



A Technique for Implementing Business Application Using Data Mining

Vinitha S¹, Dinesh P S²

¹Student, ²Asst.Prof.

^{1,2} Department of CSE, Bannari Amman Institute of Technology, Sathyamangalam.

Abstract-Information skill is now necessary in every part of our lives which help business along with enterprise to make apply of applications like decision support system, query and reporting online analytical processing, analytical analysis and business performance executive. In this paper focuses on the consequence and role of Data Warehousing and Data mining technology in industry. A Data Warehouse is a central storage area of relational database designed for query and analysis. It helps the business group to combine data from different unstable sources. These warehouses are analyzed by the newest technique known as Data Mining. In Data Mining data sets will explored to give system out of sight and indefinite predictions which can be used in future for the able judgment making. At present companies apply techniques of Data Mining that involves pattern credit mathematical and statistical techniques to search Data Warehouses and help the analyst during recognizing major trends, facts relationships and anomalies.

Index Terms: Data Warehousing, Data Mining, OLAP, OLTP, CART & CHAID.

I. INTRODUCTION

Data Warehouses are ahead vast position in Business Intelligence (BI), every group gives highest priority to maintain a shared data Warehouse. Most business applications like predictive analysis/online analytical processing, statistical, and complex query processing and critical business decisions are based on the data existing in the Data Warehouse.

Data Warehouse (DW) is a system with the purpose of, cleans, confirms and source data into a dimensional data stock up and then chains and gear querying and analysis for the purpose of decision making. Sophisticated OLAP and Data Mining tools are used to simplify worldwide exploration and complex business models. In Data Warehouse as a subject oriented, integrated, time variant and non-volatile collection of data in support of management's decision making process. BI applications in enterprises provide information for the calculated management of production by collaborating the business data and electronic data junction.

This ensures aggressive acumen and thereby helps in excellent judgment manufacture. According to B de Ville, BI refers to the technologies and use for collect storing and analyzing business data that helps the activity to make better decisions. Data Marts were used to study the data and it's a difficult task that is time strong Thus for the enhanced analysis if facts Data Mining methodologies is used. The Data Mining process comprises computer assistance analysis and extracting of huge volume of business data. Data Mining as a nontrivial mining of contained earlier strange and potentially useful in order from data.

The grouping of data warehousing and Data Mining technology has become an innovative scheme in many business areas through the mechanization of routine tasks and overview of executive procedures.

II. DEFINITION: DATA WAREHOUSE:

Data Warehouse is a warehouse of activity or business databases which provide a clear picture of recent and past operations of organizations. While it provides a logical picture of the business conditions at an exacting point of time, it is used for the able conclusion make process. It involve the advance of structure that helps the withdrawal of data during artificial ways.

Data Mining describe the process of scheming how the data is stored in order to improve the exposure and study Data Warehouse experts believe that the a mixture of stores of data are coupled and correlated to each other abstractly as well as actually. A business's data is regularly store across a integer of Databases. However, to be able to evaluate the broadest array of data, each of these databases needs to be coupled in some way. This means that the data in them need a system of being related to other significant data and that the physical databases themselves have a link so their data can be look at together used for reporting purpose.

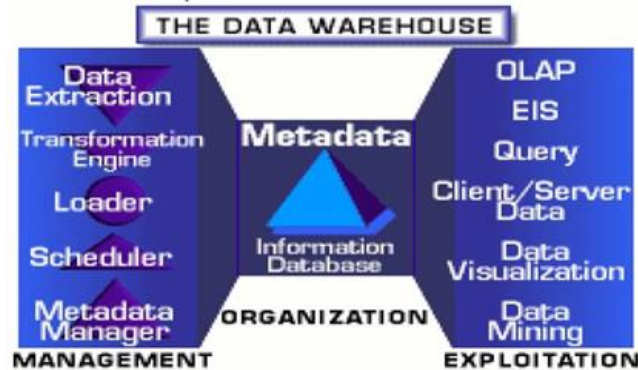


Fig 1: Data Warehouse

Multiple data stores are included by the Data Warehouses and this in a row is used by the managers for improved decision making. Data warehousing environment includes the mining of relational database, Transformation, Loading (ETL process), Online Application Processing (OLAP) engine and client analysis tools. As a business internationally the parameters and complexities involved in analysis and decision making become further complex.

