

A Novel Approach of Data Warehouse OLTP and OLAP Technology for Supporting Management prospective

B.Manivannan

Research Scholar, Dept. Computer Science, Dravidian University, Kuppam, Andhra Pradesh, India

Abstract: Data warehousing and On-Line Analytical Processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. Many commercial products and services are now available, and all of the principal database management system vendors now have offerings in these areas. Decision support places some rather different requirements on database technology compared to traditional on-line transaction processing applications. This paper provides an overview of Data warehousing, Data Mining, 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 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, Data Mining, Decision Making and Decision Support.

I. INTRODUCTION

Data warehousing provides architectures and tools for business executives to systematically organize, understand, and use their data to make strategic decisions. Data warehouse systems are valuable tools in today's competitive, fast-evolving world. In the last several years, many firms have spent millions of dollars in building enterprise-wide data warehouses. Many people feel that with competition mounting in every industry, data warehousing is the latest must-have marketing weapon—a way to retain customers by learning more about their needs. “Then, what exactly is a data warehouse?” Data warehouses have been defined in many ways, making it difficult to formulate a rigorous definition. Loosely speaking, a data warehouse refers to a data repository that is maintained separately from an organization's operational databases. Data warehouse systems allow for integration of a variety of application systems. They support information processing by providing a solid platform of consolidated historic data for analysis. According to William H. Inmon, a leading architect in the construction of data warehouse systems.

1.1 Data Warehouse

A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision making process”. This short but comprehensive definition presents the major features of a data warehouse. The four keywords—subject-oriented, integrated, time-variant, and nonvolatile—distinguish data warehouses from other data repository systems, such as relational database systems, transaction processing systems, and file systems. Let's take a closer look at each of these key features. Subject-oriented: A data warehouse is organized around major subjects such as customer, supplier, product, and sales. Rather than concentrating on the day-to-day operations and transaction processing of an organization, a data warehouse focuses on the modeling and analysis of data for decision makers. Hence, data warehouses typically provide a simple and concise view of particular subject issues by excluding data that are not useful in the decision support process. Integrated: A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and online transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures, and so on. Data Warehouse: Basic Concepts 127 Time-variant: Data are stored to provide information from an historic perspective. Every key structure in the data warehouse contains, either implicitly or explicitly, a time element. Nonvolatile: A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation, a data warehouse does not require transaction processing, recovery, and concurrency control mechanisms. It usually requires only two operations in data accessing: initial loading of data and access of data.

1.2 Differences between Operational Database Systems and Data Warehouses

Because most people are familiar with commercial relational database systems, it is easy to understand what a data warehouse is by comparing these two kinds of systems. The major task of online operational database systems is to perform online transaction and query processing. These systems are called online transaction processing (OLTP)

systems. They cover most of the day-to-day operations of an organization such as purchasing, inventory, manufacturing, banking, payroll, registration, and accounting. Data warehouse systems, on the other hand, 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 different users.

These systems are known as online analytical processing (OLAP) systems. The major distinguishing features of OLTP and OLAP are summarized as follows: Users and system orientation: An OLTP system 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: An OLTP system manages current data that, typically, are too detailed to be easily used for decision making. An OLAP system manages large amounts of historic data, provides facilities for summarization and aggregation, and stores and manages information at different levels of granularity. These features make the data easier to use for informed decision making.

An OLTP system usually adopts an entity-relationship (ER) data model and an application-oriented database design. An OLAP system typically adopts either a star or a snowflake model and a subject-oriented database design. View: An OLTP system focuses mainly on the current data within an enterprise or department, without referring to historic data or data in different organizations. In contrast, an OLAP system often spans multiple versions of a database schema, due to the evolutionary process of an organization. OLAP systems also deal with information that originates from different organizations, integrating information from many data stores. Because of their huge volume, OLAP data are stored on multiple storage media. Access patterns: The access patterns of an OLTP system consist mainly of short, atomic transactions. Such a system requires concurrency control and recovery mechanisms. However, accesses to OLAP systems are mostly read-only operations although many could be complex queries. Other features that distinguish between OLTP and OLAP systems include database size, frequency of operations, and performance metrics.

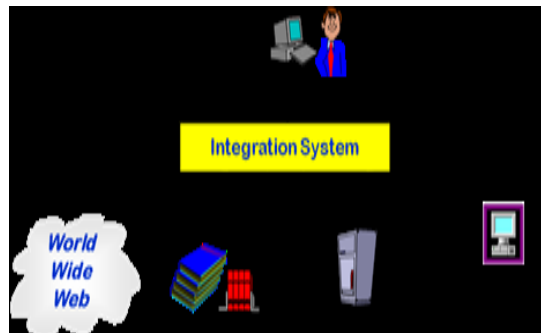


Figure 1: The above diagram shows in the integration system we are use different tools but also database management systems need the basic requirements depends of Data warehouse, meta data, OLTP and OLAP.

