites in the network. By sending and receiving of these transaction databases it reduces communication costs and time of scan partition database. In our paper we propose an implementation of association rule mining algorithm for distributed databases. The performance analysis reveals that EDMA has good running efficiency, lower communication cost and better scalability.

KEYWORDS: ARM, distributed algorithm, EDMA, distributed database, scalability.

I. INTRODUCTION

KDD is the process of automatically searching massive volumes of data for patterns using primary tools such as classification, association rule mining, clustering and so on. Data mining is really a complex topic and has its association with multiple core fields such as computer science and adds special value to rich seminal computational techniques ranging from statistics, information retrieval, machine learning and pattern recognition.

Data mining techniques are the result of a continual process of research and product development. Historical data depicts about the beginning of the business data being first stored on computers with continued improvements in data access, and more recently being used technologies allows users to navigate through their data in real time. Data mining takes this evolutionary process beyond the data access and navigation to proactive information delivery. Data mining is ready for application in the business community due to the support of three technologies that are now sufficiently mature, i.e., peta bytes of data collection, Powerful and effective multiprocessor computers and data mining algorithms. Data mining techniques can be implemented on already existing software and hardware platforms to enhance the value of existing information resources. Upon implementation on high performance client/server or parallel processing computers, data mining tools can analyse larger databases to deliver answers to lots of organizational queries.

Now a day's lot of commercial databases are growing at infinite rates. A recent survey carried out by META Group revealed that 40% of respondents are beyond the peta byte level, while 60% expect to be there by second quarter of 2016. Some of the industries like retail, these numbers can be much larger. The real need for improved computational engines can now be utilized in a cost-effective manner with parallel multiprocessor computer technology with distributed scenario. Data mining algorithms includes techniques that have existed for at least few years, but have only recently been implemented as scalable, distributed, mature, reliable and understandable tools that consistently outperform older statistical methods.

II. RELATED WORK

A. EXISTING METHOD

Many existing parallel 7 of communication, obtaining a good data layout, decomposition of data and disk I/O minimization. In case of distributed mining, synchronization is implicit with respect to message passing, hence the major goal lies in communication optimization.

The major challenge for obtaining quality performance on distributed mining is to find a good data decomposition among various nodes for good load balancing and to minimize overall communication.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

Distributed ARM algorithms aims to generate association rules from different data sets spread over various universal sites, hence they require external communications throughout the entire process.

B. PROPOSED SYSTEM

In our work we have considered the distributed database as an horizontally partitioned database where the database schema of all the partitions are considered to be the same. The distributions of the item sets in different partitions are always not identical and many items occur more frequently in some partitions than the others. As a result, many described item sets may be huge locally at some sites but not necessarily in the other sites. This skewness property poses a new requirement in the design of mining algorithm. The proposed system aims to implement the EDMA algorithm. EDMA works on an incremental approach towards data warehouse through data mart development by creating a shared framework for development. The EDMA frame work has in its scope the enterprise subject areas, dimensions, metrics, business rules, and possible data sources, all represented in a logically common Global Metadata Repository (GMR). EDMA is incrementally modified as the new development process gradually evolves the foundation for the data warehouse.

III. SYSTEM ARCHITECTURE

According to the proposed algorithm, our algorithmic framework is splitted with four modules namely administrator, customer, EDMA-DBMS and Server. The work flow details are explained below and the same is depicted in figure 1.

Administrator Module: The Capabilities of the administrator module is given below

- 1) In this module administrator can participate in each transactions and maintaining of the databases effectively.
- 2) The administrator must be authorized one to Login to the system with user name and password.
- 3) The total database can be managed by the administrator and all the customer transactions must be updated to the database.

Customer Module: The functionalities of the customer module are

- 1) Selecting the products available in the catalog and request to the administrator to process the purchase of the product.
- 2) Each customer can have their own user id otherwise they must register.

EDMA DBMS: Here the database can store all the customer details and their transactions.

- 1) Database can store all the customer details and their transactions.
- 2) The system can generate frequent transactions at the each site by using EDMA algorithm with individual support and threshold of the sites.
- 3) To distribute these frequent transactions to the different sites when they need.

Server

- 1) Get the Request from Server.
- 2) Processes each client request process one by one.
- 3) Associates synchronization method via process all the client request.

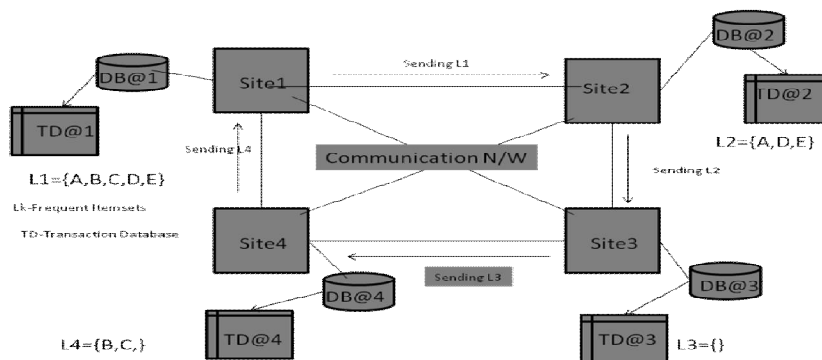


Figure 1: System Architecture

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

IV. IMPLEMENTATION RESULTS

We have used My Eclipse and rational rose as a tool to implement the algorithm with the web server as tomcat server 6.0 .For our front end design HTML, Javascript and JSP were incorporated.. The below figure 2, 3, 4, 5, 6 shows the implementation results. We used java language with windows environment to test our experimental result

Snapshot 1:

The purpose of this screen is the login of an administrator with their username and password.



