International Journal of Engineering, Management and Humanities (IJEMH)
Volume 4, Issue 5, Sep.-Oct., 2023 pp: 287-292 www.ijemh.com

The Technology of Using a Data Warehouse to Support Decision-Making in Health Care

Prachiti Sonawane
BBA BA, MIT WPU
Pune (India)

Sudeepta Banerjee
Datawarehouse professor, MIT
Pune (India)

Date of Submission: 02-10-2023 Date of Acceptance: 12-10-2023

ABSTARCT
This paper delves into the application of data warehousing technology in healthcare decision-making, particularly focusing on its utility in managing cancer-related data. Healthcare executives and medical professionals often require timely access to comprehensive health data to make informed decisions without disrupting the everyday operations of an On-Line Transaction Processing (OLTP) system. Addressing this complex challenge in healthcare decision-making, this paper proposes the adoption of healthcare data warehousing as an efficient solution. In the following sections, we elucidate key concepts such as data warehousing and On-Line Analysis Processing (OLAP). We demonstrate the transformation of data within the data warehouse into a multidimensional data cube. Furthermore, we provide a practical example showcasing the application of a healthcare data warehouse tailored for cancer-related data developed in this study. This approach empowers executive managers and doctors to analyze data from multiple perspectives with reduced query time, thereby facilitating quicker and more comprehensive decision-making.
KEYWORDS: Healthcare data warehouse, Extract-Transformation-Load (ETL), Cancer data warehouse, On-Line Analysis Processing (OLAP), Multidimensional Expression (MDX) and Health Insurance Organization(HIO).

I. INTRODUCTION

The management of data in healthcare organizations has become a challenging task due to the differing objectives, concerns, priorities, and constraints of healthcare managers. The manipulation of large amounts of health data and the use of corresponding technologies have become integral to all aspects of healthcare, including the planning, management, and delivery of healthcare services. Information plays a crucial role in the success of an organization, providing executive managers and physicians with the necessary foundation for decision-making.

Healthcare organizations typically handle significant volumes of data, including valuable information about patients, procedures, treatments, and more. However, this data is usually stored in operational databases that are not easily accessible or useful for decision makers and executives. The majority of database management systems utilized in these organizations are designed for online transaction processing (OLTP), which directly answers queries but does not cater to the specific needs of executive-level decision makers, such as what-if and what-next type queries.

Decision makers at the executive level require the ability to quickly and efficiently analyze existing health data in order to aid the decision-making process. Unfortunately, standalone databases are incapable of providing such information in a timely and effective manner. The data warehousing concept offers a robust solution for addressing data integration and information access issues. Based on online analytical processing (OLAP), this technology enables the reorganization, integration, and analysis of data, allowing users to quickly and accurately access information. Table 1 showcases the comparison between OLTP and OLAP technology, with OLAP serving as a tool for analysts to aid in planning and decision-making. Traditionally, businesses have relied on paper-based reports to make important decisions regarding performance. However, these reports are often

outdated and rely on extracting data from operational systems and combining it with other data sources. Analysts can only provide limited information at significant cost within the desired time frames. There is a need for a healthcare data warehouse designed specifically for this purpose.

OLTP databases are designed to process individual records for various operations relating to patients, procedures, treatments, drugs, etc. These databases are continually updated to support daily operations but lack the ability to provide comprehensive information. A healthcare data warehouse plays a crucial role in providing executives with information to analyze situations and make decisions. It also assists doctors in making well-informed decisions and performing their jobs effectively.

Decision-making occurs at different levels within the organization. Strategic decision-making involves decisions about the overall goals of the organization, such as the types of services to be provided and the geographical locations of operation. Tactical decision-making pertains to specific units within the organization, such as patient care services and marketing. Finally, day-to-day decisions cover areas like employee hiring, supply and medication ordering, and billing processing. Having a good system in place ensures that managers have the necessary information to make efficient decisions.

This paper focuses on the development and implementation of a prototype healthcare data warehouse for cancer diseases. Utilizing the new "data warehouse" technology, this system incorporates a large amount of analysis information necessary for healthcare decision-making.

