THE USE OF ONLINE ANALYTICAL PROCESSING (OLAP) FOR BUSINESS INTELLIGENCE

İbrahim Halil Seyrek
University of Gaziantep, Department of Business Administration

ABSTRACT

The extensive use of information technology enables organizations to collect huge amounts of data about almost every aspect of their businesses. This large amount of data has a potential to provide valuable information to organizations so that they can maintain and improve their competitive position. Today, the problem for organizations is not data collection but extraction of meaning from the data they have collected. Online Analytical Processing (OLAP) is a set of tools to extract useful information from raw data. This paper explains OLAP concepts and techniques and gives examples using a hypothetical business scenario. Suggestions for organizations are also provided.

Key words: OLAP, Business intelligence, Information systems

1. Introduction

Current business environment is very demanding from businesses. It requires organizations to develop new products and services, that is innovate continuously, adapt to changes flexibly and rapidly, be more customer-oriented, and operate in a cost-efficient manner to be competitive. In order to achieve the above mentioned objectives and cope with the challenges of fierce competition, companies need timely, accurate and high quality information. Today organizations are gathering huge amount of data from their suppliers and customers, from their operational activities and processes, and from their environment. These data are collected from internal resources like workers, machines and processes, and from external sources like stock exchanges, Internet, online databases, etc. The data collected using several types of information systems are stored in large databases or data warehouses. Although this vast amount of data has a great potential to be used in deriving useful knowledge, organizations may not have the necessary tools and skills to gain and use that knowledge. Business Intelligence (BI) systems are proposed to answer above mentioned problems by converting raw data into useful knowledge and thereby helping decision making, improving the strategic thinking and action, and satisfying the needs of organizations. BI systems can achieve these goals by exploring, integrating, aggregating and doing multidimensional analysis of the data coming from different sources and in variety of formats. These systems are used to provide useful and reliable information on several aspects of the business (Olszak and Ziemba, 2007)

Business Intelligence term was first coined in 1989 by Howard Dresner as a set of methods that support sophisticated analytical decision making in order to improve business performance. These systems are supposed to help decision makers in understanding the state of their organization (Buchanan and O'Connell, 2006). Hancock and Toren (2006) define BI as “a set of concepts, methods, and technologies designed to pursue the elusive goal of turning all the widely separated data in an organization into useful information and eventually into knowledge”. BI lets people convert raw data into useful information by providing several types of computer-based tools. These BI tools allow their users to extract, transform and load data for analysis and present the results of these analyses in the form of reports, alerts and scorecards (Davenport, 2006). BI tools are classified in several categories, like dashboards, data marts and data warehouses, data mining, scorecards and Online Analytical Processing (OLAP). In this paper we focus on OLAP.

1. Online Analytical Processing

Almost all kinds of businesses have some kind of information system to handle their daily operations and transactions. These applications which can be a departmental application like accounting and human resources or a complicated application covering all functions of a business like an ERP system are generally called Online Transaction Processing (OLTP) systems. OLTP systems are basically designed to handle daily insertions, deletions and updates of the company data and designed to optimize such update operations and simple queries like information about a customer or an order. Almost all modern OLTP systems are designed and implemented using relational databases and consist of many tables and relations normalized to allow efficient update of the data. Because of their mentioned purpose of use, the underlying structures of OLTP systems are not suitable for answering complex queries, for example a query aggregating data from several aspects of a business. These systems are generally restricted to simple and static reports which are predefined by the developers of the system. Also the execution of a complex query on a live information system may adversely affect its performance. For example, a query operating on millions of records may take too long jeopardizing the normal operation of such systems which typically handle daily business transactions and are crucial to the successful run of the business. Therefore, a different type of system design is needed for efficient execution of complex queries and supporting information requirements of users making analysis-based decisions. Online Analytical Processing (OLAP) is a general term used to refer to such systems that are basically designed to optimize complex queries and reporting operations, but not for data updates. OLAP applications are intended to provide users with more analytic capabilities which are limited in OLTP systems.

OLAP is used in several key areas in business including sales and marketing, financial management, planning and budgeting, performance measurement and quality analysis (Hart and Porter, 2004). OLAP applications are typical examples of BI technology. OLAP applications are developed especially for information rich industries like marketing, finance and consumer goods production to gain competitive advantage in a global marketplace by converting available data into useful information (Beynon and Maad, 2002). In all these different types of industries and business functions, OLAP supports its users by letting them look at their data from different perspectives, performing complex calculations quickly, aggregating data from several sources and doing all these in a user friendly manner, generally providing visual tools. In order to utilize better and appreciate its advantages, one needs to understand the basic terminology and concepts related to OLAP.

