Comparing Online Analytical Processing and Data Mining Tasks In Enterprise Resource Planning Systems

Tamer S. Abdellatif
Information Systems Department
Faculty of computer and information systems
Mansoura University
60 El Gomhuria st., Mansoura, Egypt

Mohammed Abo Elsoud
Information Systems Department
Faculty of computer and information systems
Mansoura University
60 El Gomhuria st., Mansoura, Egypt

Hesham Arafat Ali
Computers engineering and systems Department
Faculty of Engineering
Mansoura University
60 El Gomhuria st., Mansoura, Egypt

Abstract

Enterprise Resource Planning (ERP) is an environment which is often rich of data about the enterprise. Data warehouse online analytical processing techniques provided decision makers a set of useful tools to analyze report and graphically represent data of the ERP. It can be said that OLAP tools provides different summarized perspectives of the data. On the other hand, Data Mining techniques can discover previously unknown patterns of knowledge. It can be said that data mining provides a deeper look in the data. This paper provides a comparison and case-study of benefits obtained by applying OLAP or data mining techniques and the effect of integrating the both approaches in ERP.

Keywords

Data Mining (DM), Enterprise Resource Planning (ERP), Business Intelligence (BI), Online Analytical Processing (OLAP), Association Role

1. INTRODUCTION

Enterprise Resource Planning (ERP) is a set of applications for core business operations and back-end management that was originally developed for manufacturing and commercial companies [1].

The most significant factor that distinguishes ERP systems from previous generations of information systems such as MRP (Material Requirement Planning) that ERP permits organizations to integrate business processes and optimize the resources available [2].

The central transactional database of the ERP offers a rich source of data to apply analytical processing activities to gain benefits of ERP data. These analytical activities include Online Analytical Processing (OLAP) and Data Mining. Both OLAP and Data Mining approaches have different applications for business information systems such as ERP. In most cases, OLAP and data mining approaches are applied separately to accomplish different tasks in business applications [3, 4]. In the following sections, applications of OLAP and data mining on the different components of ERP systems are identified and listed for comparison. The comparison illustrates that, tasks to be performed using OLAP and data mining in ERP system components are different but yet can be integrated to accomplish new tasks [5].

The paper organized as follows; an overview of a data warehouse system as a data source to apply either OLAP or data mining operations in section 2, applying business intelligence using OLAP tasks in section 3, applying business intelligence using data mining tasks in section 4. According to sections 2, 3 and 4; section 5 provides a comparison between the tasks that can be performed in ERP systems using OLAP and data mining. Section 6 is a case study of applying and integrating both OLAP and data mining in ERP retail module.

2. HISTORICAL REVISION

The idea of ERP systems is initially started by the development of Material Requirements Planning (MRP) systems to handle the planning and schedule of complex products [6]. On next stages, MRP systems are developed to include some

operational areas of the enterprise such as (sales, operation planning and financials). Finally, ERP systems are developed to integrate all business functions of an enterprise to support decision making [7]. Figure (1) shows modules of a typical ERP system.

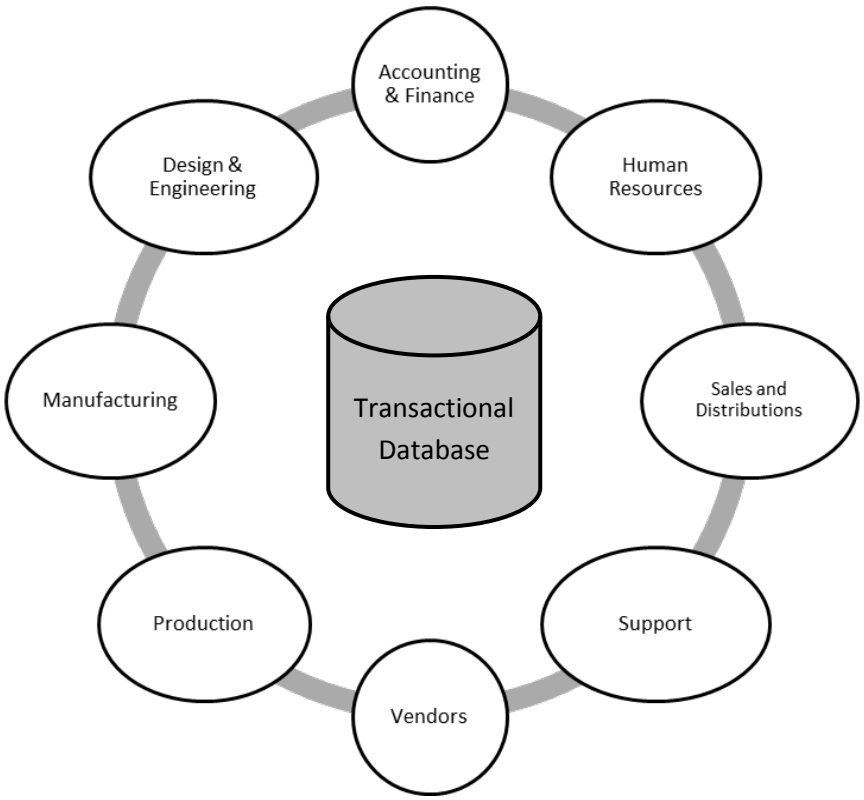

Figure (1): A typical ERP Diagram [5]