Table 1: Comparison of OLAP and OLTP (adapted from Ref[6])

CHARACTERISTICS	OLAP	OLTP
OPERATION	Analyse	Update
LEVEL OF DETAIL	Aggregate	Detail
TIME	Historical, current ,projected	Current
ORIENTATION	Attributes	Records

II. USING A DATA WAREHOUSE

The idea of data warehousing first came into being in the middle of the 1980s with the intention of allowing extensive data analysis and management reporting. A data warehouse, according to Bill Inmon, is a set of subject-specific, integrated, time-variant, non-volatile data that supports management decision-making. In order to allow querying and analysis for decision-making, Ralph Kimball defines a data warehouse as a system that harvests, cleans, conforms, and delivers source data into a dimensional data store.

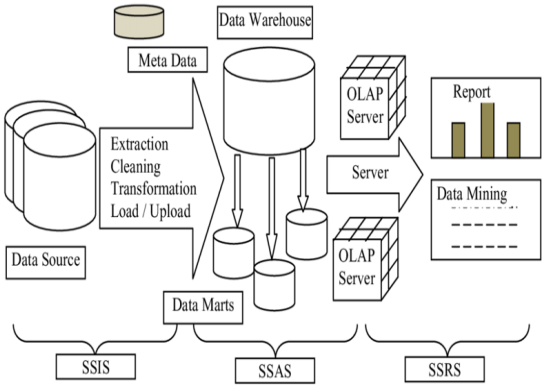
In addition to its extensive use in banking, finance, consumer goods, retail distribution, and demand-based manufacturing, data warehouses have also gained prominence in non-commercial sectors such as healthcare, government, military, education, and research. A data warehouse is typically a dedicated read-only database system that integrates data from multiple databases and other information sources. It is separate from an organization's transactional databases (OLTP databases). Its distinguishing features include a longer time horizon, multiple subject-oriented and uniformly defined databases, non-volatile data that are updated in planned cycles, and optimization for complex queries from decision makers and applications.

Transactional databases focus on answering "who" and "what" type questions and are not well-suited for answering "what-if," "why," and "what-next" type questions due to their lack of organization for analytical processing. Data warehouse architecture describes the components of the warehouse and how they fit together. Data is sourced from various channels, transformed within a staging area, and then integrated and stored in the production data warehouse for further analysis. Accuracy in the extraction of data from multiple sources is crucial as errors and anomalies are likely when integrating different datasets into the warehouse.

Therefore, tools are needed to extract, clean, integrate and upload data. The data is then stored and managed in a warehouse that provides various multidimensional views of the data to various user interface tools such as query tools, report writers, analyzers and data mining tools.

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 4, Issue 5, Sep.-Oct., 2023 pp: 287-292 www.ijemh.com

The architecture of data warehousing is shown in Figure 1:



The reason for selecting SQL Server 2008 to develop the Healthcare Data warehouse is its comprehensive capabilities. It encompasses relational database management service, as well as integrated service, analysis service, and reporting service. The integration service assists in integrating data from diverse data sources, ensuring efficient data extraction, transformation, and loading.

The analysis services offer functionalities of OLAP and data mining, enabling in-depth analysis and future trend prediction. The reporting services offer a range of data reports and graphical representations of the analysis results.

III. DATA WAREHOUSE TECHNOLOGY

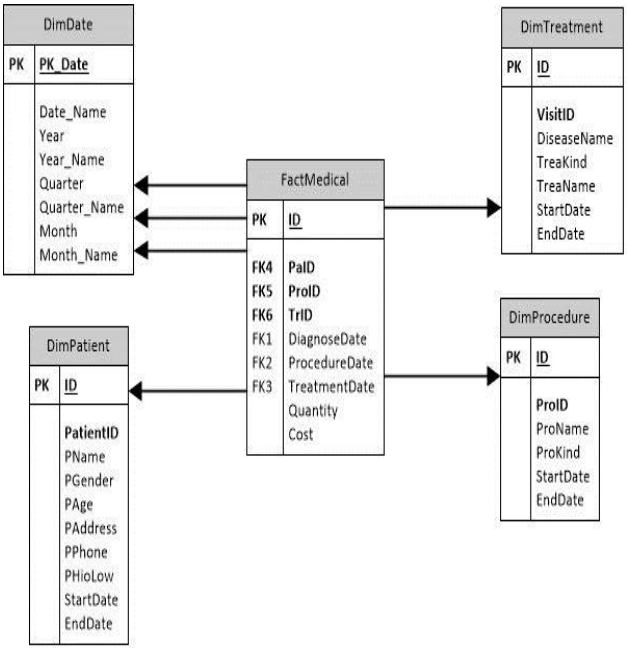
3.1 Design data warehouse for Cancer diseases.

