An Analysis on OLAP in Data Warehouse

1st Bhavesh L Dhandhukiya
[1]Assistant Professor
[1]Shree Swaminarayan College of Computer Science, Bhavnagar, India

2nd Nirav K Dave
[2]Assistant Professor
[2]Shree Swaminarayan College of Computer Science, Bhavnagar, India

3rd Sagar B Vala
[3]Assistant Professor
[3]Shree Swaminarayan College of Computer Science, Bhavnagar, India

Abstract – Online Analytical Processing is a tool for building a data warehouse using data repository and Entity relationship modelling techniques. It provides various models for processing data and analyzing data for multidimensional techniques. It also used for multidimensional analysis with complex calculations.

Index Terms –OLAP, MOLAP, ROLAP, OLTP, DOLAP, E-R Modeling

I. Introduction of OLAP

In today's data ware housing environment, with such tremendous progress in analysis tools from various vendors, you cannot have a data warehouse without OLAP. It is unthinkable. Therefore, throughout this chapter, look out for the important topics. First, you have to perceive what OLAP is and why it is absolutely essential. This will help you to better understand the features and functions of OLAP. We will discuss the major features and functions so that your grasp of OLAP may be firmed up. There are two major models for OLAP. You should know which model is most suitable for your computing and user environments. We will highlight the significance of each model, learn how to implement OLAP in your data warehouse environment, investigate OLAP tools, and find out how to evaluate and get them for your users. Finally, we will discuss the implementation steps for OLAP.

II. DEMAND FOR ONLINE ANALYTICAL PROCESSING

The top-down and bottom-up approaches for building a data warehouse. In the top-down approach, you build the overall corporate-wide data repository using the entity-relationship (E-R) modelling technique. This enterprise wide data warehouse feeds the departmental data marts that are designed using the dimensional modelling technique. In the bottom-up approach, you build several data marts using the dimensional modelling technique and the collection of these data marts forms the data warehouse environment for your company. Each of these two approaches has its advantages and shortcomings. You also learned about a practical approach to building a conglomeration of super marts with conformed and standardized data content. While adopting this approach, first you plan and define the requirements at the corporate level, build the infrastructure for the complete warehouse, and then implement one supermart at a time in a priority sequence. The supermarts are designed using the dimensional modelling technique. As we have seen, a data warehouse is meant for performing substantial analysis using the available data. The analysis leads to strategic decisions that are the major reasons for building data warehouses in the first place. For performing meaningful analysis, data must be cast in a way suitable for analysis of the values of key indicators over time along business dimensions. Data structures designed using the dimensional modelling technique support such analysis. In all the three approaches referred to above, the data marts rest on the dimensional model. Therefore, these data marts must be able to support dimensional analysis. In practice, these data marts seem to be adequate for basic analysis. However, in today's business conditions, we find that users need to go beyond such basic analysis. They must have the capability to perform far more complex analysis in less time. Let us examine how the traditional methods of analysis provided in a data warehouse are not sufficient and perceive what exactly is demanded by the users to stay competitive and to expand.

III. Need for Multidimensional Analysis

Let us quickly review the business model of a large retail operation. If you just look at daily sales, you soon realize that the sales are interrelated to many business dimensions. The daily sales are meaningful only when they are related to the dates of the sales, the products, the distribution channels, the stores, the sales territories, the promotions, and a few more dimensions. Multidimensional views are inherently representative of any business model. Very few models are limited to three dimensions or less. For planning and making strategic decisions, managers and executives' probe into business data through scenarios. For example, they compare actual sales against targets and against sales in prior periods. They examine the breakdown of sales by product, by store, by sales territory, by promotion, and so on.

IV. For effective analysis

your users must have easy methods of performing complex analysis along several business dimensions. They need an environment that presents a multidimensional view of data, providing the foundation for analytical processing through easy and flexible access to information. Decision makers must be able to analyse data along any number of dimensions, at any level of aggregation, with the capability of viewing results in a variety of ways. They must have the ability to drill down and roll up along the hierarchies of every dimension. Without a solid system for true multidimensional analysis, your data warehouse is incomplete. In any analytical system, time is a critical dimension. Hardly any query is executed without having time as one of the dimensions along which analysis is performed. Further, time is a unique dimension because of its sequential nature—November always comes after October. Users monitor performance over time, as for example, performance this month compared to last month, or performance this month compared with performance the same month last year. Another point about the uniqueness of the time dimension is the way in which the hierarchies of the dimension work. A user may look for sales in March and may also look for sales for the first four months of the year. In the second query for sales for the first four months, the implied hierarchy at the next higher level is an aggregation considering the sequential nature of time. No user looks for sales of the first four stores or the last three stores. There is no implied sequence in the store dimension. True analytical system must recognize the sequential nature of time.

