

A study of application of Data Mining in Demonstrating Business Intelligence

A.Mohammed Khalid Kaleem, B .Mohammed Iquebal Akhter

Abstract-- Business intelligence (BI) has been referred to as the process of making better decisions through the use of people, processes, data and related tools and methodologies. Data mining is the extraction of hidden stating information from large databases. It is a powerful new technology with large potential to help the company's to focus on the most necessary information in the data warehouse. This study gives us an idea of how data mining is applied in exhibiting business intelligence thereby helping the organizations to make better decisions.

Keywords-- Business intelligence, data mining, database, information technology, management information system

I. INTRODUCTION

Data mining, is the extraction of hidden stating information from large databases, is a powerful new technology with large potential to help the company's to focus on the most necessary information in the data warehouse [Quinlan J R, 1993]. The tools of Data mining predict the future behaviours and trends and allow the business to make knowledge driven and proactive decisions. The prospective and automated analysis offered by data mining moves beyond the analysis of past events provided by contemplating tools of decision support systems. The tools of data mining can answer the questions of business that were traditionally too much consumption of time to solve. The tools of data mining clean the databases for the hidden patterns, and find the stating information that the experts may miss because it lies outside the expectations. Most of the companies already refine and collect the large quantities of data. Techniques of data mining is implemented rapidly on the hardware platforms and existing software to enhance the value of the consisting resources of information that is integrated with the new systems and products as they are brought online [Fayyad U, 1996].

II. FOUNDATIONS OF DATA MINING

The figure below shows the knowledge extraction process from the available data by application of data mining techniques [Srikant R, 1997]:

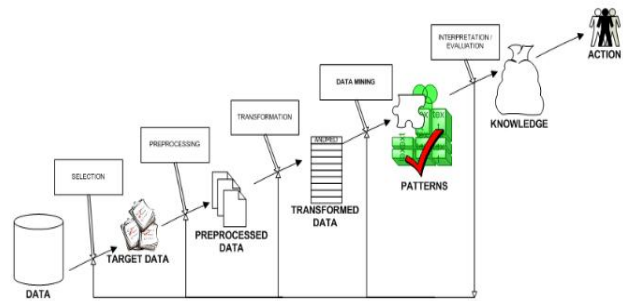


Figure 1: Knowledge Extraction Process

Data mining techniques are a result of long process of product and research development. This evolution started when the business data was 1st stored in the computers, continued with the improvements in access of data and recently, the generated technologies allow the users to navigate through their data in real time. Data mining takes the evolutionary process beyond the contemplating access of data and navigates to proactive and prospective delivery of information [Berry M, 1997]. Data mining is ready for the application in the community of business because it is supported by 3 technologies that are now adequately developed as the:

- Powerful multiprocessor computers.
- Massive collection of data.
- Algorithms of Data mining [Liu B, 1996]

III. INTRODUCTION TO BUSINESS INTELLIGENCE

Business intelligence is used to refer the number of company activities, which may undertake to collect information about their competitors or their market. Some areas are always included under the heading of business intelligence are: industry analysis, competition analysis and market analysis. Some people also consider the industrial espionage to operate for collection of information purposes to

be a form of business intelligence. In most cases, the company will create their own dedicated group of business intelligence or hire outside agency. The business intelligence group will collect information from inside the company about how the company is performing and where the improvements are made. A business intelligence group then looks for outside sources, which include the public records of other business in the same sector, customer survey information and analysis of market by third parties. A business intelligence group will drive further into particular competitors, both by examining the business model and public information, in some cases an industrial spy is used to collect the information [Weiss S M, 1998].

The systems of business intelligence are contrasted to more classic forms of collection of information by their interdepartmental focus and their general overview towards the performance of business. They are also different in their use of advanced techniques and technology to crunch and mine the data in most optimal manner. A business intelligence group change the analysis of market have a strong understanding of the specific sector of the market in which business operates, their lack of same detailed understanding about the inner management of the company and particular competitors make their information useful. In the model of business intelligence all the various forms of improvement of business are tied together so that the communication is easy and quick and each segment helps to inform the other segments to be more valuable than they would be their own [Shafer J C, 1996].

