ISSN No. 0976-5697

Volume 2, No. 1, Jan-Feb 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Study on Comparison and Transformation of OLTP into OLAP

Nitesh Vyas
Asstt. Professor
, Department Of Computer Science
Shri Vaishnav Institute Of Management,
Indore,India,
Nitesh.V25@Rediffmail.Com

Abstract: OLAP is an important feature of Business Intelligence. It is a process of converting new data into business information through multi-dimensional analysis. Database design for OLTP is in normalized form. OLTP databases provide real time accesses to its data which is being updated by other transactions. OLTP is abbreviation of On-Line Transaction Processing. This paper illustrating a theoretical procedure to convert OLTP data into OLAP data and also comparison of OLTP and OLAP. For this transformation Merge Data, Scrub Data (Cleansing data), Aggregate Data, Organize Data in Cubes is used. ER modeling is used for normalizing the OLTP database design. Dimensional modeling is used for de-normalizing the ROLAP/MOLAP design. The concept of ETL : extract, transform, and load is used for transformation of data.

Keywords: OLAP, OLTP, ETL and Data warehouse.

INTRODUCTION

OLAP - is the ability to conduct data analysis within the context of the database. OLAP data is organized hierarchically and stored in cubes instead of tables. Relational database systems have been the backbone of business applications. The more complex business requirements became, the more we focused on the so-called transactional processing part and designed the database structures accordingly. These systems are called OLTP Online Transactional Processing) system. OLAP is part of the business intelligence. Business Reporting for sales, Management Reporting, Financial Planning, Budgeting and Forecasting applications were increasingly moved out to separate systems for better performance. These systems are called OLAP (Online Analytical Processing) systems. OLTP and OLAP, are based on the relational theory. For OLAP systems, data is organized in star schemas. OLTP is the necessary prerequisite for OLAP. Column storage is necessary for both OLTP and OLAP. In an organization there is a vast data. Data is located in different places and in many incompatible formats. Data is a vital part for an organization. It should be in a reportable format for taking decision.

II. DEFINITIONS

A. (Olap) On-Line Analytical Processing

It is an important feature of Business Intelligence. It is a method of analyzing data in a multi-dimensional format. On Line Analytical Processing is a database technology that has been specially designed to deal with high performance querying and reporting. OLAP data is organized hierarchically and stored in cubes instead of tables. On-Line Analytical Processing - OLAP - is the ability to conduct data analysis within the context of the database. It is a process of converting new data into business information through multi-dimensional analysis.

B. OLTP: On-Line Transaction Processing

OLTP databases provide real time accesses to its data which is being updated by other transactions. OLTP databases contrasts earlier databases that allowed updation in batch modes as all updations were provided in one go In

case of OLTP there is no such boundation and all updations can be done anytime and other users will automatically see your changes. Application databases are OLTP (On-Line Transaction Processing) systems where every transaction has to be recorded, and super-fast at that.

C. ETL

ETL stands for Extract, Transform and Load, which is a process used to collect data from various sources, transform the data into a database. The need to use ETL arises from the fact that in modern computing business data resides in multiple locations and in
many incompatible formats. For example business data might be stored on the file system in various formats (Word docs, PDF, spreadsheets, plain text, etc), or can be stored as email files, or can be kept in a various database servers like MS SQL Server, Oracle and MySQL for example. Handling all this business information efficiently is a great challenge and ETL plays an important role in solving this problem. ETL is software that enables businesses to consolidate their disparate data while moving it from place to place, and it doesn't really matter that that data is in different forms or formats. The data can come from any source.

D. Data Warehouse

A data warehousing (or data mart) system is the backend, or the infrastructural, component for achieving business intelligence. Data warehouse is a collection of data

designed to support management for decision - making. A Data Warehouse is a, Time variant Database, Integrated Database, Subject oriented Database, Non-Volatile Database and a Data Warehouse is a historical database. Hence it is also called Read Only Database.Data warehousing deals with all aspects of managing the development, implementation and operation of a data warehouse or data mart including meta data management, data acquisition, data cleansing, data transformation, storage management, data distribution, data archiving, operational reporting, analytical reporting, security management, backup/recovery planning, etc. The cost justification for data warehousing technology, however, is not straightforward. A data warehouse is concerned with the creative and analytical aspects of a business, with a focus on applications that extract and manipulate data. As such, there is no simple, equivalent cost-justification calculation.

III. COMPARISON BETWEEN OLTP AND OLAP

In OLTP use operational data as data source and for business tasks we use snapshot of business processes and queries run by user. The database we use in OLTP is normalized form. Due to smaller database speed is very fast.

In OLAP we use consolidation data from various sources. It provides multi-dimensional views for business activities in planning and decision making. In this queries are run by system to update the aggregated data. The database we use in OLAP is de-normalized. Due to large database speed is issue. Therefore it adopts : star, snowflake or fact constellation mode of subject-oriented database design.

