

Integrating OLAP/SOLAP With E-Business: A New Conceptual Platform

Oualid (Walid) Ben Ali¹ Samar Mouakket¹

¹ College of Business Administration, Department of MIS, University of Sharjah
Sharjah, United Arab Emirates

Email: OAli@Sharjah.ac.ae; samar@Sharjah.ac.ae

Abstract

E-business domains, such as e-commerce, e-government, e-learning, e-banking, have been considered for long time as killer domains for several data analysis techniques in order to analyze the data stored in their websites' databases. Most of the researchers have been focusing on using the Data Mining (DM) technique to analyze E-business websites' databases. The DM technique has shown interesting results, but this technique presents some restrictions concerning the database to be analyzed (structure, volume, etc.) and concerning the level of expertise of the users who will interpret the analysis results. In this paper, we show that E-business domain can also provide all the right ingredients for successful application of other analysis techniques such as On Line Analytical Processing (OLAP). In addition, we propose to use a new and a promising technique called Spatial OLAP technique in order to analyze the spatial components. All our work, presented in this paper, is wrapped up in one conceptual framework which integrates E-business, OLAP and SOLAP together for easier data (non-spatial and spatial) analysis which leads to better decision making.

Keywords: E-business, E-commerce, website database, data-mining, data analysis, OLAP, SOLAP, GIS, decision-making, conceptual framework

1. Introduction

E-business domains (ex. e-commerce, e-government, e-learning, e-banking) are growing fast, and with this growth companies are willing to spend more on improving the online experience by analyzing the data behind their websites. Based on our literature review, most of the existing e-business companies have been using primitive measures to analyze their data; others have been using more sophisticated analysis techniques such as data mining (DM) to generate useful data and knowledge for decision makers [8][9]. The DM technique has shown interesting results, but it presents some restrictions concerning the size and the structure of the data to be analyzed, as well as the level of expertise of the end-users who will be interpreting the analysis results. With the growing pressure to make e-business companies more profitable, additional analysis techniques are usually required to analyze the data behind the website which are becoming more and more complex. In our work we propose two techniques which are OLAP and SOLAP (for spatial data) which will be integrated with e-business in order to analyze complex databases with hybrid data (spatial and non-spatial). Thus, in this paper, we propose a conceptual framework which integrates OLAP and SOLAP analytical techniques with e-business for more easy, advanced and

deep analysis of e-business websites' databases. The outline of the paper is presented as follow: In section 2, we present our review of the literature concerning e-business, DM, OLAP and SOLAP concepts. In Section 3, we present in detail our conceptual framework which integrates e-business, OLAP with SOLAP concepts for better and advanced analysis of hybrid data. In Section 4, we explain the importance of our model and finally section 5 is the conclusion and suggestions for future research.

2. Literature Review

2.1. E-Business

With the emergence of the Internet, a new channel has become available for all parts of the business process, such as shopping, searching for services and product information, transaction, payment, and delivery (of digital products). Thus, e-business is a broader definition of e-commerce that includes not just the buying and selling of goods and services, but also servicing customers, collaborating with business partners, and conducting electronic transactions within an organization [15]. In general, e-business has changed the face of most business functions in competitive enterprises. It has enabled on-line transactions, and facilitates the generation of large-scale real-time data. The climb in online transactions to consumers may reflect compelling advantages of internet business. Significant benefits to consumers of doing business on the internet versus in other retail venues include the vast array of alternative products/services available and the quicker access to alternatives. Along with benefits as a vehicle for purchasing/using, the internet offers consumers a powerful

means for searching out product/services information before making business.

2.2. Data Mining and E-Business

The Internet's information-rich, interactive nature can increase business efficiency by improving the availability of product/services information, enabling direct multi-attribute comparisons, and reducing information-search costs [1]. With data pertaining to various views of business transactions being readily available, many studies have applied different analytical tools, such as DM to make sense out of these data sets. Kohavi and Provost [9] present a survey of some of the recent approaches and architectures where DM has been applied in one domain of E-business fields which is e-commerce. The author explains how DM is applied to improving the services provided by e-commerce based enterprises. Kohavi and Perkh [8] present ten supplementary analyses that E-business web sites should review after running the common reports provided by most web analytics tools. These recommended analyses are based on the researchers' experience in mining data from multiple retail e-commerce sites over the last few years. Their analyses cover a wide range of topics from bot analysis to web site real estate usage analysis, and from analysis of migratory customers to geographical analysis. They have found that while some analyses require architectural support to collect the right data, the effort required is moderate in most cases and should result in significant Return-On-Investment (ROI). They suggest that some analysis, such as customer migration, can benefit directly from a wide customer signature, and many analyses may lead to insight that can be investigated through the use of customer signature.

