A Different Perspective on Business Intelligence and Data Warehousing

Rajiv kumar, Mridul
Shobhit Institute of Engineering and Technology (Deemed to be University), Meerut
Email Id- Rajiv.kumar@shobhituniversity.ac.in, mridul@shobhituniversity.ac.in

ABSTRACT: The data warehouse is subject-oriented and includes nonvolatile and time-variant data. This information aids in the decision-making process of an organization's analysts. The transmission of reliable, relevant information to the right decision makers within the required period to enable successful decision-making is referred to as business intelligence. Different academics discuss the effect of metadata on data warehousing and business intelligence in this paper.

KEYWORDS: Business Intelligence, Data Mining, Data Warehouse Model, Data Warehouse.

1. INTRODUCTION

A data warehouse, is a system that collects and consolidates data from source systems on a regular basis into a dimensional or normalized data storage. It often stores years of data and is used for corporate intelligence or other analytical purposes. It is usually updated in batches, rather than every time a transaction in the source system occurs [1]. The Data Mart is a subset of the data warehouse and is described as a collection of historical data in an electronic repository that is not used in the organization's everyday activities. Instead, business intelligence is created from this data. The data in the data mart generally pertains to a single department within the company. The Fact Table is the principal table in a dimensional model that stores the business's numerical performance measures. We aim to keep all of the measurement data generated by a business process in one data mart.

A dimension table should always be used in conjunction with a fact table. The textual descriptions of the business are contained in the dimension tables. Dimension tables in a well-designed dimensional model include numerous columns or characteristics. The rows in the dimension table are described by these characteristics. The Dimension tables are narrow in terms of row count (typically less than 1 million), but they are vast in terms of column count, with numerous huge columns. The fact table's entry points are dimension tables. The dimensions are used to create the data warehouse's user interface. An OLAP database is a system for storing, organizing, and querying data that is especially developed to serve business intelligence applications. The Extract, Transformation, and Load (ETL) system is a collection of procedures for cleaning, transforming, combining, de-duplicating, archiving, conforming, and structuring data for usage in a data warehouse.

1.1. Data Warehouse Concepts:

Data warehousing is the act of gathering data for storage in a controlled database, where the data is subject-oriented and integrated, time-variant, and nonvolatile to aid decision-making. Data from a corporation's many activities is reconciled and kept in a single repository, referred to as a data warehouse, from which analysts extract information that helps them make better decisions. To offer information, the data may then be aggregated or processed, as well as sliced and diced as needed. William Inmon and Ralph Kimball are two well-known writers in the realm of data warehouse architecture, however their approaches to several aspects of data warehousing differ [2], [3]. Inmon has a top-down strategy, whereas Kimball takes a bottom-up method. The majority of data warehouse practitioners use one of the two techniques.

A Data Warehouse, is a subject-oriented, integrated, time-variant, non-volatile data collection used to assist decision-making processes. The term "subject oriented" refers to a data warehouse that focuses on high-level business entities and organizes data by subject. The term "integrated" refers to data that is stored in uniform formats, naming standards, variable measurements, encoding structures, physical data characteristics, or domain restrictions. In a data warehouse, for example, there is only one coding scheme for ethnicity, but an organization may have four or five. The term "time-variant" refers to warehouses that give access to a larger amount of more comprehensive data over a longer period of time and that the data is tied to a certain point in time, such as a month, quarter, or year.

The warehouse data is non-volatile in the sense that once it is entered into the database, it is seldom, if ever, altered. Updates or refreshes of the data in the warehouse are done on a periodic, incremental, or complete refresh basis. Finally, the term "nonvolatile" refers to data that does not alter. A data warehouse is a collection of all the company's data marts. The dimensional model is always used to store data. Data warehousing is a subset of data marts. The data marts are focused on achieving business goals for various divisions inside the company. The data warehouse is a dimension of the data marts that has been confirmed. A data mart is a subset of a data warehouse. The data warehouse is made up of all the data marts, each of which uses a star schema (or a family of star schemas) to represent a business process in an organization. The fundamental difference between Kimball's and Inmon's approaches is that Kimball's conformed dimensions are de-normalized, but Inmon's core database model is highly normalized.