A. Importance of Business Intelligence in Organisations

When the companies are seeking cut in cost to regain the margins of profit the intelligent system of software provide better insight of the critical statistics of the data of company, which may be proven as a useful tool to build the future strategies of gain in profit. A company cannot analyze the data within the organization but can perform research market by collecting external data from web interface to analyze the trend with the help of software of Business intelligence. The software for Business Intelligence is the future of any business because it is not only the software but its intelligent system for the business provide total information of the business like personal data, analyze database data, supply chain information, sales marketing activity and customer database etc [Ferias A, 1998].

Applications of Business Intelligence are one of the systems and tools that play the major role in the process planning of business strategy. In addition, it allows the corporation to analyze, store, collect and access the necessary corporate data, which is helpful in the strategy of decision-making. The software for Business Intelligence is covering the areas of business like market research, customer, market segmentation, statistical analysis, profiling, products profitability and customer support etc. Business Intelligence system uses the data from the data mart or data warehouse of business. The corporation or an organization can use the

software solution of Business Intelligence for different purposes [Jain A K, 1988].

- Handle the consumer better
- To balance the expense streams and revenue
- Market Research
- Altering levels of staff
- Customer support
- To forecast the sales

Applications of Business Intelligence are used for different support of Management Information System. Business Intelligence helps top-level management for making strategies as the application of Business Intelligence provides some functions [Muller F, 1998]:

- Strategic planning process for an organization
- It performs optimal solution gathering, planning, data mining and data warehousing, financial analysis, etc [Mehta M, 1996].
- It provides critical Decision Support System for organization
- Gathers data from web interface along with Business Intelligence platform
- It runs against all Enterprise Resource Planning data sources

Typical software for Business Intelligence follows certain steps to carry out the meaningful information of data for an organization to use it for future decision-making and prediction.

- Analysis
- Planning
- Prediction
- Execution of reports
- Gathering of legends by AS

Software for Business Intelligence has not become very useful and important only for small organization but also for big software organizations like Microsoft, which admires and supports the support given by the applications of Business Intelligence [Chan P, 1993].

IV. NEED FOR APPLYING DATA MINING IN BUSINESS INTELLIGENCE

Data mining is a process by which raw data or the computer programs analyze large amounts of information. The computer programs employ different types of criteria to decide which information is important, to show the trends and to sort the information. In business intelligence data mining is an essential tool, due to the fact that the understanding trends help the managers to improve the share of market by capitalizing certain trends and avoids negative trends [Augural R, 1993]. Examples of this are:

- Sales Analyzing as per Date: A business use techniques of data mining to maximize the sales by increasing the

product availability that sells more during days of the week.

- Analyzing Website Traffic: Website owners analyze patterns of website traffic to determine which advertisements is more effective, based on overall success of the site, click patterns and time spent on each page. This allows the owners of site to remove the ads which improves the overall profitability of website ineffective and increases the effective ad campaigns [Haussler D, 1996].
- Analyzing patterns of Foot-Traffic: Owners of Casino uses analytic tools to find the patterns in choice of slot machines. If certain type of style of machine is picked up, regardless of placement, it is likely to be placed again. If certain location is favoured, regardless of the machine, the management uses that information to change the placement of the machines, and ensure that more traffic of gambling goes to more expensive machines. Retail establishment analyzes the patterns of foot-traffic in the store, compared with sales of various items in each location, to make decisions of placement of products [Chen M S, 1996].

The below figure shows the value created to an organization by the application of data mining in Business Intelligence:

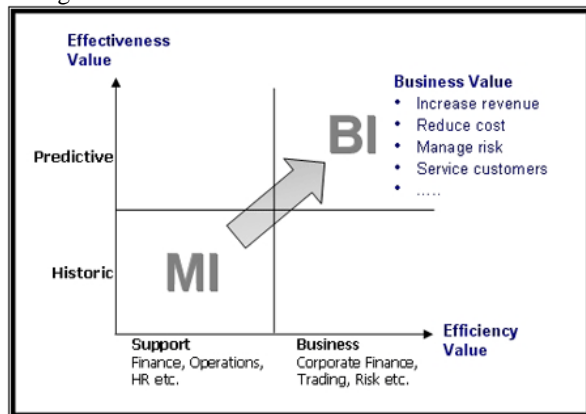


Figure 2: Value created by Business Intelligence

A. The process of using Data Mining for Business Intelligence

Business intelligence is information about company's past performance that is needed to predict the future performance of company. It reveals the evolving trends from which the company may get profit. Data mining allow users to put large amount of available information in data warehouses and it also put the process in which the gems of business intelligence are found [Hand D J, 1998].

