

Building a Diabetes Data Warehouse to Support Decision making in healthcare industry

Abubakar Ado¹, Ahmed Aliyu², Saifullahi Aminu Bello³, Abdulra'uf Garba Sharifai⁴, Abubakar Sulaiman Gezawa⁵

¹(Computer Science, Liaoning University of Technology, China)

²(Mathematical Sciences, Bauchi State University, Gadau, Nigeria)

³(Computer Science, Kano State University of Technology, Nigeria)

⁴(Computer Science, Liaoning University of Technology, China)

⁵(Computer Science, Bayero University Kano, Nigeria)

Abstract : data warehousing did not find its way easily and readily into healthcare and medicine, not like others financial institutions, Healthcare presents unique challenges for the architect of a data warehouse and the information needs for Health data warehouse were fundamentally different and complicated. When it come to health care, data warehousing is completely stagnant and still there is a need for data warehouse in healthcare. This paper presents a proposed architecture for Health care data warehouse for Diabetes diseases which could be use to monitor Diabetes disease, measure cost of infections and to detect prescription errors in addition it can also be used by healthcare executive managers, doctors, physicians and other health professionals to support the capture, healthcare process and analysis of data, and offer the potential of radically altering the practice and delivery of healthcare and medical research.

Keywords: Health care data warehouse, Extract-Transformation-Load (ETL), Diabetes Data warehouse, On-Line Analysis Processing (OLAP), Medical History, Business Analysis, Decision Support System (DSS).

I. introduction

The healthcare industry is one of the world's largest, fastest-developing and most information-rich industries [1]. Due to the growing of health care sector there is a need for health industries to open towards adoption of extensive healthcare decision support systems. Information technology in health care is still a topical subject, has stimulated developments in physician order entry, decision support systems and shared patient records [2]. Information is one of the most factors to an organization success that executive managers or physicians would need to base their decisions on, during decision making. Healthcare organizations typically deal with large volumes of data containing valuable information about patients, procedures, treatments and etc. These data are stored in operational databases that are not useful for decision makers or executives [3]. Despite all the efforts, many health care organizations still have stand-alone information systems that do not communicate with each other. More importantly, clinical information systems (CIS) such as electronic patient records, are often designed to support hands-on care for individual patients, but are not well suited for analyses on an aggregated level, for example on groups of patients with the same disease[4].

Warehouse is a central repository in which data from multiple source is brought together, integrated, manage and accessible to multiple users. In a fully realized data warehouse, data is quality controlled, fresh timely and readily retrieved for any purpose. Data warehousing required that the many disparate systems that held unintegrated data become integrated. In many circles the older unintegrated systems were called "legacy systems." The process of integration for legacy systems was never clean and easy. But organizations found that the ability to look at corporate data in an integrated manner was worth the pain and trouble of integration of the data. In a short while, data warehousing began to be a corporate necessity. Today data warehousing is conventional wisdom in information processing [5]. The clinical data warehouse is a place where healthcare providers can gain access to clinical data gathered in the patient care process. It is also anticipated that such data warehouse may provide information to users in areas ranging from research to management [6]. Clinical data warehouse can facilitate efficient storage, enhances timely analysis and increases the quality of real time decision making processes. Such methodologies share a common set of tasks, including business requirements analysis, data design, architecture design, implementation and deployment [7, 8].

Diabetes is a defect in the body's ability to convert glucose (sugar) to energy. Glucose is the main source of fuel for our body. When food is digested it is changed into fats, protein, or carbohydrates. Foods that affect blood sugars are called carbohydrates. Carbohydrates, when digested, change to glucose. Examples of some carbohydrates are: bread, rice, pasta, potatoes, corn, fruit, and milk products. Individuals with diabetes should eat carbohydrates but must do so in moderation. Glucose is then transferred to the blood and is used by the cells for energy. In order for glucose to be transferred from the blood into the cells, the hormone - insulin is

needed. Insulin is produced by the beta cells in the pancreas (the organ that produces insulin). In individuals with diabetes, this process is impaired. Diabetes develops when the pancreas fails to produce sufficient quantities of insulin – Type 1 diabetes or the insulin produced is defective and cannot move glucose into the cells and occurs most frequently in children and young adults, although it can occur at any age. Type 1 diabetes accounts for 5-10% of all diabetes in the United States. There does appear to be a genetic component to Type 1 diabetes, but the cause has yet to be identified. – Type 2 diabetes. Either insulin is not produced in sufficient quantities or the insulin produced is defective and cannot move the glucose into the cells, is much more common and accounts for 90-95% of all diabetes. Type 2 diabetes primarily affects adults, however recently Type 2 has begun developing in children. There is a strong correlation between Type 2 diabetes, physical inactivity and obesity. [9].

