www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 4 April, 2013 Page No. 1291-1297

Improving Decision Support Systems with Data mining, Data warehousing, OLAP and OLTP Technologies

Ms. Ramteke M. A*, Prof. Dhande S. S
* Dept. of Computer Science & Engineering
Sipna College of Engineering & Technology
Amravati, India.
mona.ramteke@gmail.com

Associate Professor
Dept. of Computer Science & Engineering
Sipna College of Engineering
& Technology Amravati, India.
sheetaldhandedandge@gmail.com

Abstract- *Rapidly changing markets and global competition require up-to-date, high-quality and complete information for decision support. Data mining, data warehousing, on-line analytical processing (OLAP), on-line transaction processing (OLTP) are important elements to support decision making process which has increasingly become a focus of the database industry. This paper provides a review of data mining, data warehousing, OLAP and OLTP technologies, with an emphasis on the components of Data warehouse architecture. A "data warehouse" is an organization-wide snapshot of data, typically used for decision-making. The use of data mining to facilitate decision support can lead to an improved performance of decision making. OLAP technology is used to perform complex analysis of the data in a data warehouse. OLTP technology is used to perform updates on operational or transactional systems. Online Transaction and Processing helps and manages applications based on transactions involving high volume of data. Decision Support System is a computer based information system designed to facilitate the decision making process of semi structured tasks. DSS systems and warehouses are typically separate from the on-line transaction processing (OLTP) system. Unlike a traditional relational database model or On-Line Transactional Processing (OLTP) system, an OLAP system is optimized to provide data to end users in a meaningful format through a Decision Support System (DSS) Application.*

Keywords- Data mining, Data warehouse, OLAP, OLTP, Decision Support System

1. Introduction

The data warehouses are supposed to provide storage, functionality and responsiveness to queries beyond the capabilities of today's transaction-oriented databases. Also data warehouses are set to improve the data access performance of databases. Traditional databases balance the requirement of data access with the need to ensure integrity of data. In present day organizations, users of data are often completely removed from the data sources. Many people only need read-access to data, but still need a very rapid access to a larger volume of data than can conveniently by downloaded to the desktop. Often such data comes from multiple databases. Because many of the analyses performed are recurrent and predictable, software vendors and systems support staff have begun to design systems to support these functions. Currently there comes a necessity for providing decision makers from middle management upward with information at the correct level of detail to support decision-making. Data warehousing, online analytical processing

(OLAP) and data mining provide this functionality. A decision support system (DSS) is a computer-based information system that supports business or organizational decision-making activities. DSSs serve the management, operations, and planning levels of an organization and help to make decisions, which may be rapidly changing and not easily specified in advance.

2. Concept of Data Warehousing

2.1 Definition

A data warehouse is however usually the "driver" and dominant component for a Data-driven DSS. A data warehouse is an organized collection of large amounts of structured data. It is a database designed and intended to support decision making in organizations. It is batch updated and structured for rapid online queries and managerial summaries of its contents. According to Bill Inmon (1993), who is often called the "father" of data warehousing, "a data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data". Ralph Kimball (1996), another data warehousing expert, says "A data warehouse is a copy of transaction data specifically structured for query and analysis". A data warehouse is often the component that stores data for a Data-driven DSS.

2.2 Need of Separate Data Warehouse

- Missing data: Decision support requires historical data, which operational databases do not typically maintain
- Data consolidation: Decision support requires consolidation (aggregation, summarization) of data from many heterogeneous sources: operational databases, external sources.
- Data quality: Different sources typically use inconsistent data representations, codes, and formats which have to be reconciled.

2.3 Data Warehouse Architecture

A data warehouse is defined as a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management's decisions. More generally, data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker, such as executive, manager, and analyst, to arrive at better and faster decisions. Data warehouses provide access to data for complex analysis, knowledge discovery, and decision-making. They support high performance demands on an organization's data and information. It provides an enormous amount of historical and static data.

Data Warehouses (DW) integrate data from multiple heterogeneous information sources and transform them into a multidimensional representation for decision support applications.