Data marts hold a second copy of the data from the centralized data warehouse tables, but Kimball's dimensions are not copies of the conformed dimensions, but the dimension table itself. The data warehouse bus is a set of conformed dimensions. Because these two solutions reflect distinct data warehousing philosophies, there is no right or wrong answer. In fact, most companies' data warehouses are closer to Ralph Kimball's concept. Because most data warehouses began as a departmental project, they were originally known as a data mart. They only become a data warehouse as more data marts are added later.

Some of the DW features include the following.

- It is subject-oriented.
- It is non-volatile.
- It enables the integration of a number of different application systems. By combining previous data, it aids in the processing of information.
- Data is kept in a structured manner that allows for querying and analysis.
- The information is summarized. Unlike transaction-oriented systems, DWs often do not store as much information.

1.2. Statements - Business Intelligence:

There are many different methods to defining Business Intelligence in the literature (BI). IT suppliers, news organizations, and business consultants all have different perspectives on the topic. A few instances are shown below. They should be able to demonstrate the basic notion of business intelligence when used together. The Gartner Group defines BI as a process of translating data into information and then transforming that information into knowledge following a journey of discovery [4]. Vriens and Philips defined BI as the process of gathering and processing data in order to support an organization's strategy [5].

D. Tijd defines business intelligence (BI) as “any programs that enable the analysis and reporting of corporate data in order to improve decision making and company direction” [6]. Decision makers want accurate data that has been filtered from all raw data the firm has collected in the past. The key goal is to turn these unprocessed data into useful, actionable knowledge. Common transactional software automates day-to-day tasks like creating invoices and registering them in the system. BI, on the other hand, takes a step backwards to offer a comprehensive perspective of these transactions. Figures from the past are not published in great detail; rather, they are collected, evaluated, and connected with one another in order to anticipate future actions. Business intelligence systems fall into three main categories: reporting, including OLAP, and data mining. BI tools are further divided into reporting, OLAP, and data mining by Aronson et al. They split BI tools into these three groups as well in this paper. Most definitions seem to agree that business intelligence should help define a company's basic orientation by evaluating and reporting data.

2. DISCUSSION

2.1. Business Intelligence Architecture:

2.1.1. Operational Applications vs. Business Intelligence Applications:

The two primary components of BI applications, as well as their relationship with operational applications, are depicted in Figure 1. The two primary BI components are reporting and data mining. OLAP, in my perspective, sits in between reporting and data mining. As a result, OLAP and data mining are discussed separately in this research paper.

