Jigna J. Jadav, Mahesh Panchal / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 2,Mar-Apr 2012, pp.1147-1151

Association Rule Mining Method On OLAP Cube

Jigna J. Jadav*, Mahesh Panchal**
*(PG-CSE Student, Department of Computer Engineering,
Kalol Institute of Technology & Research Centre, Gujarat, India)
** (Head and Associate Professor, Department of Computer Engineering,
Kalol Institute of Technology & Research Centre, Gujarat, India)

ABSTRACT
Data mining is use to discover a knowledge and hidden pattern from data[15]. OLAP (Online Analytical Processing) is a tool of data mining and data warehouse that performs different operations on data that is store in multidimensional database but the limitation of OLAP is that it is not capable to explain relationships between data that resides in data cube. So that's why OLAM is used it is also known as OLAP mining that takes advantages of both OLAP and data mining and gives accurate results so it is consider as a business intelligence (BI) .different mining techniques are there that can be apply on OLAP cube in paper we will how to apply association rules mining method on data cube.

Keywords – **OLAM, OLAP cube, data mining, multidimensional database ,association rules mining**

I. INTRODUCTION
Data mining is the principle of picking out relevant information from data. It is usually used by business intelligence organizations, and financial analysts, to extract useful information from large data sets or databases Data mining is use to derive patterns and trends that exist in data. Typically, these patterns cannot be discovered by traditional data exploration. The goal of this technique is to find accurate patterns that were previously not known by us. Once these patterns are found they can further be used to make certain profitable decisions for development of their businesses in future. Organizations like retail stores, hospitals, banks, and insurance companies currently using mining techniques.

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail data. This paper represents different association rules from multidimensional data, what is OLAP technology and its operations different types of OLAP, how data is stored in fact table and dimensional tables, What are the measures, what is the concept hierarchy, what is OLAM and advantages of it, how to find frequent patterns from data cube with the use of association technique on data cube to get different accurate frequent patterns.

II. OLAP(ONLINE ANALYTICAL PROCESSING)

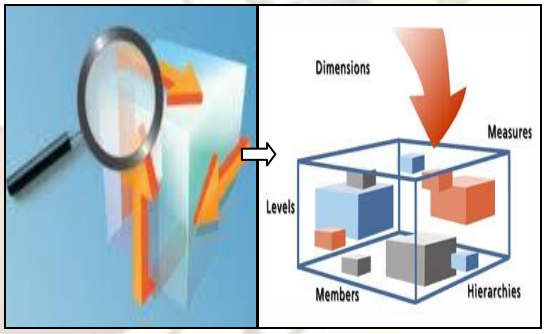


Figure 1.1 features of OLAP

OLAP -online analytical processing is the tool of data warehouse which gives different perspectives of data ,perform different operations on data .OLAP stores summarized data according to numeric measure attribute..Certain attributes the attributes whose values are of our interest like count, sales, budget, salary etc.

Main types of OLAP are discovered.
1. ROLAP 2.MOLAP 3.HOLAP

ROLAP-Relational online analytical processing

It is the fastest growing OLAP technology style, data resides in a relational database, where the base data

and dimension tables are stored as relational tables. This model permits multidimensional analysis of data as this enables users to perform a function equivalent to that of the traditional OLAP slicing and dicing features.

MOLAP- multidimensional online analytical processing

this is the more traditional way of OLAP analysis. In MOLAP, data is stored in a multidimensional cube. Instead of in the relational database.

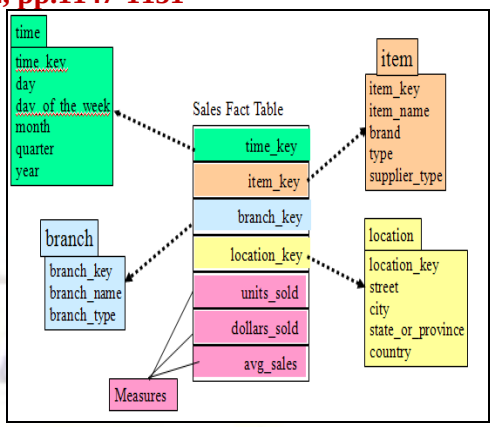
HOLAP –Hybrid online analytical processing

HOLAP is mixture of MOLAP and ROLAP into a single architecture. This tool tried to bridge the technology gap of both products by enabling access or use to both multidimensional database (MDDB) and Relational Database Management System (RDBMS) data stores. HOLAP systems stores larger quantities of detailed data in the relational tables while the aggregations are stored in the pre-calculated cubes. the advantages of this system are better scalability, quick data processing and flexibility in accessing of data sources.