Some common vendors of ERP systems are SAP, ORACLE, Microsoft, and PeopleSoft. Being established in 1972, SAP Company is one of the oldest ERP system providers around world [22]. In 2004, SAP released the first release of NetWeaver (SAP business intelligence application) . NetWeaver business intelligence applies OLAP techniques and four data mining techniques (Decision Tree, Scoring, Clustering, and Association) [23.
, 24]. NetWeaver, applies each business intelligent technique separately with no ability of integration. In this paper, the author integrates between OLAP and data mining techniques in order to answer customer requirements that cannot be answered by applying OLAP or data mining separately.

3. Business Intelligence and Data Warehousing

Business Intelligence is providing decision makers with valuable information and knowledge by leveraging a variety of sources of data as well as structured and unstructured information. The information and data could reside within or outside the organization, could be structured in different ways, and could be either quantitative or qualitative [8].
In some instances, this activity may reduce to calculations of totals and percentages, graphically represented by simple histograms, whereas more elaborate analyses require the development of advanced optimization and learning models [3].
Traditional Database systems do not satisfy the requirements of data analysis necessary for BI. They are optimized to support the daily operations of an organization and their primary concern is to ensure the fast access of data in the presence of multiple users, such as ERP system's database [9]. Data warehouses are used as a data source for On-line analytical processing (OLAP) and machine learning [10]. Figure (2), shows a typical architecture of business intelligence and data sources to create data warehouse to apply OLAP and data mining.

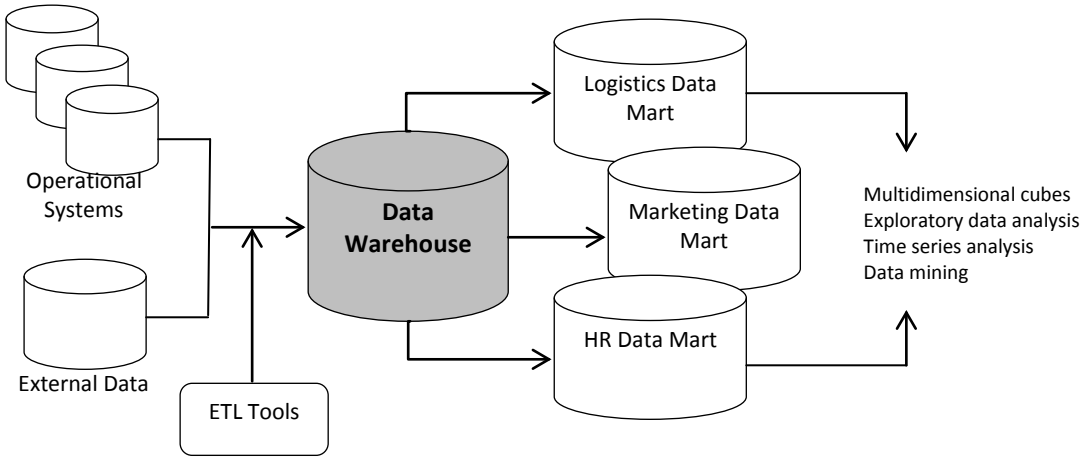

Figure (2): Business Intelligence Architecture [3]

A data warehouse is a collection of subject-oriented, integrated, non-volatile, and time-variant data to support decision making and BI [11]. In large data warehouse environments, many different types of analysis can occur. Data warehouse can be enriched with advance analytics using OLAP (On-Line Analytic Processing) and data mining. Rather than having a separate OLAP or data mining engine, they can also be integrated [12, 19]

4. BUSINESS INTELLIGENCE USING OLAP

On-line Transaction Processing (OLTP) applications are developed to meet day-to-day database transactional requirements and operational data retrieval needs of the entire user community. On the other hand, the data warehousing tools are developed to meet the information exploration and historical trend analysis requirements of management or executive user communities. [13, 14, 15]

OLAP is the whole set of tools aimed at performing business intelligence analyses and supporting decision-making processes We can therefore assume that the function of a data warehouse is to provide input data to OLAP applications [3]. Table (1) is a comparison between OLTP and OLAP showing the different tasks that can be achieved by each.