1.3. OLTP vs. OLAP

We can divide IT systems into transactional (OLTP) and analytical (OLAP). In general we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

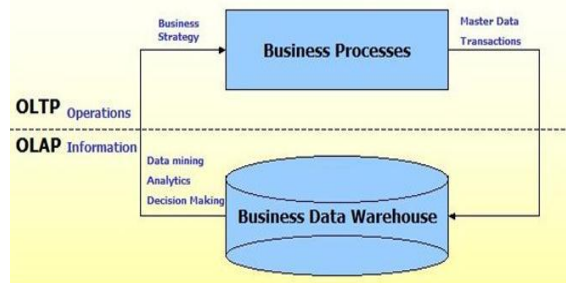


Figure 2: The above diagram shows the OLTP and OALP operations performance and how is difference both with Business processes and Business Data Warehouse

1.4 OLTP [On-line Transaction Processing] is characterized by a large number of short on-line transactions (INSERT, UPDATE, and DELETE). The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second. In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).

1.5 OLAP [On-line Analytical Processing] is characterized by relatively low volume of transactions. Queries are often very complex and involve aggregations. For OLAP systems a response time is an effectiveness measure. OLAP applications are widely used by Data Mining techniques. In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas

II. ARCHITECTURE FOR DATA WAREHOUSE AND END-TO-END PROCESS

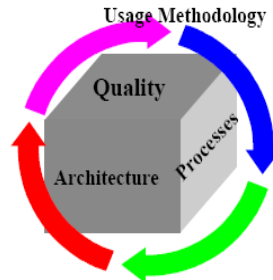


Figure 3. The different viewpoints for the metadata repository of a data warehouse.

Figure 4 shows a typical 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.

2.1 Data Warehouse vs. Prepared DBMS

OLTP (on-line transaction processing)

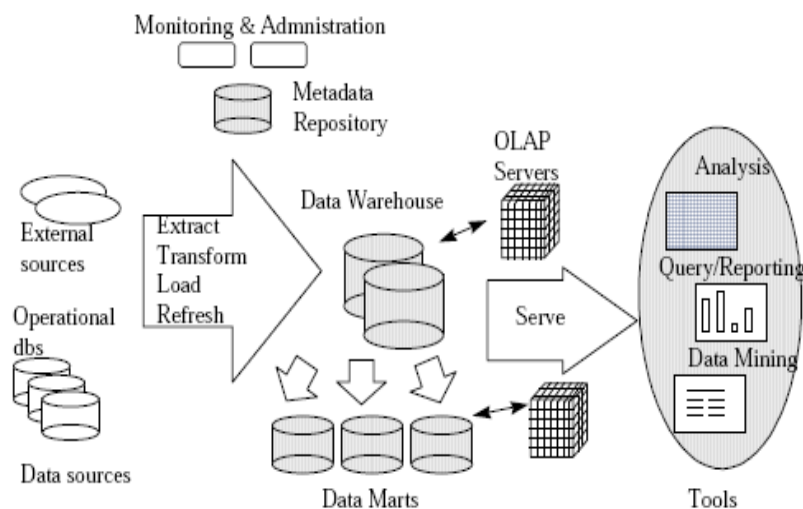


Figure 4: OLTP

- 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.

2.2 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.

2.3 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)

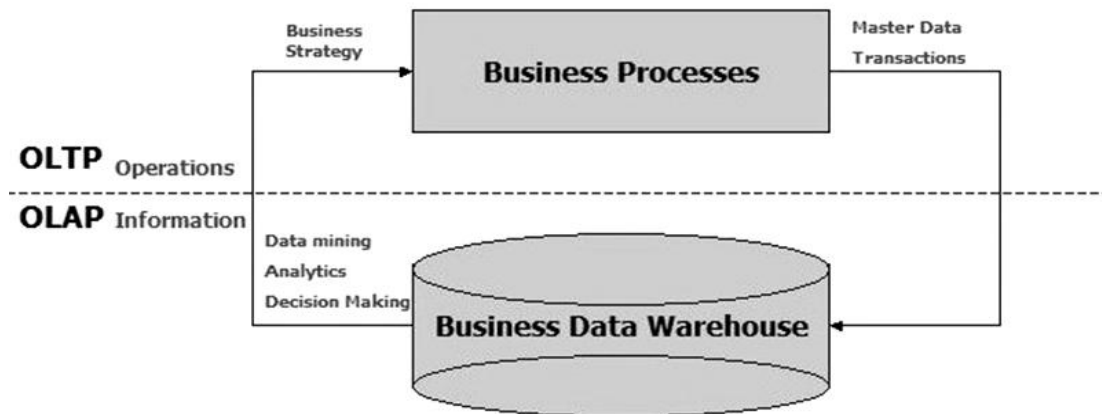


Figure 5: The above diagram shows the OLTP and OALP operations performance and how is difference both with Business processes and Business Data Warehouse

