

Flexible Data Warehouse Parameters: Toward Building an Integrated Architecture

Mustafa Musa Jaber, Mohd Khanapi Abd Ghani, Nanna Suryana, Mohammed Aal Mohammed, and
Thamir Abbas

Abstract—Clinical databases have gathered a huge amount of information about patients and their medical conditions. Relationships and patterns within this data could provide new medical knowledge. Thus it is a difficult task focusing to model a data warehouse, very often, into internal structures and implementation applications. The aim of its project is to find the parameters of medical data warehouse which will be able to construct a flexible framework by analyze the approaches of data warehouse architectures and compare them regarding the cost and integrity. The proposed parameters may assist in achieving continual access to the common data warehouse. The important consideration, however, is that the clinical data record should contain not only longitudinal health summary information but also be used for business intelligence tool such as data mining and OLAP.

Index Terms—Medical data warehouse, data warehouse architecture.

I. INTRODUCTION

Ref. [1] Defines data warehouse as the heart of the architected environment, and is the foundation of all decision support system (DSS) treatment. The work of the DSS analyst in the environmental data warehouse is much easier than in the environment of the classical heritage because there is a single integrated source of data (the data warehouse) and because the granular data in data warehouse is easily accessible.

The overall architecture of a data warehouse cannot adapt to the entire data store the creation of design data such as data modeling, normalization, and their attributes that make it easy to measure the effectiveness of treatment, captured the relationship between causation and systemic diseases and conditions [2].

It is not always necessary to create a data warehouse for analysis of OLAP. Data stored in operating systems, such as Point-Of-Sales, in the types of databases called OLTPs. OLTP, Online transaction process, the database does not have structural differences from the perspective of any other database. The main difference and the only difference is how data is stored [3].

The current medical architecture is containing a huge number of information across various departments, laboratories and related administrative processes, which are time consuming and laborious tasks to separately access and

integrate reliably. The current classification of the medical data representation in different institutions and other medical environments are facing difficulties in managing and retrieving the data contents. This study gave efforts to represent new medical data warehouse architecture to be a timely [2] solution for present-day health care requirements.

II. RELATED WORKS

A number of studies have been carried out with an aim of designing frameworks, models and/or theories explaining data warehouse implementation. The Table I shows the different between models and architectures.

III. DATA WAREHOUSE ARCHITECTURE

Data warehouse makes users gain access to wide reserves of integrated operational data to monitor trends in business, promotion of anticipation, making strategic decisions and planning efforts. Data warehouse is a process that enables organizations to extract object-oriented, time-varying, integrated, Non-Volatile corporate data from their information by traditional database management systems (DBMS) belongings is generally organized [4]. In contrast to the DBMS, which is usually used to represent storage online transaction processing (OLTP), data are for on-Line Analytical Processing (OLAP), decision making, and report generation [1]. In traditional relational databases the data are theoretically organized in one-dimensional view. Each data record has information that is organized in "fields" [5].

The data warehouse contains data about the goals of the organization and users who use high levels of management to support strategic decisions. However, such decisions can be taken at the lower levels of the organization associated with specific areas of business. In this case requires only a fraction of the data contained in the data warehouse is required. This subset usually contained in a data mart (which has a similar structure of a data warehouse) or may have its own separate platform.

A. Multi-Layer Architecture of Data Warehouse

Data warehouse (DW) play a central role in DSS, it provides vital business information to progress the decision-making process and to build a DW still difficult and complex because of design snags and the different of technologies [6]. The factors of DW are represented as multi-layer architecture as shown in Fig. 1. The different techniques and approach have been projected for the design of each DW layer [7]. The methods are not dealing with the whole DW model framework which customized the

Manuscript received March 25, 2014; revised August 2, 2014.

The authors are with Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia (e-mail: {mustafa, khanapi, nsuryana, thamir}@utem.edu.my).

development process of ETL layer (Extraction-Transformation-Load) processes, DW, repository, and so on.

B. Data Warehouse Architecture for Gis Applications

Geographic Information System (GIS) is a spatial decision support system. These systems allow users to explore and analyze geographic problems, in order to produce visual results that will help them make more informed decisions [1]. In other words, GIS is a system that allows users to generate interactive queries, analyze the spatial information, editing data, maps, and to show the results of all operations [8]. It can be used for research, resource and asset management, mapping, marketing and planning routes, among other [9].

Ref. [9] is creating a data warehouse (DW) for GEO-data to facilitate the generation of the report. A data warehouse is appealing as the main storehouse of historical data organization and is optimized for exposure and analysis. Fig.