OLTP	OLAP
Only current data is available (old data is replaced by current data by updating)	Both current and historical data available (current is appended to historical)
Short transactions (single granularity or more)	Long database transactions
Online update/insert/delete transactions	Batch update /insert/delete transactions
High volume of transactions in a given period	Periodic refreshing
Concurrency control and transaction recovery	No transaction and therefore no recovery upon failures
Largely online ad hoc query, requiring low level of indexing	Largely predetermined queries requiring high level of indexing
Can have errors or missing data	Validated and complete data

Use normalized relational database design	Use multidimensional model

Table (1): Comparing OLAP and OLTP [14]

OLAP offers a set of operations to analyse data cubes (Dimensional Database). OLAP operations include: [7]

- **Roll-up:** operation transforms detailed measures into summarized ones. This is done when one moves up in the hierarchy or reduces a dimension, i.e. one aggregate to all level.
- **Drill-down:** operation of decreasing the level of aggregation moving from one general level to a detailed level in a hierarchy.
- **Slice:** operation of selecting one dimension of a cube resulting in a sub-cube.
- **Dice:** operation of selecting two or more dimensions of a cube resulting in a sub-cube.

5. Business Intelligence Using Data Mining

Data Mining is the process of exploration and analysis to discover meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories [15].

On the most important applications targeted by Data Mining is Business Intelligence. Using data mining, businesses may be able to perform effective market analysis, compare customer feedback, identify similar products, retain highly valuable customers and make smart business decisions [16].

Data Mining uses many several predictive and statistical methods in order to explore and analyse data. Such methods include association rule, linear regression, neural networks, regression trees, cluster analysis and classification trees… [15].

Figure (4) shows the Cross-Industry Standard Process of Data Mining (CRISP-DM) defines six standard phases for DM process [17].

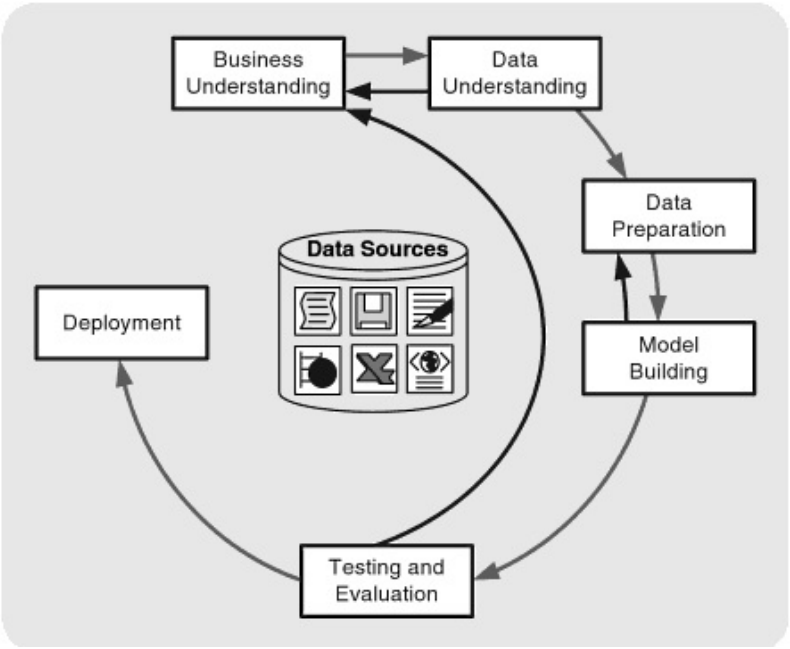

Figure 4: The Cross-Industry Standard Process of Data Mining (CRISP-DM) [17]

Operations that can be accomplished by data mining are categorized in the following classes [18] listed in table (2)

Anomaly detection	(Outlier/change/deviation detection) - The identification of unusual data records that might be interesting or data errors and require further investigation.
Association rule learning	(Dependency modelling) – Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits.
Clustering	The task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
Classification	The task of generalizing known structure to apply to new data. For example, an email program might attempt to classify an email as legitimate or spam.
Regression	Attempts to find a function which models the data with the least error.
Summarization	Providing a more compact representation of the data set, including visualization and report generation.

Table 2: Common operations performed by data mining [18]

6. OLAP VS. DATA MINING IN ERP SYSTEMS

Traditionally, OLTP of ERP systems integrates all day-to-day transactions of an enterprise in the way which enables users reduce operating cost, verify consistency of transactions, facilitate day-to-day management and provide detailed or summarized reports. Converting the OLTP data sources into data warehouse enables to apply OLAP and data mining techniques to provide more analytical reports and provide answers of miscellaneous questions to decision makers [10].

OLAP is a way to look at these pre-aggregated query results in real time. However, OLAP itself is still simply a way to evaluate queries which is different from building models of the data as in data mining. Data mining tools model data and return actionable rules, OLAP allows users to compare and contrast measures along business dimensions in real time [20]. While reporting and OLAP are informative about past facts, only data mining can help ERP customers predict the future of business. As shown in table (3), OLAP and data mining can answer different types of decision makers' questions.

OLAP	Data Mining
What was the response rate to companies' mailing?	What is the profile of people who are likely to respond to future mailings?
How many units of company's new product did it sell to existing customers?	Which existing customers are likely to buy company's next new product?
Who were my 10 best customers last year?	Which 10 customers offer me the greatest profit potential?
Which customers didn't renew their policies last month?	Which customers are likely to switch to the competition in the next six months?
Which customers defaulted on their loans?	Is this customer likely to be a good credit risk?