1.1 Data warehouse Model

A data warehouse is a system that extracts, cleans, conforms, and delivers source data into a dimensional data store and then supports and implements querying and analysis for the purpose of decision making [11]. It collects and stores integrated sets of historical data from multiple operational systems and feeds them to one or more Data Marts, which are data structures that are optimized for faster access. It may also provide end-user access to support enterprise views of data. Today, data warehouses are not only deployed extensively in banking and finance, consumer goods and retail distribution and demand-based manufacturing, it has also become a hot topic in noncommercial sector, mainly in medical fields, government, military services, education and research community etc. A Data Ware can potentially provide numerous benefits to an organization with quality improvement, and decision support by enabling quick and efficient access to information from legacy systems and linkage to multiple operational data sources.

An essential infrastructure requirement was an information management system designed to detect, track, and report the occurrence of transfusion-blood transmitted disease. Data are stored and managed in the warehouse and data marts which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools [12].

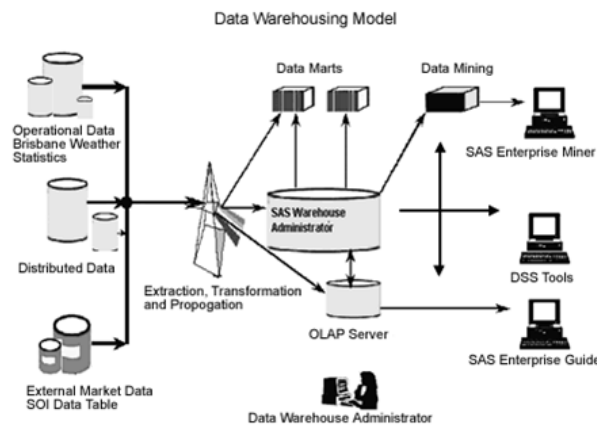


Figure 1.1: Example Data Warehousing Model [primary source: Marempudi, (2001)]

The Data Warehouse is Subject-Oriented, Integrated, Time-Variant, Non-Volatile data in support of management decisions [7, 8].

- **Subject-oriented:** means that all relevant data about a subject is gathered and stored as a single set in a useful format. Information is presented according to specific subjects or areas of interest.
- **Integrated:** refers to data being stored in a globally acceptable fashion with consistent naming conventions, measurements, encoding structures, and physical attributes, even when the underlying operational systems store the data differently.
- **Non-volatile:** means stable information that does not change each time an operational process is executed. Information is consistent regardless of when the warehouse is accessed.
- **Time-variant:** means that the data warehouse contains a history of the subject, as well as current information. Data warehouse data represents long-term data from five to ten years in contrast to the 30 to 60 day time period of operational data.

II. Building Data Warehouse For Diabetes Diseases

This is similar to the linear value chain model, but occurs in a data warehouse where the related fact tables that share common dimensions are not arranged in a linear order, where as many as 10 separate entities are processing patient encounters, and are sharing the information with one another. Many healthcare processes involve a series of patient visits or a series of outcomes. The modeling of outcomes associated with these types of healthcare processes is different from and not as well understood as the modeling of standard industry environments. The treatment is measured or generated by all the processes and organizations around the circle as show in Fig 2 below.

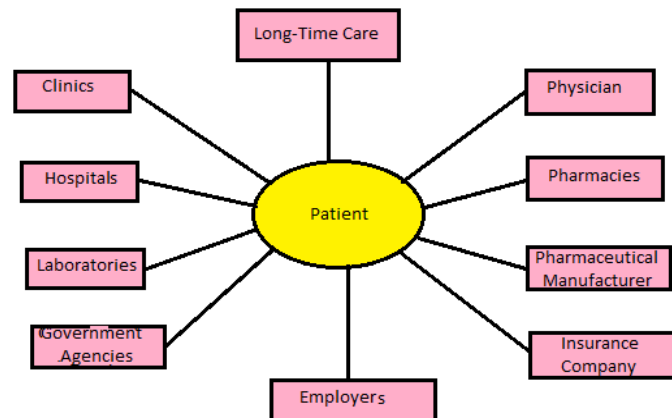


Figure 2.1 health care value Circle

2.1 Data Warehouse Model For Diabetes

A data warehouse model for Diabetes is shown in Fig 2.1 The characteristic models for the Diabetes data warehouse are:

Gender: This model contains the gender of the diabetic patient.

