International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-1, January 2022
https://doi.org/10.55524/ijircst.2022.10.1.20
Article ID IRP221016, Pages 112-116
www.ijircst.org

Influence of Data Mining and Data Warehouse on Strategic Planning: A Review Paper

Pankaj Saraswat [1], and Madhav Singh Solanki [2]

[1, 2] Assistant Professor, Department of Computer Science Engineering, Sanskriti University, Mathura, Uttar Pradesh

Correspondence should be addressed to Pankaj Saraswat; pankajsaraswat.cse@sanskriti.edu.in

ABSTRACT: Users can access their data in today's reporting environment, but it does not address all of their issues. The people have access to the statistics, but they cannot ensure the data's truthfulness or the speed with which it is returned. Data warehousing addresses the aforementioned issues by providing technology that allows users or decision makers to analyze large amounts of data in a short period of time. Data warehousing allows users to extract information in real time, which assists them in making decisions. Many businesses wish to utilize the information for additional reasons. As a result, methods for mining fresh information from data warehouses have developed. Data collection and data excavation provide the groundwork for developing choice provision and decision-making information system tools that track an organization's progress concerning its objectives. Data mining and warehousing are technologies that allow a user in the business sector or government to analyze large amounts of data and make choices that benefit the whole organization. This article uses appropriate graphics to explain the impression, benefits, and drawbacks of data mining and warehousing. The duties and responsibilities of data warehousing organizational members are also addressed in this article. To sum up, we're attempting to demonstrate how “Data Mining (DM) & Date Warehouses (DW)" may be utilized in companies, how data can aid making decisions, and how managers can conduct extra precise, meaningful, and consistent analysis using their data.

KEYWORDS: Data Marts, Data Mining, Data Warehousing, Knowledge Discovery, Staging Layer.

I. INTRODUCTION

The idea of DM & DW are gaining traction as a corporate statistics governing utilities that is anticipated to reveal information structures that may guide choices in uncertain situations. By establishing an enterprise-extensive cohesive databank of abridged, historic data, a DW aids commerce investigation and decision-making [1,2]. It combines data from a variety of disparate and incompatible sources. A data warehouse enables a management to conduct more substantial, accurate, and consistent analysis by converting data into useful information. Figure 1: Problem in Decision Making [3]. The DW is not a traditional database in the sense that we interpret the word. The major distinction is that conventional databases often store operational-type, transactional-type data, and that many judgement backing systems place an excessive amount of pressure on databases by interfering with day-to-day operations [4]. Of course, a data warehouse is a database, but it includes summarized data. A data warehouse is a database that is kept apart from the company's operating databases. Read-only data is stored in a warehouse. Data mining is also known as database knowledge discovery or knowledge discovery. Data mining is a powerful new technique that may help businesses concentrate on the most essential information in their data warehouses by extracting hidden predictive information from huge datasets. Data mining software forecasts future trends and behaviors, enabling companies to make informed, proactive choices. Data mining technologies may provide answers to business problems that were previously too time consuming to address. They search databases for hidden patterns, uncovering predicting data that specialists may overlook since it falls outside of their usual scope.

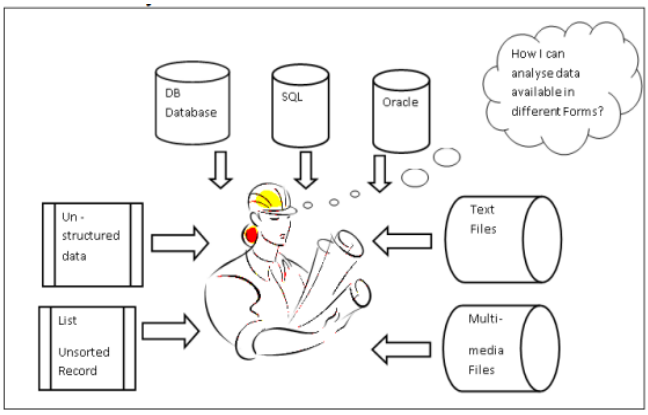

Figure 1: Problem in Decision Making [3]

A. Data Warehousing (DW)

A DW is a set of linked databases that help a DSS work. It's a gathering of subject-oriented databanks that work together to serve the DSS functions, with every data unit being nonvolatile and pertinent to a certain point in time [5]. To make data warehouse initiatives effective and produce a return on investment, several roles and responsibilities will need to be accepted. It is further suggested that the subsequent responsibilities be done full-time by devoted people as much as feasible by technical professionals, and that each accountable individual undergo specialized Data Warehouse training. The DW team must head the institute in accepting their responsibilities, resulting in a business partnership. Management must also turn these instructions into concrete strategies and ensure that the workforce follows through. The entire team with their each and every members along with their duties to make the data warehouse successful and beneficial to users and organizations are listed in Table 1 [6].

