International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Design & Develop: Data Warehouse & Data Mart for Business Organization

Farhad Khoshbakht[1], Atena Shiranzaei[2], S. M. K. Quadri[3]

Submitted: 01/11/2022 **Accepted**: 03/02/2023

Abstract: Data warehouses (DW) are the foundation of business intelligence (BI) data storage. Business organizations utilize DW to help decision-making processes as large and complicated data sets must be examined and analysed. The analytical processing technology presupposes that data are presented as straightforward DM (data marts) with a well-defined set of facts and data analysis dimensions (star schema). Despite the widespread adoption of data warehouse technology and concepts, it becomes complex for the designers to identify and extract DM from an information system. This study strategy uses three fundamental steps for the methodology: top-down analysis, bottom-up analysis and integration. Though the method is not fully mechanical, it offers more direction than earlier methods to DW and DM designers.

Keywords: *DW, DM, BI, Top-down analysis, Bottom-up analysis and Integration*

1. Introduction

A set of tools and technologies known as a BI (Business Intelligence) system enhances managerial decision-making efficiency and offers significant commercial value. BI generates current data for tactical and strategic decision-making (Shanks & Bekmamedova, 2012a). BI is the process of gathering, examining, and analysing data to transform it into actionable knowledge and gain a competitive edge. Businesses use BI technology to maximise the value of their data storage. Almost every aspect of the business is covered by BI, including problem-solving, loyalty and customer value, up-selling and cross-selling, customer churn, segmentation, cost control, revenue and profitability, and any other area of performance improvement.

Data warehouses are the foundation of data storage in business intelligence (Wang et al., 2016). A DW is storage for historical and current business data, which helps corporate executives to get the data they need to evaluate critical business processes. Such analysis is used in the organization's decision-making process. By following a process of identifying the data items that are required to be taken into the data warehousing system based on their utility in the decision-making activity of the organisation, data in the DWs come from the operational system of the organisation as well as from data sources that are external to the organisation. Sales departments, for instance, use DWs to research the purchasing habits of their clients and then adapt their commercial and distribution tactics as necessary (Faisal et al., 2017).

The DW element, a database optimised for analytical processing, has the primary objective of storing and analysing historical information. DWs are often configured as distinct, problem-driven components known as "Data Mart (DM)," which each is committed to the investigation of a specific issue. This organisational structure helps standardise data analysis and facilitates simpler usage patterns.

A DM of organisation, sometimes called "a star schema", is relatively simple; facts which are being studied form the star's centre, while other information specify the parameters along which analysis of data is feasible around the star (Ramesh et al., 2020a). The sales made by a company are the facts that make up the archetypical situation, and dimensions enable research to be performed on the sales based on product, client, point-of-sale, time of sale, as well as other factors. The architecture of the DW is more complicated, and the information of the DM could potentially be loaded from intermediary repositories, which are typically known as "operational data stores." A basic framework of the enterprise-wide data system, or at the very least the applicable schema segment, is the starting point for DM architecture. To enhance the data interoperability at the operational level, an integrated schema is built even before DWs are developed.

1Department of Computer Science Jamia Millia Islamia (A Central University), New Delhi, India 110025
1f.khoshbakht630@gmail.com
2Department of Computer Engineering, Faculty of Industry and Mining (Khash), University of Sistan and Baluchestan, Zahedan, Iran 2ashiranzaei@eng.usb.ac.ir
3Department of Computer Science Jamia Millia Islamia (A Central University), New Delhi, India 110025
3quadrismk@jmi.ac.in

DM can help departments make tactical decisions, whereas DWs provide a comprehensive picture of all organisational data for strategic business decisions (Božič & Dimovski, 2019). It is the core idea of our recommended method that DW design should be guided by business demands.

2. Dw And Bi Working Within Business Organization

Data Sources: Sources from which data will need to be retrieved and uploaded to the warehouse (or its subcategories, "DM," which contain data for particular business activities of departments) are the first aspect of data warehousing that has to be handled. This usually entails identifying the important stakeholders and the reports that must be directed toward the DW. It will be largely self-explanatory. Using the CRM for marketing or the ERP for accounting, respectively. Some will be more difficult to spot and may involve less obvious aspects of data that may need to be reported, such as customer phone calls or email correspondence.

