A Survey of Data Warehouse and OLAP Technology

P.Kavitha
Research Scholar, Department of Computer Science,Bharathiar University, Coimbatore, India.
Assistant Professor, Sree Arumugham Teacher Training,College – B.Ed, Tholudur, Tamil Nadu, India.

Dr.G.N.K.Suresh babu
Associate Professor, GKM College of Engineering and Technology,
Chennai, Tamil Nadu, India.

Abstract:- The structure of data warehouses involves Cleaning of data, Integration of data, and Transformation of data and data can be viewed as an important preprocessing step for KDD Process. Moreover, data warehouses provide online analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data generalization and data mining. Many other data mining functionalities, such as association, classification, and prediction and clustering can be integrated with OLAP operations to enhance interactive mining of knowledge at multiple stages of concept. Hence, the data warehouse has become an increasingly important platform for data analysis and on-line analytical processing and will provide an effective platform for data mining. Therefore, data warehousing and OLAP form an essential step in knowledge discovery process.

Keywords:- Data Warehouse, Data mining, OLAP, OLTP, Data Warehouse Server, Client, query processing, Normalization, Schema, etc.

I. INTRODUCTION

Defined in many different ways, but not thoroughly. A decision support database that is maintained separately from the organization's prepared database systems. Support information processing by providing a solid platform of combined, chronological data for study. A data warehouse is a subject related, integrated, time variant and nonvolatile collection of data in support of management's decision making process -William.H.Inmon. Organized Data is used in day-to-day needs of company. Informational Data is Supports other functions such as Planning and forecasting. Data mining tools often contact data warehouses rather than ready data. Data warehousing: The process of constructing and using data warehouses. Data warehousing is a collection of technique, and tools used to support knowledge employees-managing director, directors, managers, and analysts—to conduct data analyses that help with performing decision-making processes and improving information resources.

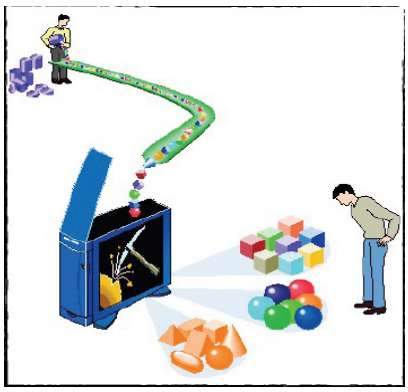
OLAP: On-Line Analytic Processing Starts with "summarizing" the data before it is possible to execute the queries (to receive a report) , this is building "the cube" , this can take a long time , both more efficient response for analysis queries, Data (summarization) is presented as cubes and sub cubes.

II. DATA WAREHOUSE

Data warehouse technology includes: data cleaning data integration on-line analytical processing (OLAP) analysis techniques with functionalities such as summarization, consolidation, and aggregation ability to view information from different angles. We are data rich, but information poor.

The large quantity of data, joined with the need for powerful data analysis tools, has been described as a data rich but information poor situation Data collected in large data repositories become "data tombs" data archives that are seldom visited Important decisions are often made based not on the information-rich data stored in data repositories, but rather on a decision maker's intuition The decision maker does not have the tools to extract the valuable knowledge embedded in the vast amounts of data Often systems rely on users or domain experts to manually input knowledge into knowledge bases. In this procedure is prone to biases and errors, and is extremely time-consuming and costly simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data, usually automatically gathered.

III. DATA MINING



Data mining is also known as knowledge Discovery in databases, it has been defined as the non trivial extraction of implicit formerly unknown, and potentially useful information from data. Data mining is a process of exploration an analysis, by semi automatic means of large quantities of data in order to discover meaningful models and rules.

Key Features of Data Warehouse Subject-Oriented

- Structured around major subjects, such as purchaser, product, sales.
- Focus on the modeling and analysis of data for decision makers, not on daily operations or transaction processing.
- Provide a simple and concise view about particular subject issues by excluding data that are not useful in the decision support process.

Integrated

- ❖ Constructed by integrating multiple, heterogeneous data sources i.e. Relational database, flat files, and online transaction records.
- ❖ Data cleaning and data integration system are applied. Ensure consistency in identification conventions, encoding structures, attribute measures, among different data sources. For example, Hotel price: currency, tax, brunch covered, etc. When data is moved to the warehouse, it is altered.

Time Variant

- ❖ The time vision for the data warehouse is significantly longer than that of prepared systems.
  - Prepared database: current value data.
  - Data warehouse data: provide information from a chronological perspective (e.g., past 5-10 years)
- ❖ each key structure in the data warehouse
  - Contains an element of time, explicitly or implicitly
  - But the key of prepared data may or may not contain “time element”.

Non-Volatile

- ❖ In fact separate store of data transformed from the prepared environment.
- ❖ Prepared update of data does not occur in the data warehouse environment.
  - Does not require transaction processing, improvement, and concurrency control machine
  - Involve only two operations in data accessing: Initial loading of data and access of data.

Prepared database systems and Data Warehouse

OLAP

- ❖ OLAP: reporting data, visualizing data, interaction with views of the data.
- ❖ OLAP provides summary data and generates rich computation. For example, OLAP answers questions like "How do sales of mutual funds in North America for this quarter compare with sales a year ago? What can we calculate for sales next quarter? What is the trend as exact by percent change?"

OLAP and data mining can balance each other. For example, OLAP might pinpoint problems with sales of mutual funds in a certain region OLAP can be used to track the net income.

