IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Comparative Analysis and Transformation Techniques between OLTP and OLAP

Dr.Pradeep Kumar Jatav

International Institute of Professional studies

DAVV,Indore

Abstract

OLAP is an integral part of Business Intelligence, which assists in the process of converting raw data into meaningful information by using multidimensional analysis. On the other hand, OLTP systems are designed in a normalized manner to process real-time transactions. This paper theoretically presents a technique for the conversion of OLTP data to OLAP data by using ETL processes. ER modeling is used in OLTP, whereas dimensional modeling is used in OLAP.

Introduction

OLAP assists in analyzing and interpreting data within the database environment. OLAP differs from traditional systems that store data in tables by storing data in multidimensional cubes. Traditional business applications were based on relational databases that supported the processing of day-to-day transactions, referred to as OLTP (Online Transaction Processing). But as business operations advanced, activities such as sales reporting, budgeting, financial planning, and forecasting were moved to a separate OLAP (Online Analytical Processing) system to improve analytical capabilities.

Though OLTP and OLAP are relational in nature, OLAP is usually relational in a star schema. OLTP systems are considered the main data source for OLAP. Organizations produce a huge amount of data in different forms and sources. It is essential to structure and organize the data in a proper manner so that it can be utilized effectively for reporting purposes.

OLTP, OLAP, and data mining, together, provide a holistic approach to data processing and analysis. The transactional data collected from OLTP systems is processed using ETL techniques and stored in OLAP data warehouses, where data mining algorithms are used for decision support. This paper will theoretically explore the significance of OLTP, OLAP, and data mining in today’s data-driven world, focusing on their relevance in facilitating efficient operations and strategic decision-making. Some important definitions are there.[1,2,12,13,15]

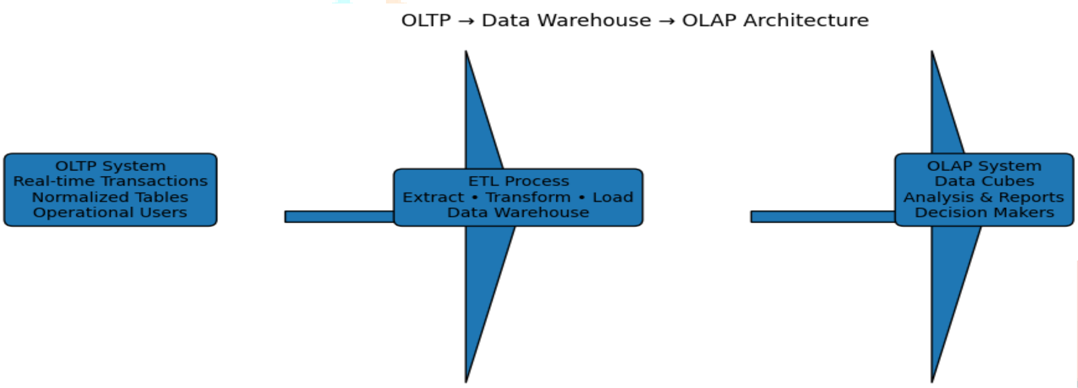
1 Data Mining
Data mining is the process of discovering meaningful patterns, correlations, and knowledge from large datasets using statistical and machine learning techniques (Fayyad, Piatetsky-Shapiro & Smyth, 1996).

2 Data Warehouse
A data warehouse is a centralized repository that stores integrated, subject-oriented, time-variant, and non-volatile data to support decision-making and analytical processing (Inmon, 1996).

3 OLTP (Online Transaction Processing)
OLTP refers to systems designed to manage and execute large volumes of real-time transactional data efficiently while ensuring data consistency and concurrency (Gray & Reuter, 1993).

4 OLAP (Online Analytical Processing)
OLAP is a category of systems that supports multidimensional analysis of historical data to enable complex queries, reporting, and decision-making (Codd, Codd & Salley, 1993).