Data contact portion which is on hand in the form of goods is the most observable part of a Data Warehouse project. Data warehousing process involves change of data from new format to a dimensional data store which consume a greater take of effort, time and fixed cost.

Since the execution of a data warehousing is expensive and grave there are a number of data extraction and data cleaning tools and load and bright utilities are available for the same. One of the most important point of the Data Warehouse is data combination.

A. Data warehousing on Facebook content

A huge model of data warehousing is what Facebook does. Facebook gather all your data such as your groups, your friends, your likes etc. All these data are stored into one vital storage area .Although Facebook is storing all these in sequence into divide databases, they store the most related and large in order into one central aggregate database. This is as of many reason like to make sure that you see the nearly all related adds that you are most likely to click on or the friends that they propose are the most important to you.

B. Significance of Data Warehousing.

Data Warehouse is a subject leaning time variant, volatile and non-volatile set of data. Data inclusion, Online Analytical Processing (OLAP) and Data cleansing are a part of the data warehousing skill. It provides a entire and constant data store from multiple sources which can be easily unstated and used in business applications.

Some of the function areas which included are Integration of data diagonally the enterprise, historical data and Quick decisions on current. That provides ad-hoc information for loosely-defined system Manage & control businesses solve imagine investigation.

C. Data Warehousing: method

The route of centralized or aggregating data from various source into one common storehouse is implemented by with Data Warehousing. Data warehousing occur before than Data Mining take set.

Data warehousing involves an exacting engineering phase, where no business users are implicated. Data stored in special databases are combined into single inclusive and easily available database are used on data warehousing. This is existing on business professionals or managers who use the data for Data Mining and form forecasts. Data is feed from a multiplicity of special sources into the Data Warehouse which is again changed, reformatted, summarized and used for executive decision making. The process of data warehousing acts as a teaching to identify the business requirements, build up the business plan and produce Data Warehouse also include project management, set up and bundle up actions.

D. Data Warehouse: planning

Data Warehouse design is based on the variety of business process related with a group. Some other consideration while going for the arrangement of a Data Warehouse include Meta data management, extent of query condition, data model plenty security and operation of full technology. Metadata is data about data which is store also as a unstructured or semi-structured form. These outline data are very useful in Data Warehouses. For example simple Data Warehouse inquiry can be used to recover January sales. Data Warehouse architecture can be shown with the materialize view in Oracle 9i as below.

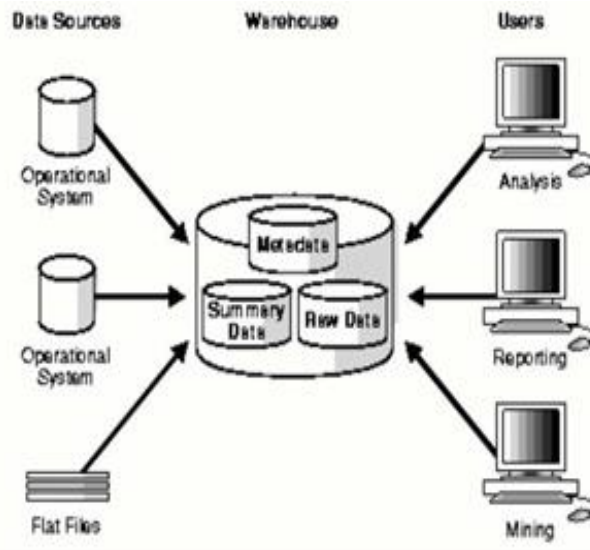


Fig 2: Data Warehouse planning

III. DATA WAREHOUSE TO DATA MINING

It is needed to choose plenty Data Mining algorithms for making Data Warehouse more useful. Data mining algorithms are used for transform data into business information and improving decision making process. Data Mining is a set of methods used for data scrutiny formed with the aim to find out exact trust associations and rules linked to data with making them out in the new advanced level excellence information. Data Mining gives ending that show the interdependence and relations of data. These dependences are largely based on various numerical and statistical relations. Data are collected from inside database and converted into various documents, information list etc. which can be further used in decision making processes. After choose the data for analysis, Data Mining is applied to the right rules of behavior and patterns. That is the reasons why Data Mining is also known as “extraction of data “data archeology” or “pattern investigation.