Data mining is a valuable method for extracting values from a data warehouse for usage in judgement backing. When data warehousing becomes a regular component of a business, new methods to utilize the data will be sought. At some point, data warehousing and data mining will face a number of additional difficulties, including: 1. Regulatory restrictions may limit the capacity to integrate different data sources. 2. Unstructured data, which is difficult to store, is likely to be included in these diverse sources. 3. Data may be accessed from practically "anywhere" thanks to the internet. This just adds to the inequality. Today's task is just to create DW & DM systems that are dependable, simple for consumption, and help people make better decisions. DM & DW will become more useful in sectors and businesses as the quantity of data grows. Data mining will aid in the discovery of new high-quality goods, the prediction of advantages from high-quality data, and the optimization of sales resources such as personnel and marketing.

Table 1: Illustrates the members and their roles [6].

Member	Role
Manager/Director	The data warehouse manager or director ensures support for the data warehouse program at the highest levels of the organization and understand high level requirements of the business. Manager staff the team and ensure adherence to a set of guiding principles for data warehousing.
Project Manager	Project managers delivers commitments on time. Project managers maintains highly detailed plan and caring about progress on it. Project manager matching team member's skills and issue list of tasks to them.
Chief Architect	The Manager/Director of data warehouse will need to rely on a Chief Architect position, as one of his/her direct reports, to work on complex issues of architecture, modeling, and tools. Chief Architect would have significant interface with the internal clients and increase their confidence in the data warehouse organization. Chief Architect should have great knowledge of business.
End User	Data warehouse is made to meet end users requirements. Data warehousing is used to answer the end users queries and generate reporting. End user receive ID and password on the data warehouse system and provide feedback to the data warehouse team like performance, functionality, data quality, metadata quality and completeness.
Database Administrator	Data warehouse group is the placement of the database administration function and the division of roles and responsibilities between the support group and the user community. Database administrator has many responsibilities like database maintenance, backup and recovery, data replication, Performance Monitoring and Summary table creation.
Application Programmer Specialist	The Data Warehouse Application Programmer is responsible for applying transformation rules as necessary to keep the data clean and consistent. Application responsibilities has many responsibilities like sourcing the data from operational systems, applying the business transformation rules.

The data is categorized according to how the users refer to it.

1. Data Sources:
   It is gathered from external as well as internal operating systems (OS) that are not linked.
2. Integrated:
   All naming convention and value representation discrepancies have been eliminated.
3. Big Data:
   Time series data sets are usually very big.
4. Metadata:
   Information about information is preserved.
5. Non-volatile:
   Data is kept in a read-only format that does not alter with time.
6. Not regularized:

DW data sometimes may be redundant, and it often is.

7. Summarized:
   Operational data is mapped into a format that may be used to make decisions.
8. Time Variant:
   Data are usually time series rather than current data.

II. DISCUSSION

A. Data Warehousing: Structural Design and Functioning

The term "data warehouse" refers to a databank that is utilized for reportage and investigation. It is a location where data is kept by combining several databases. It has the ability to store both current and historical data. New predictions may be made using previous and current databases. Figure 2 depicts several data warehouse compounds. A data warehouse is a databank that is mainly utilized for data processing and reporting. It is a central data source that is produced by combining information from one or more distinct sources. Data is received from operating systems and stored in the warehouse. The staging layer very effectively collects the data in raw form from each of the many available data systems source and stores it. The integration layer brings different data sets together by converting data from the staging layer and putting it in an operational data store database. A data mart is a tiny form of DW that focuses on a particular topic. For better performance in usage, data warehouses may be split into data marts. One or more data marts may be used to build a bigger and more sophisticated corporate data warehouse. A Data Warehouse helps business users save time and produce reports more rapidly. Business users may access all of these information in one location and make fast choices. Business users will save time by not having to gather data from numerous sources. Businesses may query data themselves with the assistance of data warehousing, saving money and time.