What were sales by region last quarter?	What are expected sales by region next year?
What percentage of the parts produced yesterday is defective?	What can be done to improve throughput and reduce scrap?

Table 3: Questions that cab be answered by OLAP vs. DM at ERP systems

The following table shows the difference between how can OLAP and data mining enhance an ERP solution

ERP Module	**OLAP Sample Tasks**	**Data Mining Sample Tasks**
Accounting & Finance Management	Compare and visualize cost /profit of company based on period or branch	Forecast total company profits based on historical data. Predicting Cash Flow Predicting overall profit/lose
Human Resources Management	Compare and visualize salaries, rewards and punishments overtime Compare employee evaluation overtime	Select candidate employee based on historical data
Vendors & Purchase Management	Find purchases totals based on product, category, vendor and time. Analysing cost of products' supply chains' Comparing purchases cost of different vendors for the same products	Determine best arrangement and quantities of purchase orders. (Purchase what of who and what amount)
Production Management	Analysing cost of products of several bills of material (BOM). Analysing manufacturing schedule variances Analysing defects compared to product, BOM, manufacturing line.	Applying Classification /Clustering technique to designs given design parameters to find out if design may result in unacceptable defect percentage in final products
Customer Relationship Management	Analyse sales according to customers' attributes (age, gender, location, marital, …) Analyse customers' response at marketing packages and offers Analysing issues reported by customers based on product, customer location, response time, …	Identify customers' behaviour patterns. Find people in similar life stages and may behave in the same way.

Sales & Distribution Management	Compare total sales based on period or branch Find the effect of a given offer on total sales Compare sales based on items Visualize customer purchases history	Determine what items sold together more probably for P.O.S. Determining customer behaviour over selling web sites. Effectively segment customers into manageable groups. Focus marketing efforts on prospects more likely to purchase. Forecast sales for a given period of time. Discover which customers will response to a given offer

Table 4: Comparing tasks that can be performed by OLAP vs DM at ERP systems

7. Case Study Integrating OLAP and Data Mining for Retail Module in ERP

6.1. About case study

Hyper Mart is a retail market running an ERP solution based on a relational database (SQL Server 2005). Aiming to manage marketing offers and bundles in a more profitable manner, the customer required to identify association between products, expected profit of each associated set of products according to available stock. Customer required that user should be able to view results into reports and charts.

Apriori Association Rule data mining is a suitable DM technique to analyse sales basket and identify associations among products. On the other hand, Association rule can neither answer questions about expected profit of each associated set of products, nor offering outputs into reports and graphics.
Integrating association rule mining with OLAP can represent an appropriate approach to fulfil customer's requirements listed in table 5.

Requirement	Priority
Total product sales for a time period	High
Total product category sales for a time period	High
Total product sales for a time period	High
Total sales of a specific store for a time period	Low
Total sales for a time period	Low
Total profits for a time period	High
Sales over stores	Low
Products of a specific category sales over stores	Low
Identify association among single products	High
Identify association among product categories	High

Provide report of associated products for a given product with a given minimum confidence	High
Provide a report of associated products with expected profit of the whole associated set.	High
Creating a chart of top 100 associated set of products comparing association confidence, and profit.	High

Table 5: Business intelligence requirements of retail module in ERP

Case study methodology

1. Create retail sales data cube.Apply OLAP methods to see which question to answer using OLAP.
2. Apply data mining Aprior Association Rule data mining to see which question to answer using data mining.
3. Save data mining outputs to a multi-dimensional database (Cube) on `which OLAP methods can be applied to integrate with data mining to answer customer requirements.

6.2. Expected results of applying OLAP

Applying OLAP to Retail data source of an ERP solution includes extracting, transforming and loading retail data into a dimensional database in Figure(5) and (6)