Age: The patient's age can be referred as the age itself, or it can be aggregate to age groups, which are important and interesting from a medical point of view.

Region: The patient's address can be analyzed, regarding language area, canton, zip code or town.

Physician: It is important to know the physician in charge for a patient.

MedicalHistory: This dimension contains the diseases the patients suffered in the past (dyslipidaemia, hypertonia, apoplexy, angina pectoris, cardiac insufficiency) and if there are diseases in his family significant for his disease (apoplexy or diabetes). Dependent of what analysis should be made and which information is needed, this dimension can be more detailed. It can contain information about the disease period and severeness, the treatments the patient is doing, the result and possible complications of the treatments. This dimension can be very complex, and it should be defined in close collaboration with the medical experts that will use this information.

DiabetesType: The patient can suffer from different diabetes types: Type I, Type II or, sometimes, gestational diabetes.

YearDiabetesDiagnosis: The year of the diabetes diagnosis, and thereby the duration of Diabetes, is an important issue in analyzing the health evolution and parameters of a patient.

Role: The parameters can be measured by the patient himself, patients relative, the Physician/nurse in charge, medical lab staff etc.

MeasurementType: This dimension contains information concerning the way the parameters were measured: automatic, manually etc.

MeasurementPlace: This dimension contains information concerning the place where the measurement has been made: medical practice, medical lab, home etc.

Parameter: Diabetic patients have to measure several parameters, with different frequency (see table 1). The parameters measuring frequency, shown in table 1 is the minimum necessary. It is desirable that a diabetes patient measures his parameters more often, e.g., blood sugar and blood pressure several times a week, or even daily.

Time: It is necessary that the physician or the patient can see the evolution of his health status over time. Interesting time aggregations are hour, day intervals (morning, midday, afternoon, evening, night), day, week, month or year.

Other dimension could be necessary or interesting dependent on the information and analysis needs, and also on the availability of data.

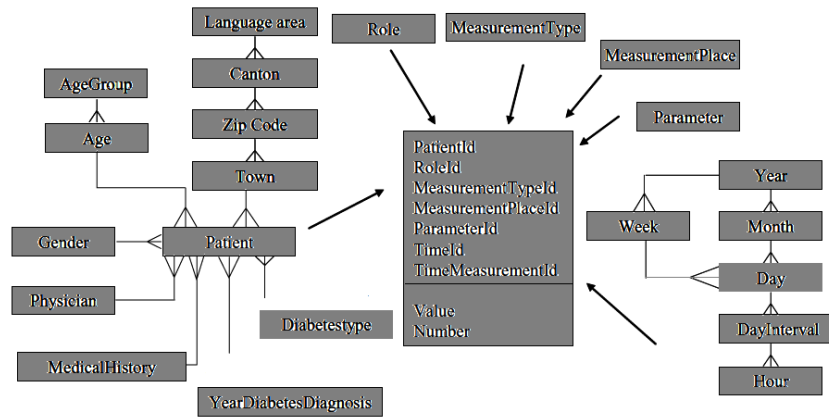


Figure 2.1: A Data Warehouse Model for Diabetes.

2.2 Business Analysis

The data warehouse enables several analyses. Both the physician and the patient can look to several statistics. The patient may be interested to see if certain behavior as nutrition- and movement program, diabetics training or drug therapy leads to improvements of patient health status.

Business Analysis is the discipline of identifying business needs and determining solutions to business problems. Solutions often include a systems development component, but may also consist of process improvement, organizational change or strategic planning and policy development [13]. The business analysis stage, consist of business process analysis as shown in the Fig 2.2 bellow.