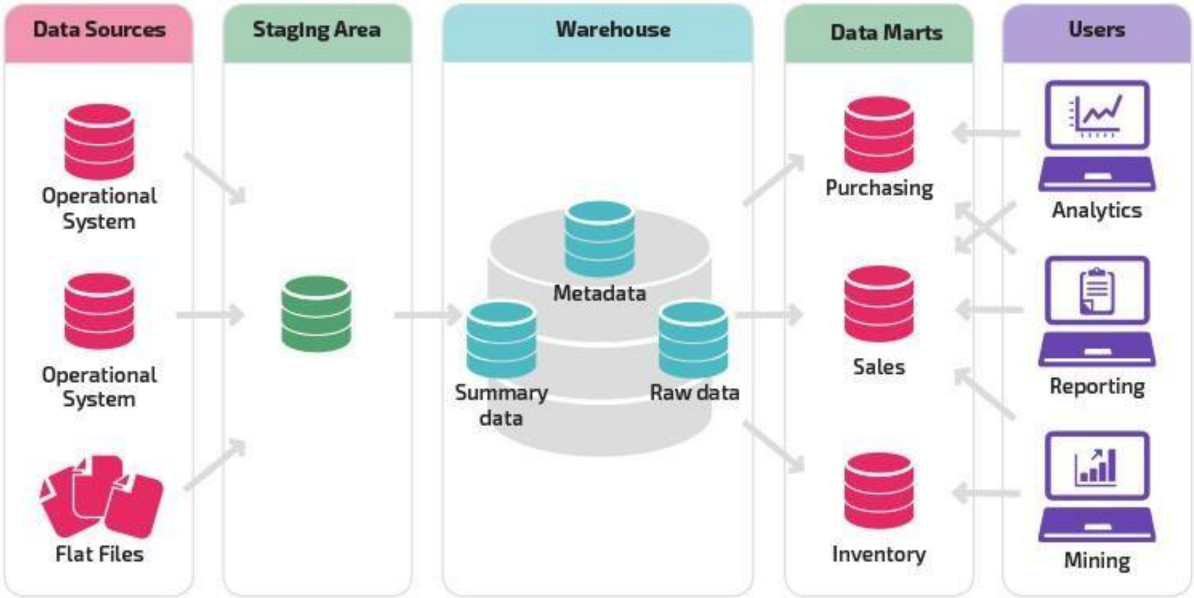

Fig. 1: BI Framework (Sharda et al., 2015)

DW: It's time to extract and load the necessary data into the DW once it has been determined what data is required. In order to consolidate data from multiple sources into a single repository, the "extract, transform, and load (ETL)" process is required. ETL is crucial because, in addition to extracting the information required for the DW, it also cleans the information to guarantee data quality and consistency across all databases, regardless of the source or system from which the information originated (Shanks & Bekmamedova, 2012b). Data extraction to a "staging area," which will contain the data in raw form, is the fundamental tenet of ETL.

Data Processing: After the given data is altered, it is put through data processing. Data processing entails taking the raw data and making sure that it is prepared for end users to use for analytical reasons. Data processing entails separating useful data from useless data, filtering, eliminating duplicates, validating, and adjusting for consistency.

BI: The data is prepared for BI algorithms to evaluate once it has been correctly processed and stored in the DW. The information will then be further transformed into data that is actionable and simple for decision makers to grasp by BI software, which will take the data from warehouses and sift it for insights (Gudfinnsson et al., 2015).

DM: By providing data in a format which can be analysed by end-users, these are the warehouse's "retail outlets." These are often designed to meet the requirements of a particular user group or decision-making process. They could be virtual or "real". DM can be implemented using either OLAP or conventional relational DBMS.

End Users: The system must provide end users with this information in a way that allows them to take appropriate action after the BI solution has processed the data to create the desired reports for them (Lechtenbörger, 2001). A crucial component of data warehousing and BI is making sure that end users are receiving the information they require in an easily understandable format.

3. Literature Review

According to Brannigan et al. (2016), to gain real business benefits, BI is a decision-making facilitator since it is crucial to the production of information for operational and strategic business decision-making. It was also mentioned by Sharda et al. (2015) that BI is a solution which may aid firms in making informed decisions.