1.1. OLAP concepts

Data for OLAP is stored in special databases which have a different structure than OLTP databases. An OLAP database has a special schema suitable for OLAP operations. This schema, which is usually called as star schema, is not fully normalized as in OLTP databases and has a dimensional structure. In this dimensional modeling of the data, there are two basic types of tables which are fact tables and dimension tables. Fact tables are used to store numeric values called measures. Measures are fields like sales amount of a product, unit price for an item or cost of a purchased material which are numeric fields that can be used in calculations. Measures are the values that the user wants to see in summary or in detail during the analysis of the data. Dimension tables, on the other hand, are tables used to view the measures from different angles, like time, product category or sales region. Dimension tables are related to fact tables through some common attribute in a foreign key relationship forming a shape like a star as shown in Figure-1. Dimensions allow users to navigate the data in fact tables from different aspects, letting them either sum up the values or drill down into the detail. Because of the multi-dimensional structure of OLAP schema, in OLAP databases there is a special storage structure which is called a cube representing multidimensionality. Cubes are basically a combination of fact and dimension tables. Although cubes in geometry are three dimensional, OLAP cubes can have several dimensions. But the term cube is used symbolically to differentiate this multidimensional structure from two dimensional row-column structures of database tables and spreadsheets.

The physical storage of OLAP databases is also different from OLTP databases. There are three different approaches. The first is relational OLAP (ROLAP) which uses a relational database engine to store data. Multidimensional OLAP (MOLAP) method, on the other hand, doesn't use relational engine but uses a distinct structure for dimensional modelling of the data. Finally, Hybrid OLAP (HOLAP) uses a combination of both techniques to store OLAP data (Riordan, 2005). Regardless the types of physical implementation, OLAP systems are expected to give fast responses to complex queries involving millions of records and a lot of calculation. In order to give a fast response to such queries, OLAP systems make calculations and store the values before the queries are executed. Therefore, during query execution, the results are unexpectedly fast considering the number of records included in the calculations.

OLAP users need a query language to retrieve values and make the calculations using several measures from fact tables. The standard language for querying relational databases has been Structured Query Language (SQL). SQL is so common that it is used by virtually any type of database system. Unfortunately, SQL is not suited for answering problems that are supposed to be solved by OLAP. Therefore OLAP tools have their own query languages. One such language is MDX, (Multidimensional Expressions). MDX is developed by Microsoft and now adopted by several other vendors in their OLAP products. MDX is standardized and has become a widely accepted OLAP query language. Although MDX is similar to SQL in some respects, it has special clauses and features making it easy to make multidimensional view and querying of the underlying data. MDX is of course very important for OLAP developers, but for users of OLAP applications the user interface of the applications usually have visual tools so that the users do not need to write the specific MDX queries. The application writes the query under the hood freeing the user from the complexities of writing an MDX query. On the other hand, more experienced users may benefit from learning this query language to perform customized and complex analyses.

2. An Example OLAP Application

In this part of the paper, in order to clarify some of the OLAP concepts and how it can be used in a typical business, a sample OLAP application will be discussed for a fictitious company, which we call as company X. We will assume that Company X is a manufacturing company, producing several types of candies and chocolates. In a business, like company X, there are several functions like purchasing, inventory management, production planning and scheduling, production, sales and marketing, distribution of products, accounting and financial management. For all these functions and processes covering more than one function, there are several questions managers need to answer. For example, a manager may want to know how much of some type of raw material is purchased in the past for several time periods to see seasonal fluctuations and make better forecasts of future purchases. Another manager may need to look at the sales of different product groups across different store types, different sales regions and different time periods. Also, how much revenue gathered from the sales of different products can be of interest. Accounting department may want to find out the relative effects of several factors on correctly determining the costs of goods sold. Another concern can be determining the relative profitability of several products and based on that deciding whether to increase or decrease the production of some products.

Questions like these are very important for any business and can be difficult to answer without a system like OLAP. For company X, we will give possible designs of two cubes; one for purchasing and another for sales to answer some questions of interest similar to above ones.

