DIFFERENTIATING THE ROLE OF ONLINE ANALYTICAL PROCESSING IN BUSINESS INTELLIGENCE

Wang, Jiangping
Webster University

Kourik, Janet L.
Webster University

Maher, Peter E.
Webster University

ABSTRACT
Business intelligence (BI) encompasses an environment to capture, integrate, transform, and provide decision support data to end users. Within such a system, online analytical processing (OLAP) enables data from a data warehousing environment to be made available to users in a usable format, thus providing strategic information for decision making. OLAP supports business decision making and business intelligence. In contrast to data warehousing, OLAP provides the channel that connects the online user and online data. Through this channel the user is connected with the information they need to perform various analytical activities including drill down and roll up, slice and dice, and visualizing data in various ways. OLAP tools support many kinds of multidimensional data analyses such as statistical and ratio computation, aggregation, comparison, and forecasting. Interest in OLAP is increasing because it puts more powerful tools online to deliver the right kind information to the right user. This paper describes unique characteristics of OLAP, its role in business intelligence, and value to business. Challenges and future directions will also be discussed.

INTRODUCTION

Business data is a vitally important, yet commonly unrealized asset for organizations. Throughout the years, operational data has often been collected and stored, becoming a potentially invaluable resource to business. However, operational data from transactional systems needs to be transformed into decision-support data before it can be easily used by executives for decision making. The business value in the operational data is, in a way, hidden from use until this transformation.

Business intelligence and analytics often become the competitive differentiators for an enterprise (Liberatore and Luo 2010, Watson and Wixom 2007). It can also greatly enhance enterprise knowledge management and customer relationship management (Cody et al. 2002, Micu et al. 2009). Yet, businesses find significant barriers to using analytics more effectively. Wailgum suggests that the technology is available but issues involving people interfere with using analytic technology (Wailgum 2010). This position is further supported by KPMG International's estimates that businesses use as little as 20% of the resources and potential within the organization's current analytical tools (KPMG International 2009). It can then be inferred that learning more about BI analytics holds great promise for enterprises.

The great need for business to understand more about business intelligence and analytic techniques is confirmed through research. MIT Sloan Management Review and the IBM Institute for Business Value recently conducted an extensive survey on analytics within organizations. The breadth of the survey is impressive as it was conducted in over 100 countries and across 30 industries. Respondents included analysts, managers and executives. A significant finding was that 38% of businesses identified the primary barrier to using BI more effectively "…as their own lack of understanding about how to use analytics." (LaValle et al. 2010)

Business intelligence is a broad framework that includes both data warehousing and information delivery systems enabling the hosting of decision support data and process analytics. Data warehousing stores decision-support data in specially designed architectures that differ from operational systems. A data warehouse (DW) by definition integrates data from multiple operational systems and is typically organized around a specific subject such as customer. In contrast to operational systems, the data warehouse stores historical data in summary form. Further, once the data are loaded in the data warehouse it remains stable, unlike an operational system in which data continuously changes.

The data warehousing environment provides some decision-support data within architectures such as the star schema. However, a data warehouse can be complemented by sophisticated online tools that offer intensive decision support to end users. In the online tool, data are pre-processed and delivered to end users to more rapidly support decision making, business modeling, and operations research.

THE BUSINESS INTELLIGENCE FRAMEWORK

Data scattered across various operational systems cannot be used directly for supporting effective high level decision making. As described earlier, decision support data are integrated, subject-oriented, time-variant, and nonvolatile. The data, stored in the DW, are snapshots of the business operation at a given point in time and are not subject to direct updates from the transactions in operational systems. Therefore operational data must be integrated and transformed into decision support data before it can be used within the BI systems. To facilitate top management's view of the business the data also has to be aligned with core business subject areas, such as customer, product, and region.

Data warehousing is part of the BI framework. Decision support data are stored in a data warehouse where specially designed schema orientates the data to support further processing for decision making. Extraction, transformation, and loading (ETL) processes transform operational data to decision support data before it is stored in a data warehouse and is ready to be used by system analysts. However, the data in a data warehouse is still not sufficient to be directly used for effective decision support. The barriers range from performance, complexity, hierarchical support and mapping from data to visualization.