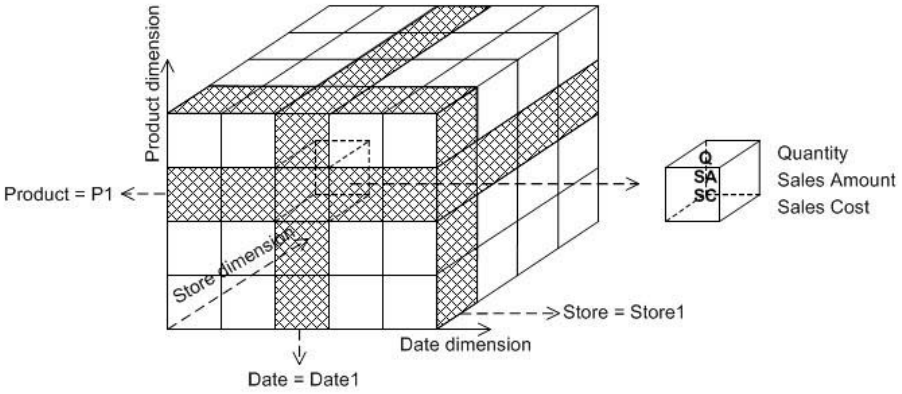

Figure 5: Data cube of ERP retail data mart

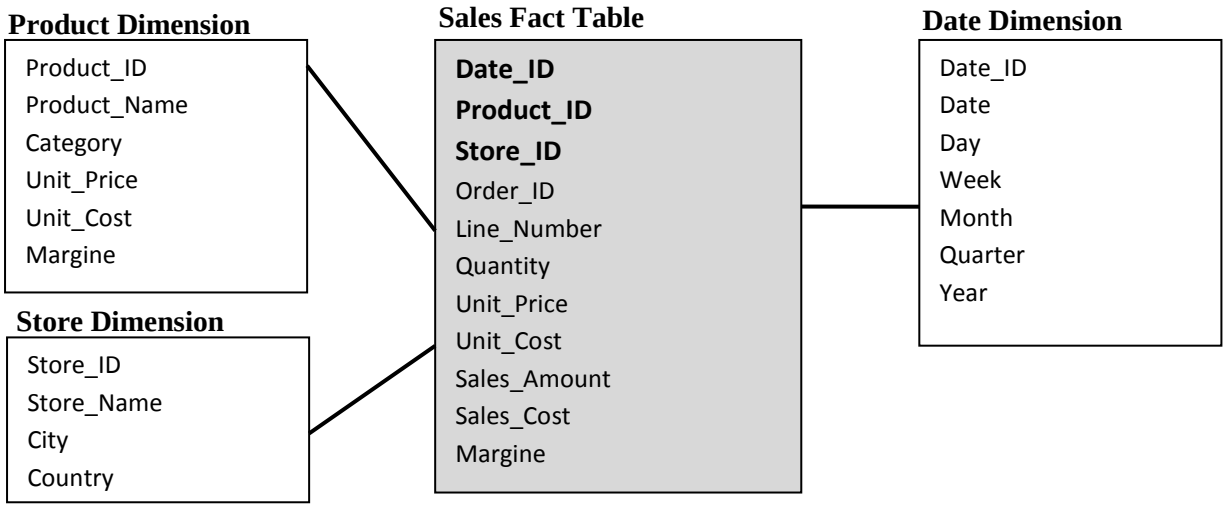

Figure 6: Dimensional data model of retail data mart

Using given operations several analytical reports can be generated as the following table (6), Figure (7) shows visualized output dash-board of applying OLAP methods to the cube of retail sales.

Total product sales for a time period (Month, Quarter, Year...)	Roll-up
Total product category sales for a time period (Month, Quarter, Year...)	Roll-up
Total product sales for a time period (Month, Quarter, Year…)	Drill-down
Total sales of a specific store for a time period (Month, Quarter, Year...)	Dice
Total sales for a time period (Month, Quarter, Year...)	Roll-up
Total profits for a time period (Month, Quarter, Year...)	Roll-up
Sales over stores	Slice
Products of a specific category sales over stores	Dice

Table 6: Tasks that can be performed by OLAP on ERP's retail data mart

Figure 7: OLAP application results

6.3. Expected results of Data Mining

One of the most powerful data mining techniques for retail analysis to find associations between items or products is using Apriori Association Rule mining.

An association rule is an expression of the form $X \Rightarrow Y$, where X and Y are sets of items and have no items in common. This rule means that given a database of transactions D where each transaction $T \in D$ is a set of items. $X \Rightarrow Y$ denotes that whenever a transaction T contains X then there is a probability that it contains Y too. The rule $X \Rightarrow Y$ holds in the transactions set T with confidence c if c% of transactions in T that contain X also contain Y [21]. Figure (8) shows the result of applying Apriori Association Rule to retail sales transactional data with minimum support of 3000 for any frequent item set and confidence of 5%.

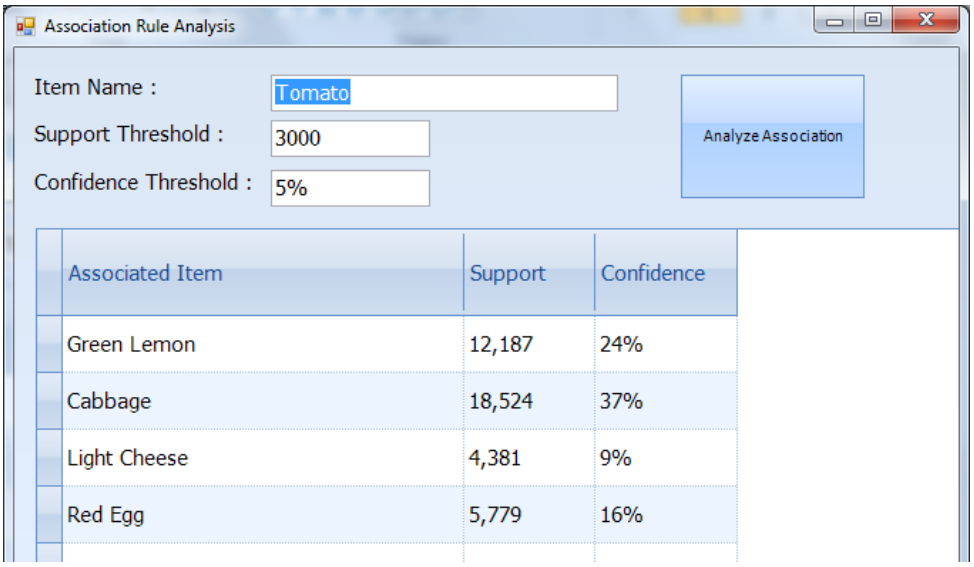

Figure8: Association rule data miming application results

Although association rule is a realistic and stable mining technique to analyse association between products, it still has some shortages:

