ISSN: 2321-7782 (Online)
Volume 3, Issue 8, August 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Investigation into Data Warehouse to Meet the Growing Demand of Information Analysis

Jyoti[1]
CBS Group of Institution,
Fathepuri, Jhajjar
India

Rajiv Munjal[2]
(Assistant CSE, Department),
CBS Group of Institution,
Fathepuri, Jhajjar, India

Upasna[3]
Ganga Institute of technology and management,
Jhajjar, India

Abstract: ***Data Warehouses are developed to meet the growing demand for information analysis that could not be met by operational systems. This is because the processing load of reporting affects their response time and is not optimized for strategic decision making. It enables the organization to make use of an enterprise wide data store to link information from diverse sources. The information is now accessible to decision makers for strategic analysis which includes trend analysis, forecasting, competitive analysis &targeted market research.***

Keywords: DWH, EDW, DATAMART, OLAP, OLTP

I. INTRODUCTION

In computing, a data warehouse (DW or DWH), also known as an enterprise data warehouse (EDW), is a system used for reporting and data analysis. DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data and are used for creating analytical reports for knowledge workers throughout enterprise. Examples of reports could range from annual and quarterly comparisons and trends to detailed daily sales analyses. Data stored in warehouse is uploaded from operational systems (such as marketing, sales, etc., shown in figure to right). Data can pass through an operational data store for additional operations before it is used in DW for reporting.

II. TYPES OF SYSTEM

a) Data Mart

A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as sales, finance or marketing. Data marts are often built and controlled by a single department within an organization. Given their single-subject focus, data marts usually draw data from only a few sources. Sources could be internal operational systems, a central data warehouse, or external data.[1] Denormalization is norm for data modeling techniques in this system.

b) Online analytical processing (OLAP)[8]

OLAP is characterized by a relatively low volume of transactions. Queries are often very complex and involve aggregations. For OLAP systems, response time is an effectiveness measure. OLAP applications are widely used by Data Mining techniques. OLAP databases store aggregated, historical data in multi-dimensional schemas (usually star schemas). OLAP systems typically have data latency of a few hours, as opposed to data marts, where latency is expected to be closer to one day.

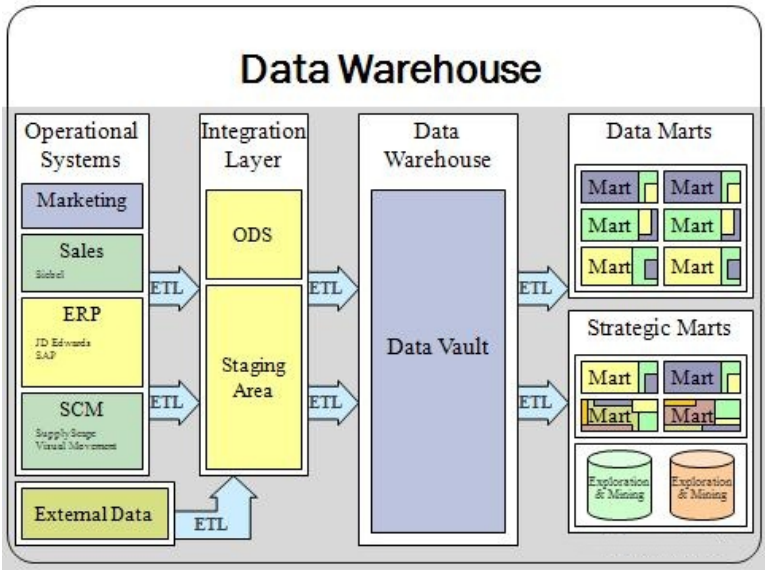

Fig 1. Data Warehouse

c) Online Transaction Processing (OLTP)

OLTP is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE). OLTP systems emphasize very fast query processing and maintaining data integrity in multi-access environments. For OLTP systems, effectiveness is measured by number of transactions per second. OLTP databases contain detailed and current data. schema used to store transactional databases is entity model (usually 3NF). Normalization is norm for data modeling techniques in this system.

d) Predictive analysis

Predictive analysis is about finding and quantifying hidden patterns in data using complex mathematical models that can be used to predict future outcomes. Predictive analysis is different from OLAP in that OLAP focuses on historical data analysis and is reactive in nature, while predictive analysis focuses on future. These systems are also used for CRM (Customer Relationship Management).

III. NEED OF DATA WAREHOUSE

- **Immediate information delivery:** It reduces time for processing of a request for example sales report can be formulated on a daily basis which enables business analyst to exploit opportunities.
- **Integration of data:** It combines data from multiple sources into a single cohesive unit.
- **Provides an insight into future:** It stores a large amount of historical information that enables decision makers to analyze prevailing trends in market.
- **Enable decision makers to look data in different ways**: It provide tools for manipulation of data and facilitate users to drill down into detail data with click of mouse.

IV. BENEFITS OF A DATA WAREHOUSE [2]

- With data warehousing, you can provide a common data model for different interest areas regardless of data's source. In this way, it becomes easier to report and analyze information.
- Many inconsistencies are identified and resolved before loading of information in data warehousing. This makes reporting and analyzing process simpler.

» The best part of data warehousing is that information is under control of users, so that in case system gets purged over time, information can be easily and safely stored for longer time period.

V. DATA WAREHOUSE USER [4]

» These are people whose job involves analyzing data to draw meaningful conclusions & make decisions based on large masses of data. They want information for following reasons:-

» Profitability Growth: information from data warehouse provides support to plan for profitability, growth and to assess results when plans are executed

» Strategic Marketing: It guides process of business growth. data warehouse offers great potential for strategic marketing by providing users with information regarding up selling and cross selling.

» Customer Relationship Management: data warehouse contains all information about customers which provides an opportunity to executives and managers to learn their customers individually from information available in data warehouse.

» Corporate Purchasing: data warehouse contains corporate wide information about different vendors and customers purchasing patterns. classification of user in based upon two perspectives their computing proficiency and their job function

» Casual user or Naïve user:-It uses data warehouse occasionally & needs a very intuitive information interface

» Regular user: They use computing option but cannot create reports on own.

» Power user: They create report and execute queries.

» Users based upon job classification:

» Executives and managers: They need information for strategic decision.

» Technical analysts: Perform complex analysis and perform drill down and roll up operations.

VI. TOOLS FOR DATA WAREHOUSE IMPLEMENTATION

This section presents a brief description of tools used in different phases of data warehouse construction. Barker (2000) conducted a research that, as a result presents four groups of data warehousing tools:

» Analysis Tools: used to study current operational database systems, to identify requirements, primary data source for information during acquisition and building data model (E.g. Bachman Analyst Tool). Development Tools: used during code generation for information acquisition, data cleansing, data integration and loading (E.g. Oracle Tool).

» Implementation Tools: used for cleansing, consolidating and loading data, some of these tools can be developed in-house. Data acquisition tools used to gather and clean data (E.g. IBM and Oracle) and information store tools used to load data into data warehouse (E.g. Oracle and Prism)

» Delivery Tools: assist in data conversion, data derivation, data loading and reporting on delivery platform. Data loader converts data from host computer to delivery platform (E.g. Oracle Developer and Structured Query Language), data glossary describes in business terms what data is on data warehouse (E.g. Lotus Notes) and querying and reporting are on-line and batch

» Reporting facilities. research presented by Baker shows that it is difficult to choose a tool for data warehouse implementation. Each data warehouse phase uses a specific tool. We need to choose carefully tool to use; otherwise we will comprise process inside phase. Different vendors have different tools for different phases of data warehouse

implementation. It seems that, based on literature review, Oracle constitutes a good tool to use during data warehouse implementation. Domenico (2001) presents a research on data warehousing area. In implementation part, he uses Oracle Data Mart Suiteâ

He chose Oracle tool based on following reasons:

» It is a tool that is easy to learn (he took three months to implement project).

» The tool supports life cycle of data mart, from project until end user results analysis.

He argues that it is important to look to infrastructure within organization where data warehouse will be implemented. Oracle environment was already implemented at Universidade do Oeste de Santa Catarina – Brazil, so it was easy to familiarize with tool.

VII. SCOPE OF RESEARCH

Study on data warehousing revealed that there are two main architectural approaches to data warehousing. An approach called as relational modeling builds data warehouse around a big central database and relies on normalized relational data model in storage of data to a relational database. other architectural approach called as dimensional modeling relies on a distributed architecture and use of a multidimensional data model in storage of data to a relational or multidimensional database. We have chosen dimensional model approach for data warehouse design. As we know that data warehouse is a collection of data from disparate sources, we have chosen ER model as our source furthermore, we have seen that there are various algorithms available for conversion of ER model to dimensional model.

References

1. Reema Thareja, Data warehousing, Oxford,2009.
2. Inmon W.H., Building Data Warehouse, John Wiley, New York(2nd edition ,2000)
3. Rilson,F., and Freire, J., “DWARF: AN Approach for Requirements Definition and Management of Data Warehouse Systems”, Proceeding of 11th IEEE International Requirements Engineering Conference 1090-9, September 08 – 12,2003.
4. Husemann, B., Lechtenborger, J., Vossen, G., “Conceptual Data Warehouse Design” Proceedings of International Workshop on Design and Management of Data Warehouses (DMDW’2000) Stockholm, Sweden, June 5-6, 2000
5. Golfarelli, M., Rizzi, S,” Designing Data Warehouse: Key Steps and Crucial Issues” in Journal of Computer Science and Information Management, Vol. 2. No.3, 1999
6. Naveen Prakash, Anjana Gosain,”An approach to engineering requirement of datawarehouses, Requirement enginnering”,Vol 1.No.13:49-72 ,2008
7. M. Golfarelli, D. Maio and S. Rizzi, Conceptual design of data warehouses from E/R schemes, Proc. Hawaii International Conference on System Sciences, Kona, Hawaii (1998), 334-343
8. Moody, D.L.and Kortink, M.A.R.,"From ER Models to Dimensional Models: Bridging theGapbetween OLTP and OLAP Design", Journal of Business Intelligence, 8, 2003.
9. Moody, D.L. and Kortink, M.A.R.,” From ER Models to Dimensional Models: Advanced Design Issues”, Journal of Business Intelligence, 8, 2003.
10. Chen,P.P.,”The entity relationship model :toward a unified view of data”, ACM transactions on database system,1,1,march 1976,9-37.