OLTP

OLTP applications are used by bank, airlines, insurance company, and other businesses that give online users direct access to information. The OLTP is an application procedure unit of work, called transactions. A single transaction might request a bank balance; another might update that balance to reflect a deposit.

An application can be simulated to run in parallel on a tightly coupled multiprocessor so that it cans concurrently procedure of multiple transactions. The quantity of parallelism depends on the system programmer, the transaction manager, and the hardware that the OLTP application runs on. An automatic teller machine (ATM) for a bank is an example of a commercial transaction processing application.

Online transaction processing increasingly requires support for transactions that span a network and may include more than one group. For this reason, new online transaction processing software uses client or server processing and brokering software that allows transactions to run on different computer platforms in a network.

Data Warehouse vs. Prepared DBMS

OLTP (on-line transaction processing)

- Main task of traditional relational DBMS.
- Everyday operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.

OLAP (on-line analytical processing)

- Major task of data warehouse system.
- Data analysis and decision making.

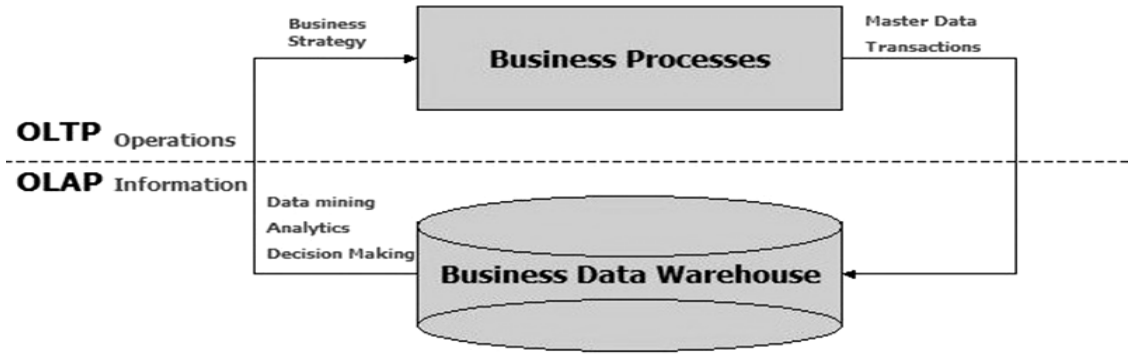
Distinct features (OLTP vs. OLAP):

- User and system orientation: consumer vs. market.
- Data contents: present, detailed vs. chronological, consolidated.
- Database design: ER and application vs. star and subject.
- View: present, local vs. evolution included.
- Access model: update vs. read-only but complex queries.

Features between OLTP and OLAP

In frequent we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

OLTP (On-line Transaction Processing)



OLTP is characterized by a large number of short on-line transactions (INSERT, UPDATE, and DELETE). The main importance for OLTP systems is put on very fast query processing, maintain data integrity in multi-access

environments and an effectiveness measured by number of transactions per second. In OLTP database there is complete and present data, and schema used to store transactional databases is the entity model (usually 3NF).

OLAP (On-line Analytical Processing) OLAP is characterized by relatively low volume of transactions. Queries are often very difficult and involve aggregations. For OLAP systems a response time is a good organization measure. OLAP applications are widely used by Data Mining procedure. In OLAP database there is aggregated, chronological data, stored in multi-dimensional schemas (usually star schema).

Separation of Data Warehouse

High performance for both systems

- DBMS— tuned for OLTP: entrée methods, indexing, synchronize control, enhancement
- Warehouse—tuned for OLAP: difficult OLAP queries, multidimensional view, and consolidation.

Different functions and data

- Missing data: Decision support requires chronological data which prepared DBs do not typically maintain
- Data consolidation: DS requires consolidation (aggregation, summarization) of data from various sources
- Data quality: different source typically use inconsistent data representations, codes and formats which have to be reconciled.

IV. CONCLUSION

Finally the separation of operational databases from data warehouses is based on the different structures, contents, and users of the data in these two systems. Decision support requires chronological data, whereas operational databases do not typically maintain chronological data. In this paper, the data in operational databases, through abundant, usually far from complete for decision making. However, many vendors of operational relational database management systems are beginning to optimize such system to support OLAP queries. As this paper continuous the separation between OLTP and OLAP systems is expected to decrease.

REFERENCES:

[1] Jiawei Han and Micheline Kamber : Data Mining Concepts and Techniques
[2] Surajit Choudhary: Data Warehousing and OLAP technology.
[3] Umeshwar Dayal: An overview of data warehousing and technology.
[4] http://www.dwinfocenter.org
[5] http://www.carolla.com/wp-dw.htm
[6] http://system-services.com/dwintro.asp
[7] Data Warehousing- Wikipedia
[8] http://research.microsoft.com/pubs/76058/sigrecord.pdf
[9] http://www.cs.sunysb.edu/~cse634/presentations/DataWarehousing-part-1.pdf
[10] http://lambda.uta.edu/cse6331/spring03/papers/dw1.pdf
[11] http://www.123helpme.com/data-wharehouse-paper-view.asp?id=164499
[12] http://www.1keydata.com/data warehousing/dimensional.html
[13] http://docs.oracle.com/cd/B10501_01/server.920/a96520/concept.htm
[14] http://www.peterindia.net/DataWarehousingView.html
[15] Jiawei Han and Micheline Kamber : Data Mining Concepts and Techniques Morgan Kaufmann publishers, San Francisco, USA,
[16] First Indian Reprint, 2002.