A Comparative Study of OLTP and OLAP Technologies

[1]Jatinderpal Singh, [2]Anshul Sood

[1,2]Dept. of Computer Science & IT, Guru Gobind Singh Khalsa College Sarhali (Tarn Taran),
Punjab-India

Abstract

This paper provides an overview of Data warehousing and comparative study of OLAP, OLTP technologies. The data warehouse supports On-Line Analytical Processing (OLAP) who's functional and performance requirements are different from those of the On-Line Transaction Processing (OLTP) applications supported by the operational databases. Data warehousing and On-Line Analytical Processing (OLAP) are essential elements of decision support. OLTP is a class of program that facilitates and manages transaction-oriented applications. An OLAP system is used for data analysis by knowledge workers, including managers, business executives and market analysts. Data warehousing and OLAP have emerged as leading technologies that facilitate data storage, organization and retrieval.

Keywords

Data Warehousing, OLAP, OLTP

I. Introduction

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains chronological data derived from transaction data, but it can include data from other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources. It is a collection of methods, techniques, and tools used to support knowledge workers, directors, managers, and analysts to conduct data analyses that help performing decision-making processes and improving information resources [2]. A data warehouse is a "subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organizational decision making[1]. It is a type of computer database that is responsible for collecting and storing the information of a particular organization. The goal of using a data warehouse is to have an efficient way of managing information and examine data. Data warehouses are not optimized for transaction processing, which is the domain of OLTP systems. Data warehouses usually strengthen historical and analytic data derived from multiple sources. A data warehouse generally stores many months or years of data to support historical analysis. The data in a data warehouse is typically loaded through an extraction, transformation, and loading process from one or more data sources such as OLTP applications, mainframe applications, or external data providers.
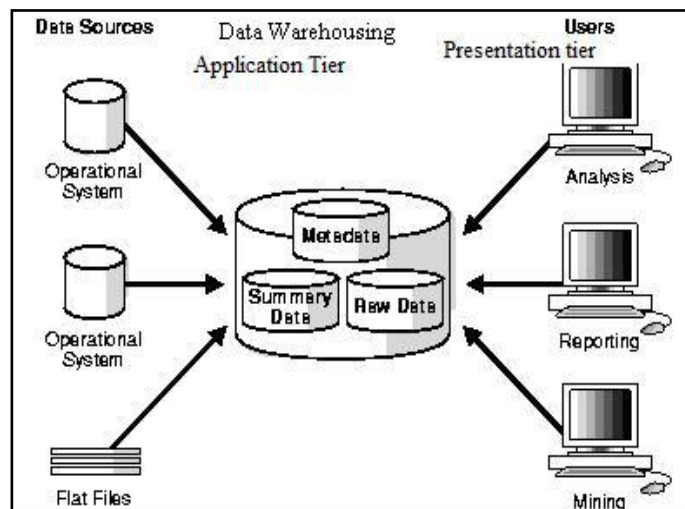

Fig. 1: Architecture of a Data Warehouse

In fig. 1. The Data Warehouse is built on three-tier architecture. Each tier can be managed and balanced independently, increasing flexibility. The three-tier architecture is made of the following layers

A. Data Sources Tier

It Stores the Data Warehouse data. It Handles the Data Warehouse ETL process where all the source data is extracted, scrubbed, altered and loaded in the Data Warehouse.

B. Application Tier

This is where all the business intelligence is built.

C. Presentation Tier

Graphical user interface allowing:- Reporting, Analysis.- Querying.
The data in a data warehouse have the following characteristics:

1. Subject Oriented

Information is presented according to specific subjects or areas of interest, not simply as computer files [6]. Data is manipulated to provide information about a particular subject. e.g., around customers, sales, or items produced.

2. Integrated

A single source of information for understanding multiple areas of interest [6]. All of the data about the subject are combined and can be analyzed together.

3. Time Variant

Historical as well as current information are maintained in detail form.

4. Nonvolatile

The data are read only, not updated or changed by users [6]. Information is consistent regardless of when the warehouse is accessed.

5. Accessible
The primary purpose of a data warehouse is to provide readily accessible information to end-users.

II. Data Warehouse vs. Operational DBMS
The data warehouse is distinctly different from the operational data used and maintained by day-to-day operational systems.

OPERATIONAL DATA	DATAWAREHOUSE DATA
Data is application oriented	Data is subject oriented
It is more detailed and updated frequently	It is summarized and not updated frequently
Run repetitively, and serves the clerical or the lower level community	Run heuristically ,and serves the managerial community
Requirements for processing understood before initial development	Requirements for processing not completely understood before development
It is compatible with the Software Development Life Cycle	It has completely different life cycle
Immediate response required when entering a transaction	Immediacy response is not required
Transaction driven	Analysis driven
The control of update a major concern in terms of ownership	The control of update no issue
It is managed in its entirety	It is managed by subsets
There is no redundancy in data and have high availability.	Data is redundant and have low availability
Small amount of data used in a process	Large amount of data used in a process

III. OLAP and OLTP
OLAP(Online Analytical Programming) contains the history of OLTP data, which is, non-volatile ,acts as a Decisions Support System and is used for creating forecasting reports. It is computer processing that enables a user to easily and selectively extract and view data from different points of view [5]. OLAP can be used for data mining or the discovery of previously undiscovered relationships between data items. In fig. 2, OLAP database does not need to be as large as a data warehouse, since not all transactional data is needed for trend analysis. It deals with Historical Data or Archival Data. Historical data are those data that are archived over a long period of time. Example: If we collect last 10 years data about flight reservation, the data can give us much meaningful information such as the trends in reservation. This may give useful information like peak time of travel, what kinds of people are traveling in various classes (Economy/Business) etc.OLTP is nothing but Online Transaction Processing, which contains a normalized tables and online data, which have frequent insert/ updates/delete. It is a class of program that facilitates and manages transaction-oriented applications, typically for data entry and retrieval transactions in a number of industries, including financial institutions, airline industry, supermarkets, and manufacturers [7]. It deals with operational data. Operational data are those data involved in the operation of a particular system. Example: In a banking System, you withdraw amount through an ATM. Then account Number, ATM PIN Number, Amount you are withdrawing, Balance amount in account etc are operational data elements.

