Hardware Improved Association Rule Mining with Hashing and Pipelining Model

K. Uthaya Kumar[1], S. Keerthana[2], L. Suba Shree[3], Dr. M. Thangamani[4]

[1, 2, 3] Student, Kongu Engineering College, Tamilnadu, India. Email: k.uthayakumar03@gmail.com

[4]Assitant Professor, Dept. of CSE, Kongu Engineering College, Tamilnadu, India. Email: manithangamani2@gmail.com

Abstract - The Apriori-based association rule mining in hardware, one has to load candidate item sets and a database into the hardware. Since the capacity of the hardware architecture is fixed, the number of candidate item sets or the number of items in the database is larger than the hardware capacity, the items are loaded into the hardware separately. The time complexity of those steps that need to load candidate item sets or database items into the hardware is in proportion to the number of candidate item sets multiplied by the number of items in the database. Too many candidate item sets and a large database would create a performance of bottleneck. Hash-based and Pipelined (abbreviated as HAPPI) architecture for hardware enhanced association rule mining. Significantly outperforms the previous hardware approach and the software algorithm in terms of execution time.

Keywords – Pipeline, hash table, Association rule mining, hash-based architecture and candidate item sets.

I. INTRODUCTION

Hash Table is a data structure which stores data in an associative manner. Hash table, stores the data in an array, each data value has its own distinctive index value. Access of data becomes very fast if we know the index of the preferred data. Pipelining is the process of accumulating instruction from the processor through a pipeline. It allows accumulating and performing instructions in an ordered process. Pipelining is a method where multiple instructions are extend beyond during execution. Association rule mining finds a interesting correlation relationships among a large set of data items[1-3]. It first determines repeated item sets satisfying user-defined minimum support, and then from which produces strong association rules to satisfy user defined minimum assurance.

II. LITERATURE REVIEW

Preethi et al.[4] used hash table filter to decrease the candidate item sets and also detect the duplicate records in the large files. Hashing And Pipelining solves the Performance bottleneck problem and acquires good scalability in [5]. In hardware architecture, it use the hash table filters to reduce the number of candidates item sets [6]. Another hardware module used is trimming filter. Trimming filter is used to reduce the items from each transaction. The items which are having the minimum support count are trimmed from the transaction. An efficient algorithm for mining association rules [7] that is faster than the previously proposed partition algorithms approximately m times where m is the number of stages in pipeline. Firstly items are kept [8] in systolic array then items which are not in close proximity with each other are trimmed or removed from the filter then put into hash table filter so that duplication of items get removed so in this way. It solves our bottleneck problem.

III.MATERIALS AND METHODS

In the proposed system contain two themes in first hash-based system received the data and accumulated into hardware and it is compared with systolic array. It is maintained in hash table. The secondly pipeline system pruning process for selecting candidate item set. Implementation is the state where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and giving confidence on the new system for the users that will work efficiently and effectively. The system is implemented only after thorough testing is done and if it is found to work according to the specification. Implementation involves careful planning, investigation of the current system and is constraints on implementation, design of methods to achieve changeover, and evaluation of the changeover methods apart from planning. Two major tasks for preparing the implementation are educating, training the users and testing the system.

Implementation Plan Preparation: The implementation process begins with the preparation of plan for implementation. According to this plan other activities are carried out. A plan discussion has been made regarding the equipment, resources and how to test the activities a clear planner prepared for the activities.

Equipment Acquisition: According to the above plan the necessary equipment have to be acquired to implement the new system, which would include all the requirements for installing and maintaining .Net framework, VB.net, SQL server,

Program Code Preparation: One of the most important development activities is coding or programming. The system flowcharts and other charts are converted into modular programs. They have to be compiled, tested and debugged.

User Training and Documentation: Once the planning has been completed the major effort in the computer department is that the user department must consist of educated and trained staff as the system becomes more complex. The success of the system depends upon how they are operated and used the system.

The quality of training the personnel is connected to the success of the system. Implementation depends upon the right people being trained at the right time. Education involves creating the right atmosphere and motivating the user. Staff education should encourage the participation of all the staff.

Changeover: Changeover is the change of moving over from the old system to the new computerized system that is done all the files have to be converted to the new format. The accuracy of the conversion is of utmost importance both to user confidence in the system and to effective operation. When the files have been set up on the computer, the changeover can take place. There are several possible methods of doing this. E.g. direct changeover, parallel running, pilot running, and staged changeover.

Direct Changeover: The direct changeover is the complete replacement of the old system by new, in one move. When direct changeover is planned, system tests and training should be comprehensive and changeover itself is planned in detail.

Parallel Running: Parallel running or operation means processing current data by both the old and new systems to cross check the results. he old system is kept alive and operational until the system has been proved for at least one system cycle, using full live data in the operational environment of place, people, equipment and time. It allows the result of the new system to be compared with the old system before the acceptance by the user. Parallel operation does not allow much time or learning and testing activities.

Staged Changeover:A staged changeover involves a series of limited size direct changeovers. The new system being introduced piece by piece. A complete start, a logical section is committed to the new system while the remaining parts or sections will be processed by the old system. The direct changeover is applied where the entire system is implemented directly after it has been developed.

IV. RESULTS AND DISCUSSION

System attributes are prodname, prodprice, Quantity, category and Dealer. The proposed system get the all attribute values from the user for performance analysis. It is shown in Figure 1. Figure 2 represent the attribute values as in the database format. All completed transaction can viewed by Figure 3. Proposed system perform pattern matching using association rule and collect meaningful data to reduce the number of candidate item sets and items from customer database concurrently.