Similarly, OLAP is the part of a BI framework that brings decision support data to the end user so that they may access information quickly and easily to support effective data analysis. OLAP plays a unique role by serving as the bridge between users and data. It presents data in front of end user facilitating multidimensional hierarchical analysis (Mansmann and Scholl 2006, Mansmann and Scholl 2007, Vinnik and Mansmann 2006). As such OLAP enables accessing, managing, analyzing, and presenting decision support data from a data warehousing environment to support business intelligence. As major components of the BI framework, both DW and OLAP are described below in greater detail.

THE ROLE OF DW IN STORING DATA

Data in a data warehousing environment cannot be used directly for effective decision making. The typical star schema in data warehouses organizes data around facts, which are the numeric data used to convey the operational performance of a business. These facts, or measures, are placed in fact tables that represent the lowest level of aggregation in data. Additionally, dimension tables provide the categories by which decision makers would like to see or categorize data for analysis. Unique identifiers for each row of the fact table are based on the categories described in the dimension tables. Used together in the fact table, as a composite primary key, they uniquely identify each row of the fact table. In this way, relationships between the fact table and the dimension tables are established.

Data warehouses store high volumes of data representing the integrated and granular historical measures. Having historical data enables time-sensitive analysis and projections to be performed. A significant amount of calculation time would be used when aggregates at any level above the lowest level in each dimension are needed. During such calculations unacceptably low performance occurs. In addition, it is cumbersome to sort out hierarchies whenever calculations along a hierarchical structure, such as drill down and roll up, are needed.

THE ROLE OF OLAP IN ACCESSING DATA

To enable quick and easy access to the business data for better analysis, mappings from data warehouses to OLAP are necessary. OLAP systems focus on objects including cubes, measures, dimensions, hierarchies, and hierarchical levels, which provide mechanisms for accessing and analyzing the data in a data warehouse. In this mapping, data flows from DW objects, such as fact and dimension tables, to OLAP objects, such as cubes, dimensions, and hierarchies, as depicted in Figure 1. The data are transformed during each flow between the objects. Through this mechanism various client tools can be applied by users to perform decision making activities.

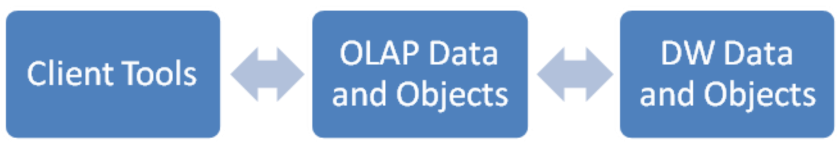

Figure 1. OLAP Maps from Client to DW

OLAP systems enable decision makers to interact with the data quickly and effectively. It provides preprocessed aggregates, where pre-calculated and derived data are stored in aggregate structures in the cube. This organization is supported by an OLAP engine that updates aggregated data periodically. Aggregates at all hierarchical levels along all dimensions are quickly available by joining the facts to hierarchy and level keys, greatly improving the efficiency of the system. In this way, OLAP connects clients to decision support data and maps business measurements, through OLAP cubes and measures, to numerical facts in data warehouse fact tables. It also maps business subject areas through OLAP dimensions and hierarchies to dimension tables in a data warehouse. Further, it maps business descriptive attributes through dimensions and hierarchical attributes to data warehouse dimension table attributes. Figure 2 shows objects in each environment in this mapping.