The architecture consists of various interconnected elements:

- *Operational and external database layer* – the source data for the DW
- *Information access layer* – the tools the end user access to extract and analyze the data
- *Data access layer* – the interface between the operational and information access layers
- *Metadata layer* – the data directory or repository of metadata information

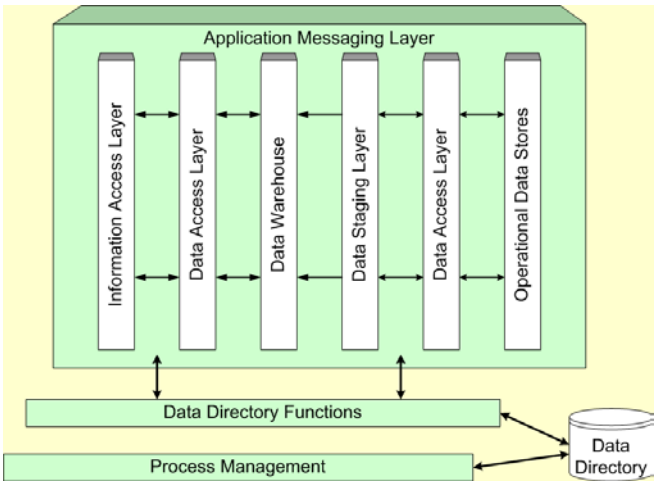

Figure1: Components of the Data Warehouse Architecture

Additional layers are:

- *Process management layer* – the scheduler or job controller
- *Application messaging layer* – the "middleware" that transports information around the firm
- *Physical data warehouse layer* – where the actual data used in the DSS are located
- *Data staging layer* – all of the processes necessary to select, edit, summarize and load warehouse data from the operational and external data bases. [11]

3. Concept of Data Mining

The ultimate goal of data mining is to assist the decision making. Decision-makers can analyze the results of data mining and adjust the decision-making strategies combining with the actual situation. Data Mining is the extraction or "Mining" of knowledge from a large amount of data or data warehouse. To do this extraction data mining combines artificial intelligence, statistical analysis and database management systems to attempt to pull knowledge form stored data. Data mining is the process of applying intelligent methods to extract data patterns. This is done using the front-end tools. Data mining is a methodology designed to perform knowledge-discovery expeditions over the database data with minimal end-user intervention. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. For example, if certain books are rarely used by members of a particular library, while the same books are frequently used at other libraries then it is appropriate to transfer these books to respective libraries to ensure its effective use. Such knowledge could only be discovered through sharing experiences of librarians or by capturing the knowledge through database and integrating them as done when building data warehouses. Decision support tools assist users in discovering knowledge. Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for development of their businesses.

4. Concept of OLAP and OLTP

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. OLAP stands for Online Analytic Processing and used in decision support systems usually runs on data warehouse. In contrast to OLTP, OLAP queries are complex, touch large amounts of data, try to discover patterns or trends in the data.

4.1 Data warehousing and OLTP

When transactional data is no longer of value to the operational environment, it is removed from the database. If a business is without a decision support facility, the data is archived and eventually destroyed. However, if there is a decision support environment, the data is transported to some type of interactive medium commonly referred to as a data warehouse.

A more precise definition is given by W. H. Inmon (1996). Specifically," A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision making process."

As the definition implies, significant variations in structure and intent exist between an OLTP database and a data warehouse. A few major differences are:

- The data in warehouse is subject oriented and based on one or more central themes, OLTP databases are process oriented and organized so as to maximize data update and retrieval.
- An OLTP system supports data processing, collection, and management. A data warehouse stores data to be reported on, analyzed, and tested.
- OLTP deals with data necessary for the efficient day-to-day operation of a business or organization. Data records in a transactional database are subject to multiple access and constant update. In contrast, the data in a warehouse exists in part because the data is no longer use to the OLTP environment. The majority of data in a warehouse is historical and not subject to change (read-only).