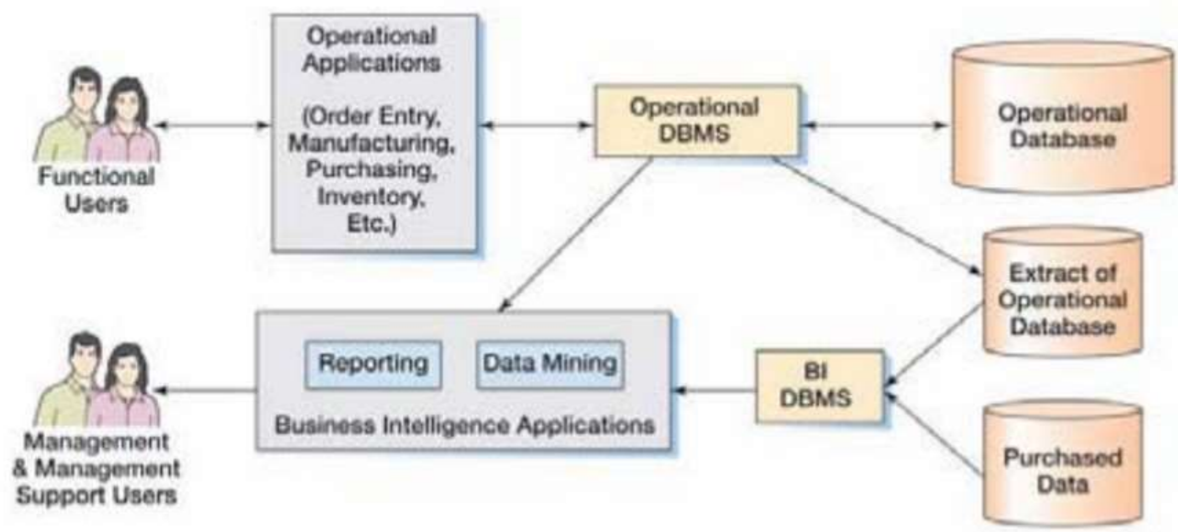

Figure 1: Relationships between Operational and Business Intelligence Applications are shown [6]**.**

On the one hand, the operational database management system allows operational business applications such as order entry, production, and buying to read from and write to the operational database (DBMS). For example, entering orders into a corporate system is mostly done on a company's operational level and does not necessitate high-level judgments. According to this plan, managers at the tactical and strategic levels should rely on business intelligence tools to make better decisions. This distinction is significant since the major objective of this research is to determine what corporate level(s) BI contributes to.

On the other hand, BI applications may only read data directly from the operational database via the operational DBMS, as long as basic reporting and/or small databases are used. The BI DBMS reads data from extractions of this operational database as well as data acquired from other data providers. BI software may then use this data to generate reports as well as do sophisticated analysis. The subdivision of these components will be discussed in further detail in the following paragraphs.

2.1.2. Need for a Data Warehouse:

Reading straight from the operational database may cause problems for sophisticated BI systems operating on big databases. Errors can arise when values are missing or in the wrong format, which can slow down the DBMS and its applications. As a result, a new database must be set up and readied for BI usage, which is an extraction from the operational database. Extraction, transformation, and loading are the three key phases in the data warehousing process (ETL). Extraction programs use a model to obtain data from a range of diverse operational databases. This model, as well as the specification of the source data pieces, are described in the metadata. For example, metadata comprising sales data in integer format produced by salespeople in a specific region defines a model explaining regional sales performance. It's worth noting that utilizing indexes speeds up the extraction process. Data transformation is sometimes required to guarantee that all data in the data warehouse is consistent.

2.1.3. Data Warehouse Data Model:

Data modeling process has three levels: high-level modeling (called the ERD, entity relationship level) that includes entities, attributes, and relationships, mid-level modeling (called the data item set) that includes data organized by department, and low-level modeling (called the physical model) that optimizes for performance [2]. Following the creation of the high-level data model, the midlevel model is built. A midlevel model is built for each key topic area or item specified in the high-level data model. After that, each region is evolved into its own midlevel model. The physical data model is produced by expanding the midlevel data model to incorporate the model's keys and physical properties. The physical data model now resembles a collection of tables, referred to as relational tables.

2.1.4. DW Modeling Techniques:

As it pertains to data modeling for the data warehouse, Ballard described database warehouse modeling as the process of creating a model for the data in order to store in the DW [7]. Entity Relationship (ER) modeling and dimensional modeling are two data modeling approaches that are useful in a data warehousing context. ER modeling creates a data model of a specific area of interest by combining two basic concepts: entities and their connections. Attributes, which might be characteristics of entities or relationships, are also included in detailed ER models. Measures, facts, and dimensions are the three main ideas used in dimensional modeling. In the context of database tables, dimensional modeling is useful for reflecting the needs of business users. Measures are numbers that can be put together and calculated.