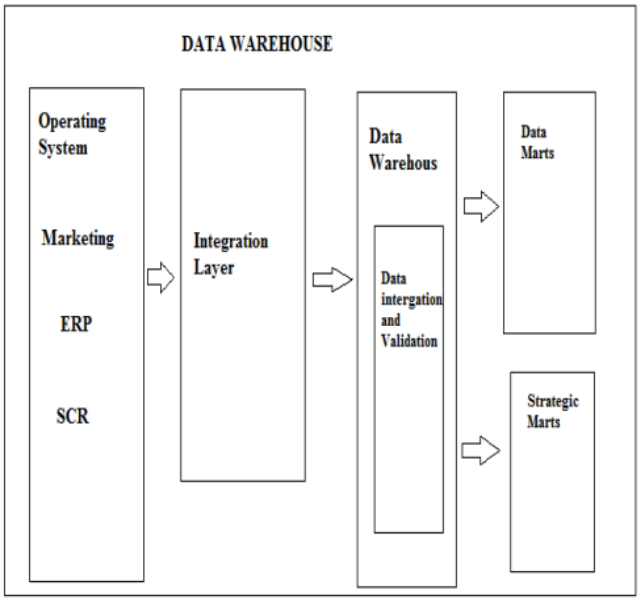

Figure 2: Components of Data warehouse [7]

B. Data Mining

Data mining programs are available for mainframe, client/server, and PC platforms in a variety of sizes. Data base mining, often known as data mining, is a method of using current data to discover new facts and connections. Issue examination, data abstraction, data cleaning, guidelines creation, yield examination, and review are all stages in the data mining process [8,9]. Flat files retrieved from on-line collections of files, data warehouses, or other data sources are common data mining sources. Data, on the other hand, may be obtained from virtually any source. DM is often an repetitive or loop based process including these stages, regardless of the data source. The stages of data mining are as follows:

a. Uniqueness Identification of the Objective

Make sure you know what you want to get out of your analysis before you start. Know what the data mining's business objective is ahead of time. Determine if the objective is quantifiable or not.

b. Selection of the Data

After you've established your objective, the following step is to choose the data that will help you achieve it. This might be a subcategory of your DW or a DM with product-specific data. It's possible that it's your customer information file. The possibility of the facts to be mined should be segmented as much as feasible. Here are a few important factors to consider: 1.Is the data up to date and relevant to the objective of profession? 2. Is the facts steady, in the sense that the mined characteristics will remain identical even after the examination?

c. Accumulation of the Data

After the author gathered the information, they need to determine which characteristics to convert into useable forms. Take into account the opinions of domain experts/creators as well as data consumers. Create methods for dealing with missing data, noise, and outliers. If required, choose between a log and square transformation. Regulate the data's circulation frequencies.

d. Evaluation of the Data

Examine the arrangement of the information. What is the database's content and composition? What is the dataset's general state and distribution?

e. Choice of Appropriate Tools

Business goals and data structure are two key aspects to consider when choosing the right data-mining technology. Both of these methods should get you to the same utility. To answer the questions, no one tool is recommended.

f. Indented Solution Preparation

Answers are first found out of topics like: What format choices are available? What is the resolution's objective? What is mainly needed by the end-users in terms of reports, graphs, and code?

g. Desired Model Preparation

The DM process may now commence. The user divides data into sets, builds the model, and evaluates it. This step involves the creation of classification rules, decision trees, clustering sub-groups, scores, code, weights, and evaluation data/error rates.

h. Findings Checking and Validatation

The findings of the study should be shared and discussed with the business customer or domain expert. Ensure that the results are accurate and relevant to the company's goals. Find out the answers to a variety of questions, such as "Do the results make sense?"

i. Reporting the Finding

Prepare a final report for the customer or business unit. The report should include information on the data preparation, tools utilized, test results, source code, and rules followed throughout the data mining process. This report aids in decision-making and contributes to the organization's development.

j. Combine components to integrate the solution

All interested end-users should be informed of the results. You may end up integrating the analysis' findings into the company's operational processes. Although data mining technologies automate database analysis, if you're not cautious, they may lead to incorrect results and conclusions. Individual transactions in operational databases may be mined using data mining. Data mining is used in both the commercial and public sectors, including banking, insurance, pharmaceutical manufacturers, health care providers, and retailing, to decrease costs, improve research, forecast the efficacy of a treatment or medication, and boost sales. Data mining is used to forecast future trends, consumer purchasing patterns, and make better decisions. Data mining increases income and reduces expenses for businesses. Data mining is also utilized in market analysis and fraud detection. However, data mining has a number of drawbacks. Data mining raises concerns about privacy and security. Data mining may be expensive when it comes to implementation. Data mining raises concerns about privacy and information abuse [10]. Data mining can't guarantee flawless findings, can't explain why something happens, and can't fix issues with your data.

B. Role of Data Warehousing and Data Mining In Decision-Making

A data warehouse's purpose is to aid in data-driven decision-making. To assist with specific kinds of choices, data mining may be used in combination with a data warehouse. To be effective, data warehousing and data mining need a knowledgeable user who can provide accurate data and a professional who can draw impartial conclusions from the result [11]. If the user provides inaccurate or insufficient information, the output will be harmed, and the prediction will be untrustworthy. The organization's decision-making relies heavily on data warehousing and data mining. Data warehousing assists in decision-making by providing answers to numerous questions for the business and the user. There are many kinds of organizational inquiries, including tactical, strategic, and update questions.

A tactical query is a database procedure that tries to figure out what the best course of action is at the moment. A tactical query gives information to rank and file components in the field that need to react rapidly to a set of unfolding events, while a strategic query provides information to make long-term business decisions. Tactical searches often provide a limited number of results. The result set is often fewer than a dozen rows long. Typically, the result set is intended to fit into a single display window. A strategic query is a database activity that tries to figure out what occurred, why it occurred, and/or what will occur next. It usually requires access to a large quantity of detailed data from the warehouse and may vary in complexity from basic table scans to multi-way joins and sub queries. Report creation, OLAP, decision support, ad-hoc, data mining, and other applications produce strategic questions. A database procedure that changes the state of a database is known as an update query. Teradata offers a collection of bulk load tools that may be used to quickly load huge amounts of data into the database.

III. CONCLUSION

The term Data Mining and Warehousing refers to the current shift in corporate practices. To determine their own business trends, all small and large industries gather and use data from different sources. Organizations that understand their competitors' strengths and shortcomings enhance their progress toward the objective and grow their commercial empire. A DW is a business elucidation, not a technological solution. Data warehousing and data mining must continuously overcome unknown hurdles in order to assist organizations in making decisions and enhance their reputation. DM aids in protecting and treating data into digestible pieces, while warehousing aids in evaluating data and putting it in a format that facilitates trend comparison, data analysis for business forecasts, and decision making. In a nutshell, data warehousing and data mining include accurately converting data from different source systems into a common format, assisting the company in making sound business decisions, and assisting in the expansion of the corporate empire. Consistency and data quality are improved by using a data warehouse. Because all of the data from the different departments has been standardized, each department will provide findings that are consistent with the others. It is informative and well-organized. One of the most useful features of data warehouses is that they allow data from many sources to be integrated in one place.

REFERENCES

[1] Moscoso-Zea O, Andres-Sampedro, Luján-Mora S. Datawarehouse design for educational data mining. In: 2016 15th International Conference on Information Technology Based Higher Education and Training, ITHET 2016. 2016. Zdenka P, Petr S, Radek S. Data analysis: Tools and methods. In: Recent Researches in Automatic Control - 13th WSEAS

International Conference on Automatic Control, Modelling and Simulation, ACMOS'11. 2011.
[2] Mawilmada PK. Impact of a data warehouse model for improved decision - making. Star. 2011;
[3] Gardner SR. Building the data warehouse. Commun ACM. 1998;
[4] Fang Y. A DSS assistant model for college counselors based on data mining. In: Proceedings - 2017 International Conference on Smart Grid and Electrical Automation, ICSGEA 2017. 2017.
[5] Arif M, Zahid S, Kashif U, Ilyas Sindhu M. Role of leader-member exchange relationship in organizational change management: Mediating role of organizational culture. Int J Organ Leadersh. 2017;
[6] Domingues MA, Soares C, Jorge AM, Rezende SO. A data warehouse to support web site automation. J Brazilian Comput Soc. 2014;
[7] Data mining in healthcare decision making and precision. Database Syst J. 2016;
[8] Lei X-F, Yang M, Cai Y. Educational data mining for decision-making: a framework based on student development theory. In 2017.
[9] Liu L. Privacy preserving data mining for numerical matrices, social networks, and big data. Diss Abstr Int Sect B Sci Eng. 2016;
[10] Kłodawski M, Lewczuk K, Jacyna-Gołda I, Zak J. Decision making strategies for warehouse operations. Arch Transp. 2017;