2.1. Purchasing cube

As a manufacturing company, company X buys several raw materials from different suppliers to be used in the production of candies and chocolates. The purchasing cube is expected to support queries regarding the purchasing of these raw materials. The cube basically consists of a purchasing fact table and dimension tables about materials, suppliers and time. Purchasing fact table consists of fields like the amount of material purchased and the unit price of the material. The table also contains key fields to connect rows in the fact tables to rows in the dimension tables. In the fact table of purchasing cube, key fields are material ID, supplier ID and the date of purchase. These fields also exist in the corresponding dimension tables. Dimension tables let the user look at the measured and calculated data through different angles. In the purchasing cube, the material dimension contains information about the raw material purchased by the company. It contains a key material ID field to uniquely define the purchased material and also join that dimension to the fact table. Also there are other fields describing the raw material like its name, unit of measure, the category of raw materials it belongs to, etc. The raw material can be categorized based on the several factors. This category information is put into some other table (e.g. material category table) and the raw materials in the material dimension table are linked to that category table through a foreign key field in the dimension table. This enables the user of the cube to browse the fact data based on different raw materials and different raw material categories. The supplier dimension provides information about the supplier from which the company acquires the materials it needs for the production. Other than analyzing purchasing data based on individual suppliers, the cube may provide information about different supplier types. Suppliers may be classified based on several factors like suppliers from different cities and countries if there are any or any other custom criteria that the company uses to classify suppliers. Finally time dimension lets the user of the purchasing cube to explore the data on different time periods providing a historical analysis of the purchasing information. The time dimension can be used to summarize data on daily, weekly, monthly, quarterly or yearly basis. A sample look at the purchasing data can be seen on Figure-2. This figure shows the values of the cube dimensions in a hierarchical tree structure and the values of measures in a table. This is a very simplistic view of the purchasing cube due to space limitations. In a real application, the user can customize the view of the data by changing the places of several dimensions, putting them in rows or columns. Also the user may increase or decrease the level of detail the table shows by expanding or contracting the values in the table.

2.2. Sales cube

The sales cube for company X consists of a sales fact table containing sales related measures and some dimensions to view the summary and detail data on these dimensions. The fact table contains sales quantity, sales amount, discount percentage for the item sold, cost of goods sold and profit amount as measures. It also has some key attributes to relate to dimension tables. The cube has product, store and time as dimensions. The product dimension table contains information about the products that the company sells to its customers. It has fields like product ID and product name and it has a relation to a product family table to allow analysis based on product groups. These product groups can be formed based on the several custom criteria defined by the users allowing the creation of several custom product families. Company X sells its products to stores of different types, so customers of the company are stores and so customers are represented in store dimension. The store dimension table has basic attributes about the store, like store ID, store name and store type, and connected to a region table to determine the sales region of the store and to allow further analysis based on sales regions. Finally, time dimension table holds information about the date the sales is done and related time information like the fiscal period information. Based on the data in time dimension, sales figures on a monthly basis or quarterly basis can be calculated easily. The dimension tables are related to fact table through common fields between fact and dimension tables. The data in dimension tables can be used to define hierarchies which allow looking at fact data at several levels. For example, product hierarchy can be used to look at the sales data at the product or product family level. In a similar fashion, a store hierarchy lets the user to navigate and drill down sales data at different region level and specific store level. Figure-3 shows a sample look at the sales cube.

2.3. Other cubes

In addition to purchasing and sales cubes mentioned above, other type of cubes can be created for analysis purposes. For example, an inventory cube can be created to make inventory analysis, to understand the material flow into the company and to get a clear picture of stock levels. Another potential cube which can be useful for managers is an accounting cube. Accounting cube can be used to look at the accounting data by different fiscal years. The cube can contain several measures related to accounting, such as assets, liabilities, and income data. These measures can be viewed through several dimensions like an account dimension or a time dimension. A summary report can be easily generated to grasp the general situation of the company or the user can drill down into specific details. Accounting cube can also let managers do several types of custom analyses thereby enabling management accounting.

3. Conclusion

In any business and especially those operating in information rich and turbulent markets, managers need timely, accurate and high quality information for effective decision making. Business Intelligence is an umbrella term that refers to a set of technologies to help such effective decision making. Business Intelligence tools help managers in understanding the current situation of their business, in determining the trends in the market segment they are operating, in identifying the possible causes of the results they got, and in making several other types of analyses. In that respect OLAP as a BI technology can greatly help in analyzing and understanding the data about several aspects of a business. Therefore managers should seriously consider utilizing OLAP in their decision making processes. Whether as part of a larger application or a standalone product, managers should have these applications in their toolbox.

