Introduction on Data Warehouse with OLTP and OLAP

Arpit Parekh

Maharaja Krishnakumar Sinhji Bhavnagar University, Department of Computer Application, Shree Swaminaranyan Naimisharanaya College of Management and IT, Bhavnagar, India
aparekh5888@gmail.com

Abstract: *Data warehouse, online transaction processing system (OLTP) and on-line analytical processing (OLAP) are basically main need of the database collection in business, corporate fields and many areas. Nowadays many services, products and new techniques are available and offering many ideas in the DBMS. This paper is research about the data warehousing with OLAP and OLTP with the basic need and main parts of the business database management. The main here consider that Data warehouse usage, process, Data warehouse with Meta data, online transaction processing system (OLTP), on-line analytical processing (OLAP) and also OLTP vs. OLAP with advantages and disadvantages [5].*

Keywords: Meta data, online transaction processing system, on – line analytical processing, Data Warehouse

1. Introduction

OLTP, OLAP, Meta data and Data warehouse are essential elements of supports system, which has increasingly become a focus of the database industry. Many commercial products and services are now available, and all of principal database management system providers now offerings in these areas. Decision support places some different requirements on database technology compared to traditional on – line transaction processing applications [2].
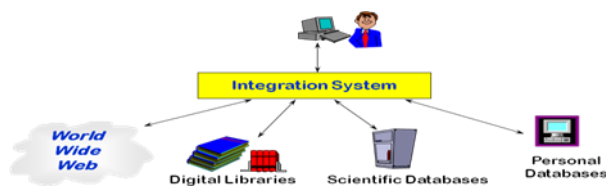

Figure 1: The above diagram shows in the integration system we are use different tools but also database management systems need the basic requirements depends of Data warehouse, meta data, OLTP and OLAP.

1.2 Metadata

Meta data is "data about data" or Meta data is "the data used to define other data". It specifies source, values, usage and features of DWH data and defines how data can be changed and processed at every architecture layer. Meta data is stored in a Meta data repository which all the other architecture components can access. Meta data is a critical need for using, building, and administering the data warehouse. For end-users, metadata is like a roadmap to the data warehouse contents. A Meta data repository is like a general-purpose information directory that includes several enhancing functions [6].

According to Kelly, a tool for Meta data management should:
- Allow administrator to perform system administration operations, and manage security.
- Allow end users to navigate and query Meta data.
- Allow end users to uses a GUI.
- Allow end users to extend Meta data.
- Allow Meta data to be imported / exported into from other standard and formats.

1.3 OLTP: The online operational Database System that performs online transaction and query processing is called **on – Line transaction Processing (OLTP) systems.** Ex. "Day to day" operations of organizations, such as purchasing, inventory, manufacturing, banking, payroll registration, and accounting [5].

1.4 OLAP: Data warehouse serve users and knowledge in the role of data analysis and decision making. Such systems can organize & present data in various formats as per needs of different users. These systems are known as **on – line analytical processing (OLAP) systems [5].**

1.5 Data Warehouse: According to Barry Devlin, a single, complete and consistent store of data obtained from a variety of different sources made available to end users in a way they can understand and use in a business context. [1]

2. Research Methodology
2.1 Data Warehouse Usage

- Data warehouses and data marts are used in a wide range of applications.
- Business executives use the data in data warehouses and data marts to perform data analysis and make strategic decisions.
- In many areas, data warehouses are used as an integral part for enterprise management.
- The data warehouse is mainly used for generating reports and answering predefined queries.
- It is used to analyze summarized and detailed data, where the results are presented in the form of reports and charts.
- Later, the data warehouse is used for strategic purposes, performing multidimensional analysis and sophisticated operations.
- Finally, the data warehouse may be employed for knowledge discovery and strategic decision making using data mining tools.
- In this context, the tools for data warehousing can he categorized into *access and retrieval tools, database reporting tools, data analysis tools, and data mining tools [3]*.