V. Fast Access and Powerful Calculations

Whether a user's request is for monthly sales of all products along all geographical regions or for year-to-date sales in a region for a single product, the query and analysis system must have consistent response times. Users must not be penalized for the complexity of their analysis. Both the size of the effort to formulate a query or the amount of time to receive the result sets must be consistent irrespective of the query types. Let us take an example to understand how speed of the analysis process matters to users. Imagine a business analyst looking for reasons why profitability dipped sharply in the recent months in the entire enterprise. The analyst starts this analysis by querying for the overall sales for the last five months for the entire company, broken down by individual months. The analyst notices that although the sales do not show a drop, there is a sharp reduction in profitability for the last three months. The analysis proceeds further when the analyst wants to find out which countries show reductions. The analyst requests a breakdown of sales by major worldwide regions and notes that the European region is responsible for the reduction in profitability. Now the analyst senses that clues are becoming more pronounced and looks for a breakdown of the European sales by individual countries. The analyst finds that the profitability has increased for a few countries, decreased sharply for some other countries, and been stable for the rest. At this point, the analyst introduces. Another dimension into the analysis. Now the analyst wants the breakdown of profitability for the European countries by country, month, and product. This step brings the analyst closer to the reason for the decline in the profitability. The analyst observes that the countries in the European Union (EU) show very sharp declines in profitability for the last two months. Further queries reveal that manufacturing and other direct costs remain at the usual levels but the indirect costs have shot up. The analyst is now able to determine that the decline is due to the additional tax levies on some products in the EU. The analyst has also determined the exact effect of the levies so far. Strategic decisions follow on how to deal with the decline in profitability.

VI. OLAP is the Answer

Users certainly need the ability to perform multidimensional analysis with complex calculations, but we find that the traditional tools of report writers, query products, spread sheets, and language interfaces are distressfully inadequate. What is the answer? Clearly, the tools being used in the OLTP and basic data warehouse environments do not match up to the task. We need different set of tools and products that are specifically meant for serious analysis. We need OLAP in the data warehouse. In this chapter, we will thoroughly examine the various aspects of OLAP. We will come up with formal definitions and detailed characteristics. We will highlight all the features and functions. We will explore the different OLAP models. But now that you have an initial appreciation for OLAP, let us list the basic virtues of OLAP to justify our proposition.

- Enables analysts, executives, and managers to gain useful insights from the presentation of data.
- Can reorganize metrics along several dimensions and allow data to be viewed from different perspectives.
- Supports multidimensional analysis. Is able to drill down or roll up within each dimension.
- Is capable of applying mathematical formulas and calculations to measures.
- Provides fast response, facilitating speed-of-thought analysis.
- Complements the use of other information delivery techniques such as data mining.
- Improves the comprehension of result sets through visual presentations using graphs and charts.
- Can be implemented on the Web.
- Designed for highly interactive analysis

Guidelines for an OLAP system:

VII. OLAP Characteristics

Let us summarize in simple terms what we have covered so far. We explored why the business users absolutely need online analytical processing. We examined why the other methods of information delivery do not satisfy the requirements for multidimensional analysis with powerful calculations and fast access. We discussed how OLAP is the answer to satisfy these requirements. We reviewed the definitions and authoritative guidelines for the OLAP system. Before we get into a more detailed discussion of the major features of OLAP systems, let us list the most fundamental characteristics in plain language. OLAP system.

- let business users have a multidimensional and logical view of the data in the data warehouse,
- facilitate interactive query and complex analysis for the users,
- allow users to drill down for greater details or roll up for aggregations of metrics along a single business dimension or across multiple dimensions,
- provide ability to perform intricate calculations and comparisons,
- present results in a number of meaningful ways, including charts and graphs.