Figure 2: Administrator Login Page

Snapshot 2:

The customer can select the products from catalog, and then the administrator can fill the details as shown in the screen. After click on submit button.



Figure 3: Purchase order Details

Snapshot 3:

This screen purpose is the customer is unregistered then registered. Fill the following Customer Registration Form then click on submit.

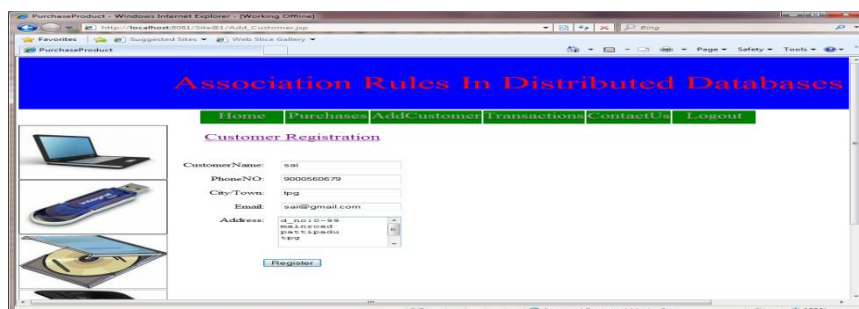


Figure 4: Add New Customer Page

Snapshot 4:

The purpose of the screen is

- Display total transactions of each customer.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

- Display total frequently purchases products.
- It display a link, when click on the link it can receive frequent transactions at different sites.

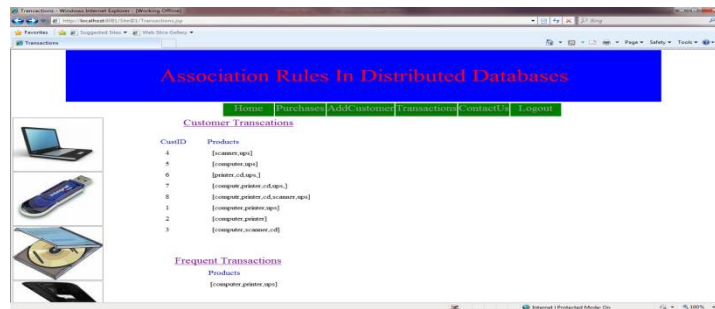


Figure 5: Customer Transactions Page

Snapshot 5:

- The purpose of this screen is to display the receiving Frequent Transactions at different sites.
- Also displays transactions.

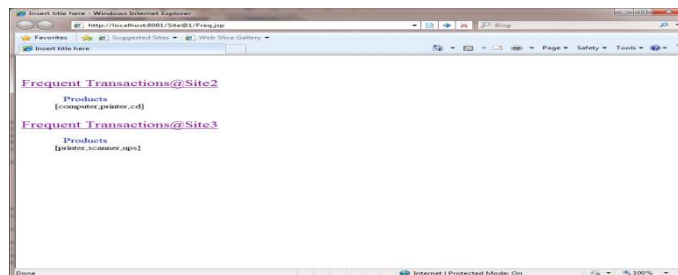


Figure 6: Receiving Frequent Transactions Screen

V. CONCLUSION

In our discussed paper, we implemented an efficient EDMA algorithm. EDMA generates fewer messages than FDM, so we can conclude EDMA is more scalable, also as per our investigation EDMA generates fewer messages than FDM, especially when the number of participating sites was five, as in the previous experiment. On the other hand, CD broadcasts its candidate support counts to all other sites and subsequently receives support counts of others, broadcasting numerous messages when we increase number of sites. However there are certain limitations to our system i.e Only the permanent administrator can access the system, System works in all platforms and its compatible environments, advanced techniques are not used to check the authorization.

REFERENCES

- [1] A. Schuster and R. Wolff , "Communication-Efficient Distributed Mining of Association Rules," Proc. ACM SIGMOD Int'l Conf. Management of Data,ACM Press, 2001,pp. 473-484
- [2] D.W. Cheung, et al., "A Fast Distributed Algorithm for Mining Association Rules," Proc. Parallel and Distributed Information Systems, IEEE CS Press,1996,pp. 31-42;
- [3] Kimball, R., Ross, M.: The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling. 2nd edn. John Wiley & Sons, New York (2002)
- [4] M.J. Zaki and Y. Pin , "Introduction: Recent Developments in Parallel and Distributed Data Mining," J. Distributed and Parallel Databases , vol. 11, no. 2, 2002,pp. 123-127.
- [5] Ma, Y., Liu, B., Wong, C.K.: Web for Data Mining: Organizing and Interpreting the Discovered Rules Using the Web. SIGKDD Explorations, Vol.2 (1). ACM Press, New York (2000) 16-23
- [6]M. Z. Ashra_, D. Taniar, and K. A. Smith, "A Data Mining Architecture for Distributed Environments", IICS 2002, Vol 2, pages 27-38, 2002.
- [7] Albert Y. Zomaya, Tarek El-Ghazawi, Ophir Frieder, "Parallel and Distributed Computing for Data Mining", IEEE Concurrency, 1999.



ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

International Journal of Database Theory and Application Vol. 4, No. 3, September, 2011 88

[8] The Grid Security Infrastructure. The Globus Project. [Online]. Available: <http://www.globus.org/security>.(last accessed on 23.2.2011).

[9] The Globus Resource Specification Language RSL v1.0. The Globus Project. (available Online at: http://www.globus.org/gram/rsl_spec1.html)

[10] M.Cannataro and D. Talia, "KNOWLEDGE GRID Architecture for Distributed Knowledge Discovery", CACM, Vol. 46, No. 1, pp. 89-93, 2003.