2.2 Need for Data Warehousing:
- Industry has huge amount of operational data.
- It is a platform for consolidated historical data for analysis.
- It stores data of good quality so that knowledge worker can make correct / strategic decisions.
- Better business intelligence for end-users.
- Reduction in time to locate, access, and analyze information.
- Consolidation of disparate information sources.
- Strategic advantage over competitors.
- Faster time-to-market for products and services.
- Replacement of older, less-responsive decision support systems.
- Reduction in demand on IS to generate reports.[3]

2.3 The Process of Data Warehouse Design [5]:
A DWH can be built using a Top – Down Approach, Bottom - Up Approach or combination of both.
Top – Down Approach:
The top – down approach start with the overall design and planning. It is useful in cases where the technology is mature and well known, and the business problems that must be solved are clear and well understood.
Bottom – Up Approach:
The bottom up approach starts with experiments and prototypes. This is useful in the early stage of business modeling and technology development. It allows organizations to move forward at considerably less expense and to evaluate the benefits of the technology before making significant commitments.
Combined Approach:
In the combined approach, an organization can exploit the planned and strategic nature if the Top – down approach

while retaining the rapid implementation and opportunistic application of Bottom up approach.
The design and construction of data warehouse may consist of the following steps:
- ✓ Planning
- ✓ Requirements study
- ✓ Problem Analysis
- ✓ Warehouse Design
- ✓ Data integration
- ✓ Testing and
- ✓ Deployment of the data warehouse.

2.3 Software systems can be developed using two methodologies:

Waterfall method and Spiral method:
The waterfall method performs a structured and systematic analysis at each step before proceeding to the next, which is like a waterfall, falling from one step to next.
The spiral method involves the rapid generation of increasingly functional systems, with short intervals between successive releases. This is considered a good choice for data warehouse development, like for data marts, because the turnaround time is short, modifications can be quickly, and new designs and technologies can be adapted in a timely manner.
So, the warehouse Design process consists of following steps:
1.] Business process model:
For example, orders, invoices, shipments, inventory, account administration and sales. If the business process us organizational and involve multiple complex object collections, a data warehouse model should be followed.
2.] Choose the grain of the business process:
The grain is the fundamental atomic level of data to be represented in the fact table for this process, for example, individual truncations, individual daily snapshots, and so on.
3.] Choose the dimensions:
That will apply to each fact table record. Typical dimension are time, item, customer, supplier, warehouse, transaction type, and status.
4.] Choose the measures:
That will populate each fact table record. Typical measures are numeric additive quantities like dollars sold and units sold [5].

2.4 The major feature between OLTP & OLAP [5]:
1. **User and system orientation:**
   Clerks, clients and information technology professionals use OLTP systems and it is customer_ oriented whereas the OLAP is market-oriented used by knowledge workers, analysis and managers.
2. **Data contents:**
   An OLTP system manages current data and is not used for decision making purposes. An OLAP manages large amounts of historical data with facilities for Summerton and aggregation.
3. **Database design:**
   An OLTP system usually adopts an Entity-relationship model and an application oriented

database design. An OLAP system uses a star or a snowflake model.

4. **View:**

An OLTP system works on the current data within organization, without using historical data in different organizations. An OLAP systems deal with information that originates from different organizations, adding information from many data stores. Because of their huge volume, OLAP data are stored on multiple storage media.

5. **Access Patterns:**

The Access Patterns of an OLTP system consist of Short, atomic transactions. Such a system requires concurrency control and recovery mechanisms. Accesses to OLAP systems are read only operations.

2.5 OLAP

- Logically, OLAP servers present business users with multidimensional data from DW or data marts, without concerns regarding how or where the data are stored.

- So an OLAP Server is a high capacity, multi user data manipulation engine specifically designed to support and operate on multi-dimensional data structure.

- However, physical architecture and implementation of OLAP servers must consider data storage issues.

- OLAP stands for On-Line Analytical Processing.

- The key feature is "Multidimensional", the ability to analyze metrics in different dimensions such as time, geography, gender, product, etc.

- Implementations of a warehouse server for OLAP processing include the following

- MOLAP (Multidimensional OLAP) and ROLAP (Relational OLAP) and Hybrid OLAP (HOLAP) [5].

2.6 OLAP Types:

1. Relational OLAP (ROLAP) servers:

- These are intermediate servers that stand in between a relational back-end server and client front-end tools.

- They use a *relational or extended-relational DBMS* to store and manage warehouse data, and OLAP middleware to support missing pieces.

- ROLAP servers include optimization for each DBMS back end, implementation of aggregation navigation logic, and additional tools and services.

- ROLAP technology tends to have greater scalability than MOLAP technology.

- The DSS server of Micro strategy, for example, adopts the ROLAP approach [3].

Advantages:

- Can handle large amounts of data: The data size limitation of ROLAP technology is depends on data size of the underlying RDBMS. So, ROLAP itself places no limitation on data amount.

- Can leverage functionalities inherent in the relational database: RDBMS already comes with a lot of functionalities. So ROLAP technologies, (works on top of the RDBMS) can control these functionalities.