2 shows the proposed architecture of a data warehouse for GIS applications. The processes of data warehousing is used to design and develop data warehouses for exposure and to effective DSS, data warehouse design and development, has already paying attention of several researchers.

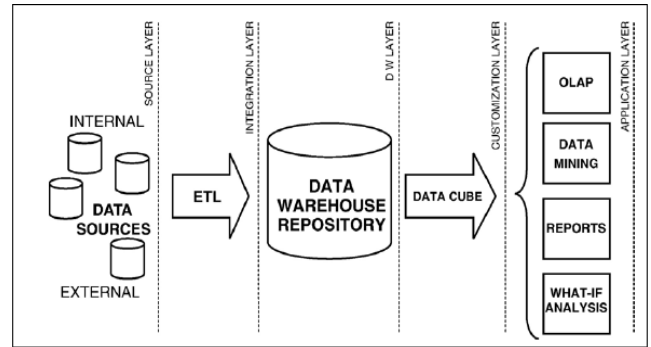


Fig. 1. Multi-layer architecture of data warehouses customization.

TABLE I: COMPARING BETWEEN DATA WAREHOUSE ARCHITECTURES AND MODELS

Study	Study Problem	Solution	Method	Finding
[4]	The retrieving process of data components based relational database has been found as a serious problem in the semantics web which needs to re-manage the structure relation.	This study proposed the using of data mining structure in demonstrating of the data elements to help the user in building a better multidimensional model for complex data.	Methodological framework for the process of multidimensional modeling of complex data. This method focused on customizing the data source elements from the complex database.	He founding of deploying this method is a relational solution for the data elements for the integration of complex data into relational (or XML-native) databases.
[5]	The using of the data base application for the whole system during the data processing.	This study presented automates entity relationship model into data warehouse logical model to generate semi-star schema by using artificial neural networks.	Entity-relationship diagram (ERD) to SS Schema Conversion and neural network techniques have been applies in this study for providing a semi automation technique to generate SS from ERD. This semi-automation requires five steps to be followed to get an SS schema as the output.	The obtained result presents the comparable of the existing techniques and gives better results than previous techniques.
[6]	Integrating data warehouse system warehouse based on multi access.	Domain-specific conceptual design issue based on low-grained, time segmented data into the traditional warehouse.	This study used ontologies classification of data warehouse domain, as well as of the healthcare and pharmacy domains, this one helps to simplify the data process matching between the federation and the component warehouses.	Providing data warehouse domain ontology, this domain simplified ontology process in the healthcare and pharmacy area.
[7]	Difficulties in the bridge table with all possible combinations of diagnoses.	Proposed a model resembles to the star schema to inherit its easy understanding and multidimensional aspects in the medical records.	This study applied an object-relational databases, this relation organize data in the relational tabular structures, which also integrate complex objects.	Presenting a dimension tables, based on five star schema classifications for the medical data relation.

A process for applying data warehousing techniques to serve GIS applications is included in the process as shown in Fig. 2 these include ESRI data source, preparing and Handling the Data, My SQL GIS Data Warehouse and Data Marts.

C. Data Warehouse Architecture for Network/Web Services

The main objective of Gems Data Warehouse (DW) is for

Automated Booking System (ABS) to make reports on performance, dependability and compare it with the system level agreements (SLAs) that the ABS has arranged to afford its corporate clients. An SLA is an agreement between the supplier and the client (usually a company) on the quality of service that must be delivered. [10] Developed system architecture for the DW, the main purpose of the system is to report service on performance and dependability. Fig. 3 shows the proposed architecture.

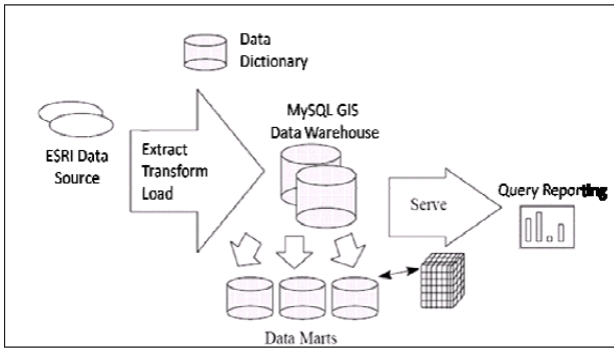


Fig. 2. Data warehouse architecture for GIS applications.