A. Example of Data Mining: scheme recognition of credit card usage

Credit card companies will aware you when they think your credit card is illegally used by someone other than you. Companies will have a history of the customer’s purchase and know only where the acquire have been made. If a purchase is made in a city far gone from where you live, the companies will put an aware to promising scheme since their Data Mining shows that you don’t normally make

purchases in that city. Companies can either stop the card for that transaction or put a flag for disbelieving activity.

B. Data Mining procedure

Data Mining provides ways to make best use of data during fast automation Data Mining software use modeling techniques to make a demonstration that is a set of examples or a mathematical relationship based on data from position where the answer is known and then apply the same model to other situation where answers are hidden.

The 3 main stages drawn in in the process of Data Mining are:

1) Exploration: Data preparation, attack and change are occupied in this stage. A separation of records will be special to reduce the number of variables to a convenient range. This depends on the complexity of analysis of graphical and arithmetic data.

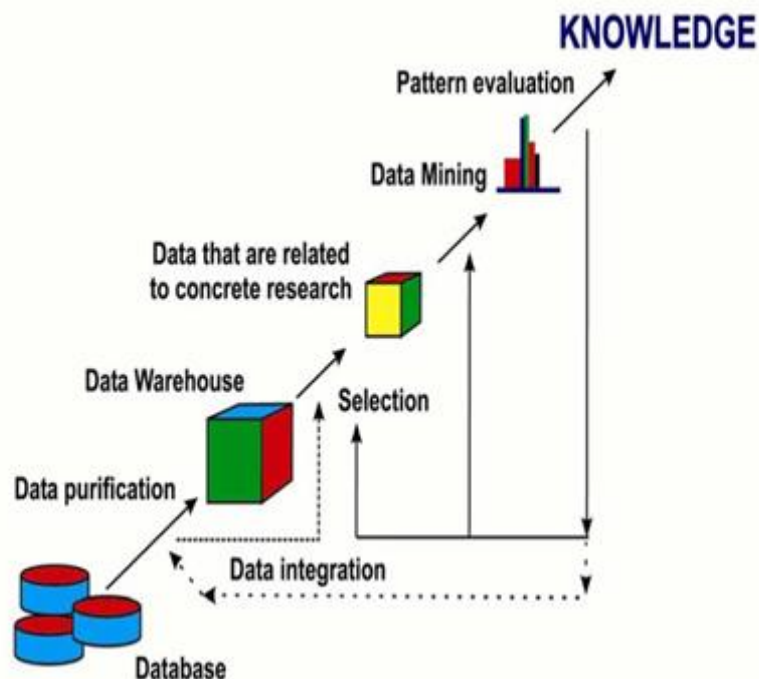


Fig 3: The procedure of knowledge improvement from database by using Data Warehousing and Data Mining technology.

2) Model building and support: In this stage the best reproduction will be taken base on their analytical show. Various techniques used for judgment of model comprise bagging, boosting, stacking and Meta learning.

3) Reliant :In this final stage the best model is select and it is functional to the new data sets to produce guess of the likely ending. One simple example for this is the online shopping site doing e-commerce contact through credit card deploy neural networks and Meta learner to name fraud.

Data Mining process involve use of various technique and methods. Most common techniques are:

1) organization: Stored data will be grouped into different classes. This allows locating data into pre-determined groups.

2) Clustering: Facts items are grouped into cluster of similar groups. It may be of It may be of hierarchical or non-hierarchical.