4.2 Data warehousing and OLAP

Data warehousing developed, despite the presence of operational databases due to following reasons:

• An operational database is designed and tuned from known tasks and workloads, such as indexing using primary keys, searching for particular records and optimizing 'canned queries'. As data warehouse queries are often

complex, they involve the computation of large groups of data at summarized levels and may require the use of special data organization, access and implementation methods based on multidimensional views. Processing OLAP queries in operational databases would substantially degrade the performance of operational tasks.

• An operational database supports the concurrent processing of multiple transactions. Concurrency control and recovery mechanisms, such as locking and logging are required to ensure the consistency and robustness of transactions. While and OLAP query often needs read-only access of data records for summarization and aggregation. Concurrency control and recovery mechanisms, if applied for such OLAP operations, may jeopardize the execution of concurrent transactions.[10]

• While operational data stores address the integration issue of disparate transactional databases, they are not suitable for analytic processing to support strategic decision-making. According to Singh (1998), tactical decisions are based largely on data in the operational data store, whereas, long-term strategic decisions require the historical trend analysis that is only available from the data warehouse. It further stated that an ODS does not need to have the historical data that a data warehouse must store for use by strategic decision makers.

5. Definition and characteristics of decision support system

In 1998 Turban defines a decision support system as "an interactive, flexible and adaptable system, exclusively designed to offer support in solving unstructured or semi-structured managerial problems, aiming to improve the decisional process. The system uses data (internal and external) and models, providing a simple and easy-to-use interface, thus, allowing the decision maker control over the decision process. The DSS offers support in all decision process's stages". [6]

Some of the most important characteristics of the DSSs are: uses data and models; enhances the learning process; grows the efficiency of the decision making process; offers support in the decision making process and allows the decision maker control over the entire process; offers support in all stages of the decision making process; offers support for decision makers in solving structured or unstructured problems; offers support for a user or for a group of users etc.

5.1 Components of Decision Support System

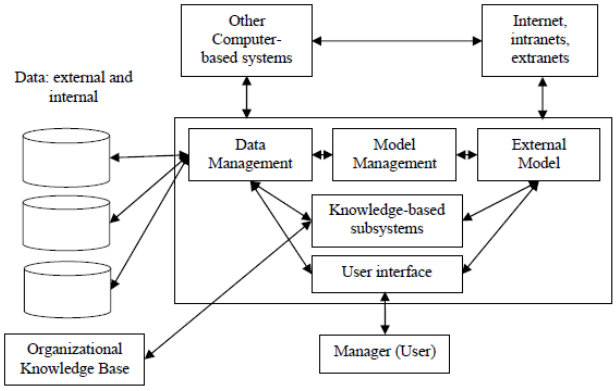

Figure 2: A schematic view of DSS

A DSS application can be composed of following subsystems:

1. Data Management subsystem: The database management subsystem includes a database, which contains relevant data for the situation and is managed by software called the database management system (DBMS). The database management subsystem can be interconnected with the corporate data warehouse, a repository for corporate relevant decision-making data.
2. Model Management subsystem: The model base gives decision makers access to a variety of models and assist them in decision making. The model base can include the model base management software (MBMS) that coordinates the use of models in a DSS. This component can be connected to external storage of data.
3. Knowledge-based Management subsystem: This subsystem can support any of the other subsystem or act as an independent component. It provides intelligence to augment the decision maker's own. It can be interconnected with the organization's knowledge repository, which is called the organizational knowledge base.
4. User Interface subsystem: The user interface, also called the dialog management facility, it allows users to interact with the DSS to obtain information. The user interface requires two capabilities; the action language that tells the DSS what is required and passes the data to the DSS and the presentation language that transfers and presents the user results. The DSS generator acts as a buffer between the user and the other DSS

components, interacting with the database, the model base and the user interface.

5.2 Decision Making and Problem Solving Process

A Problem occurs when a system does not meet its established goals or does not work as planned. Problem solving may also deal with identifying new opportunities. Problem solving is the most critical activity a business organization undertakes. Problem solving begins with decision making.