A partnership between the business units as well as IT as well as an information & analysis usage culture is one of the seven major aspects which have been selected to evaluate BI for Williams and William (2007). BI/DW technical readiness as well as the procedure of creating a decision-making culture are two other areas which have been selected.

Research on data warehouse design was taken into consideration by Hurtado & Gutierrez (2007). They dealt with Data Warehouse design issues in an effort to enhance the Data Warehouse logical design procedure. They provided a collection of schema modifications along with certain application-specific tactics and guidelines. Moody et al. (2010) did take into consideration the study issue of designing data warehouse. A technique for creating dimensional models from entity relationship models was put out by them. The process is broken down into three steps. Several categories are created in the first stage by categorising the data model's entity types. The second stage is identifying any hierarchies that are present in the model. In order to create dimensional models, these hierarchies must be collapsed and transaction data combined.

Data warehouse design was taken into consideration by Lechtenbörger (2001). In order to undertake schema design for data warehouses in a manner comparable to that of traditional databases, they have introduced normal forms for conceptual data warehouse schemata. The suggested work, which emphasises summarization, is based on generalised multidimensional normal form.

According to Adela and Diaconita (2009), several facets of the development of "business intelligence systems (BIS)" have been made evident. These include the system's design, lifecycle, modelling approaches, and assessment methods for the system's performance. The authors have highlighted that BI systems can gather, integrate internal & external data, analyse and provide presentable information (key performance indicators) to the executive management to support strategic planning, forecasting and tracking of business performance. The authors have highlighted that BI systems enable managers to view data holistically, provide new insights affecting business processes, and improve the information quality for strategic decisions.

He et al. (2015) explained how to gather opinions, perspectives, and data about a company, industry, product, and consumer that are available on the Internet and web portals and how to analyse and prepare insights for decision-making and competitive advantage. The author has mentioned the dashboard of sentiment analysis for Wal-Mart employees. However, best practises for applying the market intelligence framework have not been offered.

Ramesh et al. (2020b) defined the notion of BI tools including DW, data mining, and OLAP technologies, as well as their fundamental properties. The writers have emphasized several advantages that can be attained by adopting BI tools in the company, including growing sales, profits, customer satisfaction, cost savings, and market share. The authors have emphasized that knowledge management is one of the fundamental requirements for establishing a lasting competitive advantage for firms. A BI tool may assist in effectively and efficiently constructing knowledge about clients and partners.

Wieder et al. (2013) noted that the quality of the BI system's data impacts the managerial decision-making process. As a result, organisations should start a data quality management program to ensure high-quality data in the BI system. BI's operational and strategic effects on the value chain have been distinguished by Fink et al. (2017), who have also established a thorough measuring tool for performance effects.

4. Objectives

Based on the literature, many cases where data warehousing design and development fail to deliver effective results for various reasons. Following are the objectives:

- To design and develop data warehouses and data marts for business organization.
- To help business managers in formulating and solving business problems through decision making.

5. Materials & Methods

The analysis from the “top-down”, the analysis from the “bottom-up”, as well as the assimilation of the results are the 3 main steps of the technique. This approach makes use of the two complementing approaches to data mart design: “top-down” analysis highlights user requirements, while “bottom-up” analysis highlights the semantics of the operational databases that are currently in use. The last stage combines these two points of view and produces a workable solution that most accurately reflects the user's objective and is backed by the available data sources.

Top-Down Evaluation: The University of Maryland created (Goal/Question/Metric) GQM as a methodical methodology, serves as the foundation for the data warehouse design method (Rouhani et al., 2016). A process model, a collection of forms, and a set of rules make up GQM. The process model outlines the various tasks that must be completed to design and run a measurement programme. There are a lot of similarities to be found in the processes of building a measurement programme and constructing a data warehouse. One method for creating a data warehouse is developed through the utilisation of a number of GQM principles and concepts, which are then adjusted.

