International Research Based Journal
Vol (7)-Issue(2),Mar,12, 2023,ISSN 2348-1943

A Study of Health Data Warehousing and Health Data Warehousing Process

Dr. Neha Bhardwaj
Assistant Professor,
Pt. C.L.S. Government College, Karnal,India

Abstract-
The purpose of a data warehouse is to extract data from operational systems and transform it into formats suitable for data analysis. Supposed to operational systems in which data is deleted when it is no longer needed by a particular application, data warehouses retain data over time. It is the non-volatility of the data that makes historical analysis possible. As opposed to operational systems which store the most recent version of the data, data warehouses keep track of it, including a history of the changes that took place. Time-variance enables trends analysis is over time. Finally, the purpose of the data is to improve management by gaining a better understanding of the enterprise .
Keywords: Data warehouse, Data Acquisition.

1. Introduction

A data warehouse is a "centrally managed and easily accessible copy of data collected in the transaction information system so far corporation. These data area aggregated, organized, catalogued and structured to facilitate population-based queries, research and analysis" Best known as the "father of data warehousing," Bill Inmon characterizes the data warehouse as "a subject-oriented, integrated, non-volatile and time-variant collection of data in support of management's decisions." The data is not organized to support specific applications such as laboratory or imaging systems, but rather by subjects, i.e. patients, and is therefore subject-oriented. The data originates in multiple operational systems, and is integrated both by definition and content.

Several attributes characterize a data warehouse. A data warehouse is large by definition, i.e. it contains up to thousands of terabytes of data or more. The system offers a historical perspective, i.e. it can cover up to 30 years or more. The technology integrates data from several transaction information systems, i.e. data is collected from source systems such as billing, registration or scheduling. The analysis provided by a data warehouse spans across multiple business processes, e.g. the data pertaining to a billing system will be compared against the data contained in a scheduling system. A data warehouse provides an explorative approach. It offers insight into areas that have not yet been investigated and issues that have not yet been anticipated. The output of the data warehouse takes the form of reports and metrics. From the collection of the data to its transformation for query purpose sand the production of reports, data warehouses require the intervention of specialized staff .The key role of a data warehouse is to provide decision-maker swith the compelling business intelligence that enables them to understand problems, discover opportunities, and measure performance. To effectively play

this role, the data warehouse must integrate the internal and external data acquired over time and translate it into current conditions. In doing so, the data warehouse is the instrument that enables decision makers to locate and apply relevant data, and helps them to predict and measure the impact of their decisions over time.

2. Health Data Warehousing Process

In order to become a real organizational asset leveraged throughout the organization, data must be properly identified and inventoried. It must be extracted, organized, combined, stored and managed in a secured manner. Based on user requirements and reporting expectations, a master data model is established as the foundation for the warehousing effort which, as shown on Figure1, encompasses four functions.

Data Acquisition. A data warehouse acquires its data from the organization's operational systems as well as from systems external to the organization such as suppliers and regulatory institutions. Not only must the data be extracted from the source systems, but it must be cleaned and transformed to conform to the standardized architecture, and it must be loaded into the data warehouse. Known as extract- transform-load or ETL, this function is at the core of data warehousing. Equally central is the establishment of robust metadata, i.e. the comprehensive documentation of the data and all processes related to the data warehouse including data models, a data dictionary and ETL load statistics which must be made readily available to the user community (Adelman, 2003; Imhoff, Galemmo, & Geiger, 2003; Inmon, Imhoff, & Terdeman, 1999; Sakaguchi & Frolick, 1997).

Data Warehouse Population. In order to be presented to users in a uniformand consistent manner, the data that flows in the data warehouse must follow a consistent and logical process. Based on the organization's requirements and experience, the storage of the data usually follows the dimensional approach, which stores data in a form similar to both its true dimensionality and the form needed at the time of reporting, or the relational approach, which relies on relational database management principles (Inmon et al., 1999; Saka guchi & Frolick, 1997).

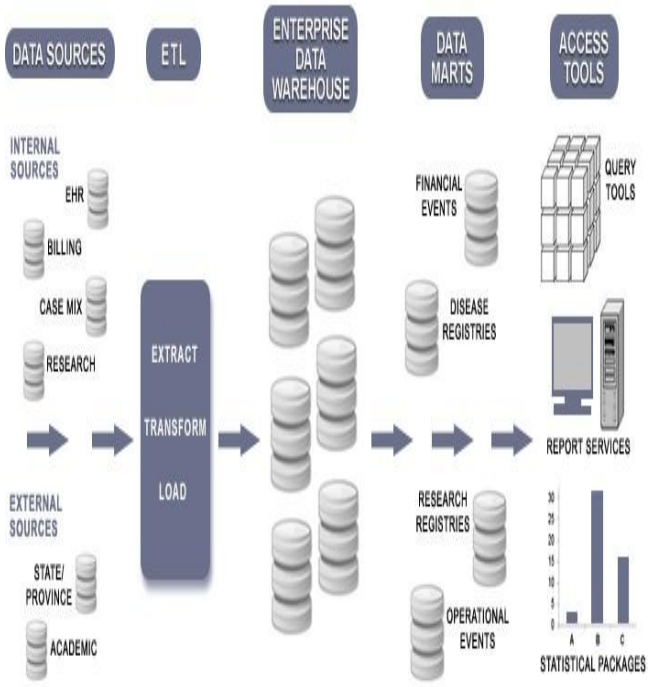

Figure1–Health Data Warehouse Architecture

Data Marts Creation. To better address the needs of specific users, smaller subsets of data are drawn from the data warehouse in the form of custom-designed databases and offer a multidimensional view of the data such as by

service or location as well as overtime. Such subsets enable a better understanding and greater probing of the data, and offer faster responses to queries (Hristovski, Rogac, & Markota, 2000).