5. ETL is an acronym for Extract, Transform, and Load. It is a procedure that can be used to extract data from different sources and transform the data into a database. The requirement for using ETL comes from the fact that in modern computing, business data is stored in different places and in many different formats.[15]



2. Literature Review

Inmon (2005) described a data warehouse as an integrated, subject-oriented, time-variant, and non-volatile data repository designed to support decision-making, forming the basis for contemporary enterprise-level data warehouse design and development. In a similar vein,

Chaudhuri and Dayal (1997) offered one of the first comprehensive reviews on data warehousing and OLAP technology. They highlighted the need to integrate data from multiple sources to support effective analysis. Their research also highlighted the need for robust ETL (Extract-Transform-Load) processes and the use of multidimensional data models for efficient analysis.

Gray et al. (1997) introduced the data cube operator, which facilitated easier aggregation and multidimensional analysis in OLAP systems. This contributed greatly to improved query efficiency and faster decision-making processes.

Kimball and Ross (2013) offered dimensional modeling techniques and star schema designs that improved query speed and made OLAP systems more accessible. Their work, together with other contributions, formed the conceptual and structural basis for modern data warehouse systems.

Codd et al. (1993) formally defined the term OLAP and proposed a set of guidelines for developing systems capable of handling multidimensional data analysis and providing fast query responses.

Sarawagi, Agrawal, and Megiddo (1998) investigated the discovery-driven analysis process in OLAP data cubes, demonstrating how analytical systems can discover hidden patterns and meaningful trends in large amounts of data.

Vassiliadis and Sellis (1999) analyzed logical models for OLAP databases and assessed various schema designs for multidimensional analysis. Their research contributed to the development and enhancement of more efficient OLAP system architectures.

Hellerstein and Stonebraker (2005) investigated database system architecture in depth, pointing out the importance of concurrency control, indexing, and logging in ensuring high performance and data consistency in OLTP systems. Later, Harizopoulos et al. (2008) investigated OLTP workloads, pointed out the most critical performance bottlenecks in traditional databases, and proposed new system designs for more efficient high-volume transaction processing.

Abadi et al. (2013) investigated column-oriented database systems for OLAP operations and compared their performance with traditional row-oriented systems typically used in OLTP. Later, Stonebraker et al. (2018) proposed the concept of Hybrid Transaction/Analytical Processing (HTAP), which refers to the integration of transactional and analytical processing in a single system. Along the same lines, Plattner (2014) proposed new database architectures that can efficiently support both OLTP and OLAP workloads simultaneously.

ETL tools provide a necessary link between OLTP and OLAP systems. Vassiliadis (2009) provided a comprehensive discussion on ETL techniques, describing the importance of extraction, transformation, and loading for effective data integration. Rahm and Do (2000) discussed data cleaning approaches necessary for maintaining high-quality data in data warehouses, concentrating on activities such as standardization, duplicate removal, and validation. These two studies together emphasize the importance of high-quality data for precise and trustworthy analysis.

Armbrust et al. (2015) presented Spark SQL as a framework for scalable data analysis and processing, enabling fast OLAP analysis on distributed data. Zaharia et al. (2016) presented Apache Spark as a unified engine for big data processing, supporting both batch and real-time processing. These two frameworks have enabled organizations to develop scalable and high-performance analytical systems.

In general, the literature suggests that OLTP systems are optimized for real-time transaction processing, while OLAP systems are optimized for complex analytical queries. Data warehouses serve as an intermediary layer that integrates data through ETL processes and offers a structured environment for analysis. The combination of OLTP and OLAP systems through hybrid architectures and advanced data processing frameworks is an ongoing area of research in database and information systems.

3. Comparative Analysis of OLTP and OLAP Systems

OLTP to OLAP transformation is critical for the development of a data warehouse for business intelligence. OLTP systems are responsible for handling day-to-day business transactions, such as banking, ordering, and inventory management. These systems are highly normalized and designed for high concurrency and fast transaction processing, but are not ideal for complex analysis and historical reporting. For analytical purposes, data is migrated from OLTP systems to OLAP systems using the ETL (Extract, Transform, Load) process. Data is extracted from operational systems such as ERP and CRM, transformed to clean, integrated, validated, and standardized data, and finally loaded into a data warehouse using multidimensional models such as star or snowflake schema. [1,2,12,13]