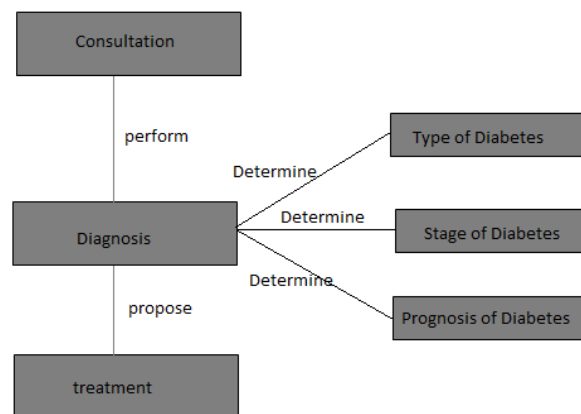


Figure 2.2 Diabetes data warehouse use case diagram

From the case diagram there are three main functional models in the process which are consultation, Diagnosis and treatment. At the consultation phase, Patient go to doctor for consultation when certain symptoms are noticed. But the doctor and the pathologist will undergone a series of test on the patient to determine the type, stage and the prognosis of the diabetes. if you are diagnosed with diabetes, the doctor will also run blood tests to check for auto antibodies that are common in type 1 diabetes. These tests help the doctor distinguish between type 1 and type 2 diabetes. The presences of ketones — byproducts from the breakdown of fat — in the patient urine also suggest type 1 diabetes, rather than type 2. The doctor will proceed with a number of tests to see how advanced the Diabetes is and how depth it has spread and also the doctor will look into other different factors to determine the prognosis of the disease. At the final phase the doctor and oncologist will counsel the patient and suggest the best the best treatment options available, based on the above investigation, type, the stage of the disease and some prognostic factors. Below are some treatment normally used to cure Diabetes:

Type 1 diabetes: Treatment includes:

- Taking several insulin injections every day or using an insulin pump.
- Monitoring blood sugar levels several times a day.
- Eating a healthy diet that spreads carbohydrate throughout the day.
- Regular physical activity or exercise.
- Regular medical checkups.

Type 2 diabetes: Treatment includes:

- Weight Loss Surgery
- Natural Remedies
- Non-Insulin Diabetes Injectables
- Oral Diabetes Medications

2.3 Architecture Design

Data warehouse architecture is a description of the components of the warehouse, with details showing how the components will fit together [7]. The Data Warehouse Architecture provides an integrated data warehouse environment while delivering incremental solutions. The architectural design focuses on the application of a centralized data warehouse, data marts, individual marts, metadata repositories, and incremental solution architectures. Different data warehousing systems have different structures. Some may have an ODS (operational data store), while some may have multiple data marts. Some may have a small number of data sources, while some may have dozens of data sources. Since a data warehouse is used for decision making, it is important that the data extracted from multiple sources should be corrected. It is inevitable that when different data are integrated into the data warehouse, there is a high probability of errors and anomalies. Therefore, tools for data extraction, data cleaning, data integration and finally data load are required. Data are stored and managed in the warehouse which presents multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools [13]. In view of Figure 5 below shows the overall propose Diabetes data warehouse architecture. The data is store in the production data warehouse for analysis while imported from several data sources and transformed them within a staging area before it is integrated.

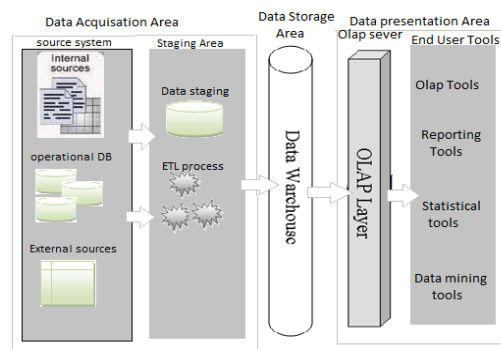


Figure 2.3 Diabetes Data Warehouse Architecture

The Figure above shows the three major areas of the Diabetes data warehouse architecture which includes Data acquisition, Data storage and Information delivery.

Data Acquisition: The objective of the Data Acquisition process is to identify, extract, transform, and transport the various source data necessary for the operation of the data warehouse. Data Acquisition is performed between several components of the warehouse, including operational and external data sources to data warehouse, data warehouse to data mart, and data mart to individual marts. In the architecture data acquisition covers the entire process of extracting data from the database, medical files such as medical records, other tests and results then moving all the extracted data to the staging area and preparing the extracted data for loading into data warehouse repository. This stage includes two functions which are data extraction and data transformation.

- Data Extraction Select data from multiple sources that is the database and medical files and determine the types of data(filters to be applied to the Diabetes data warehouse)
- Data transformation transform extracted data to data for data warehouse repository. Clean data, de-duplicate, and merge data from Database and medical files.

Data storage: This is where the transformed and cleansed data sit. Based on scope and functionality, 3 types of entities can be found here: data warehouse, data mart, and operational data store (ODS). This stage covers the process of loading the transformed data from the staging area into the data warehouse repository, All functions for transforming and integrating between the database and medical files are completed in the data staging area. The function and services for this area are the following:

- Load a health care data into data warehouse tables.
- Optimize the loading process.
- Perform incremental loads at regular prescribed intervals.

- Support loading integrated data into multiple tables

Data presentation: This refers to the information that reaches the users. This can be in a form of a tabular or graphical report in a browser, an emailed report that gets automatically generated and sent every day, or an alert that warns users of exceptions, among others. Usually an OLAP tool and or a reporting tool is used in this layer. The data presentation component makes it easy for the doctors and decision makers to access the information directly from the data warehouse. The function and services for this stage are the following:

- Doctors and decision makers can browse data warehouse content.
- Enable queries of aggregate tables for faster results and analysis.
- Provide multiple levels of data granularity.
- Monitor doctors and decision making access to improve service and for future enhancements.