3) Regression: This method uses a statistical data set to develop a best fit mathematical formula. This method can be used to feed new facts sets and get a better forecast This is suitable considered for incessant quantitative data.

4) Connection: It is a rule $X \rightarrow Y$ such that X and Y are data items sets.

5) In order pattern matching: It allows expect behavior patterns and trend based on the in order rule $A \rightarrow B$ which implies that event B will all the time be follow by A.

C. Next construction Data Mining Techniques

Data Mining uses black box advance to discover data and open understanding using Exploratory Data Analysis (EDA) techniques. The techniques used in Data Mining are a merge of statistics, database explore and artificial intelligence. Next making Data Mining techniques include artificial neural networks, judgment trees, instruction rules and inherent algorithms.

1) Reproduction neural networks: This system uses non-linear predictive model to allow education through training. Computers are skilled to think, act and take result similar to humans. These models are quite composite to use even by the experts because it is packed as a complete solution. It determine important forecast for a model.

2) Rule orientation this system enables knowledge discovery and unsubstantiated learning. It extract useful pattern from database based on accuracy and arithmetic significance. Deduction will be more correct and has better logic by neural network.

It creates confident confusions to choose the best regulation from a collection of rules. usually rule instruction is used on databases with many columns of double fields or fields with higher cardinality in order to assemble the suitable patterns for making a better guess a bottom – to – top ensue is chosen.

2) Judgment trees: Decision tree is a Data Mining technique where tree shape structures are representing the set if result generating rules for a data set classifications. The early node or the top node is known as the root. Depending results of test, the search partitioned into two or more nodes. It is a rapid Data Mining technique since its essential take away or no pre-processing of business data. It is used for both search and guess using Classification And Regression Trees (CART) and Chi Square Automatic Interactions detection (CHAID). CART generate two system split from data set segmentation which needs less research of facts than CHAID which generate a multi-way split. Rules are jointly exclusive and quite extensive

3) Genetic algorithms: This Data Mining is based on optimized technique n inheritance and natural selections, combination and alteration. Genetic algorithms are used in patterns appreciation either as classifier or as an optimization tool. According to Chuck Kelly (2002), genetic algorithms hold the endurance of the fittest using heuristic functions even affectation the troubles.

IV. TRANSPORTATION FOR IMPLEMENTING DATA WAREHOUSE AND DATA MINING

Data Warehouse and Data Mining function are quite removed in mass and storage space capacity. Enterprise applications range from 10 GBs to advanced. Data Warehouse is a very bendy explanation that can explore database more capably than any other Online Transaction Processing (OLTP) environment. The major gain of this is that the client does not have to hold data of relational model and complex query languages.

A. Data ware house effecting phases.

According to Barry D & Addison – Wesley, 1997 Data Warehouse execution phases include:

- 1) Analysis of existing situations: This is an extremely important phase in the Data Warehouse aim since at this phase a possibility of recognition and key of the problems can be seen. Since the users will have better facts about the problems than the designer their judgment is very central for a good warehouse design.
- 2) Selecting the most proper data for study from the offered data: instead of using the total OLTP database, the data split which include all the exciting data linked to the focus will be special
- 4) Filtering data exciting for analysis: data study does not need all the data. Because of this the filter of data will be done linked to certain time or some explicit area.
- 5) Extracting data in production database: after falling and filtering of data, data are being extracted in staging folder from which the Data Warehouse is person built. Data Transformation Services (DTS) package is printed in SQL server 2000. Enclose writing is very significant in Data Warehouse accomplishment because post can be arranged to function by design so that the users can clean and driven data.
- 6) Selecting fact table, dimensional tables and proper schemas: entity-relationship model regularly used in the propose of relational databases. This is suitable for OLTP. A Data Warehouse require a short ,subject leaning plan that facilitates online data analysis. The simplest scheme is a lone table scheme which consists of disused fact table. Data Warehouse contains a large central fact table contain the bulk of data with no job loss and a set of smaller height tables.
- 7) Selecting quantity take of aggregations and warehouse models: the next step in scheme Data Warehouse is selecting power .It needs considered size that are attain from.
- 8) Various calculation operations with other size: Data Warehouse solutions also use aggregations through which they solve the uncertainty very fast. The increasing of the take of combined data speed up the user exact query.
- 9) Creating and using the cube: the cube is being created on either client or server computer. Basic factors for selecting the place for cube's storehouse are size of the cube, performance of the clients and server's computers, and throughput of the system. The cube created can be used by the support of various client apparatus.