Designing the structure of a data warehouse is distinct from designing operational systems in terms of queries and computation. Operational systems rely on simple pre-defined queries. In contrast, data warehousing environments involve more complex queries that join multiple tables and requires more computation time and flexibility. This has led to the emergence of a new approach to data modeling, with the multidimensional or data cube model being widely adopted. When designing front end tools, database design, and query engines for OLAP (Online Analytical Processing), it is crucial to consider the multidimensional view of the data.

There are two common modeling techniques, known as "Star Schema" and "Snowflakes Schema", used to represent multidimensional data. In this context, the "star" schema is preferred due to its clarity, convenience, and rapid indexing capabilities. In simple terms, a star schema is a specific type of database design that supports analytical processing and includes a set of denormalized tables. The star schema consists of two types of tables: the fact table (the central table) and dimension tables. The fact table contains keys and measurements, while the dimension tables are directly linked to the fact table.

Overall, designing the data warehouse structure requires consideration of the unique requirements and complexities of data warehousing environments, and the star schema is a commonly used approach due to its efficiency and effectiveness. MDX queries then use to query from fact and dimension tables within the star schema, with constraints on the data to return required information.

Cancer star schema :



3.2 Cancer data warehouse development

The Cancer Data Warehouse has been developed based on the previous star schema data model. The development process includes the following steps:

1. Data extraction: Databases and files are just two examples of the sources from which data is retrieved. In particular, we took information out of an Arabic-created Access database. Several tables, including the "Patient, Procedure, Ticket, and Visit" tables, served as the source of this data.

After being transformed, these tables were loaded into dimension and fact tables.

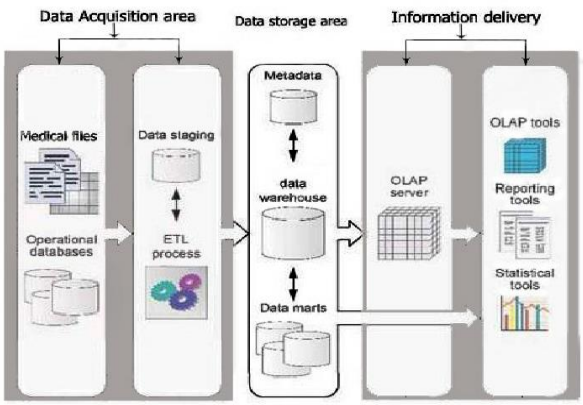
2. Data transformation and cleaning: To maintain consistency, the extracted data underwent a

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 4, Issue 5, Sep.-Oct., 2023 pp: 287-292 www.ijemh.com

transformation procedure before being imported into the data warehouse. To do this, built-in transformations from SSIS were used. The extracted data has to be mapped to the repository's preferred format for the cancer data warehouse. Additionally, data conversion procedures were used to convert columns from one data type to another and Arabic column names were changed to English names. Additionally, a fuzzy look-up was employed to fix typos in the data rows.

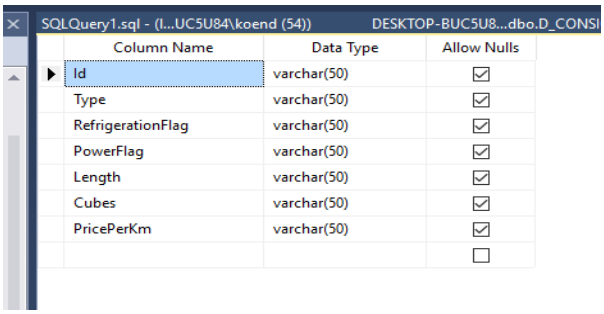
3. Data loading: The data is put into the data warehouse after being cleansed and converted. The data flows are processed by a slowly varying dimension adapter for the dimension load tasks after being sent through a sort editor to eliminate any duplicate rows. However, before being processed by a slowly changing dimension adapter, the data flows that feed the fact tables do not pass via a sort adapter. The ETL (Extract, Transform, Load) process for the data is made up of these three processes taken together.