- Analysing association between products, user must identify the set of products that to be checked, or identify minimum threshold of support of confidence.
- On retails database with so many products, it becomes impractical to define which products to be checked if associated.
- Association rule mining results is not suitable for visualizing.
- Comparing support and confidence of several products association is not easy.
- User cannot get further important information related to associated products such as the available stock quantity of each product, profitability of each frequent product set.

Integrating OLAP with data mining can resolve the above shortages which result in a better approach for analysing ERP database for decision support.

6.4. Integrating OLAP and Data Mining

In this section, we present alternatives for coupling OLAP and data mining techniques so that they can benefit from each other's advances for the ultimate objective of efficiently providing a flexible answer to data mining queries addressed either to a bidimensional (relational) or a multidimensional database. In particular, we investigate two techniques:

(1) Defining new operators similar in spirit to online analytical processing (OLAP) techniques to allow "data mining on demand" (i.e., data mining according to user's needs and perspectives). The implementation of OLAP-like techniques relies on three operations on lattices, namely selection, projection and assembly. A detailed running example serves to illustrate the scope and benefits of the proposed techniques. [4]

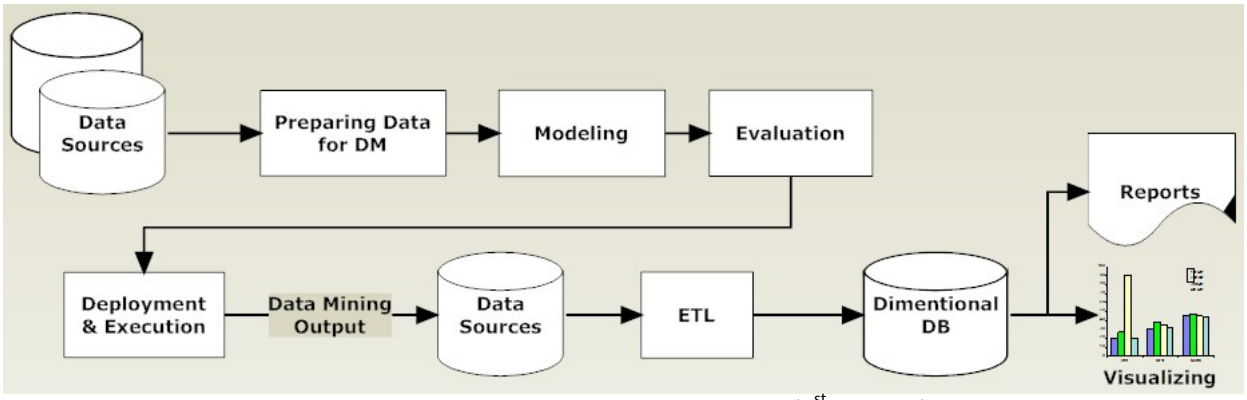

Figure 8: Integrating OLAP and Data Mining (1st Method)

(2) Multi-dimensional data mining integrates core data mining techniques with OLAP-based multidimensional analysis. It searches for interesting patterns among multiple combinations of dimensions at varying levels of abstraction. [16]

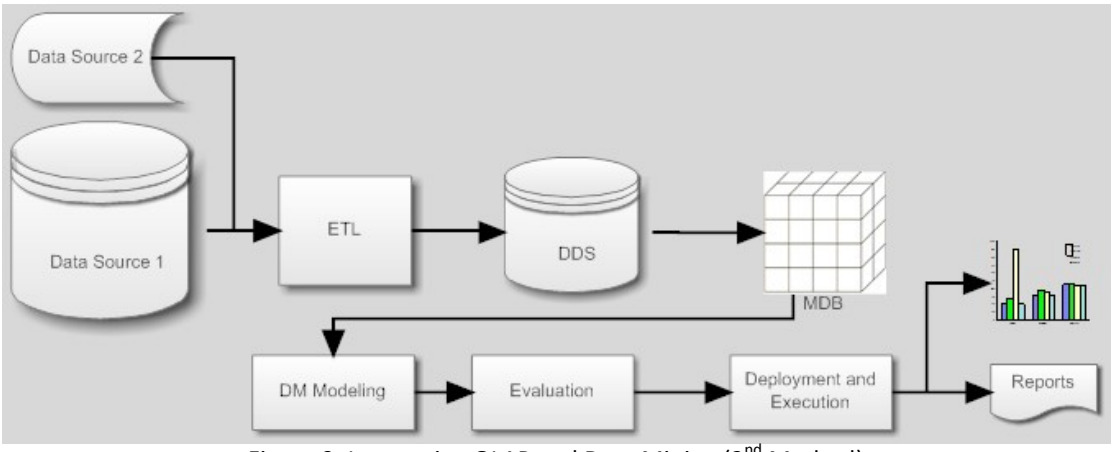

Figure 9: Integrating OLAP and Data Mining (2nd Method)

6.5. Applying Method (1) to case study

In this section, integration between Data Mining and OLAP is done according to method (1). Output of association role data mining that is applied to retail database are stored into a multi-dimensional database (MDB) as shown in figure (11, 12) . Storing data mining outputs to an MDB allows applying OLAP operations (Roll-up, Drill-down, Dice and Slice).