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

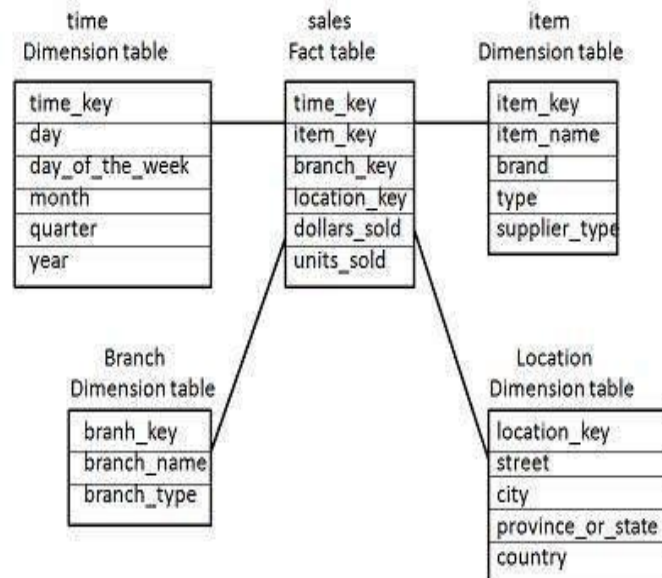
- Missing data: Decision support requires chronological data which prepared DBs do not typically maintain
- Data consolidation: DS requires consolidation (aggregation, summarization) of data from various sources
- Data quality: different source typically use inconsistent data representations, codes and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

II. DATABASE DESIGN METHODOLOGY

The multidimensional data model described above is implemented directly by MOLAP servers. We will describe these briefly in the next section. However, when a relational ROLAP server is used, the multidimensional model and its operations have to be mapped into relations and SQL queries. In this section, we describe the design of relational database schemas that reflect the multidimensional views of data. Entity Relationship diagrams and normalization techniques are popularly used for database design in OLTP environments. However, the database designs recommended by ER diagrams are inappropriate for decision support systems where efficiency in querying and in loading data (including incremental loads) are important. Most data warehouses use a *star schema* to represent the multidimensional data model. The database consists of a single fact table and a single table for each dimension. Each tuple in the fact table consists of a pointer (foreign key – often uses a generated key for efficiency) to each of the dimensions that provide its multidimensional coordinates, and stores the numeric measures for those coordinates. Each dimension table consists of columns that correspond to attributes of the dimension. Figure 3 shows an example of a star schema.

3.1 Star Schema

- Each dimension in a star schema is represented with only one-dimension table.
- This dimension table contains the set of attributes.
- The following diagram shows the sales data of a company with respect to the four dimensions, namely time, item, branch, and location.
- There is a fact table at the center. It contains the keys to each of four dimensions.
- The fact table also contains the attributes, namely dollars sold and units sold.



Star schemas do not explicitly provide support for attribute hierarchies. *Snowflake schemas* provide a refinement of star schemas where the dimensional hierarchy is explicitly represented by normalizing the dimension tables, as shown in Figure 4. This leads to advantages in maintaining the dimension tables. However, the denormalized structure of the dimensional tables in star schemas may be more appropriate for browsing the dimensions.

Fact constellations are examples of more complex structures in which multiple fact tables share dimensional tables. For example, projected expense and the actual expense may form a fact constellation since they share many dimensions.

VI. 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 database, 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. In Database management system multiple requirements are available and include new business techniques. In above discussion we see that the Business Database management and Data warehousing depended on the performance of Meta data, OLTP and OLAP performance. In the data warehousing move toward, information is requested, processed, and merged continuously, so the information is readily available for direct querying OLAP and analysis at the warehouse.

V. 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] Data Warehousing in the Real World; Sam Anahory & Dennis Murray; 1997, Pearson
- [16] Data warehousing System; Mallach; 2000, McGraw
- [17] Data Mining – Concepts & Techniques; Jiawei Han & Micheline Kamber, Jain Pei, Morgan Kaufmann publication; Third Edition; Page: 125 to 130, 151,152, 164, 165
- [18] Metadata Standards and Metadata Registries: An Overview" (PDF). Retrieved 2011-12-23.
- [19] <http://datawarehouse4u.info/OLTP-vs-OLAP.html>