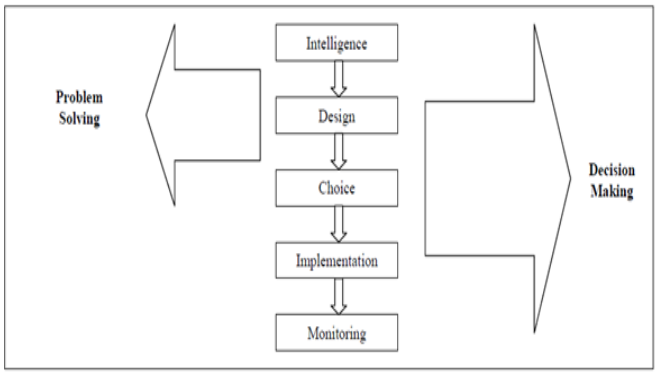

Figure 3: Decision making and problem solving process

6. Decision Support System with data warehouse

A data warehouse is defined as a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management's decisions. More generally, data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker, such as executive, manager, and analyst, to arrive at better and faster decisions. Data warehouses provide access to data for complex analysis, knowledge discovery, and decision-making. They support high performance demands on an organization's data and information. It provides an enormous amount of historical and static data.

The data in a data warehouse have the following characteristics:

- *Subject-orientation*: Whereas classical database design emphasizes the function of the transactions, a subject-oriented representation of data emphasizes the structuring of data according to the core-business areas and the strategic goals of a company, e.g., products, customer, turnover, ...
- *Integration*: The change from function- to subject-orientation leads to a modification of database organizations across several database systems. The result is an integrated store of data which includes all relevant aspects.
- *Time-variance:* The term 'time-variance' has two different but related meanings: From a more static point of view, the data contains time aspects, e.g., in form of time-stamps, time related attributes etc. Dynamically speaking time-variance relates to the strategy of ensuring ongoing refreshment of data.
- *Non-volatility:* Once a data record is brought into the data warehouse it remains unchanged. Refreshing data does always translate to adding new data,i.e. all data accesses are read-only transactions.

A data warehouse can also be viewed as a process for gathering, storing, managing, and analyzing data. Following figure displays the key components of the warehousing process.

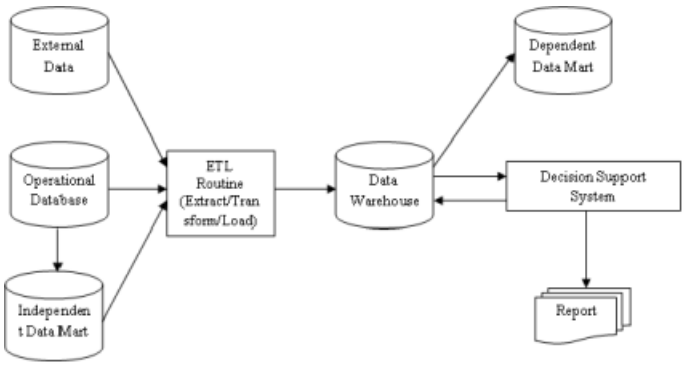

Figure 4: A data warehouse process model

Figure4 shows the data entering the data warehouse from three main sources. External data represents items such as economic indicators, weather information, and the like that are not specific to the internal organization. An independent data mart is a data store that is similar to a warehouse but limits its focus to a single subject. An independent data mart is structured using operational data as well as external data sources. The data stored in a data mart can be loaded into the central data warehouse for use by other faces of the organization. Regardless of the external data source, the process of moving data into the warehouse will likely involve some low level procedural programming.

Figure4 illustrates that prior to entering the warehouse, data is processed by an ETL(extract,

transform, and load) routine. The primary responsibilities of the ETL process include: extracting data from one or more of the input sources shown in figure1, cleaning and transforming the extracted data as necessary, and loading the data into warehouse. Data transformations are often used to resolve granularity issues, correct data inconsistencies between multiple operational databases, and to timestamp individual records. Once transformed and cleaned, the data enter the warehouse where it is stored in a relational or multi relational format.