III. Conclusion

Health care data warehouse have become more popular in recent years by means of incorporating modern computer and telecommunication technologies into health system and objectives to facilitate real time analyses. In this paper we use the technology of OLAP and ETL in Designing Diabetes data warehouse and the implementation tools are made on Microsoft Sql Server 2008, Sql server integration service 2008 (SSIS), Sql server analysis service 2008 (SSAS), Sql server Reporting service 2008(SSRS) and C # language. By using the diabetes data warehouse we can analyse the diabetes, the cost of treatment for these diseases, Death rate with respect to the type of Diabetes and the impact of a particular drug on the disease, provide the patients with accurate and reliable ways of caring for themselves, It allow the health care provider to make an informed healthcare decision, It will also allow the health care provider detect and subsequently diagnose the problem earlier and avoiding problems later. This will also help the health care provider improve the care they provide to their patients and also the system will identify the blood sugar levels and blood pressure and manage the treatment by recommending and monitoring food consumption, physical activity, insulin dosage, etc. in short our Diabetes data warehouse is able to assist executive managers and doctors by providing accurate and timely information for Healthcare decision-making This will help the health care provider improve the care they provide to their patients.

References

- [1]. KFF. (2006). "Comparing Projected Growth in Health Care Expenditures and the Economy", Snapshots: Health Care Costs Retrieved 22.10.2007, from <http://www.kff.org/insurance/snapshot/chcm050206oth2.cfm>
- [2]. Dr.Osama E.Sheta and Ahmed Nour Eldeen, BUILDING A HEALTH CARE DATA WAREHOUSE FOR CANCER DISEASES, International Journal of Database Management Systems (IJDMS) Vol.4, No.5, October 2012
- [3]. [3] Dr. Osama E.Sheta and Ahmed Nour Eldeen, THE TECHNOLOGY OF USING A DATA WAREHOUSE TO SUPPORT DECISION-MAKING IN HEALTH CARE, International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013.
- [4]. Marleen de Mul, Peter Alons, Peter van der Velde, Ilse Konings, Jan Bakker, Jan Hazelzet, (2010) Development of a clinical data warehouse from an intensive care clinical information system CMPB, 2010.07.002 .
- [5]. W H Inmon, DATA WAREHOUSING IN THE HEALTHCARE ENVIRONMENT, Inmon Data Systems, www.inmondatsystems.com.
- [6]. Sen, A. and Jacob, V. S. (1998): Industrial Strength Data Warehousing, Communication of the ACM, **41**(9), 28-31.
- [7]. Inmon, W. (2002): Building the Data Warehouse, 3rd edition, Wiley-New York.
- [8]. Kimbal, R., Reeves, L., Ross, M. AND Thronthwaite, W. (1998): The Data Warehouse Lifecycle Toolkit, Wiley- NY.
- [9]. Diabetes Research wellness foundation, what is Diabetes?. www.diabeteswellness.net
- [10]. Daniel ries, richard winter, Health care data warehouse in the government, 2012 winter corporation, combridge MA . www.wintercorp.com
- [11]. Ralph Kimball, Joe Casertam. (2004), The Data Warehouse ETL Toolkit, John Wiley, 1996.
- [12]. Surajit, Chaudhuri and Umeshwar, Dayal, An Overview of Data Warehousing and OLAP Technology, Appears in ACM Sigmod Record, March 1997
- [13]. Kathleen B Hass, Richard Vander Horst (2008) .From Analyst to Leader: Elevating the Role of the Business Analyst.
- [14]. Marempudi, A. (2001),Data Warehousing for better decisions. Intellibusiness, www.datawarehousing.com
- [15]. Andreea Ionas, Andreas Meier, Ciprian Pop, Marco Savini, Henrik Stormer, Using a Data Warehouse Approach for Mobile Patients, [andreea.ionas|andreas.meier|ciprian.pop|marco.savini|henrik.stormer]@unifr.ch
- [16]. Elham Mohammed Thabit Abd Alameer, Using REAL TIME DECISION SUPPORT SYSTEM FOR DIABETES MANAGEMENT, Journal of Kerbala University , Vol. 10 No.2 Scientific . 2012.
- [17]. E. Thomsen, G. Spofford, D. Chase (1999), Microsoft OLAP Solutions, Wiley, New York