4. Multidimensional Data Modeling: Cubes, dimensions, and hierarchies are fundamental ideas on which the multidimensional data model is built. While dimensions are the cube's named edges that display certain data, the cube itself symbolizes the complete dataset. How these dimensions are arranged is determined by hierarchies. The database design is in line with the user's preferred paths for accessing and browsing the cancer data warehouse thanks to this dimensional modeling strategy. Additionally, cubes, dimensions, measurements, hierarchies, levels, and cells are among the OLAP structures in this paradigm.



In summary, the Cancer Data Warehouse has been developed using a star schema data model. The ETL process extracts, transforms, and loads data from multiple sources, and the resulting multidimensional data model provides a consistent and user-friendly design for the cancer data warehouse.

These, taken together, define the logical structure of an OLAP database. Measures are the data that we wish to analyze, while dimensions define the organization of these measures. Our Cancer data warehouse may contain a Fact Medical table that has fields for Patients, Time, Treatments, Procedures, Cost and Quantity. If so, we will generally analyze Cost and Quantity by warehouse, Patients, Time, Treatments and Procedures. In this case, Cost and Quantity will be our measures, and warehouse, Patients, Time, Treatments and Procedures will each be a dimension. The dimension contains the elements that called members. The major operations that could be done on OLAP cubes are Selection, Roll-up, Drill-Down and Slice, through which we can view data from all perspectives and all levels .



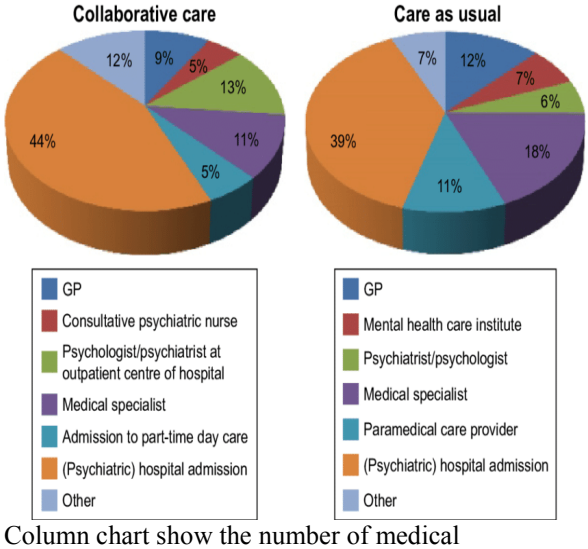
IV. IMPLEMENTATION

The Cancer data warehouse was utilized at the Elsharkiya Branch of a Health Insurance Organization in Egypt. This allowed for validation of the prototype properties and showcased the system's capabilities in a real-life application scenario. One particular example that was highlighted was the cost of treatment for Cancer diseases. The process for making other healthcare decisions is similar. The Cancer data warehouse offers various features, including an interface designed for novice users, as depicted in Figure 6. The main window contains a menu bar and tools bar that provide access to all other application forms. In this main window, all elements are thoroughly explained. The first menu, titled "Create Cancer DW," incorporates submenus for creating a Star Schema, Fact Table, Dimension Table, and Relationship. The second menu, labeled "Extract - Transform - Load," includes a submenu for executing the ETL Package. The third menu, named "Multidimensional Data," offers a submenu for creating and deploying Multidimensional Data. The fourth menu, "Information Delivery," contains

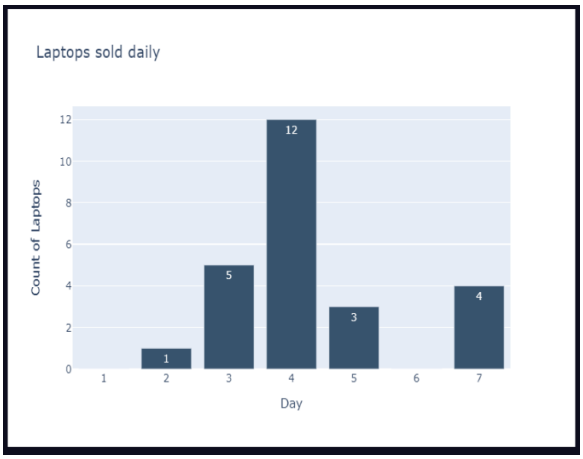
International Journal of Engineering, Management and Humanities (IJEMH)
Volume 4, Issue 5, Sep.-Oct., 2023 pp: 287-292 www.ijemh.com