IV. Need of Data Warehousing and OLAP
OLAP technology enables data warehouses to be used effectively for online analysis, providing rapid responses to iterative complex analytical queries. OLAP's multidimensional data model and data aggregation techniques organize and summarize large amounts of data so it can be evaluated quickly using online analysis and graphical tools. The answer to a query into historical data often leads to subsequent queries as the analyst searches for answers or explores possibilities. OLAP systems provide the speed and flexibility to support the analyst in real time.
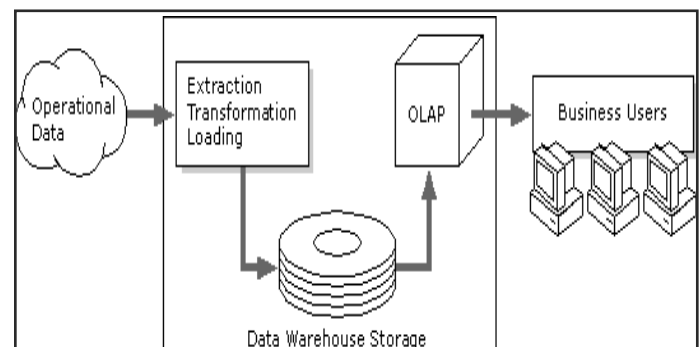

Fig. 2: OLAP Functioning

Moreover, processing OLAP queries in operational databases would substantially degrade the performance of operational tasks. An operational database supports the concurrent processing of multiple transactions. 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.

V. OLTP Versus OLAP

A. Major Distinguishing Features

1. 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 [6].

2. 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 [6].

3. 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 [6].

4. 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 [6]. Because of their huge volume, OLAP data are shared on multiple storage media.

5. Access Patterns

Access patterns of an OLTP system consist mainly of short, atomic transactions. Such a system requires concurrency, control and recovery mechanisms [6]. But, accesses to OLAP systems are mostly read-only operations, although many could be complex queries.

OLTP	OLAP
Users are Clerk, IT professional	Users are Knowledge workers
The purpose of data is to control and run fundamental business tasks, and day to day operations	The purpose of data is to help with planning, problem solving, and decision support
Data base design is Application-oriented Highly normalized with many tables	Data base design is Subject-oriented Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Data is Current, up-to-date Detailed, flat relational Isolated	Data is Historical, Summarized, multidimensional Integrated, consolidated
It is used repetitively	It is used in Ad-hoc fashion
Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Access is Read/write	Lots of scans
Relatively standardized and simple queries Returning relatively few records	Its queries are often complex involving aggregations
Records accessed are Tens	Records accessed are Millions
Users are Thousands	Users are Hundreds
Space Requirements can be relatively small if historical data is archived. Data base size is 100MB-GB	Space Requirements larger due to the existence of aggregation structures and history data; requires more indexes than OLTP. Data base size is 100GB-TB
Metric is Transaction throughput	Metric is Query throughput, response
Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Processing Speed is very fast	Processing Speed depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method

VI. Characteristics of OLTP Versus OLAP

	OLTP	OLAP
System scope	Single business process	Multiple business subjects
Data sources	One	Many
Data model	Static	Dynamic
Data volume	Small/medium	Large
Update	Immediate	Periodic
Historical data	Not available	Available
Response times	Can be in minutes	1 second
Users	Front office	Managers/ Executive

VII. Conclusion

Data warehouses provide OLAP tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data mining. 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. OLAP enables the organization as a whole to respond more quickly to market demands. Market awareness, in turn, often yields improved revenue and profitability.

It is critical to ensure that transactional and analytical systems are separated from one another so that performance and functionality are maintained. Overall, OLAP and OLTP systems each serve a distinct purpose and when used correctly, can be an extremely powerful tool in establishing or maintaining a successful business.

References

[1] Inmon, W.H.,"Building the Data Warehouse", Second Edition, New York: John Wiley & Sons., 1996.

[2] Kimball, R.,"The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses", New York: John Wiley & Sons., 1996.

[3] Kachur, R.,"Data Warehouse Management Handbook", Paramus: Prentice Hall, 2000

[4] Adelman, S., Moss, L.,"Data Warehouse Project Management", Boston: Addison-Wesley, 2000.

[5] Chaudahari, Surajit, Dayal, Umeshwar,"An overview of Data warehousing and OLAP technology", SIGMOD Record, 1997.

[6] Han, J, Kamber, M,"Data Mining Concepts and Techniques", Morgan Kaufmann, 2001.

[7] Han, Jiawei, Kamber, Micheline,"Data Mining: Concepts and techniques", Academic Press, 2001.

[8] Adelman, S., Moss, L.,"Data Warehouse Project Management", Boston: Addison-Wesley, 2000.

[9] Witten, Ian H., Frank, Eibe,"Data mining: Practical machine learning tools and techniques", Academic Press, 2000

[10] Codd, E. F., Salley C. T.,"Providing OLAP to User-Analysts: An ITMandate", Manchester: Hyperion Solutions Europe, E. F. Codd & Associates, 1993.