B. Data Mining Implementations

Microsoft Decision Tree (MDT) algorithms are based on prospect of a variety of attributes and it is when calculation is necessary. These algorithms also generate rules. MDT also enables the user to consider a huge amount of Data Mining troubles.

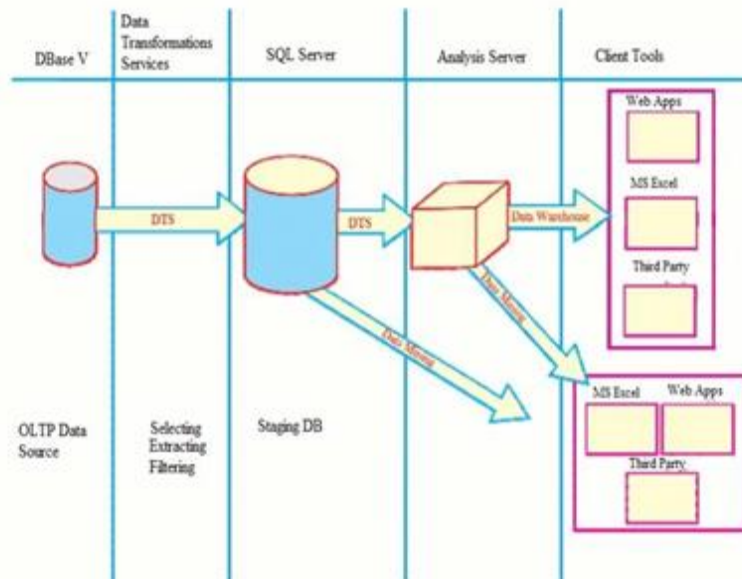


Fig 4: Scheme of Data Warehouse and Data Mining performance

Database size with query complexity is the 2 grave technical drivers for Data Mining. New hardware architectures like Massively Parallel Processors (MPP) are used which can linkage hundreds of rate processors to attain better routine. Data Mining is at present firmly used in different industries. All the main database vendors are with various Data Mining techniques in their platform. Some of them are:

- 1) SQL server: it is Microsoft database stand that allow Data Mining during the use of clustering with classification algorithms.
- 2) SAS, SPSS and S-PLUS are higher statistical packages for implement Data Mining algorithms.
- 3) Darwin: is an Oracle Data Mining set for implement organization and decision trees, K-nearest neighbors, regression analysis, clustering algorithms and neural networks.

V. DATA WAREHOUSE AND DATA MINING: APPLICATION AREAS IN INDUSTRY

Data warehousing and Data Mining has gain enhanced status in multiple areas of business to evaluate the huge databases speedily which would be too composite and time overriding .Some of these use areas are planned below.

- 1) Administration: for searching activist shape and threat assessment.
- 2) Backing: investigation and forecasting of business concert for stock and link analysis.
- 3) Banking: toward learn underwriting, credit approval etc.
- 4) Shortest marketing: for identifying projection that are built-in out mailing list so as to obtain main reaction time.
- 5) Medicine: for drug analysis, diagnosis, worth control and epidemiological studies.
- 6) Modern: for enhanced excellence control and protection.
- 7) Churn investigation: to predict clients who are likely to give up the business and travel to a participant business.
- 8) Sell segmentation: to spot customer's common individuality and actions that purchases the same goods of a company.
- 9) Tendency investigation: to investigate the difference among the customer's behavior over successive months.
- 10) Deception exposure: to identify the fraud users in telecommunication commerce as well as credit card treatment.
- 11) Web promotion: for advertisement and personalization opportunities.