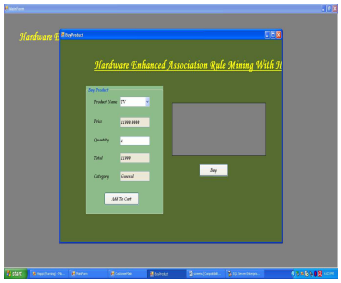

Fig.1.Product add to the cart

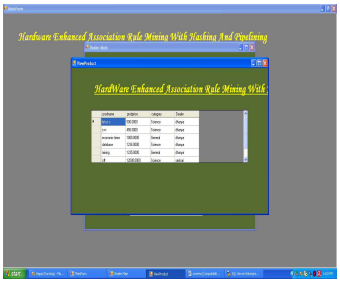

Fig.2.User input view

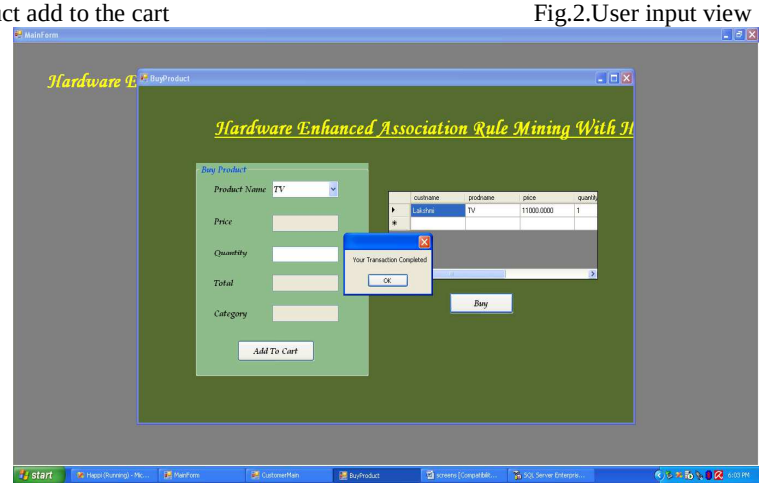

Fig.3.Hardware enhanced hashing and pipelining

V. CONCLUSION

The HAPPI architecture for hardware-enhanced association rule mining. The bottleneck of a priori-based hardware schemes is related to the number of candidate item sets and the size of the database. To solve the problem, to apply the pipeline methodology in the HAPPI architecture to compare item sets with the database and collect useful information to reduce the number of candidate item sets and items in the database simultaneously. HAPPI can prune infrequent items in the transactions and reduce the size of the database gradually by utilizing the trimming filter, HAPPI can effectively eliminate infrequent candidate item sets with the help of the hash table filter. Every application has its own merits and demerits. The project has covered almost all the requirements. Further requirements and improvements can easily be done since the coding is mainly structured or modular in nature. Changing the existing modules or adding new modules can append improvements. Further enhancements can be made to the application, so that the future enhancement is to develop the application through website and useful manner than the present one and the future work, are going to increase the clock frequency of the hardware architecture will try to optimize the bottleneck module .

REFERENCES

[1] Z.K. Baker and V.K. Prasanna, "Efficient Hardware Data Miningwith the Apriori Algorithm on FPGAS," Proc. 13th Ann. IEEESymp. Field-Programmable Custom Computing Machines (FCCM),2005.

[2] M. Hahsler, C. Buchta Hornik, "Selective Association Rule Generation," Computational Statistic, vol. 23, no. 2, pp. 303-315, Kluwer Academic Publishers 2008

[3] Xiaoqi Gu, Yongxin Zhu, Shengyan Zhou, Chaojun Wang, Meikang Qiu, and Guoxing Wang. 2016. A Real-Time FPGA-Based Accelerator for ECG Analysis and Diagnosis Using Association-Rule Mining. ACM Transactions on Embedded Computing Systems 15, 2 (2016), 25.

[4] V.B. Nikam and B.B. Meshram. 2014. Scalable Frequent Itemset Mining using Heterogeneous Computing: ParAprioriAlgorithm. International Journal of Distributed and Parallel Systems 5, 5 (2014), 13.

[5] Ke Sun and Fengshan Bai, Mining weighted association rules without preassigned weights, *IEEE Transactions on Knowledge and Data Engineering*, (20) 4 (2008), pp. 489-495.

[6] M.Preethi, P.K.Mangaiyarkarasi, "Hardware Magnified Association Rule Mining for Hash Table Filter using MD5 Algorithm", International Journal of Advanced Research in Computer Engineering & Technology, voi.6, no.12, 2017.

[7] Mamatha Nadikota, Satya P Kumar Somayajula,Dr. C. P. V. N. J. Mohan Rao" Hashing and Pipelining Techniques for Association Rule Mining", Mamatha Nadikota et al, / (IJCSIT)International Journal of Computer Science and Information Technologies,vol.2,no,4,2011.

[8] Mr. Praveen S Patil ," Hardware-Enhanced Association Rule Mining with Hashing and Pipelining", International Journal of Combined Research and Development Vol. 1, no. 1,

[9] Kanhaiya Lal, N.C.Mahanti," Mining Association Rules in Large Database by Implementing Pipelining Technique in Partition Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.4, June 2010.

[10] Krishan Rohilla1, Shabnam Kumari2, Reema3," Data Mining based on Hashing Technique", International Journal of Trend in Scientific Research and Development,vol.1,no.4, May-Jun 2017.