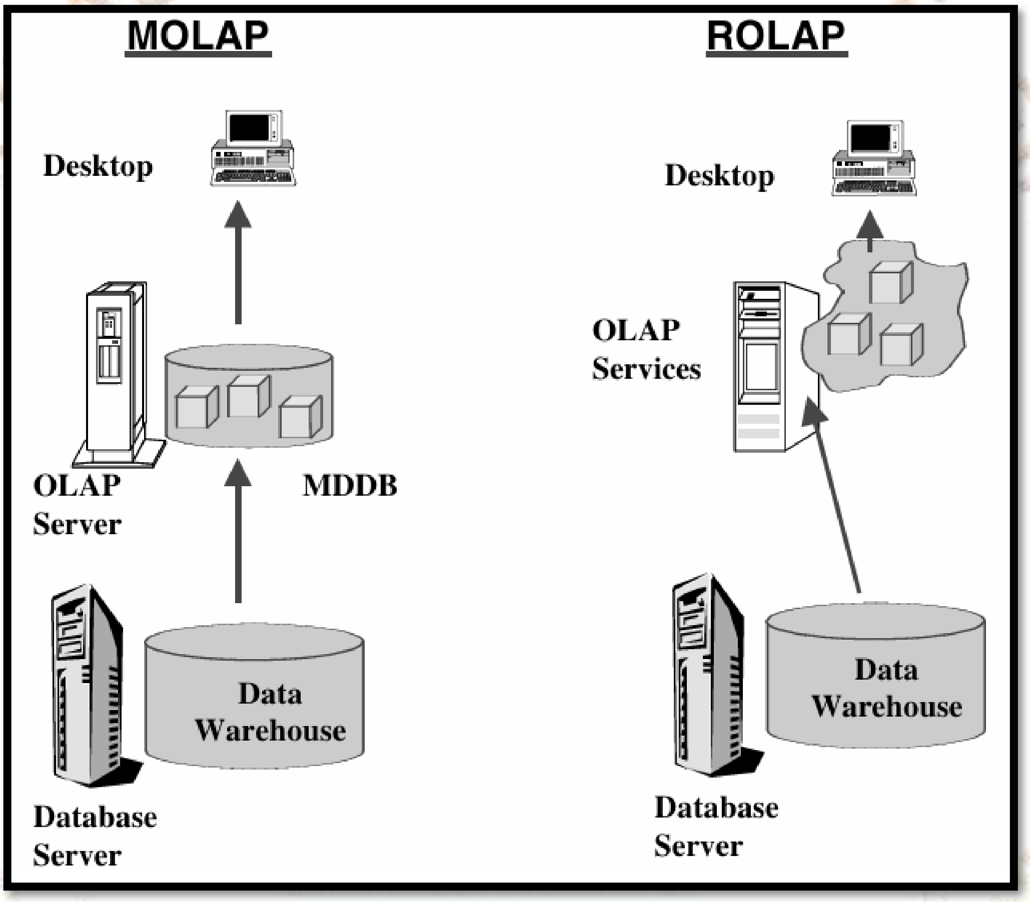
Uses and Benefits

After exploring the features of OLAP in sufficient detail, you must have already deduced the enormous benefits of OLAP. We have discussed multidimensional analysis as provided in OLAP systems. The ability to perform multidimensional analysis with complex queries sometimes also entails complex calculations. Let us summarize the benefits of OLAP systems:

- Increased productivity of business managers, executives, and analysts
- Inherent flexibility of OLAP systems means that users may be self-sufficient in running their own analysis without IT assistance
- Benefit for IT developers because using software specifically designed for the system development results in faster delivery of applications
- Self-sufficiency of users, resulting in reduction in backlog
- Faster delivery of applications following from the previous benefits
- More efficient operations through reducing time on query executions and in network traffic
- Ability to model real-world challenges with business metrics and dimension