Data mining is not a framework or intelligence tool. Business intelligence is drawn typically from an enterprise data warehouse, which is used to uncover and analyze the

information about past performance in an aggregate level. Business intelligence and Data warehousing provides a method for the users to foresee future trends from analyzing past patterns in the organizational data. Data mining is more spontaneous, that allows increased insight beyond data warehouse. Implementation of data mining in an organization will serve as a guide to uncover inherent tendencies and trends in the historical information. It also allows for classifications of data, statistical predictions and groupings [Mitchell T M, 1997].

Most companies deduce, gather and refine large quantities of data. Techniques of data mining is implemented rapidly in the hardware platforms and existing software to enhance the value of existing resources of information and can be integrated with new systems and products as they become a part of the system. When it is implemented on high performance parallel processing computer or client/server, tools of data mining analyze large databases to deliver the answers to different types of twist questions [Madigan D, 1996].

Software of data mining allows the users to analyze large databases to solve the problems of business decision-making. The tools of data mining predict future behaviours and trends, allows the business to make knowledge driven and proactive decisions. Data mining is an extension of statistics with few machine learning twists and artificial intelligence thrown in. Like statistics, data mining is not a solution for business it is a technology [Stafford B, 1997].

V. CASE STUDY- NETEZZA PERFORMANCE SERVER

An infrastructure of the Business intelligence company is a great challenge with the demand for analyzing and storing information. It is a technology of patchwork that has piled up for different reasons over a long period. Many flavours of the Database Management System software, is implemented partially as middleware strategy or, as a collection of various assorted disk arrays, mid-tier Symmetric Multiprocessing server and myriad applications of end-user depending on various communication and database standards. Costly system administrators and databases that are fighting to keep up with the demands of user hold this technology of patchwork together. As much information is added to existing systems, they are becoming unreliable and decreasing dramatically. The initiatives of online such as analyzing and capturing click stream data, threatens to defeat the infrastructure entirely. As the demands of Business intelligence and customer growth of data the Vice President of the Netezza Company, the CIO and the Customer Knowledge Management are facing costly systems administrators and DBAs make multi-billion-dollar investments in software, networking and storage and hardware to lose battle to maintain the present levels of performance [Website. Download 101].

This situation is common for 2000 companies across many industries. Over the past few decades, high profile initiatives of management strain the present infrastructure with their need for access to data across the enterprise are:

- Operations management
- Customer Relationship Management
- Enterprise Resource Planning
- Supply Chain Management
- Partner Relationship Management.

With all these stresses on the present infrastructure of Business intelligence it is not surprised that many technical pain points and business has evolved:

- Slow access of information inhibits spontaneous queries, which results in lost opportunities.
- Complex reports and queries require days or even hours to process.
- Speeds up processing, and the data are summarized and sampled limited analysis of depth.
- Collecting useful information requires training. A person who needs the data are not the same people who runs the queries. This results in lost opportunities and creates bottlenecks in the organization.
- Produces predictive models that is critical and often requires statistician's staff.
- Scalability suffers growth in databases beyond 100s of gigabytes.
- Costs increases discontinuously and unpredictably and the performance remains poor.
- Loading of data is painful and slow and requires medley of tools of Extract Transform Load tools;
- Queries run on outdated information yielding inaccurate and misleading results.
- Combining or standardizing data is difficult and time-consuming because data is held on different legacy systems and formats [Website. Download 101].
- Analysis of data marts from large warehouse data extract data is critical to achieve in an acceptable frame of time.
- Teams of system administrators and database are required to tune queries and systems to achieve acceptable performance.