Information Access. In order to meet the information needs of the end-users, a series of analysis and reporting tools must be made available that leverage the latest technology with minimal overlap. Structured Query Language (SQL) queries are used on an ad-hoc basis and involve specific interfaces. On-Line Analytical Processing (OLAP) is used to "slice and dice" large volumes of data, and provides analysts with an interface to manipulate views and levels of aggregation. Web reporting enables the selection and presentation of customized reports from web interfaces. Dashboards and scorecards utilize graphical interfaces to present key indicators and quality measures and allow drilling down into top-level measures to assess their components. Data mining techniques apply algorithms to summarize, model and cluster data with the aim of identifying novel and potentially useful correlations and patterns in data. Statistical analysis can be applied to the data, e.g. measures of central tendency, analysis of variance, regression, and time series analysis. Geographic Information Systems(GIS) provide geographic displays of data and can be used, among others, to analyze where are sources are needed for specific products and services, or to demonstrate geographic variations in distribution or consumption patterns

3. State of Health Data Warehousing Evaluation

If the anticipated benefits of data warehouses are quite significant, the system requires ample financial and technical resources, as well as qualified lab or and time. By spanning across entire organizations, it is subject to multiple individual and organizational factors. Its dependency on existing source systems renders the quality of its output vulnerable, and since its output is in the form of reports and metrics used for decision support purposes, numerous and potentially critical repercussions are associated to its use. The level of complexity is further increased when data warehousing is applied to healthcare. Medical data is more voluminous and heterogeneous than the data found in any other economic sector, and in order to cover all aspects of the care process, data warehouses must address areas as varied as clinical research, treatment effectiveness, financial analysis and customer relationship.

If the literature extols the virtues of health data warehouses, there is little evidence of their assessment. even when extended to sectors not related to healthcare, the review of the current literature returns very few publications on data warehousing evaluation. Only isolated dimensions such as data and system quality or user satisfaction have received more attention and benefit from definitions and performance measures . Whenever knowledge has been gathered by the practitioners' community, it is mostly relying on anecdotal evidence. Furthermore, very few cases can be found that describe how the assessment of the technology can be made, and these cases do not provide insight on how to systematize such assessment. In lieu of evaluation principles and methods, the concept of success often serves as a basis to the system's assessment. As discussed in, Critical Success Factors (CSFs) are an

application of such concept. CSFs are elements identified as vital in order for an organization or project to achieve its mission and reach successful targets. Failure to meet the objectives associated with CSFs results in the failure of the project or organization (Watson, Gerard, Gonzalez, Haywood, & Fenton, 1999).The research departed from CSFs. Not only does this practice reduce assessments to a dichotomous approach, but it relies on ill-defined and contentious concepts. Data warehouses can be successful if implemented within budget, but at the same time their effective use may not. The converse can equally be true, or the technology may be highly appreciated by users but nonetheless not widely used. In other words, some CSFs are identified without being empirically used and some of the factors used involve different measuring mechanisms. The notion of success has also been the object of extensive academic research. For over twenty years, DeLone and McLean have attempted to define information systems success along with the dimensions of success themselves. Their theoretical model was chosen as a foundation to the research described in this dissertation. The latter also departed from the DeLone and Mc Lean theory. If its model and the empirical studies conducted to apply it have helped establish a set of success factors and measures, a consensus is still lacking as to what the concept of success entails. In the absence of such consensus, the use of factors such as the actor for whom success is defined and rather his/her view can prevail with in an organization persists. Asa result, the assessment of information systems remains often unaddressed, and the comparison of systems across organizations remains largely impossible.

References

1. Subramanian,G.H.(1994).Areplication ofperceivedusefulnessandperceivedea seofuse measurement. *Decision Sciences*, *25*(5–6), 863–874.
2. Success.(n.d.)In*OxfordEnglishDictionary' sonlinedictionary*(2012).Retrievedfrom
3. *http://dictionary.oed.com/success*
4. Symons,V.(1991).Areviewofinformations ystemsevaluation:content,contextandproce ss.*European Journal of Information Systems*, *1*(3), 205–212.
5. Taylor,S.,&Todd,P.(1995).Understandingi nformationtechnologyusage:Atestofcompe tingmodels.
6. *InformationSystemsResearch*,*6*(2),144–176
7. Tellis, W.(1997).Introduction to case study. *Qualitative Report,* 3(2) . Retrieved from http : //www.nova.edu/ssss/QR/QR3-2/tellis1.html
8. Teng, J., & Calhoun, K. (1996). Organizational computing as a facilitator of operational and managerial decision making: An exploratory study of managers' perceptions. *Decision Sciences*, *27*(4),673–710.
9. Teo, T.S.H., & Choo, W.Y.(2001). Assessing the impact of using the Internet for competitive intelligence .*Information and Management*,*39* (1),67–83.

10. Teo ,T .S .H.,& Wong, P. K. (1998).An empirical study of the performance impact of computerization in the retail industry. *Omega—The International Journal of Management Science*, *26*(5), 611–621.
11. Torkzadeh, G., & Doll, W. J. (1999). The development of a tool for measuring the perceived impact of information technology on work. *Omega—The International Journal of Management Science*,*27*(3), 327–339.
12. Urbach, N., Smolnik, S. ,& Riempp, G. (2008) .*A methodologica l examination of empirica lresearchon information systems success: 2003 to 2007*. Proceedings of the 14th Americas Conference on Information Systems, Toronto, Ontario.