Table: 3.1 Comparisons between OLTP and OLAP

Feature	OLTP	OLAP
Full Form	Online Transaction Processing	Online Analytical Processing
Purpose	Supports day-to-day business operations	Supports analysis and decision-making
Type of Queries	Simple, short transactions (INSERT, UPDATE, etc)	Complex, long-running analytical queries
Data Nature	Current and real-time data	Historical and aggregated data
Data Volume	Small per transaction	Very large datasets
Database Design	Highly normalized schemas	Denormalized / multidimensional schemas
Users	Large number of concurrent users	Fewer users (like analysts, managers)
Performance Focus	Fast transaction processing	Fast query and analysis performance
Update Frequency	Frequent data updates	Periodic data refresh
Example	ATM transactions, order processing	Sales trend analysis, forecasting

4. Methodology for OLTP to OLAP Data Transformation

The conversion of data from OLTP systems to an OLAP environment is done in a systematic way that helps in effective analysis and decision-making. First, the data from various OLTP systems is consolidated, cleaned, and made standardized to eliminate inconsistencies, duplicates, and errors, and a common dataset is formed. Then, the ETL (Extract, Transform, Load) process is used to extract the data, transform it by aggregation, calculation, filtering, and formatting, and finally load the data into a data warehouse or OLAP system and form multidimensional structures like stars or snowflakes. This helps in the conversion of transactional data into decision-support data.

4.1 OLTP to OLAP Transformation Process

The process of transitioning from OLTP (Online Transaction Processing) to OLAP (Online Analytical Processing) is accomplished through the ETL (Extract-Transform-Load) process. In operational databases, OLTP databases are used to store real-time transactional data, such as sales, financial transactions, and customer information. However, this data is very normalized and not amenable to analysis. As such, data is extracted from OLTP databases, transformed through cleansing, integration, aggregation, and standardization, and finally loaded into a data warehouse. The data warehouse stores historical and integrated data in a multidimensional format, which is ideal for OLAP tools to execute complex queries, reporting, and decision-making. [1,2,13,15]

Figure: OLTP → OLAP Transformation

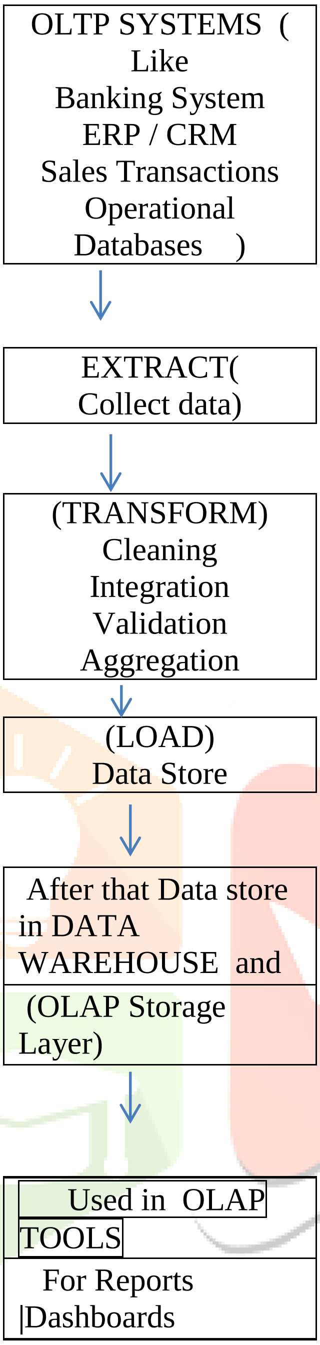


Figure: Transformation of data from OLTP systems to OLAP environment through ETL process including extraction, transformation, and loading into a data warehouse for analytical processing.

4.2 Transformation of OLTP Data into OLAP Data: Step-by-Step Process

The transformation of OLTP data into OLAP data includes a number of sequential processes that are required to transform the transactional data into a form that is suitable for analytical processing.

Step 1: Data Extraction from OLTP Systems

The transactional data is first extracted from the OLTP databases, including those that handle sales, inventory, banking, or human resources. While extracting the data, only those tables, fields, and records that are of interest for analysis are chosen. This step is planned in such a way that it does not affect the ongoing operations in any way and ensures that the OLTP systems are not impacted in any manner while the data is being extracted for further processing.