The below figure shows the multiple sources which contributes to the stresses of infrastructure of Business intelligence:

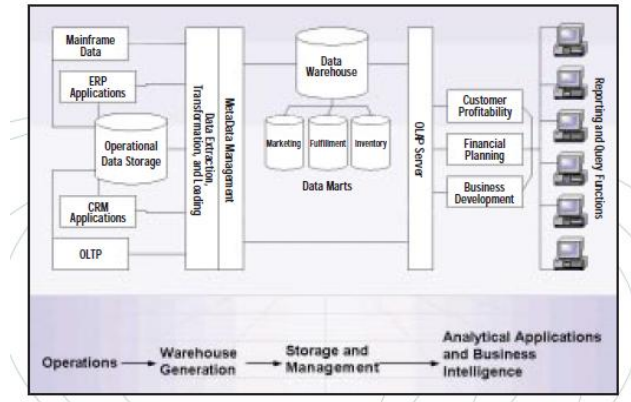


Figure 3: Existing BI Infrastructure

VI. FUTURE INFRASTRUCTURE OF BUSINESS INTELLIGENCE

Today, new platform of Business intelligence is needed to revitalize and rationalize the existing infrastructure. In today's business environment this new platform of Business intelligence provides a foundation for the growth of exponential data. It merges large parallel software and hardware and storage, which focus directly to provide optimal scalability and response times at the terabyte level. The new platform of Business intelligence eliminates yesterday's patchwork of storage, hardware and software, and enables optimized access to information. As technical and business demands continue to change and grow in the new century, the new platform of Business intelligence is designed to scale with the need, scope and performance and size of data. And it does all this in a predictable and affordable price [Website, Netezza].

The below figure shows the Netezza Performance server:

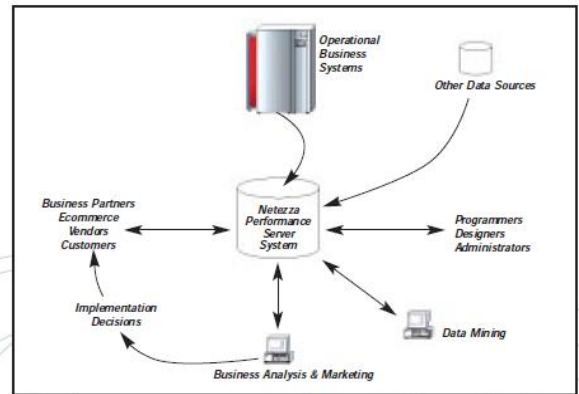


Figure 4: Netezza Performance Server Architecture

The new platform of Business intelligence offers large parallel processing, scalability and open architecture and allows growing without painful integrations of new systems or complex upgrades.

The Netezza Performance server of 7000 series is the realization of the new platform of Business intelligence. The Netezza Performance server system is an enterprise-class data warehouse appliance that delivers ease-of-use and breakthrough performance in a fraction of cost of traditional data warehouse. The Netezza Performance server appliance offers:

- 20 to 60 times the performance of existing data warehouse systems is half the price and finally affordable solution to users of business performance needs.
- Tight integration of database components, storage and server is largely parallel architecture that provides complex analysis and reports and rapid optimized execution of interactive queries.
- Extreme reliability and ease-of-use and a part-time DBA that requires to manage the system.
- Self-tuning data storage, real-time data loading and high throughput on complex and large workloads.
- Straightforward integration, through open Application Programming Interfaces with all main elements of the Business intelligence environment: Business intelligence tools and applications, Enterprise Application Integration and Extract Transform Load tools, data sources and legacy systems [Website, Download 101].

ACKNOWLEDGMENT

We are thankful to our Professors with whom we had useful discussions regarding data mining and business intelligence. Any suggestions to further improvement of this topic are most welcome.

REFERENCES

- [1] Haussler D, Mining Scientific Data, 1996.
- [2] Hand D J, Data Mining, 1998.
- [3] Mitchell T M, Mining of databases, 1997.
- [4] Madigan D, Statistical Inference and Data Mining, 1996.
- [5] Stafford B, Distributed Data Mining using Agent Based Architecture, 1997.
- [6] Chen M S, Data Mining: An Overview, 1996.
- [7] Mehta M, Fast Scalable Classifier for Data Mining, 1996.
- [8] Shafer J C, Data Mining, 1996.
- [9] Liu B, Post Analysis of data mining, 1996.
- [10] Download 101, Available at http://download.101com.com/pub/tdwi/files/BI_in_a_Real-Time_World_Sept04.pdf, accessed on 20th December 2010.
- [11] Srikant R, Mining Association Rules, 1997.