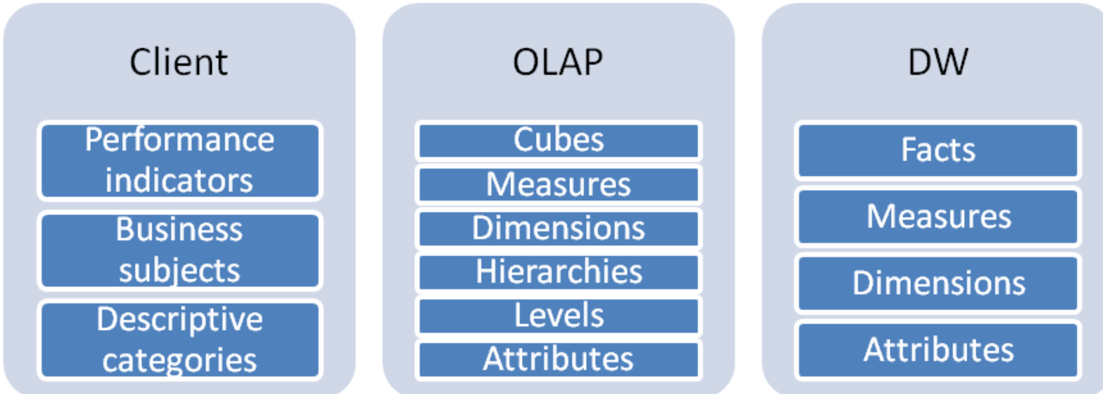

Figure 2. OLAP and DW Objects Mappings

THE ROLE OF OLAP IN MANAGING DATA

OLAP systems manage data around cubes in multidimensional databases. The cube contains measures for each unique combination of the dimension and aggregated values formed by the dimension hierarchies. A dimension may contain one or more hierarchies to assist analysis from different perspectives along the dimension, thus satisfying various needs. For example, the time dimension may have a hierarchy for calendar year and a hierarchy for fiscal year. OLAP manages relationships between cubes and dimensions, and between cubes and hierarchies, forming the edges of the cube so that analysis can be performed along either dimensions or hierarchies.

An OLAP cube contains all levels of data, from the detailed level to the top level. The lowest level of details maintained in the cube determines the lowest level of granularity of analysis that can be performed. For example, along the calendar year hierarchy in the time dimension four levels could be defined for the hierarchy from top to bottom: all years, calendar year, calendar quarter, and month. Data for each level are aggregated and stored in the cube. These aggregates at each level include the values for intersection with each level of other dimension/hierarchies defined, such as country, region, state, in the OLAP system. In this way, any time an aggregation analysis is performed, a reference to the level, or level key, will be sufficient to retrieve the previously summarized data without calculation against raw data in the underlying data warehouse.

OLAP cubes also contain both base and derived data. Base data are mappings from an underlying data warehouse such as data for sales. The derived data are analysis-specific calculated data such as "sales year-to-date" (Ytd) and "year-to-date percentage change previous year" (Ytd % Chg Pr Year), as shown in Figure 3. Therefore, a typical query does not need to perform additional calculations on the fly, since aggregations and derived values have already been calculated and stored within the cube by an OLAP engine.

Geography	Product	Time	Sales	% Chg Pr Period	% Chg Pr Year	Ytd	Ytd % Chg Pr Year
South America	Portable Music and Video	CY2007	$1,668,735	-	-	$1,668,735	-
North America	Portable Music and Video	CY2007	$3,795,936	-	-	$3,795,936	-
Oceania	Portable Music and Video	CY2007	$11,587	-	-	$11,587	-
Europe	Portable Music and Video	CY2007	$2,274,220	-	-	$2,274,220	-
Africa	Portable Music and Video	CY2007	$647,626	-	-	$647,626	-
Asia	Portable Music and Video	CY2007	$6,704,484	-	-	$6,704,484	-
South America	Portable Music and Video	CY2008	$1,777,808	6.5	6.5	$1,777,808	6.5
North America	Portable Music and Video	CY2008	$4,055,191	6.8	6.8	$4,055,191	6.8
Oceania	Portable Music and Video	CY2008	$13,437	15.9	15.9	$13,437	15.9
Europe	Portable Music and Video	CY2008	$2,450,443	7.7	7.7	$2,450,443	7.7
Africa	Portable Music and Video	CY2008	$728,666	12.5	12.5	$728,666	12.5
Asia	Portable Music and Video	CY2008	$7,238,704	7.9	7.9	$7,238,704	7.9
South America	Portable Music and Video	CY2009	$1,978,079	11.2	11.2	$1,978,079	11.2
North America	Portable Music and Video	CY2009	$4,526,968	11.6	11.6	$4,526,968	11.6
Oceania	Portable Music and Video	CY2009	$12,134	-9.7	-9.7	$12,134	-9.7

Figure 3. Base data and Derived Data

OLAP enables quick and easy information retrieval by managing multidimensional data in cubes and hierarchies. This provides the capability of users having online access to the data, and being able to interact with the data. Through such access decision makers can perform analysis, including slicing and dicing, drill down and roll up, with any salient data and can manipulate data in any desired way to reveal information. All of above analysis activities would not be possible if the user does not have online access to the data or information is returned in an unacceptable format.