Although OLAP can offer a lot of advantages, managers should take into account several factors in getting and using OLAP applications. First of all, like any other information system OLAP applications should be designed considering the needs of the users they have been developed for. They should be tailored to the needs and peculiarities of the user organization. There are several vendors trying to sell their products and usually exaggerating their capabilities and showing these products as if they are the cure for all the problems of a business. Organizations need to be aware of the unrealistic claims of the vendors. Unfortunately, experience tells us that many IS implementations, especially those that are big in size and complexity, have a bad reputation of being money drains for companies, not yielding the benefits its proponents are claiming. So, businesses need to be realistic about their expectations and should negotiate with the vendors accordingly. If the OLAP solution will be developed in-house, for example by IT personnel of the company, this project should be managed carefully. Like any other IS project, BI projects should have clear business objectives and managed properly in order to realize expected benefits. Users should be involved in the project

Finally, we should not forget that BI tools are not themselves solutions for our complicated business problems. We need people to use them effectively who will make better decisions with the help of these tools. An interest towards OLAP applications should be created in the users and they should be conceived about the benefits of using these applications. Many studies done on the user's acceptance of information systems show that one of the most important factors in user's acceptance of information systems products is the level of user's perception that the system will be useful for them and enhance their job performance. Therefore, the benefits of OLAP applications and advantages they will provide for their users should be clearly explained. Also, users need to be trained for effectively using these tools. Although most OLAP tools have user friendly interfaces and visual tools like drag and drop capabilities for making complex calculations and browsing data at several levels of detail, users still need some level of training in order to proficiently use such applications. Any tool can be useful only if it is adopted and used by the users.

References

Beynon, M. and S. Maad (2002), "Empirical Modelling of Real Life Financial Systems: The Need for Integration of Enabling Tools and Technologies", *Transactions of the SDPS*, Vol. 6, No.1, pp.43-58.

Buchanan, Leigh and Andrew O'Connell (2006), "A Brief History of Decision Making", *Harvard Business Review*, Jan. 2006, pp.32-41.

Davenport, Tomas H. (2006), "Competing on Analytics", *Harvard Business Review,* Jan. 2006, pp.99-107.

Hancock, John C. and Roger Toren (2006), Practical Business Intelligence with SQL Server 2005, NJ: Addison Wesley Professional.

Hart, Mike and Gabi Porter (2004), "The Impact of Cognitive and Other Factors on the Perceived Usefulness of OLAP", *The Journal of Computer Information Systems*, Vol. 45, No.1, pp.47-56.

Olszak, Celina M. and Ewa Ziemba (2007), "Approach to Building and Implementing Business Intelligence Systems", *Interdisciplinary Journal of Information, Knowledge, and Management*, Vol. 2, pp. 135-148.

Riordan, Rebecca M. (2005), Designing Effective Database Systems, NJ: Addison Wesley Professional.

Figure-1: Star Schema of A Typical OLAP Cube.