III. MULTIDIMENSIONAL DATABASE

A multidimensional database is a part of OLAP to allow for the efficient and convenient storage and retrieval of large volumes of data that is (1) related with each other(2) stored, viewed and analyzed from different perspectives. These perspectives are called dimensions.

Dimensional table contains all dimensions and their different attributes. Fact table contains measure that is numeric aggregated value of a particular dimensions and keys of dimensions table that explain relationships BETWEEN dimensions with no redundancy. Star schema contains fact table, dimension table.



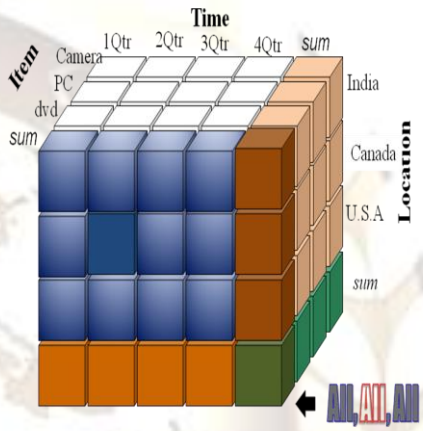


FIGURE 1.2 MULTIDIMENSIONAL DATABASE AND DATA CUBE[15]

IV. OLAM: ONLINE ANALYTICAL MINING

OLAM that is online analytical mining that apply mining technique on OLAP cube that has cleaned data, so it will give more accurate data and with different OLAP operations.

On-Line Analytical Mining (OLAM) (also called OLAP mining), which integrates on-line analytical processing (OLAP) with data mining and mining knowledge in multi-dimensional databases[15].

OLAM system will integrate OLAP and data mining and mine various kinds of knowledge from data warehouses, it is important to develop a variety of knowledge and data visualization tools.

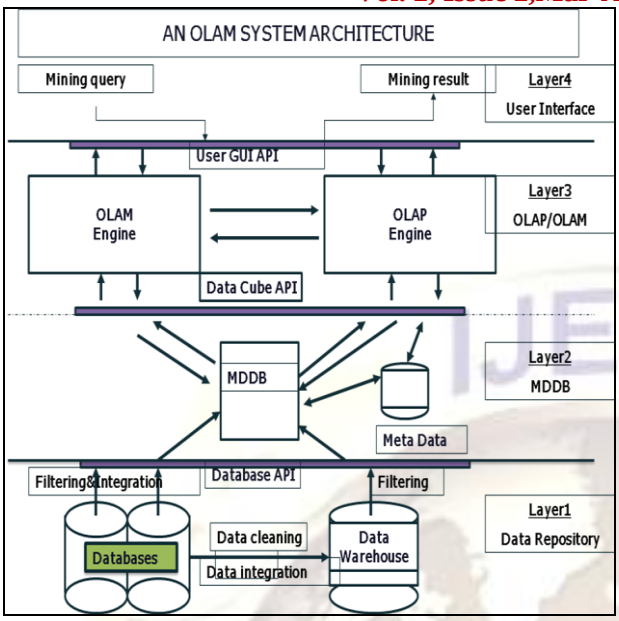


FIGURE 1.3 OLAM ARCHITECTURE[15]

V.ASSOCIATION RULE MINING

Association rule is a one kind of technique of data mining used to discover interesting patterns and correlations between data. Frequent itemsets or frequent patterns, as the name suggests, are patterns that occur frequently in data. Association is a descriptive approach to exploring data that can help identify relationships among values in a database. Association finds rules about items that appear together in an event such as a purchase transaction Two measures are used

Support of an association rule is defined as the percentage of records that contain X U Y to the total number of records in the database.

$$support(X->Y)=\frac{no\_of\_tuples\_containing\_both\_X\_and\_Y}{total\_no\_of\_tuples}$$

Confidence of an association rule is defined as the percentage of the number of transactions that contain XUY to the total number of records that contain X, where if the percentage exceeds the threshold of confidence an interesting association rule X=>Y can be generated.

$$confidence(X->Y)=\frac{no\_of\_tuples\_containing\_both\_X\_and\_Y}{no\_of\_tuples\_containing\_X}$$

V.ASSOCIATION RULES APPLY ON OLAP CUBE[16].

Three dimension tables are STUDENT, COLLEGE and ZONE all have primary key and different attributes of different dimensions. Use UNIVERSITY table as a fact table that contains all the reference keys of dimension tables and measures according to which data will be store in to data cube. With the help of UNIVERSITY relational database and SQL SERVER 2005 analysis services make a OLAP cube as display in fig1.4.here star schema is there that has one fact table and others are dimensions table.