As a rule, once data enters the warehouse it is not subject to change. An obvious exception to this rule is when errors are detected in the data. The warehouse also stores another type of data known as metadata. Metadata is technically defined as data about data. The purpose of metadata is to allow for a better understanding of the nature of the data contained in the warehouse. Two general types of metadata have been defined: structural and operational. Structural metadata emphasizes data descriptions, data types, representation rules, and relationships between data items. Operational metadata is primarily responsible for describing the quality and usage of data.

The primary function of the data warehouse is to house data for decision support. Figure1 shows that data is copied from the data warehouse for analysis by the decision support system. We also see data warehouse from the decision environment. Any data entering the data warehouse from the decision support system will be from metadata created from information gained through one or more decision support processes.

Three categories of decision support can be defined. Specifically,

1. Reporting data: Reporting is considered the lowest level of decision support. However, a reporting facility capable of generating informative reports for a variety of clientele is of utmost importance for the successful operation of any business.
2. Analyzing data: Data analysis is usually accomplished with some form of multidimensional data analysis tool.
3. Knowledge Discovery: Knowledge discovery typically takes place through data mining. However, manual data mining techniques involving repeated query and data analysis can sometimes uncover interesting patterns in data.

Besides storing data for decision support, the warehouse is a data store for creating smaller departmental warehouses known as dependent data marts. A dependent data mart is typically about a single subject and designed for a specific purpose. The data mart is likely to contain summary information showing a higher level of granularity than that of the data warehouse.

7. CONCLUSION

A data warehouse is a historical database that summarizes, integrates, and organizes data from one or more operational databases in a format that is more efficient for analytical queries. Building a data warehouse involves identifying operational sources, extracting and cleansing data, and loading the data into the warehouse. Data warehouse solutions are essential for all service- and consumer-oriented companies, e.g. from financial service (banking, insurance), retail, and consumer-industry. The benefit stems from the fact that information can be uniformly accessed from a single, consistent, integrated and up-to-date source of high-valued information. A data warehouse thus lays the foundation for a decision support environment. In addition, OLAP-facilities are an emerging technique for exploiting data on-line by offering comfortable means for manipulating data and computing new information tailored to individual needs. Therefore, OLAP is a promising way for extracting information in order to be able to make well-informed and fast decisions which are important to competitiveness and growth for small and large companies.

REFERENCES

[1] Data Mining A Tutorial-Based Primer Richard J. Roiger, Michael W. Geatz, (First Impression, 2007). Pearson Education.

[2] Irtishad Ahmad, Salman Azhar, Pranas Lukauskis "Development of a decision support system using data warehousing to assist builders/developers in site selection" *Elsevier* Automation in Construction 13 (2004) 525–

542.
[3] David Arnott , Graham Pervan "Eight key issues for the decision support systems discipline" *Elsevier* Decision Support Systems 44 (2008) 657–672.
[4] Efraim Turban, Jay E. Aronson (Fourth Indian Reprint, 2004). *Decision Support Systems and Intelligent Systems*, 6th edition. Pearson Education, New Delhi.
[5] K.P.Tripathi "Decision support system is a Tool for making better decisions in the organization" Indian Journal of Computer Science and Engineering (IJCSE) Vol. 2 No. 1
[6] Turban, E. - Decision Support Systems and Intelligent Systems, 5th ed.,Englewood Cliffs, New Jersey, Prentice Hall, 1998
[7] Chaudhuri, S & Dayal, U 1997, An Overview of Data Warehousing and OLAP Technology, SIGMOD Record 26:1, Mar, pp 65-74.
[8] http://www.olapcouncil.org
[9] 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 .
[10] G. Satyanaryana Reddy, Rallabandi Srinivasu, M. Poorna Chander Rao, Srikanth Reddy Rikkula " Data Warehousing, Data Mining, OLAP and OLTP are essential elements to support decision-making process in industries" (IJCSE) Vol. 02, No. 09, 2010, 2865- 2873.
[11] http://www.csun.edu/~hcmgt004/ DWPresSp05.ppt
[12] Kachur, R. (2000) Data Warehouse Management Handbook,Paramus: Prentice Hall.
[13] Adelman, S. & Moss, L. (2000) Data Warehouse Project Management, Boston: Addison-Wesley.