Bottom-up investigation: This method starts with an "entity-relationship (ER) model", then does a comprehensive analysis to determine which entities have the potential to be facts, and finally builds a considerable number of directed graphs using these entities as the centre nodes. Each & every graph looks like a possible star schema, with all possible dimensions revolving around the fact entity. The ER model is investigated using an algorithm that generates potential star schemas. An ER schema is viewed by the method as a graph, and each n-ary (n. 2) or many-to-many relationship is first converted into a one-to-many relationship by adding the appropriate nodes and edges (He et al., 2015). Following this, the approach provides each entity with a numerical label which is proportional to the number of additive features that it holds. Entities that are potential fact entities are provided with a label that is not null. The next step is to choose each potential fact item as the central node of a graph, which is created by taking into account all potential dimensions that can be reached from the centre through one-to-many or one-to-one links. The programme automatically generates every feasible snowflake schema and operates effectively by recycling the output of earlier computations. In the last step, snowflake schemas are changed into star schemas, which can be thought of as a denormalization of relationships.

ER schema is transformed into a "Connectivity Graph (C-Graph)" to collect all relevant information from the schema for analytical tasks in the first step. The extraction method for snowflake schemas from C-Graph is therefore illustrated (Moody et al., 2010). Lastly, every snowflake model is transformed into a star schema. Finally, the conversion of each snowflake model into a star schema is done.

Integration: The warehouse's structure is identified at the top-down phase as it develops from user requirements. The bottom-up stage provides all multidimensional schemas which can be retrieved from the present information system. Combining the 2 systems and specifying how optimal criteria might be transferred to the actual system is a crucial stage (Graydon et al., 2022). Furthermore, the integration enables users to consider unique analytic aspects which did not come from customer requirement but that the system may make accessible. Integration contains 3 actions:

• **Terminology analysis:** Before they can be put together, top-down as well as bottom-up schemas should be translated into a common language of terms. To do this, we need to figure out how GQM words are linked to system terms (Awan et al., 2021).

• **Schema matching:** After mapping the terms, possible solutions are compared to ideal solutions which depict GQM goals. When the optimal solution as well as a candidate solution share a fact, this is called a match. The indicators take into account the number and size of the facts they share.

• **Ranking & selection:** A few of the potential "star schemas" are ultimately chosen, taking into account both the priority of the aim and the metrics of the matching (Paradza & Daramola, 2021).

6. Result & Discussion

The phase 2 (bottom-up) is more structured as well as strict, and it gives rise to an integration stage in which a formal indicator may be used to match requirements with possible solutions. The 1st phase (top-down), on the other hand, is more casual and predicated on conversation. In the "top-down" phase, the necessities of the users are boiled down and put together. This activity is, by its very nature, a complicated procedure that includes a lot of people. Over time, data is acquired and sorted from sources which are mostly casual and difficult to comprehend. Alternatively, since the 2nd and 3rd phases start with formal and structured data (the schema), they can be supplemented by a rigorous and systematic method. GQM was chosen as the top-down phase's guiding methods because it is a goal-oriented way to find out what needs to be done. In contrast to other goal-oriented techniques, the GQM's concentration on collecting qualitative & quantitative features of a complicated event provides a significant advantage.

7. Conclusion & Future Scope

Although the method is not fully mechanical, it offers significantly more direction than earlier methods to data warehouses and data marts designers. Identification and classification of the enterprise data model's entities that are important for decision making still require careful investigation. Process also encourages the creation of independent data warehouses with uniform design principles based on a shared enterprise data model.

Issues and problems with data warehouse design that were talked about in this last section will be looked at in more detail in future work. We would also like to add to the approach performed in this paper by taking more requirements into account during the integration phase. More work needs to be done in the future to learn more about this topic and come up with a better series of principles as well as suggestions to help the person who designs the data warehouse in a more effective as well as comprehensive way.

