ISSN:0975-9646

OLAP, OLTP for DECISION-MAKING PROCESS using Data Mining and Warehousing

Chinmay R. Deshmukh, Prof. R.R.SHELKE

Dept.Of Comp.Sci.& Engg.,Amravati University
H.V.P.M. COET , Amravati.

Abstract — **One may claim that the exponential growth in the amount of data provides great opportunities for data mining. In many real world applications, the number of sources over which this information is fragmented grows at an even faster rate, resulting in barriers to widespread application of data mining. A data warehouse is designed especially for decision support queries.**
Data warehousing is the process of extracting and transforming operational data into informational data and loading it into a central data store or warehouse. The idea behind data mining, then is the " non trivial process of identifying valid, novel , potentially useful, and ultimately understandable patterns in India". Data mining is concerned with the analysis of data and the use of software technique for finding patterns and regularities in sets of data.
OLTP is the customer-oriented service and is used for transaction and query processing by clerks, clients and IT professionals. An OLAP system is market oriented and is used for data analysis by knowledge workers, including the managers, executives and analysts.
Data warehousing , OLAP have emerged as leading technology that facilitate data storage , organization and then, significant retrieval.
Data warehousing provides the means to change raw data into information for making effective business decision – the emphasis on information , not data. The data warehouse is the hub for decision support data.
The data warehouse supports on-line analytical processing (OLAP), the functional and the performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases. Data warehouses provide on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities , which facilitates effective data mining.
This paper also explains partition algorithm to discover all requirements sets from the data warehousing using the data mining. Also explained relation between operational data , data warehouse and data marts.

Keywords — **Data Warehousing, OLAP, OLTP, Data Mining , 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. (Inmon, W.H., 1992) 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 (executive, manager, and analyst) to make better and faster decisions.

2. DATA WAREHOUSING

2.1 Definition of data warehousing

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 . 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.

2.2 DATA WAREHOUSING FUNDAMENTALS

A data warehouse (or smaller-scale data mart) is a specially prepared repository of data designed to support decision making. The data in a data warehouse have the following characteristics :

1. *Subject oriented* — The data are logically organized around major subjects of the organization, e.g., around customers, sales, or items produced.
2. *Integrated* — All of the data about the subject are combined and can be analyzed together.
3. *Time variant* — Historical data are maintained in detail form.
4. *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 queries or through applications such as a decision support system or an executive information system.

2.3 Architecture and End-to-End Process

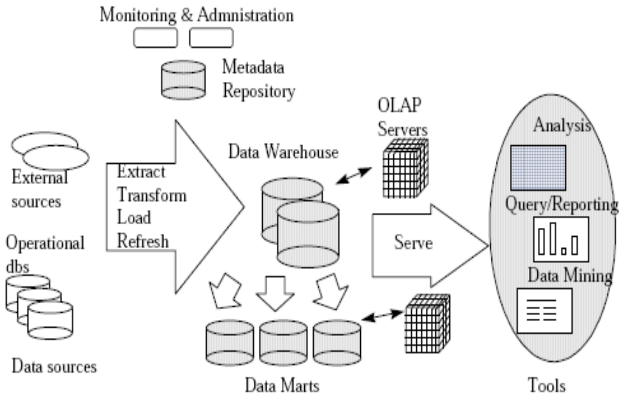

Figure 1: Data Warehousing Architecture

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.

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 Need of 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.
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.

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.

Figure 2- Data Flow

5. DATA MINING

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. The spreadsheet is still the most compiling front-end application for Online Analytical Processing (OLAP). The challenges in supporting a query environment for OLAP can be crudely summarized as that of supporting spreadsheet operation effectively over large multi-gigabytes databases. In a database application the queries issued are well defined to the level of what we want and the output is precise and is a subset of operational data. In data mining there is no standard query language and the queries are poorly defined. Thus the output is not precise (fuzzy) and do not represent a subset of the database.

6. DECISION MAKING USING A DATA WAREHOUSE

A Decision Support System (DSS) is any tool used to improve the process of decision making in complex systems. A DSS can range from a system that answer simple queries and allows a subsequent decision to be made, to a system that employ artificial intelligence and provides detailed querying across a spectrum of related datasets. Amongst the most important application areas of DSS are those complicated systems that directly "answer" questions, in particular high level "what-if" scenario modeling. Over the last decade there was a transition to decision support using data warehouses (Inmon 2002). The data warehouse environment is more controlled and therefore more reliable for decision support than the previous methods.

7. DATA WAREHOUSING INTEGRITY

Data Warehouses (DW) integrate data from multiple heterogeneous information sources and transform them into a multidimensional representation for decision support applications. Apart from a complex architecture, involving data sources, the data staging area, operational data stores, the global data warehouse, the client data marts, etc., a data warehouse is also characterized by a complex lifecycle. In a permanent design phase, the designer has to produce and maintain a conceptual model and a usually voluminous logical schema, accompanied by a detailed physical design for efficiency reasons. The designer must also deal with data warehouse administrative processes, which are complex in structure, large in number and hard to code; deadlines must be met for the population of the data warehouse and contingency actions taken in the case of errors. Finally, the evolution phase involves a combination of design and administration tasks: as time passes, the business rules of an organization change, new data are requested by the end users, new sources of information become available, and the data warehouse architecture must evolve to efficiently support the decision-making process within the organization that owns the data warehouse. All the data warehouse components, processes and data should be tracked and administered via a metadata repository.

8. 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 online 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, 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.

ACKNOWLEDGEMENT

The author thanks Prof.R.R.SHELKE of HVPM COET, Amravati for her help in synthesizing the discussions that occurred on this topic.

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] 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.

[4] Finnegan, P. & Sammon, D. (1999) Foundations of an Organisational Prerequisites Model for Data Warehousing, Proceedings of the 7th European Conference on Information Systems (ECIS 1999), Copenhagen, June.

[5] O'Donnell, P., Arnott, D., & Gibson, M. (2002) Data warehousing development methodologies: A comparative analysis, Working Paper, Melbourne, Australia: Decision Support Systems Laboratory, Monash University.