Step 2: Merge and Organize Data

The data extracted from the OLTP systems is merged into a common area where it is further processed for further analysis. During this step, data cleansing and validation methods are used to eliminate inconsistencies, eliminate duplicates, and address errors in the data. The data is then organized based on certain business areas like sales, customers, products, and transactions, which can then be easily aggregated for further analysis.

Step 3: Transformation

The transformation step is where the integrated data is transformed to make it suitable for OLAP processing. In this step, various processes like denormalization, summarization, aggregation, and the development of calculated metrics are carried out to make the data suitable for analysis. Business rules are applied to make the data formats consistent and to remove any inconsistencies. This step prepares the data to efficiently handle multidimensional queries, aggregation, and complex analysis in the OLAP environment.

Step 4: Load into OLAP System

The transformed data is then loaded into a centralized data warehouse or OLAP system for analysis. In this step, the data is arranged in a multidimensional schema like the star schema or snowflake schema, which is designed to efficiently handle fast queries, aggregation, and reporting. The OLAP environment holds integrated, historical, and consistent data, allowing users to carry out complex analysis, reporting, and gain insights that help in strategic decision-making.

Step 5: Analytical Processing

Lastly, the OLAP system provides support for multidimensional analysis, complex querying, and sophisticated reporting. The user can carry out operations such as slicing, dicing, roll-up, and drill-down analysis to analyze data from various angles. This analytical functionality helps organizations to discover trends, carry out forecasting, and produce valuable insights, thus aiding in effective business intelligence and decision-making.

This is a structured ETL process that ensures the operational OLTP data is transformed into valuable insights through OLAP. [1,2,3,4,12,13,15]

4.3. OLTP vs OLAP Cleaning Comparison

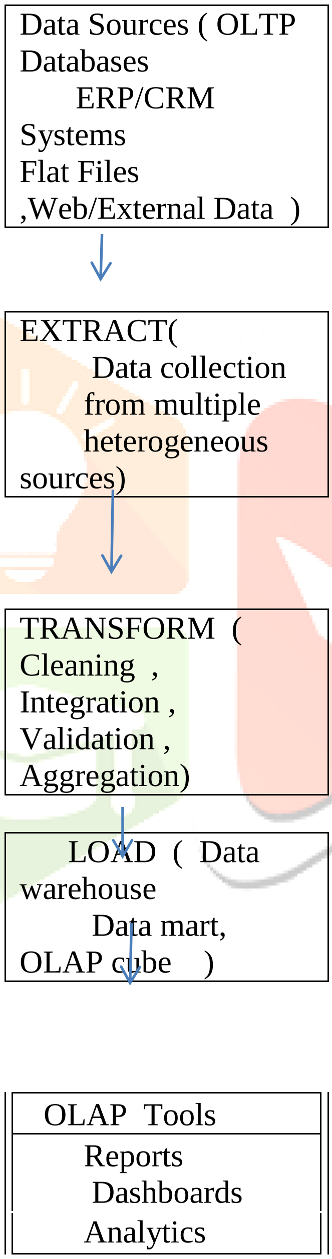
Feature	OLTP Cleansing	OLAP Cleansing
Processing type	Real-time	Batch
Tools used	APIs, validation tools	ETL & data quality tools
Data size	Small transactions	Large historical data
Goal	Correct data entry	Analytical accuracy
Speed	Very fast	Moderate
Example tools	Melissa, Loqate	Informatica, Talend

4.4 Process of ETL (Extract–Transform–Load)

ETL Process (Extract–Transform–Load)

The ETL (Extract–Transform–Load) process is a basic element of data warehousing and OLAP technology, which combines data from various heterogeneous sources into a single repository for further analysis. The process involves the following steps: extraction, where data is obtained from OLTP databases, ERP systems, flat files, web logs, and other sources. The extracted data is usually inconsistent, incomplete, and redundant. The transformation step involves data cleansing, validation, standardization, and consolidation to make it consistent and of high quality. Some of the operations involved in this step include removing duplicates, correcting errors, changing data types, applying business logic, aggregating data, and consolidating data from multiple sources into a single schema. The final step, loading, involves depositing the transformed data into a data warehouse, data mart, or OLAP cube, either completely, incrementally, or near real-time. The ETL process is essential in ensuring that high-quality data is available for analysis, reporting, and decision-making in today's business intelligence environment. [1,2,12,13,15]