2.1.5. DW Database Design Modeling:

Data modeling is divided into three tiers. They are mental, intellectual, and physical in nature. Only the first two will be discussed for the sake of this paper. Physical design is dependent on the DBMS and describes how data is actually stored. Conceptual design manages concepts that are close to how users perceive data; logical design manages concepts related to a specific type of DBMS; and conceptual design manages concepts that are close to how users perceive data. The fundamental objective of conceptual design modeling is to create a formal, comprehensive, abstract design that meets the needs of the users. The definition of structures that enable efficient access to information is part of the logical architecture of a data warehouse. The designer creates multidimensional structures based on a conceptual schema that represents the information needs, source databases, and non-functional (mostly performance) requirements. Specifications for data extraction tools, data loading procedures, and warehouse access techniques are also included in this phase. A functioning prototype for the end-user should be developed at the conclusion of the logical design process.

2.1.6. Developing Data Warehouse:

Planning the development and deployment of a standard data warehouse should be treated as an IT project; thus, the same factors that cause IT projects to fail apply to data warehouse development; thus, the need for Project Planning and following the system development life cycle. Careful planning, requirements definition, design, prototype, and execution are all required. The cycle model is divided into five stages, as shown in Fig. 2.

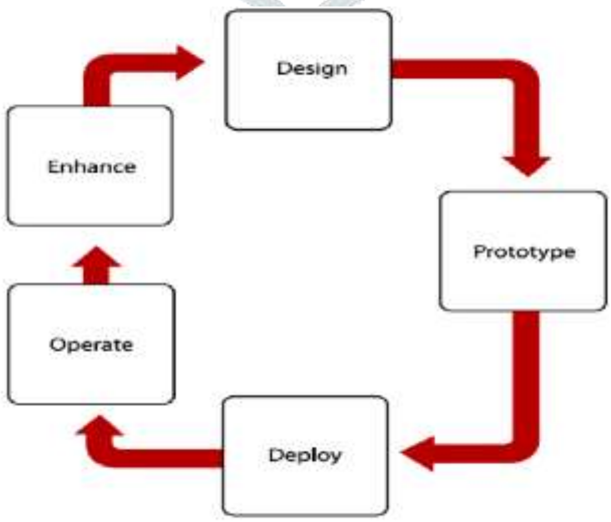

Figure 2: Illustrates the model of Data warehouse Development Lifecycle (DWLC) [8]

The Design step gathers data from both available data inventories and analyst requirements and analytical needs, as well as robust data models, and transforms it into data marts and intelligent data.

The prototype deployment step involves bringing a functional model of the data warehouse or data mart design into touch with a set of opinion-makers and a specific end-user clientele. As the design team travels back and forth between design and prototype, the goal of prototyping changes. The deploy step is where the user-approved prototype gets formalized for true production use. The operation is responsible for the day-to-day upkeep of the data warehouse or mart, as well as the data delivery services and client tools that give analysts access to the warehouse, as well as the management of ongoing extraction, transformation, and loading processes that keep the warehouse up to date with respect to authoritative transactional source systems. If the initial design and execution failed to fulfill requirements, the augmentation step transitions easily back to basic design.

2.1.7. Business Intelligence Concepts:

BI was first coined as a catch-all name for data analysis software. Meanwhile, the definition of business intelligence (BI) has expanded to include all components of an integrated decision support architecture. Data from OLTP is coupled with analytical front ends in business intelligence systems to "display complicated and competitive information to planners and decision makers." The data warehouse (DW), which combines OLTP data for analytical activities, is a key component of BI systems. BI is viewed from a management perspective as a process in which data from within and outside the company is aggregated and integrated to create information that allows for rapid and effective decision-making. The goal of business intelligence (BI) in this case is to provide an informational environment and process through which operational data obtained from transactional systems and external sources may be evaluated and “strategic” business dimensions revealed. From this perspective, terms like "intelligent company" emerge: a firm that uses business intelligence to make faster and better decisions than its competitors. The term "intelligence" refers to the act of screening, analyzing, and reporting information in order to convert a large amount of data into knowledge. The technology perspective portrays BI as a collection of tools for storing and analyzing data. The focus is on the technology that allow for the recording, retrieval, modification, and analysis of data, rather than the process itself.