THE ROLE OF OLAP IN ANALYZING DATA

Although the standard structured query language (SQL) for relational databases provides functions to perform basic aggregation operations, it is weak, from any perspective, in the complex calculations for needed analytics. OLAP needs to support complex calculations, complex analytical processing, and "what-if" analysis. Analytical operations are performed by an OLAP engine that prepares the derived data for storage as a value in the cube.

To support advanced analytical processing, an enhanced family of aggregate and analytic SQL functions have been introduced. OLAP support for an extensive set of analytic functions facilitates responses to complex business queries, thus making analyzing and reporting significantly easier. Some examples of advanced aggregation functions include ROLLUP, CUBE, GROUPING, GROUPING SETS, RANKING, and relative contribution functions, as listed in Table 1.

In addition, an OLAP engine extends SQL's new analytical capabilities and provides even more powerful analytic functions to perform hierarchy navigation and calculations involving ranking, time series, ratio, percentage, and the difference between time periods. The analytic functions enable decision makers to make advanced calculations for comparisons and identify trends. Sophisticated calculations are embedded within the cube to enhance the analytic process. These calculations often involve data from many rows and inter-row calculations. For example, a calculation may compare the current year's sales for each region and product category with sales from the same period in the previous year and two years prior. The OLAP cube structure is designed to accommodate this type of analysis.

Table 1. Examples of Advanced Aggregation Functions

FUNCTION	DESCRIPTION
ROLLUP	Group the selected rows and return a single row summary for each group.
CUBE	Group the selected rows based on the values of all possible combinations and return a single row summary for each group.
GROUPING SETS	Specify multiple groupings of data and prune the unneeded aggregates.
RANK	Calculate the rank of a value in a group of values.
RATIO_TO_REPORT	Compute the ratio of a value to the sum of a set of values.

Drill down and roll up analysis is performed along dimension hierarchical structures, with which decision makers analyze how a particular data value contributes to the whole. For example, if a report shows an unexpected low on the sales for a given product category, a drilldown is needed to examine the detailed numbers within that product category. OLAP systems maintain hierarchy views that encapsulate all of the hierarchical information of the dimension so queries can traverse seamlessly from any level to its parent/ancestor or children/descendants.

THE ROLE OF OLAP IN PRESENTING DATA

To answer complex analytical questions, results from OLAP data need to be presented comprehensively to the end users in meaningful and easily understood ways. At the same time, the values in the result set need to be dynamic, that is permit manipulation, to support further interaction from the users. Dashboards and scorecards are often used to summarize measures. They monitor key aspects of the organization operations showing and monitoring the health of the essential aspects of the business.

The presentation of analytics has to be flexible enough to support interactive report generation, including graphical and chart representations. To support multidimensional analysis, hypercube pivoting allows users to look at the data in the cube from any perspective. To support hierarchical analysis, drill down and roll up allow users to retrieve information at any level along the dimension hierarchies. Multidimensional analysis enables users to focus on specific slices of the cube to perform a more detailed analysis in the operation of slice and dice.

All of above analytic result presentations are online with the data in the OLAP cube. Consequently, users obtain almost instantaneous results when they change the cube’s perspective, or perform drill down to the next level along the dimension. It is generally not acceptable performance if the required chart or report is not displayed within a reasonable period of time.

OLAP supports various client tools in interactive reporting and chart generation. Further, it supports different types of users ranging from novices to power users with straightforward user interfaces, preprocessed reports, complex analysis tools, and even an advanced query capability to enhance the manipulation and analysis of information in the cubes. Other aspects of presenting data to end users include various capabilities in formatting data to enhance the clarity of a report and utilizing visualization tools with constraints to reveal important information that otherwise may not be readily obvious. Reports can be stored with customized filters and selected data ranges for future use. Moreover, OLAP provides a flexible environment for the creation of exploratory analysis report on top of the cube data.