The multi-dimensional data mart that contains association rule mining is consisted of one fact table that represent association between up to three products as a frequent product set P(P1, P2, P3) into two values of Support and Confidence as shown in Table(7).

Product_ID1	Int	Frequent itemset member 1
Product_ID2	Int	Frequent itemset member 2
Product_ID3	Int	Frequent itemset member 3
Support	Int	Support value of Product 1
Confidence	Float	Confidence of the frequent itemset

Table 7: Fact table design in a dimensional database of DM outputs

The product dimension that shown in table (8) contains product related information. It also contains aggregated values such as Margin and Stock_Amount which help on generating reports based on the multi-dimensional database to fulfil user's requirements.

Product_ID	Int	Product dimension identity (PK of Product dimension table)
Product_Key	Int	Products key (Unique code)
Product_Name	Varchar(100)	The name of product
Category	Varchar(100)	Category whose product imply to
Unit_Price	Float	Sale price for a single unit of the product
Unit_Cost	Float	Cost price for a single unit of the product
Unit_Margine	Float	Profit per product unit (Sales price – cost price)
Stock_Amount	Float	The amount of product in all stocks

Table 8: Product dimension table design in a dimensional database of DM outputs

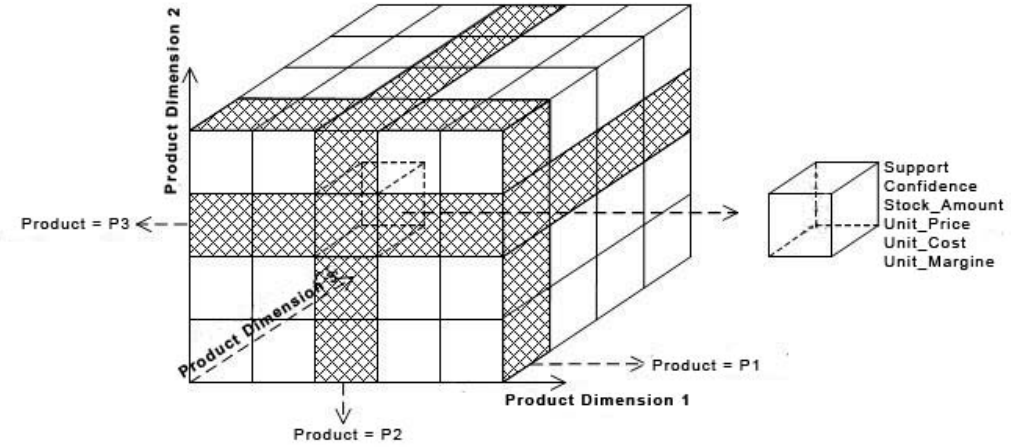

Figure 9: Data cube of association rule DM outputs

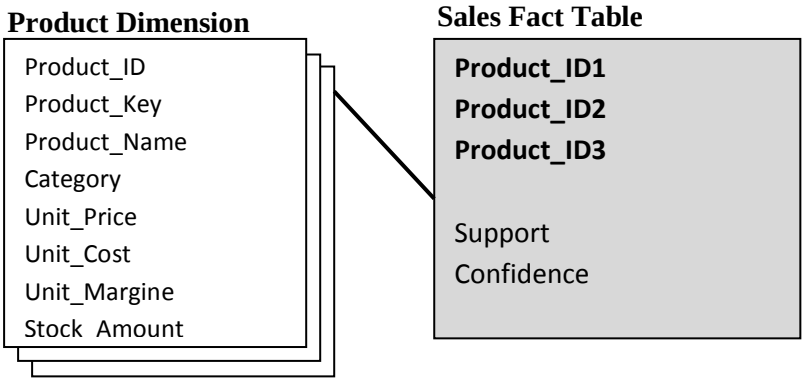

Figure 10: Dimensional database design of DM outputs

Storing association rule data mining output results to a dimensional database allows applying OLAP operations on the stored output in order to add more aspects to the results. This integration enables fulfilling customers' requirements as shown in table (9).