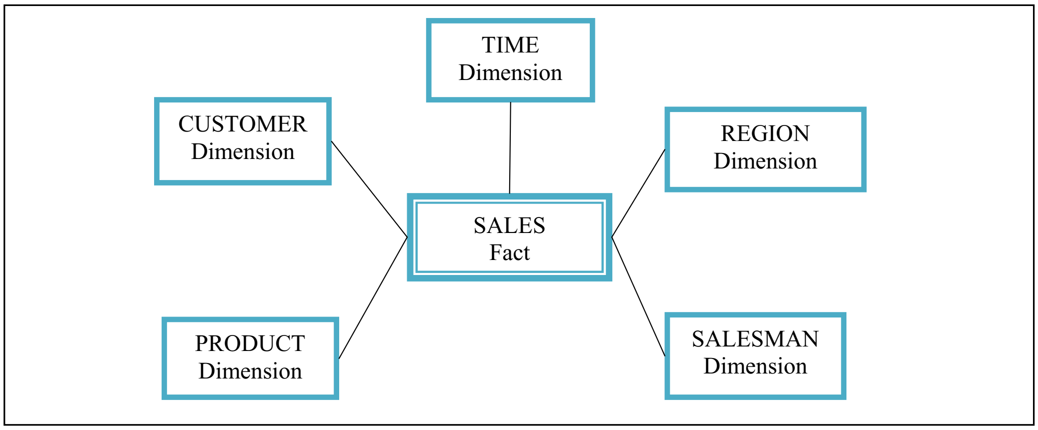


Figure-2 : A Sample View of Purchasing Cube

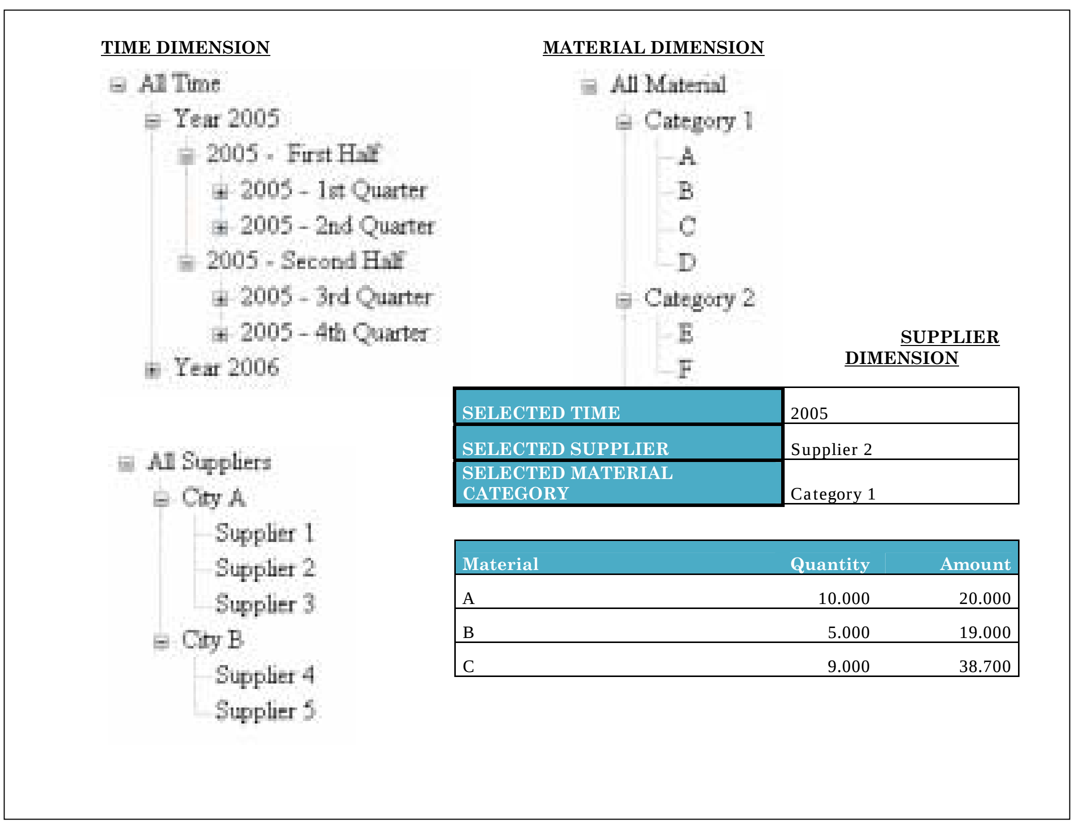
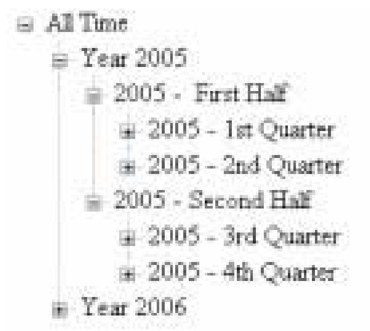
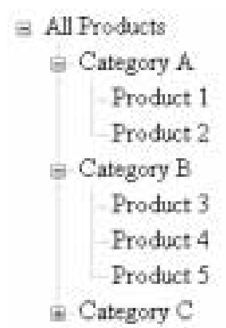


Figure-3 : A Sample View of Sales Cube

TIME DIMENSION



PRODUCT DIMENSION



STORE DIMENSION

SELECTED TIME	2005 -2nd Quarter
SELECTED STORE	Store 3
SELECTED PRODUCT CATEGORY	Category B

Product	Quantity	Amount	Discount
Product 3	25.000	15.000	2.250
Product 4	43.000	17.200	860
Product 5	17.000	13.600	2.720