VIII. OLAP MODELS

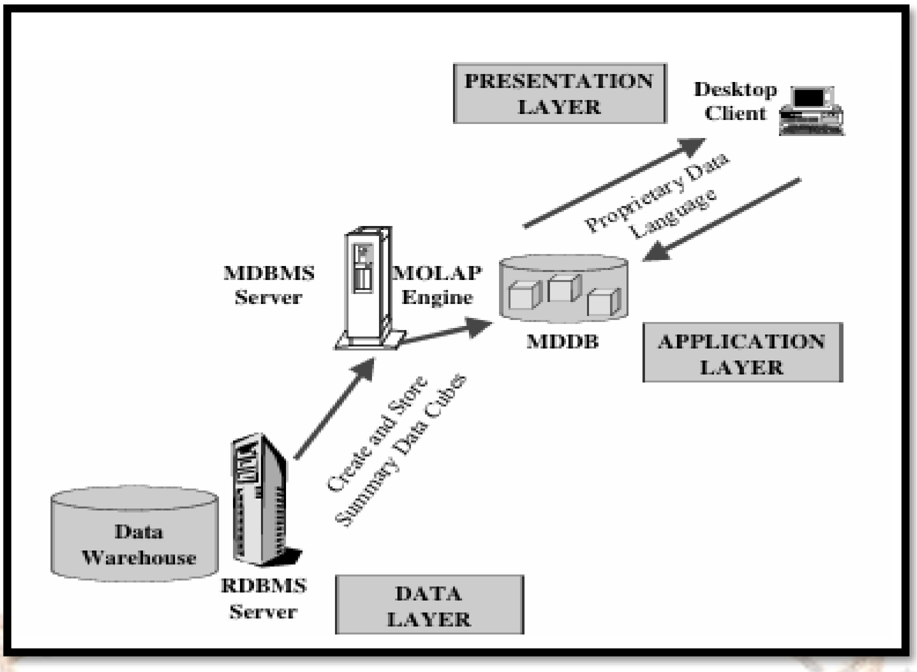
Have you heard of the terms ROLAP or MOLAP? There is another variation, DOLAP. A very simple explanation of the variations relates to the way data is stored for OLAP. The processing is still online analytical processing, only the storage methodology is different. ROLAP stands for relational online analytical processing and MOLAP stands for multidimensional online analytical processing. In either case, the information interface is still OLAP. DOLAP stands for desktop online analytical processing. DOLAP is meant to provide portability to users of online analytical processing. In the DOLAP methodology, multidimensional datasets are created and transferred to the desktop machine, re quiring only the DOLAP software to exist on that machine. DOLAP is a variation of ROLAP.



IX. THE MOLAP MODEL

The MOLAP model, data for analysis is stored in specialized multidimensional databases. Large multidimensional arrays form the storage structures. For example, to store sales number of 500 units for product Product A, in month number 2001/01, in store StoreS1, under distributing channel Channel05, the sales number of 500 is stored in an array represented by the values. The array values indicate the location of the cells. These cells are intersections of the values of dimension attributes. If you note how the cells are formed, you will realize that not all cells have values of metrics. If a store is closed on Sundays, then the cells representing Sundays will all be nulls. Let us now consider the architecture for the MOLAP model. Please go over each part of Figure. Note the three layers in the multitier architecture. Precalculated and prefabricated multidimensional data cubes are stored in multidimensional databases. The MOLAP engine in the application layer pushes a multidimensional view of the data from the MDDBs to the users.

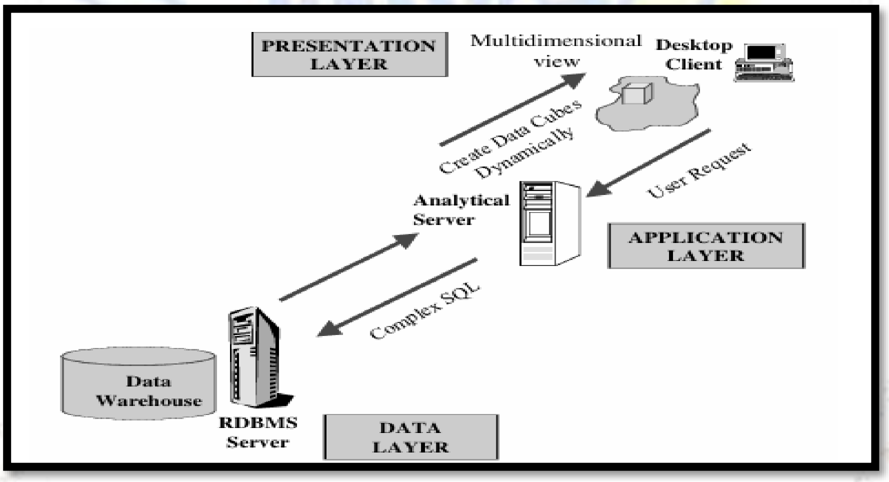
As mentioned earlier, multidimensional database management systems are proprietary software systems. These systems provide the capability to consolidate and fabricate summarized cubes during the process that loads data into the MDDBs from the main data warehouse. The users who need summarized data enjoy fast response times from the pre-consolidated data.



