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.2022.10.1.17
Article ID IRP221013, Pages 97-101
www.ijircst.org

Educational Data Mining and Data Warehouse Design Using Business Intelligence

Pankaj Saraswat[1], and Swapnil Raj[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: Business intelligence (BI) is basically a collection of programs, utilities and apps that allows you to analyze large quantities of data (Big Data). Every year, educational institutions deal with enormous amounts of Big Data. The use of business intelligence (BI) in these organizations is critical for improving processes and supporting decision-making. A Datawarehouse is the foundation of every BI project (DW). The scheme deliberations for implementing DW in any of the institutional setting are described in this article. Using educational data mining (EDM) methods, the DW shall be utilized in an information unearthing procedure to manage the statistics and facts for the study of vital accomplishment metrics. A knowledge management framework's main technical assets are a data warehouse (DW) and an enterprise architecture (EA) repository (KMF). The agenda further was created to provide order to the processes of knowledge generation, capture, transfer, and digitization. This handbook and the framework are two of the results of a private university's research effort. In addition, an illustrative example demonstrates picking of the finest technique in higher education. The stages for DW design are given in the case study. This research may help academics and practitioners who want to create a data warehouse to analyze data utilizing EDM methods.

KEYWORDS: Business intelligence, Data Warehouse, Datawarehouse, Educational data mining, Knowledge management.

I. INTRODUCTION

Every day, massive quantities of data are produced in teaching-learning organizations. The correct utilization of such data may be critical for meaningful information or knowledge generation. In research and academia, organizational knowledge may help with strategic planning and improving key performance metrics. The usage of business intelligence (BI) and related technologies such as a datawarehouse (DW) and data mining program or algorithm is suggested for transforming data into knowledge [1]. In a previous paper, a knowledge management framework (KMF) too is suggested in order to govern process of knowledge production and dissemination[2]. This paradigm recommends combining business intelligence (BI) with get-up-and-go design also referred to as enterprise architecture (EA) to capture all of an organization's knowledge aspects. People, technology, and processes are the three dimensions. The use of this KMF at educational institutions may help to improve a variety of procedures and services. In addition, the framework provides web components for visualizing data from EA and BI sources.

Many articles discuss data mining, along with the concept of educational data mining (EDM), and the data warehouse (DW), which is used to analyze and handle multidimensional data [3]. Despite this, only a few publications discuss design issues for this kind of DW. This article provides a framework for creating a DW in educational settings. In addition, the process of extracting, transforming, and loading (ETL) statistics from functional sources of information into a DW is explained. In addition, the article describes the procedures taken to create a DW in a privately owned institution as a case study. A DW is yet another of the planned KMF's main repositories, and analyzing the data stored there is a realistic method to detect flaws in institutional processes and offer feasible and suitable remedies to improve instructive and administrative enactment. In addition, this research examines the key distinctions between conventional DW design and a similar kind of DW scheme in instructive settings.

A. Background

The goal of this study is to look at the design aspects for implementing a DW in teaching-learning organizations. The state of the art of the subject, as well as the technologies and components required for the design of the DW, are discussed in this part.

a. Business Intelligence

Business intelligence (BI) is a system made up of each technological and institutional components which mainly provides past data to users for investigation and facilitates better choice building and supervision assistance with the goal of improving overall organizational performance. BI is based on a collection of tools and apps that allow for the analysis of large quantities of data (Big Data) in order to

enhance organizational policymaking and enactment. To achieve this goal, judgement producers must have right of entry to all of the establishment's facts & figures in order to evaluate the company, its needs, and its tendencies. As said by Gartner, the global business intelligence market will spread $16.9 million in 2016, up 5.2 percent from the preceding year [4]. All BI platforms, management suites, and advanced analytics solutions are included in the scope of this market. The production of facts to allow well-versed policymaking is one of the most important uses of BI. Knowledge is created by gathering data from general public, procedures, and tools and putting it into a system that can analyze it.

b. *Datawarehouse*

A data warehouse (DW) stores huge quantities of data in order to analyze and manage historical data effectively. The use of this technology provides a chance for businesses to consolidate important data and create helpful knowledge. The DW is a information fountain that is filled by combining several functional information foundations retained by various business units within an organization [5,6]. When it comes to developing a DW, experts have suggested a variety of methods. However, two of them stick out. Bill Inmon and Ralph Kimball suggested these methods. Practitioners and researchers may select which one best fits a given situation based on the project's needs.

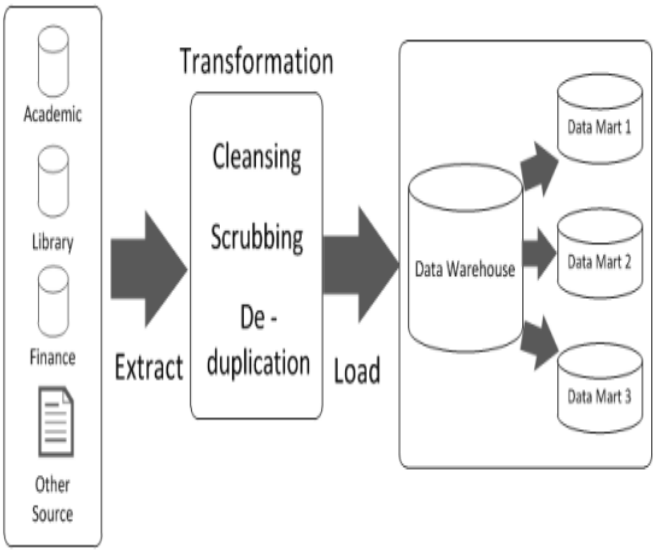

Figure 1: Bill Inmon's approach for Datawarehouse strategy [7]

The technique utilized to construct the DW is the primary distinction between these methods. As shown in Fig 1, Inmon offers a top-down method, while Kim put forward a bottom-up methodology, as seen in Fig 2. Because all of the data marts are fed from the principal warehouse of data, Inmon's approach offers a highly reliable representation of facts through the data stores. Because it considers the whole company, the top-down strategy is adaptable in supporting change management.

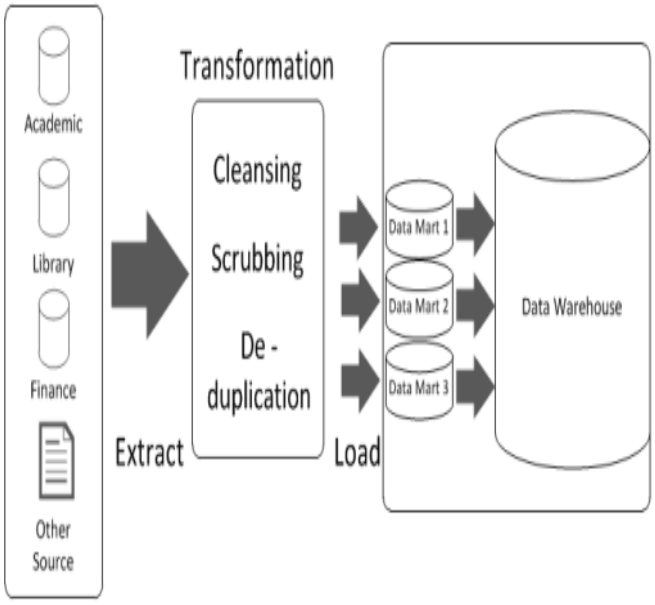

Figure 2: Ralph Kim's approach for warehouse scheme [7]

Kim's (bottom-up) approach, also known as dimensional modeling, emphasizes the worth of the DW to managers in the initial phases thru establishing information fountains (data marts) [8]. His idea is to create separate data marts for each business function, according to his study. To provision the investigative and reportage arrangements, the data put into each data mart is a kind of a duplicate of the online transaction processing (OLTP) data specially formatted aimed at systematic inquiring and reporting. These data marts may ultimately be merged to form a comprehensive corporate data warehouse. The major difference between the two methods is that, while Inmon's methodology produces a get-up-and-go DW first, Kim's method produces data marts designed for the professional entities prior to creating an enterprise DW, as seen in the figures. Table 1 lists the most significant differences between the two methods. It's difficult to determine which DW design approach is better based only on Table I. In spite of the statistic that Table 1 provides a kind of guidance on how to go about building a DW, the strategy to take is highly dependent on the institution's business goals, commercial prototype, budget, and the degree of interdependence among various commercial units. Inmon's strategy is best suited to large, stable businesses that can afford to spend more money and time building enterprise data warehouses. Furthermore, since the technique is adaptable and the data model is easily rebuilt or reorganized, this methodology may be utilized in companies with changing business contexts. When time and money are limited, however, Kimball's approach is suggested. This is due to the fact that data marts may be built incrementally.

Table 1: Variances among Inmon and Kim DW scheme [9]

	Differences	
	Inmon	**Kimball**
Source required	Yes	Yes
Staging	Yes	Yes
ETL	Yes	Yes
Data marts	Yes	Yes
Business requirements	Yes	Yes
Time attribute of data	Yes	Yes
Enterprise DW	Yes	No
Dimensional tools	No	Yes
Relational tools	Yes	No
Process oriented	No	Yes
Normalized data model	Yes	No
Complex to design	Yes	No
Continuous and discrete time frame	Yes	No
Slowly changing time frame	No	Yes

Table 2 lists further advantages and disadvantages of both methods. The following are the most significant characteristics listed in the table: Data incorporation by means of Inmon's method necessitates an organization-wide process, while data integration using Kimball's approach is done on a per-business-area basis. Another significant factor is that Inmon's approach has a greater starting cost and implementation time than Kimball's methodology. Despite the fact that Inmon's approach takes more time and requires a larger budget to begin a BI project, implementing it may end up with a powerful data storehouse for exploration that spans all of the organization's enterprise scopes. With all of these factors in mind, the BI analyst must evaluate situation earlier to determining which technique is appropriate for the DW or BI project. Regardless of the technique chosen, there are certain activities that are common to all methods and are detailed below. The following are some of the stages involved in a BI project: diagnosis, information requirement analysis, methodology selection, technical infrastructure setup, data warehouse architecture, and data ETL. After the data has been loaded, analysis requires the use of control panel and broadcasting tools. This research aids in the identification of important data for the production of knowledge. Powerful analytical tools are required for effective data analysis. Online analytical processing (OLAP) and data mining tools are the two major kinds of analysis tools. OLAP technologies offer access to business information via multi-dimensional understandings of comprehensive data in order to improve decision making. To assist strategic choices, data mining employs software methods to uncover hidden patterns and trends in big datasets.

Table 2: Inmon's vs Kimball's methodologies [9]

	Inmon	**Kimball**
Building datawarehouse	Time consuming	Takes lesser time
Maintenance	Easy	Difficult, often redundant and subject to revisions
Cost	High initial cost. Subsequent project development costs will be much lower	Low initial cost. Each subsequent phase will cost almost the same
Time	Long start-up time	Shorter time for initial set-up
Skill requirement	Specialist team	Generalist team
Data integration requirements	Enterprise-wide	Individual business areas

Researchers at educational institutions may utilize a number of implements and EDM approaches and modus operandi to excerpt data that isn't apparent at first sight. This may raise attentiveness of the institution's present condition and offer methods for predicting future trends, allowing academic and organizational performance to improve.

c. Extraction Transformation and Loading (ETL)

ETL basically refers to a procedure of dig out data from dispersed applications in business and departmental units throughout the company and importing it into the DW. The extraction of data from operational data sources is the initial stage in the process. Normally, these data sources are databases, although data may also be saved as flat or XML files. The primary goal of this phase is to transform the information into a solitary arrangement that can be used for transformation and loading. The alteration procedure necessitates data adaptation and normalization. With ETL software, this process may be automated. The program may be used to apply a set of guidelines and utilities to the data that has been extracted. This set of criteria guarantees that the data is in the proper format and free of mistakes. Data cleaning is the name for this stage of the transition. Loading is the final stage in the ETL procedure. The procedure imports the data into the warehouse that will be used as the final destination. The frequency of load is determined by the data base administrator based on the organization's needs and the criticality of the analysis. The ETL procedure is critical to a BI project's success. The quality of the data put into the DW has a significant impact on the outcomes of the data or business analysts' analysis activities.

d. Educational Data Mining

It is a growing field that permits for the discovery of information from informative settings via the development and application of DW methodologies and practices to data storehouses. The data is mostly stored in DWs or enterprise design storehouses. The use of data mining methodologies and procedures enables institutions to get a better understanding of their students, instructors, and learning styles. One of the most important aspects of the knowledge discovery process is data mining (KDP). KDP may be used in a variety of business settings. Fig 3 shows a KDP in use at teaching-learning institutes. The KDP collects information from educational settings. This information is preprocessed before being utilized for analysis. This updated information may be stored in the DW and examined using various data mining methods.

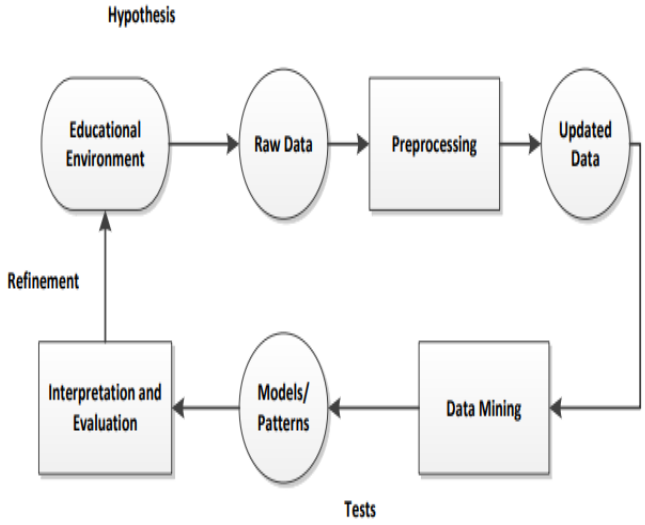

Figure 3: Information unearthing process in establishments related to education [10]

EDM examines the milieu of the organization & is a crucial module of the planned KMF on behalf of judgment production to enhance procedures, amenities, and signs including pupil dropout, qualification percentage, restructuring processes, and organizational management. Learning analytics is a subfield of MDE that is supplementary to MDE (LA). MDE attempts to find information via computerized errands backed by mortal verdict, whereas LA emphasizes information unearthing based on human judgment and depends on automation technologies.

II. DISCUSSION

A. Data warehouse Design

The following is the primary research topic that this article attempts to answer: What are the design considerations for implementing a DW in a learning environment? The study performed in order to propose a design approach for educational institutions is given in this part. The suggestions and design considerations for establishing the DW project and the ETL procedure are also discussed. The stages in this analysis are as follows:

a. Diagnosis

Definition of the organization's present condition, the scope of the BI project, and the stakeholders to be identified. The European Foundation for Quality Management (EFQM), the Open Group Architecture Framework (TOGAF), excellence model, the Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis, and business process models (BPM) are just a few examples of management frameworks that can be used for this purpose.

b. Information need analysis

Interviews are the most common method for this phase. There are multiple studies that must be carried out. First, an examination of the organization's information supply. This involves a review of all relevant sources of data that are relevant to the project's scope. Second, an examination of the objective information requirements. This necessitates an examination of the organization's business models, procedures, and strategy. Finally, a qualitative data need assessment must be completed, which contains all of the stakeholders' stated needs.

c. Selection of the methodology

The DW design approach must be chosen at this point. We suggest analyzing and matching the features of the situation with the guidelines provided in Table 2 after the project scope has been established and the needs have been clearly recognized. This stage's output is the methodology that will be used in subsequent steps.

d. Setting up the technological infrastructure

This phase requires the selection of all BI tools, databases, reporting, and data mining software. A feasibility study should be carried out before deciding on the technical infrastructure.

e. Datawarehouse design

The DW's conceptual, logical, and physical design must be modeled once the technical framework has been set up. It is highly suggested that you utilize CASE software for this. A star schema or a snowflake architecture should be developed depending on the project's objectives and approach.

f. ETL execution

The ETL of information from transactional cradles of data into the target DW is the next stage in the project. This phase involves identifying data sources (internal and external), as well as extracting data to intermediary files or straight to the destination database. A data cleaning procedure is required in both instances. Last but not least, the intermediate files must be loaded into the target DW.

CONCLUSION

The contrasts between current DW design methods are discussed in this article. We discovered that the worth manacle is the utmost significant feature that affects the

apparition of the enactment venture when comparing the variances amongst a commercial DW design and an enlightening DW strategy. In the literature, Kim and Inmon are among the supreme prominent and recommended methods. We've discussed the distinctions between both methods and when to use one over the other in this article. This article will hopefully provide as a everyday guide for experts and academics interested in implementing a DW and experimenting with EDM techniques in educational settings. Data marts may be implemented in accordance with the strategic objectives of information analysis. Furthermore, changes in the organizational environment may be implemented solely to the data marts schemas that are implicated, making change management simple.

REFERENCES

[1] Business Intelligence & Knowledge Management - Technological Support for Strategic Management in the Knowledge Based Economy. Inform Econ J. 2008;

[2] Al-Ahbabi S, Singh SK, Singh Gaur S, Balasubramanian S. A knowledge management framework for enhancing public sector performance. Int J Knowl Manag Stud. 2017;

[3] Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications. 2014.

[4] Gartner Says Worldwide Business Intelligence and Analytics Market to Reach $16.9 Billion in 2016 [Internet]. [cited 2018 Sep 10]. Available from: https://www.gartner.com/en/newsroom/press-releases/2016-02-03-gartner-says-worldwide-business-intelligence-and-analytics-market-to-reach-17-billion-in2016

[5] Agarwal AK, Badal N. A novel approach for intelligent distribution of data warehouses. Egypt Informatics J. 2016;

[6] Santoso LW, Yulia. Data Warehouse with Big Data Technology for Higher Education. In: Procedia Computer Science. 2017.

[7] Maślankowski J. The Evolution of the Data Warehouse Systems in Recent Years. Finanse. 2013;

[8] Moscoso-Zea O, Paredes-Gualtor J, Luján-Mora S. A Holistic View of Data Warehousing in Education. IEEE Access. 2018;

[9] Kimball vs. Inmon Data Warehouse Architectures [Internet]. [cited 2018 Sep 10. Available from: https://www.zentut.com/data-warehouse/kimball-and-inmon-data-warehouse-architectures/

[10] KumarYadav S, Pal S. Data Mining Application in Enrollment Management: A Case Study. Int J Comput Appl. 2012;