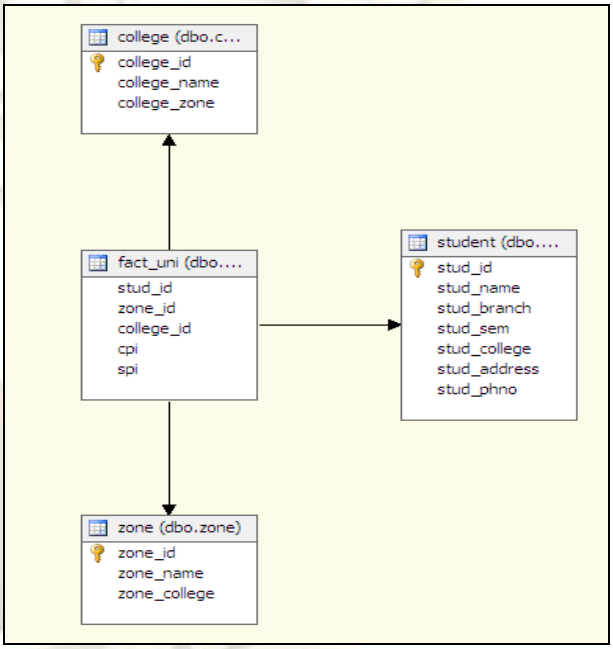


Figure 1.4 multidimensional database display with star schema

Browse the cube according to different dimensions and in OLAP cube data will be stored according to measure attribute that is a numeric attribute here CPI and SPI are consider as a measure attributes. In fig 1.5. OLAP cube is generated and different attributes in dimensions are according to student name, branch semester, college name etc.

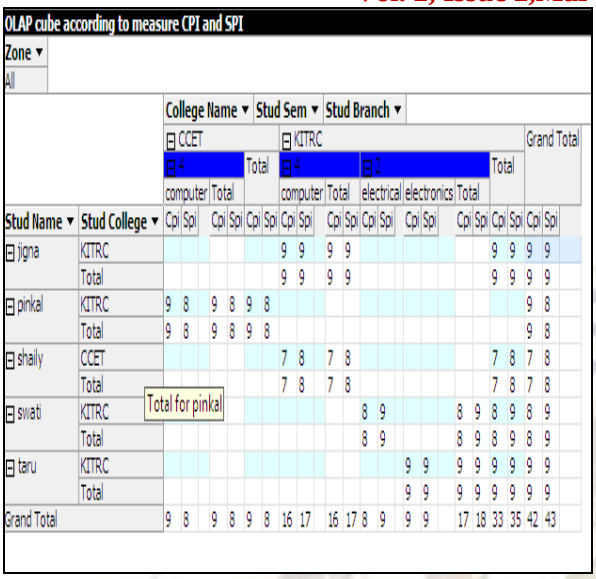


Figure 1.5 Browsing OLAP cube according to measure

Apply mining structure on OLAP cube and according to that apply association rule mining technique on data cube. It will gives the different frequent items and rules from the data cube.

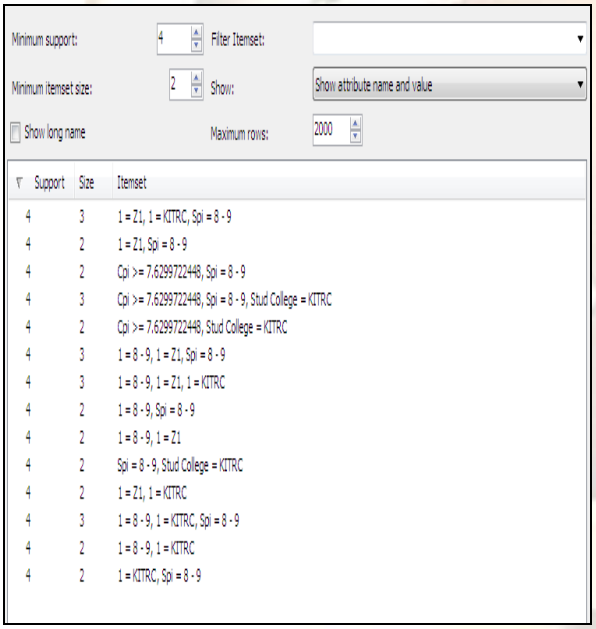


Figure 1.6 frequent items according to MIN_SUPPORT from OLAP cube

In FIG 1.6 frequent items are found according to min_sup=4

And with help of probability rules are generated in Fig 1.7. .rules represents that how many frequent items are associated with each other.