References

[1] Adela, B., & Diaconita, V. (2009). A model for Business Intelligence Systems' Development Adaptive Information Systems View project Intelligent system for trading on wholesale electricity market (SMARTRADE) View project. https://www.researchgate.net/publication/40646563

[2] Awan, U., Shamim, S., Khan, Z., Zia, N. U., Shariq, S. M., & Khan, M. N. (2021). Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. Technological Forecasting and Social Change, 168, 120766. https://doi.org/10.1016/j.techfore.2021.120766

[3] Božič, K., & Dimovski, V. (2019). Business intelligence and analytics use, innovation ambidexterity, and firm performance: A dynamic capabilities perspective. The Journal of Strategic Information Systems, 28(4), 101578. https://doi.org/10.1016/j.jsis.2019.101578

[4] Brannigan, C., Harling, M., & Townend, M. (2016). The factors that influence decision making by 8-12 year olds in Child and Adolescent Mental Health Services (CAMHS): a systematic review. https://www.researchgate.net/publication/305787345

[5] Faisal, S., Sarwar, M., Shahzad, K., Sarwar, S., Jaffry, W., & Yousaf, M. M. (2017). Temporal and Evolving Data Warehouse Design. Scientific Programming, 2017, 1–18. https://doi.org/10.1155/2017/7392349

[6] Fink, L., Yogev, N., & Even, A. (2017). Business intelligence and organizational learning: An empirical investigation of value creation processes. Information & Management, 54(1), 38–56. https://doi.org/10.1016/j.im.2016.03.009

[7] Graydon, C., Teede, H., Sullivan, C., de Silva, K., & Enticott, J. (2022). Driving impact through big data utilization and analytics in the context of a Learning Health System. In Big Data Analytics for Healthcare (pp. 13–22). Elsevier. https://doi.org/10.1016/B978-0-323-91907-4.00019-4

[8] Gudfinnsson, K., Strand, M., & Berndtsson, M. (2015). Analyzing business intelligence maturity. Journal of Decision Systems, 24(1), 37–54. https://doi.org/10.1080/12460125.2015.994287

[9] He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data. Industrial Management & Data Systems, 115(9), 1622–1636. https://doi.org/10.1108/IMDS-03-2015-0098

[10] Hurtado, C. A., & Gutierrez, C. (2007). Handling Structural Heterogeneity in OLAP. In Data Warehouses and OLAP (pp. 27–57). IGI Global. https://doi.org/10.4018/987-1-59904-364-7.ch002

[11] Lechtenbörger, J. (2001). Data Warehouse Schema Design *.

[12] Moody, D., Kortink, M., Moody, D. L., & Kortink, M. A. R. (2010). From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/

[13] Paradza, D., & Daramola, O. (2021). Business Intelligence and Business Value in Organisations: A Systematic Literature Review. Sustainability, 13(20), 11382. https://doi.org/10.3390/su132011382

[14] Ramesh, G. S., Rajini Kanth, T. v., & Vasumathi, D. (2020a). A Comparative Study of Data Mining Tools and Techniques for Business Intelligence (pp. 163–173). https://doi.org/10.1007/978-981-13-8253-6_15

[15] Ramesh, G. S., Rajini Kanth, T. v., & Vasumathi, D. (2020b). A Comparative Study of Data Mining Tools and Techniques for Business Intelligence (pp. 163–173). https://doi.org/10.1007/978-981-13-8253-6_15

[16] Rouhani, S., Ashrafi, A., Zare Ravasan, A., & Afshari, S. (2016). The impact model of business intelligence on decision support and organizational benefits. Journal of Enterprise Information Management, 29(1), 19–50. https://doi.org/10.1108/JEIM-12-2014-0126

[17] Shanks, G., & Bekmamedova, N. (2012a). Achieving benefits with business analytics systems: an evolutionary process perspective. Journal of Decision Systems, 21(3), 231–244. https://doi.org/10.1080/12460125.2012.729182

[18] Shanks, G., & Bekmamedova, N. (2012b). Achieving benefits with business analytics systems: an evolutionary process perspective. Journal of Decision Systems, 21(3), 231–244. https://doi.org/10.1080/12460125.2012.729182

[19] Sharda, R., Delen, D., & Turban, E. (2015). Business Intelligence and Analytics: Systems for Decision Support, 10th Edition.

[20] Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. Information Sciences, 367–368, 747–765. https://doi.org/10.1016/j.ins.2016.07.007

[21] Wieder, B., Ossimitz, M. L., & Chamoni, P. (2013). The Impact of Business Intelligence Tools on Performance: A User Satisfaction Paradox? Business Intelligence Research at UTS (Accounting).

[22] Williams, S., & William. (2007). The Profit Impact of Business Intelligence. Elsevier. https://doi.org/10.1016/B978-0-12-372499-1.X5000-5