VI. CONCLUSION

Data Warehouse and Data Mining technologies have light future in business application as it helps to generate new possibilities by robotic calculation of trend and behaviors in a large database. Data mining techniques help to by design discover the unknown patterns like identify strange data that stress errors generated during the data access. Data Warehouse & Data Mining technologies have grow to be a hit with a range of industry like sales & selling healthcare group economic institution and many more. These technologies have a lot of payback in unstable fields. It can be said among pleasure that these technologies help the quick study of data and thereby civilizing the value of judgment making procedure. Both Data Mining and Data Warehousing are business intellect tools that are used to turn in sequence or data into actionable knowledge. Data store expert design data storage space system that connect related data in unlike databases where as a Data mining run more meaningful and efficient queries to look up business. The huge data volume and very multipart knowledge sighting actions associated with business organization make the Data Warehouse with its OLAP and Data Mining tools a very large technology behind decision making. Thus Data Warehouse & Data Mining are very important components in business operation to gain competitive aptitude. These technologies allow arithmetic multidimensional analysis of data to assess relationships, correlation and trend in business.

REFERENCES

- i. Inmon W.H., "Building the Data Warehouse", Second Edition, JWiley and Sons, New York, 1996.
- ii. P. Bergeron, C. A. Hiller, (2002), "Competitive intelligence", in B.Cronin, Annual Review of nformation Science and Technology,zedford, N.J.: Information Today, vol. 36, chapter 8
- iii. B. de Ville, (2001), "Microsoft Data Mining: Integrated Business Intelligence for e-Commerce and Knowledge Management", Boston: Digital press.
- iv. Frawley W., Piatetsky – Shapiro G. and Matheus C., "Knowledge Discovery in Databases: An Overview", AI Magazine, Fall 1992, pgs 213-228.
- v. C. Date, (2003), "Introduction to Database Systems", 8th ed., Upper Saddle River, N.J.: Pearson Addison Wesley.
- vi. Han Jiawei, KamberMicheline, "Data Mining: Concepts and Techniques", 2nd edition, Morgan Kaufman Publishers, March 2006. ISBN 1-55860-901-6.
- vii. Oracle9i Data Warehousing Guide Release 2 (9.2), Part No. A96520-01, March 2002.
- viii. Berry, M.J.A., and Linoff, G., "Mastering data mining", The Art and Science of Customer Relationship Management, 1999.
- ix. Bhavani, T., Data Mining: Technologies, Techniques, Tools and Trends, 1999.
- x. Jiwei, H., and Micheline, K., Data Mining: Concepts and Techniques, Simon Fraser.
- xi. D. Pyle, (2003), "Business Modeling and Data Mining", Morgan Kaufmann, San Francisco, CA.
- xii. M.H. Dunham, (2005), "Data Mining – Introductory and Advanced Topics", Prentice Hall.
- xiii. Berson Alex, Smith J. Stephen, Thearling Kurt, (1999), "Building Data Mining Applications for CRM", McGraw-Hill Companies.
- xiv. Gilman Michael, (2004), "Nuggets and Data Mining", Data Mining Technologies Inc. Melville, NY 11714, (631) 692-4400
- xv. Chen, S. H, (2002), "Genetic Algorithms and Genetic Programming in Computational Finance", Boston, A: Kluwer.
- xvi. Kelly Chuck, (2002), "What is the role of Genetic Algorithms in Data Mining", Information Management: How your Business Works,Electronic Newsletter, <http://www.information-management.com/news/5755-1.html>.
- xvii. Barry, D., Data Warehouse from Architecture to Implementation, Addison-Wesley, 1997.
- xviii. Krulj, D., "Design and implementation of data warehouse systems", M Sc. Thesis, Faculty ofOrganizational Sciences, Belgrade, 2003.
- xix. Chapple Mike, "Data Mining: An Introduction", (2011), <http://databases.about.com/od/datamining/a/datamining.htm>.
- xx. Alexander Doug, "Data Mining", (2000), <http://www.laits.utexas.edu/norman/BUS.FOR/course/mat/Alex/>, electronic article.