The MOLAP Model

The ROLAP Model In the ROLAP model, data is stored as rows and columns in relational form. This model presents data to the users in the form of business dimensions. In order to hide the storage structure to the user and present data multidimensionally, a semantic layer of metadata is created. The metadata layer supports the mapping of dimensions to the relational tables. Additional metadata supports summarizations and aggregations. You may store the meta data in relational databases. Now see Figure 15-17. This figure shows the architecture of the ROLAP model. What you see is a three-tier architecture. The analytical server in the middle tier application lay er creates multidimensional views on the fly. The multidimensional system at the presentation layer provides a multidimensional view of the data to the users. When the users is sue complex queries based on this multidimensional view, the queries are transformed into complex SQL directed to the relational database. Unlike the MOLAP model, static multidimensional structures are not created and stored. True ROLAP has three distinct characteristics:

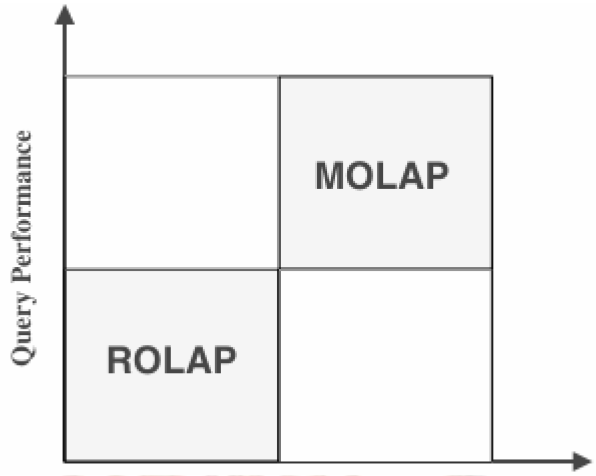
- Supports all the basic OLAP features and functions discussed earlier
- Stores data in a relational form
- Supports some form of aggregation



The ROLAP Model

X. ROLAP VERSUS MOLAP

Should you use the relational approach or the multidimensional approach to provide on line analytical processing for your users? That depends on how important query performance is for your users. Again, the choice between ROLAP and MOLAP also depends on the complexity of the queries from your users. Figure charts the solution options based on the considerations of query performance and complexity of queries. MOLAP is the choice for faster response and more intensive queries. These are just two broad considerations. As part of the technical component of the project team, your perspective on the choice is entirely different from that of the users. Users will get the functionality and benefits of multidimensionality from either model but are more concerned with questions relating to the extent of business data made available for analysis, the acceptability of performance, and the justification of the cost. Let us conclude the discussion on the choice between ROLAP and MOLAP with Figure. This figure compares the two models based on the specific aspects of data storage, technologies, and features. This figure is important, for it pulls everything together and presents a balanced case.



Complexity of Analysis

Data Storage Data stored as relational tables in the warehouse. ROLAP Detailed and light summary data available. Very large data volumes. All data access from the warehouse storage. Underlying Technologies Functions and Features Use of complex SQL to fetch data from warehouse. ROLAP engine in analytical server creates data cubes on the fly. Multidimensional views by presentation layer. Known environment and availability of many tools. Data stored as relational tables in the warehouse. MOLAP Various summary data kept in proprietary databases (MDDBs) Moderate data volumes. Creation of pre-fabricated data cubes by MOLAP engine. Propriety technology to store multidimensional views in arrays, not tables. High speed matrix data retrieval. Summary data access from MDDB, detailed data access from warehouse. Sparse matrix technology to manage data sparsity in summaries. Figure ROLAP versus MOLAP. OLAP IMPLEMENTATION CONSIDERATIONS Limitations on complex analysis functions. Drill-through to lowest level easier. Drill-across not always easy. engine. Proprietary Faster access. Large library of functions for complex calculations. Easy analysis irrespective of the number of dimensions. Extensive drill-down and slice-and-dice capabilities.

XI. REFERENCES

[1] Dr. B. Shadaksharappa: Data Warehousing and Data Mining.
[2] Davin Taniar: Data Warehousing and Analytics.
[3] Ralph kimbal . The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling