IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 3, Issue 3, June-July, 2015
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

Data Warehousing - Essential Element To Support Decision-Making Process In Industries

Ashima **Bhasin**[1], Mr Manoj Kumar[2]

[1]Computer Science Engineering Department, [2]Associate Professor, CSE

SGT Institute of Engineering & Technology, Gurgaon, India

Abstract

This paper provides an overview of Data warehousing, OLAP, OLTP technologies, exploring the features, applications and the architecture of Data Warehousing. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases. Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. OLTP is customer-oriented and is used for transaction and query processing by clerks, clients and information technology professionals. An OLAP system is market-oriented and is used for data analysis by knowledge workers, including managers, executives and analysts. Data warehousing and OLAP have emerged as leading technologies that facilitate data storage, organization and then, significant retrieval. Decision support places some rather different requirements on database technology compared to traditional on-line transaction processing applications.

Keywords: Data Warehousing, OLAP, OLTP, Decision Making and Decision Support

1. INTRODUCTION

A data warehouse is a "subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organizational decision making. Typically, the data warehouse is maintained separately from the organization's operational databases. There are many reasons for doing this. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases. Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker to make better and faster decisions. It serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. The data can be stored in many different types of databases. One data base architecture that has recently emerged is the "data warehouse" a repository of multiple data sources.Data warehouse technology includes data cleansing, data integration and online Analytical processing. OLAP stands for analysis techniques with functionalities such as summarization, consolidation and aggregation, as well as the ability to view information from different angles.

Ten years ago, Data Warehousing was largely unknown. Today, many companies are receiving considerable business value from their warehousing efforts. First American Corporation (FAC), a regional bank located in the Southeast, lost $60 million in 1990 and was operating under letters of agreement with regulators. A new senior management team developed a customer intimacy strategy with a data warehouse at the heart of the strategy. Using warehouse data, FAC was able to determine the profitability of all of their clients and products; develop programs to attract, maintain, and enhance their customer base; create profitable new product and service offerings; and redesign their distribution

channels to increase profitability and better meet customers' needs. Data warehousing helped FAC to become a profitable, innovative leader in the financial services industry. Since the introduction of the data warehouse concept in the late 1980ies data warehouse systems are now an established component of information systems landscape in most companies.

2.DATA WAREHOUSING

2.1 Definition of data warehousing

According to W.H.Inmon, a leading architect in the construction of data warehouse systems, A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process . So, data warehouse can be said to be a semantically consistent data store that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. So, its architecture is said to be constructed by integrating data from multiple heterogeneous sources to support and /or adhoc queries, analytical reporting and decision-making. Data warehouses provide on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data mining. The functional and performance requirements of OLAP are quite different from those of the on-line transaction processing applications traditionally supported by the operational databases. Data can now be stored in many different types of databases. One type of database architecture that has recently emerged is data warehouse, which is a repository of multiple heterogeneous data sources, organized under a unified schema at a single site in order to facilitate management decision-making . Data warehouse technology includes data cleaning, data integrating, and online analytical processing (OLAP) that is, analysis techniques with functionalities such as summarization, consolidation and aggregation, as well as the ability to view information from different angles. A data warehouse is defined as a "subject-oriented, integrated, time variant, non-volatile collection of data that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. In data warehouses historical, summarized and consolidated data is more important than detailed, individual records. Since data warehouses contain consolidated data, perhaps from several operational databases, over potentially long periods of time, they tend to be much larger than operational databases. Most queries on data warehouses are ad hoc and are complex queries that can access millions of records and perform a lot of scans, joins, and aggregates. Due to the complexity query throughput and response times are more important than transaction throughput. Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker to make better and faster decisions. Data warehousing technologies have been successfully deployed in many industries: manufacturing (for order shipment and customer support), retail (for user profiling and inventory management), financial services (for claims analysis, risk analysis, credit card analysis, and fraud detection), transportation (for fleet management), telecommunications (for call analysis and fraud detection), utilities (for power usage analysis), and healthcare (for outcomes analysis). This paper presents a roadmap of data warehousing technologies, focusing on the special requirements that data warehouses place on database management systems (DBMSs).

2.2 DATA WAREHOUSING FUNDAMENTALS

A data warehouse is a specially prepared repository of data designed to support decision
making. The data comes from operational systems and external sources. To create the data warehouse, data are extracted from source systems, cleaned (e.g., to detect and correct errors), transformed (e.g., put into subject groups or summarized), and loaded into a data store (i.e., placed into a data warehouse).

The data in a data warehouse have the following characteristics:

Subject oriented — The data are logically organized around major subjects of the organization, e.g., around customers, sales, or items produced.

Integrated — All of the data about the subject are combined and can be analyzed together.

Time variant — Historical data are maintained in detail form.

Nonvolatile — The data are read only, not updated or changed by users.

A data warehouse draws data from operational systems, but is physically separate and serves a different purpose. Operational systems have their own databases and are used for transaction processing; a data warehouse has its own

database and is used to support decision making. Once the warehouse is created, users (e.g., analysts, managers) access the data in the warehouse using tools that generate SQL (i.e., structured query language) queries or through applications such as a decision support system or an executive information system.

"Data warehousing" is a broader term than "data warehouse" and is used to describe the creation, maintenance, use, and continuous refreshing of the data in the warehouse.

2.3 Architecture and End-to-End Process

Figure 1 shows a typical data warehousing architecture.

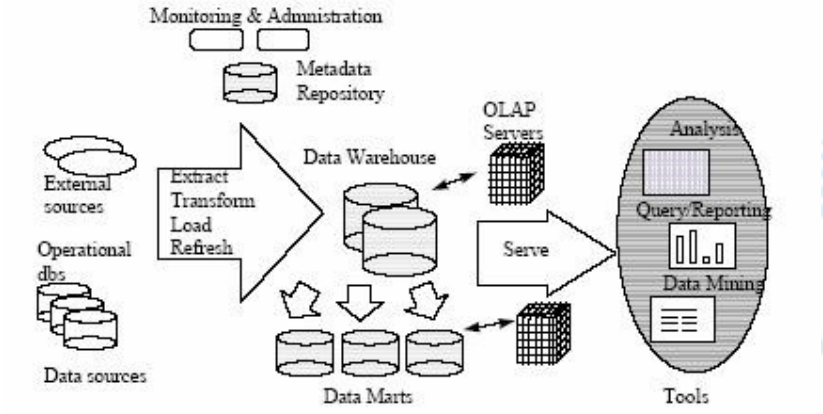


Figure 1: Data Warehousing Architecture

It includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage. In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools.

Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system. Designing and rolling out a data warehouse is a complex process, consisting of the following activities:

Define the architecture, do capacity planning, and select the storage servers, database and OLAP servers, and tools.

Integrate the servers, storage, and client tools.

Design the warehouse schema and views.

Define the physical warehouse organization, data placement, partitioning, and access methods.

Connect the sources using gateways, ODBC drivers, or other wrappers.

Design and implement scripts for data extraction, cleaning, transformation, load, and refresh.

Populate the repository with the schema and view definitions, scripts, and other metadata.

Design and implement end-user applications. Roll out the warehouse and applications.

3. OLTP AND OLAP:

The job of earlier on-line operational systems was to perform transaction and query processing. So, they are also termed as on-line transaction processing systems (OLTP). Data warehouse systems serve users or knowledge workers in the role of data analysis and decision-making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of the different users. These systems are called on-line analytical processing (OLAP) systems.

3.1 Major distinguishing features between OLTP and OLAP

Users and system orientation: OLTP is customer-oriented and is used for transaction and query processing by clerks, clients and information technology professionals.

An OLAP system is market-oriented and is used for data analysis by knowledge workers, including managers, executives and analysts.

Data contents: OLTP system manages current data in too detailed format. While an OLAP system manages large amounts of historical data, provides facilities for summarization and aggregation. Moreover, information is stored and managed at different levels of granularity, it makes the data easier to use in informed decision-making.

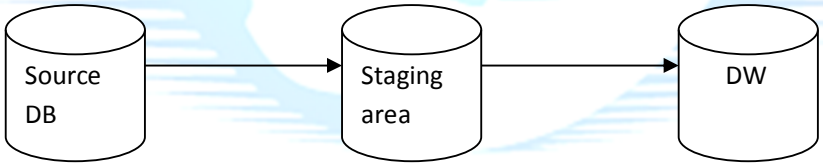
Database design: An OLTP system generally adopts an entity –relationship data model and an application-oriented database design. An OLAP system adopts either a star or snowflake model and a subject oriented database design.

View: OLTP system focuses mainly on the current data without referring to historical data or data in different organizations. In contrast, OLAP system spans multiple versions of a database schema, due to the evolutionary process of an organization. Because of their huge volume, OLAP data are shared on multiple storage media.

Access patterns: Access patterns of an OLTP system consist mainly of short, atomic transactions. Such a system requires concurrency, control and recovery mechanisms. But, accesses to OLAP systems are mostly read-only operations, although many could be complex queries.

4. DATA FLOW

The steps for building a data warehouse or repository are well understood. The data flows from one or more source databases into an intermediate staging area, and finally into the data warehouse or repository (see Figure 2).



At each stage there are data quality tools available to massage and transform the data, thus enhancing the usability of the data once it resides in the data warehouse.

5. STEPS FOR DESIGNING DATA WAREHOUSE

Designing a data warehouse is a complex process, which consists of following activities: • Define the architecture, do capacity planning and select the storage servers, database and OLAP servers, and tools.

- Integrate the servers, storage and client tools.
- Design the warehouse schema and views.
- Define the physical warehouse organization, data placement, and partitioning and access methods.
- Connect the sources using gateways, ODBC drivers or other wrappers.
- Design and implement scripts for data extraction, cleaning, transformation, load and refresh.
- Populate the repository with the schema and view definitions, scripts, and other metadata.
- Design and implement end-user applications.
- Roll out the warehouse and applications.

6. WHY OLAP IN DATA WAREHOUSE

Simply told, a data warehouse stores tactical information that answers "who?" and "what?" questions about past events. While OLAP systems have the ability to answer "who?" and "what?" questions, it is their ability to answer "what if?" and "why?" that sets them apart from Data warehouses.

• OLAP enables decision making about future actions. In contrast to Data warehouse, this is usually based on relational technology. OLAP uses a multidimensional view of aggregate data to provide quick access to strategic information for further analysis.

• OLAP and data warehouses are complementary. A data warehouse manages and stores data. OLAP transforms data warehouse "data" into "strategic information". It ranges from basic navigation and browsing (often known as 'slice and dice') to calculations, to more serious analysis such as time series and complex modeling.

CONCLUSION

Data warehouse can be said to be a semantically consistent data store that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. So, its architecture is said to be constructed by integrating data from multiple heterogeneous sources to support and /or adhoc queries, analytical reporting and decision-making. Data warehouses provide on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data mining. Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. OLTP is customer-oriented and is used for transaction and query processing by clerks, clients and information technology professionals. The job of earlier on-line operational systems was to perform transaction and query processing. Data warehouse systems serve users or knowledge workers in the role of data analysis and decision-making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of the different users. OLAP applications are found in the area of financial modeling (budgeting, planning), sales forecasting, customer and product profitability, exception reporting, resource allocation, variance analysis, promotion planning, and market share analysis. Moreover, OLAP enables managers to model problems that would be impossible using less flexible systems with lengthy and inconsistent response times. More control and timely access to strategic information facilitates effective decision -making. This provides leverage to library managers by providing the ability to model real life projections and a more efficient use of resources. OLAP enables the organization as a whole to respond more quickly to market demands. Market responsiveness, in turn, often yields improved revenue and profitability. And there is no need to emphasize that present libraries have to provide market-oriented services.

ACKNOWLEDGMENT

The success of our research is never limited to the individual undertaking the paper. It is the cooperative effort of the people around an individual that spell success. For all efforts, behind this successful research, we are highly intended to all those personalities without whom this research would ever be completed. We find no words to express our gratitude towards those who were constantly involved with us throughout our work.
I am thankful to my guide Mr Manoj (Associate Professor, SGT Institute of Engineering & Technology, MDU Rohtak)

REFERENCES

[1] Devlin, B. & Murphy, P. (1988) An Architecture for a Business and Information System, IBM Systems Journal, 27 (1), 60-80.
[2] Inmon, W.H. (1996) Building the Data Warehouse, Second Edition, New York: John Wiley & Sons.
[3] Kimball, R. (1996) The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, New York: John Wiley & Sons.
[4] Gardner, S.R. (1998) Building the Data Warehouse, Communications of the ACM , 41 (9), 52-60.
[5] Simon, A. (1998) 90 Days to the Data Mart, New York: John Wiley & Sons.

[6] Frolick, M.N. & Lindsey, K. (2003) Critical Factors for Data Warehouse Failure, Journal of Data Warehousing, 8 (1).

[7] Hwang, H.-G., Kua, C.-Y., Yenb, D. C., & Chenga, C.-C. (2002) Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in Taiwan, Decision Support Systems, In Press, Corrected Proof.