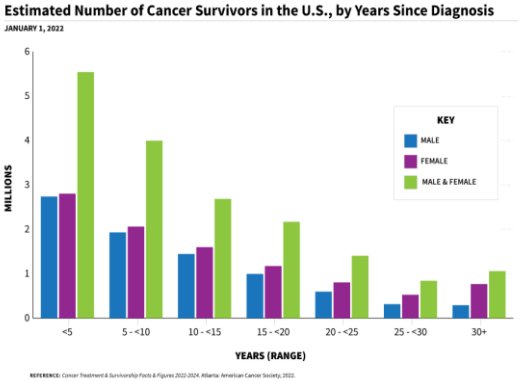
submenus for accessing Patient and Medical Procedure Information, Patient Information, Treatment Information, and generating Reports over the Web. The query results can be displayed in various formats, such as line charts, pie charts, bar charts, area charts, and report files. These results are shown directly on the interface using a combination of C# language, MDX queries, and Excel. This significantly aids managers in formulating appropriate healthcare decisions or determining an effective storage strategy for the warehouse.

Pie chart in Excel show the patient cost by HIO law:



Column chart show the number of medical procedure by patient gender :





V. CONCLUSION

This paper utilizes ETL, OLAP, and reporting technologies to create a software component. The implementation tools used include Microsoft SQL Server 2008, SSIS, SSAS, and SSRS. The cancer data warehouse is advanced in several ways. It allows for insights into factors affecting healthcare management activities, helping managers make informed decisions to improve performance. Additionally, it provides quick responses to queries and offers a multidimensional view of data, allowing users to analyze figures from different perspectives.

Overall, the cancer data warehouse effectively supports executive managers and doctors by providing accurate and timely information for healthcare decision-making purposes.

REFERENCES

[1]. Ahmad, I. (2000). Data warehousing in construction organizations. In ASCE Construction Congress-6 (pp. 1-10). Orlando, FL.

[2]. Dych, J. (2000). E-Data: Turning Data into Information with Data Warehousing. Addison-Wesley, Reading.
[3]. Poe, V., Klauer, P., & Brobst, S. (1998). Building a Data warehouse for Decision Support. Prentice Hall, Upper Saddle River.
[4]. Johns, M. L. (2002). Information Management for health professions (2nd ed.). Delmar Thomson Learning Inc.
[5]. Teh, Y. W., & Ong, S. S. (2009). Development of a Data Warehouse for Lymphoma Cancer Diagnosis and Treatment Decision Support.
[6]. Inmon, W. H. (2002). Building The Data Warehouse. Wiley Computer Publishing.
[7]. Kimball, R., & Casertam, J. (2004). The Data Warehouse ETL Toolkit.
[8]. Gray, P. G., & Watson, H. J. (1998). Decision Support in the Data Warehouse. Prentice-Hall, New Jersey.
[9]. Ahmad, I., & Azhar, S. (2002). Data warehousing in construction: from conception to application. Proceedings of the First International Conference on Construction in the 21st Century: Challenges and Opportunities in Management and Technology (pp. 739-746). Miami, Florida, USA: Florida International University.
[10]. Hadley, L. (2002). Developing a Data Warehouse Architecture.
[11]. Sheta, O. E., & Eldeen, A. N. (2012). Building a Healthcare Data Warehouse for Cancer Diseases. IJDMS, Vol 4, No 5.
[12]. Surajit, Chaudhuri, & Dayal, U. (1996). An Overview of Data Warehousing and OLAP Technology.
[13]. Mohania, M., Samtani, S., Roddick, J., & Kambayashi, Y. (2010). Advances and research directions in data warehousing technology. Retrieved from [source link].
[14]. Corey, M., & Abbey, M. (1999). SQL Server 7 Data Warehousing. McGraw-Hill, Berkeley.
[15]. Thomsen, E., Spofford, G., & Chase, D. (1999). Microsoft OLAP Solutions. Wiley, New York.
[16]. Vassiliadis, P., & Skiadopoulos, S. (2000). Modeling and optimization issues for multidimensional databases.

AUTHORS
Prachiti Sonawane (BBA BA II YEAR) MIT WORLD PEACE UNIVERSITY. MAHARASHTRA PUNE (INDIA)
CO AUTHOR :
Sudeepta Banerjee (Professor in MIT WORLD PEACE UNIVERSITY) MAHARASTRA PUNE (INDIA).