Table I: shows Comparison between OLTP and OLAP

perational	Informational
Transactional	Analytical
Database Constantly updated	Database Periodically updated
Constant performance	sporadic usage
Day-to-day operations	Depends on Managerial needs
Nature of applications is Structured	Nature of applications is Flexible
Current data	Current and historical data
Short database transactions	Long database transactions
Online update/insert/delete	Batch update/insert/delete
Normalization is promoted	Denormalization is promoted
Transaction recovery is necessary	Transaction recovery is not necessary

IV. PROCESSES FOR TRANSFORMATION OF OLTP DATA INTO AN OLAP DATA

A. Merge Data

a) Merge all the data related to specific items (products, customers, employees) from multiple OLTP systems into a single OLAP system.

b) The merge process must resolve differences in encoding between the different OLTP systems. The merge process must be able to match common employee data from both systems, perhaps by comparing employee names and addresses.

c) The merge process must also be able to convert data stored using different data types in each OLTP system to a single data type used in the OLAP system.

a. Data Cleansing: is the process of ensuring that a set of data is correct and accurate. During data cleansing, records are checked for accuracy and consistency, and either corrected, or deleted as necessary. It is a process of identifying or changing the inconsistencies and inaccuracies.

b. Inconsistent: If the data is not in a proper format.

c. Inaccuracy: It is applicable for numeric values.

B. Common Software Products For Name And Address Cleansing

*a. Trillium:*Used for cleansing the name and address data. The software is able to identify and match households, business contacts and other relationships to eliminate duplicates in large databases using fuzzy matching techniques.

b. *First Logic:*Aggregate Data: is the data that is the result of applying a process to combine data elements from different sources. The aggregate data is usually taken collectively or in summary form.

a) OLTP data records all transaction details.

b) OLAP queries typically need summary data, or data aggregated in some fashion. Systems designed to handle the queries required to discover trends and critical factors are called online analytical processing (OLAP) systems. OLAP queries typically require large amounts of data .

c) The degree to which you aggregate the data in a data warehouse depends on a number of design factors, such as the speed requirements of your OLAP queries..

C. Organize Data In Cubes

1) Relational OLTP data is organized in a way that makes some analysis processing difficult and time-consuming.

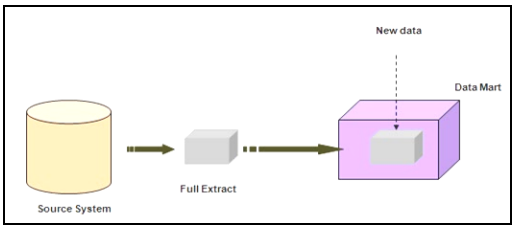
2) When OLTP data is moved into a data warehouse, it must be transformed into an organization that better supports decision support analysis. The process of building a data warehouse involves reorganizing OLTP data stored in relational tables into OLAP data stored in multidimensional cubes.

D. Process Of ETL

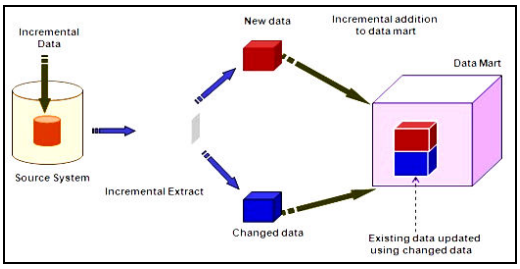
The process of extracting the data from the OLTP and legacy data sources and transforming it into the warehouse servers is called the ETL process. The ETL process has 3 main steps, which are Extract, Transform and Load.

E. Extract

The first step in the ETL process is extracting the data from various sources. Each of the source systems may store its data in completely different format from the rest. There are two types of Extraction:



a. Full Extraction



b. Periodic Extraction

F. Transform

Once the data has been extracted and converted in the expected format, it's time for the next step in the ETL process, which is transforming the data according to set of business rules. The data transformation may include various operations including but not limited to filtering, sorting, aggregating, joining data, cleaning data, generating calculated data based on existing values, validating data, etc.

G. Load

The final ETL step involves loading the transformed data into the destination target, which might be a database or data warehouse.

The process of making data available through OLAP applications typically goes through following phases :
Step I : Add new record in source database, based on Primary Key.
Step II**:** Extract the data from OLTP or legacy data sources into a staging area.
Step III: Transform the data into a OLAP system. This involves actions such as data cleaning and aggregation.
Step V: Load the data into a data warehouse.

V. CONCLUSION

This research concluded theoretical review for transforming of OLTP data into OLAP. Once the data is loaded into a data warehouse, an important part of an OLAP system is to provide facilities for decision makers to access and analyze the data in the data warehouses. ETL is an important tool for transformation of data.

VI. REFERENCES

Books

[1] Berson, A., Data Warehousing, Data Mining, and OLAP, McGraw-l 1997.
[2] 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.
[4] Adamson, C., and Venerable, M., Data Warehouse Design Solutions, Jonh Wiley & Sons, 1998.
[5] Alexis Leon and Mathews Leon : Database Management Systems : Co-published by Leon Press, Chennai and Vikas Publishing House Pvt Ltd., New Delhi – 110 014.
[6] McClain, G.R., OLTP Handbook, McGrwa-Hill, 1993.

On the Internet

[1] www.peaksoftware.com/glossary/index.html.
[2] http://msdn.microsoft.com/enrary/aa933151 (SQL.80).aspx
[3] http://articles.techrepublic.com.com/5100-10878_111032206.html
[4] .http://datastageinfosoft.blogspot.com