Requirement	Using OLAP to accomplish
Reports of associated single products	Drilling-down associated products cube
Reports of associated product categories	Rolling-up by discarding Product_Key and considering

Provide report of associated products for a given product with a given confidence threshold	Dicing associated products cube to a subcube of product and associated products
Provide a report of associated product sets with expected profit of the each associated set.	Drilling-down by considering additional details of product dimension including Unit_Margin, Stock_Amount
Creating a chart of top 100 associated set of products comparing association confidence, and profit.	Drilling-down by considering additional details of product dimension including Unit_Margin, Stock_Amount, Confidence

Table 9: Customer requirements to be accomplished by applying OLAP on DM outputs dimensional database

8. Conclusion and Future Work

Enterprise resource planning information systems are a common example of business application that results in a huge amount of data. Businesses that use ERP systems can benefit from business intelligence OLAP and data mining (DM) approaches that can apply to ERP's data in order to generate reports, charts and identify new knowledge to support decision makers.

Although OLAP and data mining can perform different tasks on ERP's data, but integrating both approaches is very useful to perform new tasks that may be required by businesses decision makers and ERP users that cannot be performed using one of the two approaches. One method of integration is applying OLAP operations to DM output results.

In future work, another method of integration can be used to integrate OLAP with DM. The method is to apply data mining operations such as (association rule mining, clustering, classification …) to the multi-dimensional database in order to perform new advanced business intelligence tasks.

Acknowledgements

The authors would like to thank all the people and institutions that have allowed them to use their own real business data which helped authors a lot to make this study.

References

[1] Jagan Nathan Vama: *ERP in practice: ERP strategies for steering organizational competence and competitive advantage* (New Delhi: Tata McGraw-Hill Education)

[2] Tin Yu Ho: *Impact Of ERP And OLAP Systems On Management Accounting Practices And Management Accountants* (Swedish School of Economics and Business Administration, 2006)

[3] Carlo Vercellis, *Business Intelligence - Data Mining and Optimization for Decision Making* (UK, Chichester: John Wiley & Sons, 2009)

[4] Robert Wrembel, Christian Koncilia: *Data Warehouses and OLAP - Concepts, Architectures and Solutions* (Pennsylvania: Idea Group, 2007)

[5] Matthew P. Stephens, Hugo X. Ramos: *Who Moved My ERP Solution?* (Journal of Industrial Technology, 2003) Vol. 19, No. 1

[6] Christian N. Madu, Chu-Hua Kuei: *ERP And Supply Chain Management* (New York: Chi Publishers, 2005)

[7] Alexis Leon: *Enterprise Resource Planning* (New Delhi: Tata McGraw-Hill Education, 2005)

[8] Rajiv Sabherwal, Irma Becerra-Fernandez: *Business Intelligence* (New Jersey: John Wiley and Sons, 2009)

[9] Elzbieta Malinowski, Esteban Zimányi: *Advanced data warehouse design: from conventional to spatial and temporal Applications* (Berlin: Springer, 2008)

[10] Dr. K.V.K.K., Prasad: *Data Warehouse Development Tools* (New Delhi: Dreamtech Press, 2005)

[11] Torben Bach Petersen: *Data Warehousing and Knowledge Discovery: 12th International Conference* (Berlin: Springer, 2010)

[12] Vincent Rainardi: *Building a data warehouse with examples in SQL Server* (New York: Apress, 2007)

[13] C.S.R. Prabhu: *Data warehousing: concepts, techniques, products and applications* (New Delhi: PHI Learning Pvt, 2004)

[14] Arshad Khan: *SAP and BW Data Warehousing: How to Plan and Implement* (Nebraska: iUniverse, 2005)

[15] Galit Shmueli, Nitin R. Patel, Peter C. Bruce: *Data Mining for Business Intelligence: Concepts, Techniques, and Applications* (New Jersey: John Wiley and Sons, 2010)

[16] Jiawei Han, Micheline Kamber, Jian Pei: *Data Mining: Concepts and Techniques* (Massachusetts: Elsevier, 2011)

[17] David Louis Olson, Dursun Delen: *Advanced Data Mining Techniques* (Berlin: Springer, 2008)

[18] Fayyad, Usama; Gregory Piatetsky-Shapiro, and Padhraic Smyth: *From Data Mining to Knowledge Discovery in Databases* (AI Magazine, 1996), Vol. 17

[19] Asem Omari: *Data Mining for Retail Website Design and Enhanced Marketing* (Düsseldorf: Heinrich-Heine-University, 2008)

[20] John Wang: *Encyclopaedia of Data Warehousing and Mining* (Montclair State University, Information Science Reference, 2009) Vol. III

URLS

[21] Oracle® Database Data Warehousing Guide "OLAP and Data Mining" <http://download.oracle.com/docs/cd/B28359_01/server.111/b28313/bi.htm> (24 Spt. 2011)

[22] Wikipedia "SAP AG" <http://en.wikipedia.org/wiki/SAP_AG> (24 Spt. 2011)

[23] Wikipedia "SAP R/3", < http://en.wikipedia.org/wiki/SAP_R/3> (24 Spt. 2011)

[24] Wikipedia "SAP NetWeaver", <http://en.wikipedia.org/wiki/SAP_NetWeaver> (24 Spt. 2011)