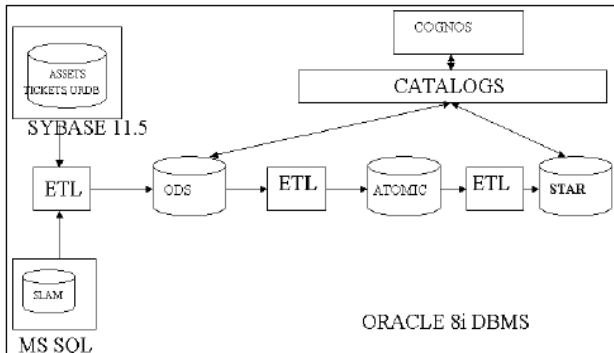


Fig. 3. Architecture of the data warehouse system.

DW system is used to reports service of the time required to produce reports is much enhanced. In addition, performance reports do not affect applications that use the source databases. An example of these reports is that the knowledge worker (executive, manager, and analyst) to generate monthly reports for clients and collect statistics such as mean time to repair (MTTR) and mean number of notes in the fault to a client company ABS [10].

D. Warehousing Architecture for Clinical Data

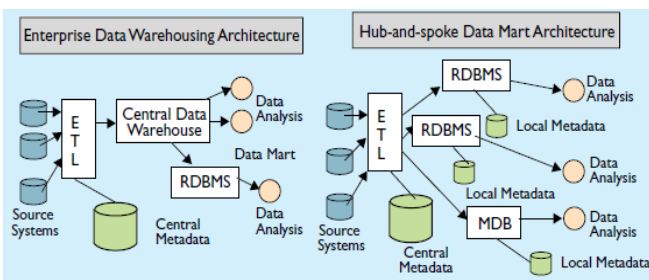


Fig. 4. Different types of DW architectures [13].

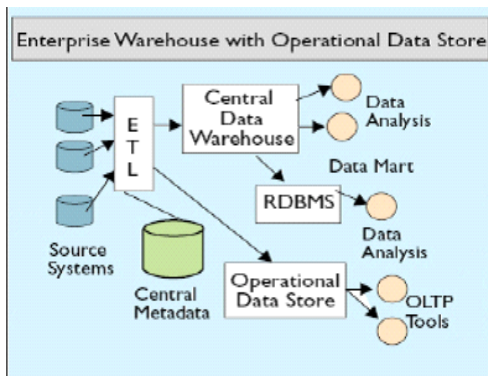


Fig. 5. Enterprise DW architecture [13].

Clinical data warehouse is a place where healthcare providers can get access to clinical data collected in the

process of patient care. It is also assumed that such a data warehouse can give information to users in areas ranging from research to management [11], [12]. In this regard, the organization of the captured design data such as data modeling, normalization, and their aspects that makes it easy to measure the effectiveness of treatment of the relationship between causality and treatment protocols for systemic diseases and conditions.

Ref. [13] did their experiment in the development of the CDW through a number of methodologies for data warehouse Fig. 4 presents [11], preserving the data aspects for function portability and allocation between societies.

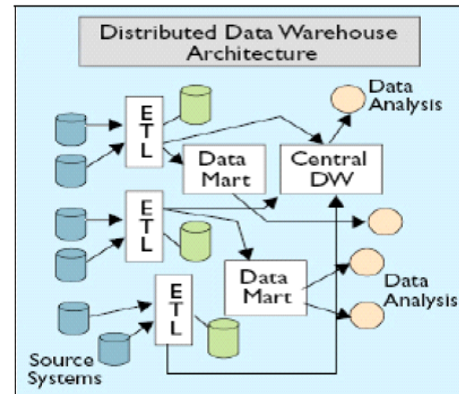


Fig. 6. Distributed DW architecture [13].

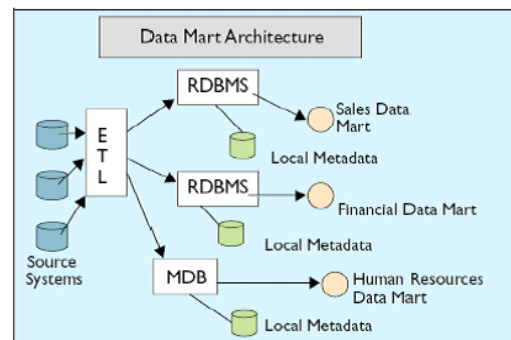


Fig. 7. Data mart architecture [13].

At the design stage, they are faced with issues such as some of the methodology for data warehouse cannot claim the proposed CDW. They experimented with all potential combinations, enterprise warehouse was implemented with intelligence Store architecture (Fig. 5), and disseminated data warehouse architecture (Fig. 6) using the SAS data warehouse administrator software module .