Disadvantages:

- **Performance can be slow**: Each ROLAP report is a SQL query (or multiple SQL queries) in the relational database, the query time can be long if the underlying data size is large.

- **Limited by SQL functionalities**: ROLAP technology relies on generating SQL statements to query the relational database, and SQL statements do not fit all needs.

- ROLAP technologies are limited by what SQL can do. ROLAP vendors moderate this risk by building the tool with ability to allow users to define their own functions.

2. Multidimensional OLAP (MOLAP) servers:

- These servers support multidimensional views of data through array-based multidimensional storage engines. They map multidimensional views directly to data cube array structures.

- Advantage of using data cube is that it allows fast indexing to pre-computed summarized data.

- With multidimensional data stores, storage utilization may be low if data set is sparse.

- In such cases, sparse matrix compression techniques should be explored. Many MOLAP servers adopt a 2-level storage representation to handle dense and sparse data sets: denser sub cubes are identified and stored as array structures, whereas sparse sub cubes employ compression technology for efficient storage utilization [3].

Advantages:

- **Excellent Performance**: A MOLAP cube is built for fast data retrieval, and is optimal for Slicing and Dicing operations.

- **Can perform complex calculations**: All calculations have been pre-generated when the cube is created. Hence, complex calculations are not only feasible, but they return quickly.

Disadvantages:

- **Limited in the amount of data it can handle**: Because all calculations are performed when the cube is built, it is not possible to include a large amount of data in the cube itself.

- The data in cube cannot be derived from a large amount of data. Indeed, this is possible.

- Only summary-level information will be included in the cube itself.

- **Requires additional investment**: Cube technology is often proprietary and does not

already exist in the organization. Therefore, to adopt MOLAP technology, chances are additional investments in human and capital resources are needed.

3. Hybrid OLAP (HOLAP) servers:

- The hybrid OLAP approach combines ROLAP and MOLAP technology, benefiting from the greater scalability of ROLAP and the faster computation of MOLAP.
- For example, a HOLAP server may allow large volumes of detail data to be stored in a relational database, while aggregations are kept in a separate MOLAP store. The Microsoft SQL Server 2000 supports a hybrid OLAP server.
- HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP.
- For summary-type information, HOLAP leverages cube technology for faster performance.
- When detail information is needed, HOLAP can "drill through" from the cube into the underlying relational data [3].

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. [7]
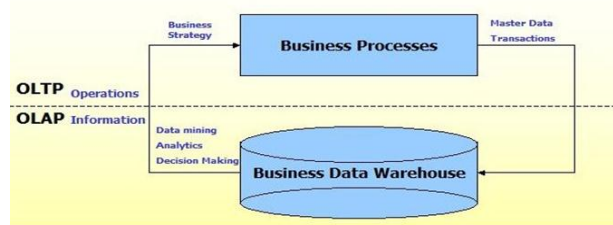

Figure 2: The above diagram shows the OLTP and OALP operations performance and how is difference both with Business processes and Business Data Warehouse [7].

3.1 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).

3.2 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 (usually star schema). The following table summarizes the major differences between OLTP and OLAP system design [7].

Concepts	OLTP for Operation system	OLAP for Data Warehouse
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	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
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Backup and Recovery	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

4. CONCLUSION
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.

5. REFERENCES

BOOKS:
[1] "The Story So Far". 2002-04-15. Retrieved 2008-09-21.

[2] Data Warehousing in the Real World;
Sam Anahory & Dennis Murray; 1997, Pearson

[3] Data Warehousing, Data Mining and OLTP;
Alex Berson, Stephen J. Smith 1997, McGraw Hill
Page: 4, 14-16.

[4] Data warehousing System; Mallach; 2000, McGraw

[5] Data Mining – Concepts & Techniques; Jiawei Han & Micheline Kamber, Jain Pei, Morgan Kaufmann publication; Third Edition; Page: 125 to 130, 151,152, 164, 165

[6] Metadata Standards and Metadata Registries: An Overview" (PDF). Retrieved 2011-12-23.

[7] http://datawarehouse4u.info/OLTP-vs-OLAP.html

[8] An Introduction to Database Systems; C. J. Date,
A. Kannnan, S. Swaynathan

WEBSITES:
[8] http://en.wikipedia.org.

[9] Google Scholar. http://scholar.google.com/scholar.

[10] My blog: www.aparekh5888.blogspot.com

Author Profile

Parekh Arpit received the Msc.it and PGDCA degrees in Information Technology from Bhavnagar University and Computer Application from Saurashtra University in 2011 and 2009, respectively. Works at shree swaminaranya Naimisharanaya College of Management and IT.