2.1.8. Data Warehouse and Business Intelligence High Level Architecture:

According to Eckerson of the Data Warehouse Institute, a research was conducted on the success factor in adopting BI, systems, and the function of the data warehouse in this process [9]. He describes the BI process as a "data refinery" in its entirety. Data from several OLTP systems is combined to create a new product called information. The transformation is handled by the data warehouse staging process. Users that have access to programs like specialized reporting tools, OLAP tools, and data mining tools may turn data into knowledge. This is included in Kimball's data warehouse [3]. The goal of the data warehouse, is to provide quick access to data within the company to end-users (primarily managers). In order to accomplish so, the organization's operational systems must be used to gather daily operational data. The OLTP system is what it's called. The data from the source systems is staged before being transmitted to the presentation servers (Kimball et al 1996). The data in the staging stage goes through four steps: extraction, transformation, loading, and presentation. The data marts, which represent business sectors in the company, are built on the presentation stage.

The data is kept in the data mart or data warehouse as star schemas, which are made up of FACT and DIMENSION tables. The entity relational diagram (ERD) used in older systems is not the same as this. There is a distinction between data warehouse and business intelligence architecture, as proposed by two well-known industry experts. Data-driven methods are recommended. This indicates that the decision-making process starts with facts and concludes with requirements. Rather than Inmon's approach, advocates for the usage of requirements-driven techniques. The data warehouse project begins with project planning to assess the organization's preparedness for a data warehouse and to establish the data warehouse team's staffing requirements. The most essential success element is a clear knowledge of business needs, and Kimball claims that this method of requirements collecting varies significantly from data-driven requirements analysis [3]. The business requirements lay the groundwork for the three parallel tracks, which are focused on technology, data, and end-user applications.

3. CONCLUSION

This paper examines data warehousing, business intelligence, and OLTP, as well as their designs, the necessity for data warehouses, and various modeling approaches and their applications in the development of data warehouses from many perspectives. The research stresses the relevance of technology in the creation of data warehouses and business intelligence systems. Agile ideals lay less focus on tools when compared to humans and relationships. However, tools that may provide lower cost, faster reaction time, and throughput solutions are driving the growth of BI into analytics and big data; for example, NoSQL technologies are becoming more widely used. Overall, DW/BI development should be viewed as a sociotechnical phenomenon that takes into account methodological, organizational, and technological concerns. As a result, DW/BI development may differ from traditional software development, which is better suited to direct implementation of the agile methodology.

REFERENCES

[1] V. Rainardi, *Building a data warehouse: With examples in SQL server*. 2008.

[2] W. H. Inmon, "Building the Data Warehouse," *A Wiley QED Publ.*, 1993.

[3] R. Kimball and M. Ross, *The Data Warehouse Toolkit, The Definitive Guide to Dimensional Modeling*. 2013.

[4] S. R. Gardner, "Building the data warehouse," *Commun. ACM*, 1998, doi: 10.1145/285070.285080.

[5] D. Vriens, "The Role of Information and Communication Technology in Competitive Intelligence," in *Information and Communications Technology for Competitive Intelligence*, 2011.

[6] "Business Intelligence." https://www.tijd.be/algemeen/algemeen/business-intelligence/5381658.html (accessed Sep. 15, 2018).

[7] C. Ballard, D. Herreman, D. Schau, R. Bell, E. Kim, and A. Valencic, "Data Modeling Techniques for Data Warehouse," *Zenithresearch.Org.in*, 2012.

[8] E. Mangia *et al.*, "Proceso iterativo de diseño del modelo multidimensional de soporte a datos clínicos de pacientes diabéticos [Iterative design process for multidimensional model to supports clinical data of diabetic patients]," *Vent. Inform.*, 2016, doi: 10.30554/ventanainform.33.1401.2015.

[9] W. Eckerson, "The Secrets of Creating Successful Business Intelligence Solutions Smart Companies in the 21st Century," *Seattle Data Warehous. Inst.*, 2010.