They chose this method with the spanning including several data marts (Fig. 7) for a variety of administration and operations management for example summaries can be performed by team leaders, the identification of clinical activities within a certain period, the factors expecting excellence of clinical results and regular observing clinical outcomes for senior management).

IV. SELECTION OF DATA WAREHOUSE ARCHITECTURE FACTORS

One of the most important aspects is selection of factors and how that particular factor does effect of the proposed architecture concluded as [12]:

A. Interdependence between the Organization Departments

The information that transferred between organizations itself could have rules and regulation depends on the organization structure. Thus it is likely that firms with high information interdependence select an enterprise-wide architecture and it will be a positive factor for the selection.

B. Top Management's Information Needs

With the purpose of providing the responsibilities of top management often requires information from lower organizational levels to monitor the progress of the workflow which needs to find out an architecture supports it. In addition, it is a positive effect on the architecture.

C. Urgency of Need for a Data Warehouse

Organization might need an urgent data of data warehouse or data mart and that urgency of the business need may dictate a fast implementation. There are architectures are more quickly implemented than others, which can influence the architecture that is selected.

D. Enduser Task

Some of the users need like a non-routine tasks which help them to provide the data in their own way. It gives a positive effect on the performance of data warehouse.

E. Compatibility of Existing Systems

One of the most important aspect is that capability of systems used in the data warehouse architecture such as platform and, distribution of data. Compatibility will have a positive effect on sharing information between the organization departments.

F. Ability of IT Staff

The building of a data warehouse can be a daunting task and implementing some data warehouse architectures. Novice user of business intelligent tool is a challenge of integration of architecture.

V. EVALUATION OF ARCHITECTURE SUCCESS

Data warehouse factors cannot be the only aspect of successful architecture based on literature there are some additional factor which is not internal factor such as information and system quality and, the ability of data integration. On the other hand, project management issues for instant, whether the implementation was on budget and on schedule. The major success categories are identified below.

A. Information Quality

Information quality contains the following dealings – information completeness, information consistency and information accuracy.

1) Information completeness

A warehouse must have almost all the decision support data that is needed. It should hold data for all essential business processes and subject areas and, provide the data that is needed by users and applications.

2) Information consistency

The creation of data warehouse is to have a single version

of the truth. It should eliminate the problem of having inconsistent data. The data warehouse should provide a single system of record for the organization.

3) Information accuracy

Data warehouse should be as accurate as its intended use demands. Queries and reports should contain few errors because of data problems. Real-world objects and events should be correctly described [12].

B. System Quality

It includes three measures – system integration, system scalability, and system flexibility.

1) System integration

Providing integrated data is an important requirement for a data warehouse. Using appropriate primary keys, a warehouse should integrate data from multiple sources, including both internal and external data.

2) System scalability

The demands on data warehouses grow over time, so they must be scalable. They should be able to handle increases in the number of users, the complexity and number of queries, and the volume of data without negatively affecting system performance.

3) System flexibility

Data warehouses should be flexible. It should be easy to add new business processes and subject areas. The warehouses should be able to adapt to new requirements quickly. They should be able to easily support future application needs.

C. Individual Impact

A data warehouse does not create value by itself. Value creation occurs when users employ the warehouse in their work. Users should be able to quickly and easily access data. They should be able to think about, ask questions, and explore issues in ways that were not previously possible. Overall, the warehouse should improve users' decision-making capabilities.

D. Organizational Impacts

Ultimately, the warehouse should have positive impacts on the organization. It should satisfy the business requirements for which it was built, facilitate the use of BI, support the accomplishment of strategic business objectives, enable improvements in business processes, lead to high, quantifiable ROI, and improve communications and cooperation across organizational units.

VI. CONTRIBUTION OF STUDY

Choosing of flexible architecture component helps decision makers to use information to weigh alternatives, analyze options, and anticipate implications and project outcomes within the context of their organization and environment to design a good architecture for medical data warehouse. On other hand, the problems which are existing in information system in medical organization which are not suitable for decision support, and they have been designed to support short transactions affecting a few records at a time will be solved.

VII. MEDICAL DATA WAREHOUSE COMPONENTS

Data warehouse architecture consists of seven components. These components offer a high level of flexibility and scalability for companies and institutions wishing to implement the decisions of the Business Intelligence [14].

This study has included the following components for designing a medical data warehouse architecture:

A. Dataset Resources

The dataset resources have been adopted from UCI of Breast Cancer data sets.

B. Data Staging Area

Data staging area is a preparation to build a data warehouse which limited to the production, cleaning, alignment and load data from legacy systems based on ETL tool.