Figure: ETL Process Architecture



9. Conclusion

The above discussion describes how the OLTP data can be transformed into an OLAP system. The OLTP system is used to store real-time transactional data, which is then further transformed into an OLAP system. The ETL process is used to extract data from the OLTP system, which is then further transformed into an OLAP system. The OLAP system is used to provide multidimensional analysis, which is further used to make business decisions. The combination of OLTP, ETL, and OLAP systems helps to transform business data into valuable insights. The conclusion based on this study provides steps: 1. insert a new record with Primary Key or foreign key in the source database, 2. extract the data from OLTP. 3. Transform the data into an OLAP system with data clearing and aggregation steps. And after 4. Load the data into a data warehouse.

References

1.	Inmon, William H. Building the data warehouse. John wiley & sons, 2005.
2.	Kimball, Ralph, and Margy Ross. The data warehouse toolkit: The definitive guide to dimensional modeling. John Wiley & Sons, 2013.
3.	Chaudhuri, Surajit, and Umeshwar Dayal. "Data warehousing and OLAP for decision support." In Proceedings of the 1997 ACM SIGMOD international conference on Management of Data, pp. 507-508. 1997.
4.	Gray, Jim, Surajit Chaudhuri, Adam Bosworth, Andrew Layman, Don Reichart, Murali Venkatrao, Frank Pellow, and Hamid Pirahesh. "Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals." Data mining and knowledge discovery 1, no. 1 (1997): 29-53.
5.	Codd, Edgar F. "Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate." http://www. arborsoft. com/papers/coddTOC. html (1993).
6.	Sarawagi, Sunita, Rakesh Agrawal, and Nimrod Megiddo. "Discovery-driven exploration of OLAP data cubes." In International Conference on Extending Database Technology, pp. 168-182. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
7.	Vassiliadis, Panos. "A survey of extract–transform–load technology." International Journal of Data Warehousing and Mining (IJDWM) 5, no. 3 (2009): 1-27.
8.	Rahm, Erhard, and Hong Hai Do. "Data cleaning: Problems and current approaches." IEEE Data Eng. Bull. 23, no. 4 (2000): 3-13.
9.	Abadi, Daniel, Anastasia Ailamaki, David Andersen, Peter Bailis, Magdalena Balazinska, Philip Bernstein, Peter Boncz et al. "The Seattle report on database research." ACM Sigmod Record 48, no. 4 (2020): 44-53.
10.	Armbrust, Michael, Reynold S. Xin, Cheng Lian, Yin Huai, Davies Liu, Joseph K. Bradley, Xiangrui Meng et al. "Spark sql: Relational data processing in spark." In Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 1383-1394. 2015.
11.	Meng, Xiangrui, Joseph Bradley, Burak Yavuz, Evan Sparks, Shivaram Venkataraman, Davies Liu, Jeremy Freeman et al. "Mllib: Machine learning in apache spark." Journal of Machine Learning Research 17, no. 34 (2016): 1-7.
12.	Mukherjee, Rajendrani, and Pragma Kar. "A comparative review of data warehousing ETL tools with new trends and industry insight." In 2017 IEEE 7th International Advance Computing Conference (IACC), pp. 943-948. IEEE, 2017.
13.	Nitesh Vyas , "A Study on Comparison and Transformation of OLTP into OLAP", in 2011,International journal of advanced Research in computer science volume2
14.	Books [1] Berson, A., Data Warehousing, Data Mining, and OLAP, McGraw-l 1997.
15.	Prabhu,C.S.R.(Ed.), Database for the Millennium 2000, Tata McGraw-Hill,1999. [3] Thomsen, E.,OLAP Solutions : Building Multidimensional Information Systems, John Wiley & Sons, 1997.