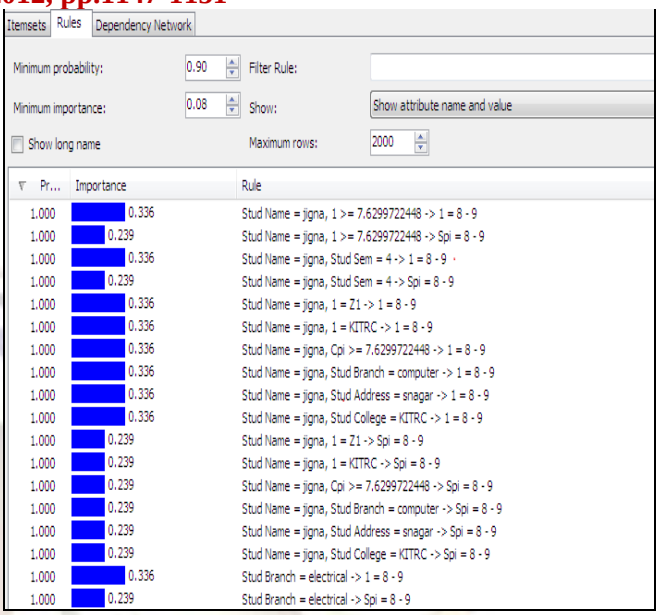


Figure 1.7 Association rules from university cube

CONCLUSION

OLAP defines different operations on data cube but it cannot give the relationship between data so that's why different authors makes research on it and combine OLAP with data mining techniques is also called OLAP mining. This paper presents OLAM that uses association rule mining method on OLAP cubes. That gives frequent items and rules from OLAP cubes. Other mining methods can also apply and gets advantage of OLAM.

REFERENCES

1. Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1993)*, Washington, D.C., USA, May,(Pg 207-216).
2. Kamber, M., J. Han, J. Chiang. Using *"Data Cubes for Meta rule-Guided Mining of Multi-Dimensional Association Rules*", Technical Report, CMPT–TR–97–10, School of Computing Sciences, Simon Fraser University, 1997.
3. Micheline Kamber , Jiawei Han Jenny , Y. Chiang , Jiawei Han , Jenny Y. Chiang .*"Metarule-Guided Mining of Multi-Dimensional Association Rules Using Data Cubes"* 3rd Int. Conf. Knowledge Discovery and Data Mining 1997

4. Chaudhari. S., & Dayal, U. (1997). *An Overview of Data Warehousing and OLAP*.
5. Technology. *SIGMOD Record*, *26*(1), 65-74.Zhu, H. (1998). *On-Line Analytical Mining of Association Rules*. Master's thesis, Simon Faster University, Burnaby, British Columbia, Canada, December. *Discovery (DaWaK 2000)*, London, UK, September, (pp 210-223).
7. Dong, G., Han, H., Lam, J. M. W., Pei, J., & Wang, K. (2001). Mining Multi-Dimensional Constrained Gradients in Data Cubes. In *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB 2001)*, Rome, Italy, September, (pg.321-330)
8. Zaki, M. J., & Hsiao, C. J. (2002). CHARM: An Efficient Algorithm for Closed Item set Mining. In *Proceeding of the 2nd SIAM International Conference on Data Mining (SDM'02)*, Arlington, VA, USA, April.
9. mieliński, T., Khachiyan, L., & Abdulghani, A. (2002). Cubegrades: Generalizing Association rules. *Data Mining and Knowledge Discovery*, *6*(3), 219-258
10. Nestorov, S., & Jukić. N. (2003). Ad-Hoc Association-Rule Mining within the DataWarehouse. In *Proceedings of the 36th Hawaii International Conference on SystemSciences (HICSS 2003)*, Big Island, Hawaii, USA, January, (pp 232-242).
11. Galina Bogdanova, Tsvetanka Georgieva"*Discovering the Association Rules in OLAP Data Cube with Daily Downloads of Folklore Materials*" International Conference on Computer systems And technologies CompSysTech' 2005.
12. Tjioe, H. C., & Taniar D. (2005). Mining Association Rules in Data Warehouses. *International Journal of Data Warehousing and Mining*, 1(3), 28-62
13. Ben Messaoud, R., Boussaid, O., & Loudcher, R. S. (2006a). Efficient Multidimensional DataRepresentation Based on Multiple Correspondence Analysis. In *Proceedings. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006)*, Philadelphia, USA, August, (pp 662-667).
14. Ben Messaoud, R., Loudcher, R. S., Boussaid, O., & Missaoui, R. (2006). *Enhanced Mining of Association Rules from Data Cubes*, In *Proceedings of the 9th ACM International Workshop on Data Warehousing and OLAP (DOLAP 2006)*, Arlington, VA, USA, November, (pp 11-18)
15. Data mining - *concept and techniques* Han & Kamber.
16.SQL SERVER 2005 books online.