C. Medical Data Warehouse

This section concerned on store any special medical technology in term of data mart relation. This section has been reported as a source of reality for all common medical data [15].

D. Data Marts

Helped to provide a conceptual view of dataset elements. This section targets its entry into data warehouse through different data ports [16].

E. Data Extraction-Transformation-Load (ETL)

Is used to extract data from different data sources and transform data to common form, then moving the data into data marts. This section helped to integrate medical data [17], [18].

VIII. COMPONENTS STRENGTH

This study has been employed more than approaches of medical dataset as data store. Technically OLAP can be used to design the architecture, for users, the hospitable administrators and top manager or/and sophisticated user can use MDW by using Microsoft SQL Server. Building the architecture can by using Microsoft Visual Studio for performing the OLE database operations. SQL can be used to make the database that stores the necessary medical dataset information for the admin.

IX. COMPONENTS BUILDING LIMITATION

Although these components provide the admin with an easier manner for view the database relations in more advance presentation, there are some significant disadvantages to the proposed architecture which addressed the following points:

- Resources no actual server can be utilized in testing the proposed application.

- The developing of Microsoft Visual Studio pages is more complex than performing the database operations because of the limitation of the contents.

The medical architecture cannot tested using local host server. But with limited financial

REFERENCES

- [1] W. H. Inmon, *Building the Data Warehouse*, John Wiley & Sons, 2005.
- [2] T. R. Sahama and P. R. Croll, "A data warehouse architecture for clinical data warehousing," *Book a Data Warehouse Architecture*, 2007.
- [3] A. Sen and A. P. Sinha, "A comparison of data warehousing methodologies," *Communications of the ACM*, vol. 48, no. 3, pp. 79-84, 2005.
- [4] R. Kimball, M. Ross, and R. Merz, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, Wiley, 2002.
- [5] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, and B. Becker, *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing Inc., 2008.
- [6] D. Perez, M. Somodevilla, and H. Pineda, "Fuzzy spatial data warehouse: A multidimensional model," in *Proc. Eighth Mexican International Conference on Current Trends in Computer Science*, 2007.
- [7] H. Schepperle and A. Merkel, "Data warehouse," in *Proc. Internationals Symposium: Data-Warehouse-System und Knowledge-Discovery*, 2004.
- [8] H. Zimmerman, *Fuzzy Set Theory and Its Applications*, Kluwer Academic Publishers, 1991.
- [9] R. Mohammad, J. B. Erwin, M. Abdel Rahman, M. Terrance, M. Keivan, and A. Reda, *Data Warehouse Architecture for GIS Applications*, 2008.
- [10] D. Wierschem, J. McMillen, and R. McBroom, "What academia can gain from building a data warehouse," *Educause Quarterly*, vol. 1, pp. 41-46, 2003.
- [11] A. Sen and A. A. Sinha, *Comparison of Data Warehousing Methodologies*, *Communication of the ACM*, vol. 48, no. 3, pp. 79-84, 2005.
- [12] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, pp. 37-54, 1996.
- [13] S. Tony and C. Peter, "Data warehouse architecture for clinical data warehousing," in *Proc. Australian Computer Society, Conferences in Research and Practice in Information Technology*, 2007, p. 68.
- [14] M. Banek, A. M. Tjoa, and N. Stolba, "Integrating different grain levels in a medical data warehouse federation," *Data Warehousing and Knowledge Discovery*, Springer, pp. 185-194, 2006.
- [15] W. Y. Lin and C. A. Wu, *An Object-Relational Modeling for Clinical Data Warehouse*, 2005.
- [16] R. Kimball and M. Ross, "The data warehouse toolkit: the complete guide to dimensional modelling," *Nachdr*, New York: Wiley, 2002.
- [17] M. Abdulameer, N. Suryana, H. A. Abdullah, and M. M. Jaber, *Convert Database Structure into Star Schema Structure for Data Warehouse*, 2011.
- [18] D. Wegener and S. Rüping, "On reusing data mining in business processes-a pattern-based approach," *Book on Reusing Data Mining in Business Processes-A Pattern-Based Approach*, Springer, pp. 264-276, 2011.



Mustafa Musa is an Iraqi student studying in Malaysia majoring in software engineering at Technical University Malaysia. This project was completed as a part of his postgraduate studies. Mustafa graduated with a master degree of information technology in May 2010, and pursuing his PhD degree of software engineering, with telemedicine as the research topic, in 2011. Mustafa has joined Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group with the view of furthering his studies in health systems internationally in the future and hereby contributing to improved healthcare in Iraq.