CHALLENGES AND FUTURE DIRECTIONS

To successfully meet the needs mentioned above, OLAP faces future challenges in its design and implementation. Achieving adequate performance in processing data is always a key challenge. Business intelligence relies on decision support data that by its very nature are voluminous. Complex calculations involving high volumes of data adds a heavy burden on computational resources and may reduce acceptable performance. Therefore new approaches will be needed in the future to work efficiently with larger and larger data sets.

In addition to traditional structured data, which can be presented in a tabular format, businesses are accumulating more and more unstructured data. Unstructured data includes components such as images, video clips, and email or online chat messages. Information derived from unstructured data is becoming a critical part of business intelligence. OLAP needs to be more sophisticated in processing, analysis, and presenting unstructured data. The automated analysis of natural language for decision support systems (DDS) is a young discipline. Searching graphical data, without manual tagging, is another area for growth. Users have demanded capabilities on not only presenting data from an historical perspective, but also performing mining on the data to further uncover underlying hidden knowledge for predictive analysis.

End users of OLAP systems have become more sophisticated, requiring even greater capabilities in more complex query, report, and analysis tools. To support complex operations, such as drag-and-drop and dimension pivoting with online data, OLAP systems are becoming more and more heavily dependent on the client software typically installed at the user's location. An OLAP client application, that supports powerful graphics generation and advanced calculation, has to be installed on the client side to access an OLAP server. Heavy client-dependency creates issues in platform compatibility, easy access, and connectivity. Shifting the work to the client is likely to require more powerful hardware and in turn more expense. As the Internet becomes more prevalent in information delivery, the demand for web-enabled OLAP is increasing. In the future, delivering OLAP over the Internet will be the mainstream method of delivery for business intelligence. The convergence of technologies in data warehousing and the Internet thin client architecture will significantly make it easier and more convenient for users but will require further development in thin client software.

Better information delivery remains the most compelling driving force for business decision making. OLAP plays unique and increasingly important roles in business intelligence. Therefore, more attention needs to be given to OLAP, and it should be made more accessible to the end users. OLAP is critical in accessing, managing, analyzing, and presenting decision support data, which affects every aspect of delivering the right information to the right users.

REFERENCES

Cody, W. F., Kreulen, J. T., Krishna, V. and Spangler, W. S. (2002). "The Integration of Business Intelligence and Knowledge Management", IBM Systems Journal, Vol. 41, No. 4, 2002, 697 – 713.

KPMG International (2009). "Survive Downturn with Data Analytics", Journal of Accountancy, February, 2009.

LaValle, S., Hopkins, M., Lesser, E., Shockley, R. and Kruschwitz, N. (2010). "Analytics: The New Path to Value", IBM Institute for Business Value and MIT Sloan Management Review.

Liberatore, M. J. and Luo, W. (2010). "The Analytics Movement: Implications for Operations Research", Interfaces, Vol. 40, No. 4, July–August 2010, pp. 313–324.

Mansmann, S. and Scholl, M. H. (2006). "Extending Visual OLAP for Handling Irregular Dimensional Hierarchies", Proceedings of 8th International Conference on Data Warehousing and Knowledge Discovery, 95 – 105.

Mansmann, S. and Scholl, M. H. (2007). "Exploring OLAP Aggregates with Hierarchical Visualization Techniques", Proceedings of the 2007 ACM Symposium on Applied Computing, 1067 – 1073.

Micu, A, Micu, AE, and Capatina, A. (2009). "Innovative Analysis of a CRM Database using Online Analytical Processing (OLAP) Technique in Value Chain Management Approach", Proceedings of the 10th WSEAS International Conference on Mathematics and Computers in Business and Economics, 305 – 310.

Vinnik, S. and Mansmann, F. (2006). "From Analysis to Interactive Exploration: Building Visual Hierarchies from OLAP Cubes", Proceedings of 10th International Conference on Extending Database Technology, 496 – 514.

Wailgum, T. (2010). "Biggest Barriers to Business Analytics Adoption: People, Not Tech" CIO Magazine, CXO Media, Inc., November 11, 2010.

Watson, H. J. and Wixom, B. H. (2007). "The Current State of Business Intelligence", IEEE Computer, September 2007, 96 – 99.