As it is shown above, DM is widely used to analyze the data behind the E-business websites. This technique has

shown interesting results, but it presents some restrictions concerning the volume and the structure of the data to be analyzed as well as the level of expertise of the end-users who will be interpreting the analysis results.

2.3. OLAP and SOLAP analysis techniques

2.3.1 OLAP

OLAP is a powerful data analysis tool because it allows users to “navigate” or “explore” different levels of summarization by interactively changing the set of dimension attributes and their abstraction levels. In other words, users can navigate from one cuboid to another interactively in order to obtain the most interesting statistics through a set of pre-defined OLAP operations (such as roll-up, drill-down, slice, and dice) [14].

OLAP (On Line Analytical Processing) has been defined as “...the name given to the dynamic enterprise analysis required to create, manipulate, animate and synthesize information from exegetical, contemplative and formulaic data analysis models. This includes the ability to discern new or unanticipated relationships between variables, the ability to identify the parameters necessary to handle large amounts of data, to create an unlimited number of dimensions, and to specify cross-dimensional conditions and expressions” [14]. Other OLAP definitions have since been proposed, including “A software category intended for the rapid exploration and analysis of data based on multidimensional approach with several aggregation levels” [2].

The multidimensional approach is based on dimensions and measures. Dimensions represent the analysis axes, while measures are the numerical attributes being analyzed against the different dimensions. A set of measures aggregated according to a set of dimensions

forms what is often called a data cube of hypercube [24]. Inside a data cube, possible aggregations of measures on all the possible combinations of dimension members can be pre-computed. This greatly increases query performance in comparison to the conventional transaction-oriented data structures found in relational and object-relational database management systems (DBMS). The common OLAP architecture can be divided into three parts: the multidimensional database, the OLAP server that manages the database and carries out the different calculations, and finally the OLAP client that accesses the database via the OLAP server.

Finally, it is commonly found in the literature that the multidimensional approach of analysis is more in agreement with the end user’s mental model of the data [4][11][25]. Based on this approach, the interface of a tool exploring the multidimensional paradigm, such as OLAP is usually very intuitive and the user can perform analysis ranging from simple to complex, mostly by clicking on the data being organized in a meaningful way [25]. This adds to the fact that the multidimensional data structure is optimized for rapid ad hoc information retrieval, which greatly facilitates the data exploration and analysis process.

2.3.2 SOLAP

In various e-business domains which involve spatial data (real state, environmental planning, precision agriculture, etc.), participating business may increase their economic returns using knowledge and data extracted from spatial databases. Analyzing this type of data is often not properly exploited and only few applications called Spatial Data Mining (SDM) have been developed in order to analyze such type of data. In our research we use the Spatial OLAP technique in order to

analyze the spatial part of the data inside the E-business databases.

Traditional OLAP offers good support for simultaneous usage of descriptive, temporal and spatial dimensions in a multidimensional analysis process. Descriptive dimensions are used to describe the data to be analyzed. The temporal ones take into account the temporal aspect of the analysis and the spatial dimensions allow for the spatial reference of the phenomena under study. However, using traditional OLAP tools, the spatial dimensions are treated like any other descriptive dimension, without consideration for the cartographic component of the data. OLAP tools present serious limitations in support of spatio-temporal analysis (no spatial visualization, practically no spatial analysis, no map-based exploration of data, etc.).

Data visualization facilitates the extraction of insight from the complexity of the spatio-temporal phenomena and processes being analyzed, as well as it offers a better understanding of the structure and relationships contained within the dataset. In the context of information exploration, maps and graphics do more than make data visible; they are active instruments in the end user's thinking process [16][17]. Without a cartographic display, OLAP tools lack an essential feature, which could help spatio-temporal exploration and analysis. A SOLAP tool remedies to this lack because they support the geometric spatial dimensions that can include geometric shapes spatially referenced on a map. This type of spatial dimension allows the dimension members to be visualized and queried cartographically.

A SOLAP system can be defined as a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic

displays as well as in tabular and diagram displays [2]. SOLAP fundamental concepts, historical perspective and extensive references can be found in [2][16][23] while SOLAP prototypes are described by [2][5][6][10][18][19][20][22] among others.

Integrating OLAP with SOLAP is geared towards decision-support as it has the following features:

- *Easy*: The ease comes from the ability to conduct analysis without having to master a query language or to understand the underlying structure of the database, which may be very complex in the particular case of spatio-temporal databases [16]. In fact, the analyst interacts directly with the data and focuses on the results of the analysis rather than on the procedures required by the tool to perform the analysis process.

- *Rapid*: It is rapid because data are pre-aggregated, computation time is reduced and it is possible to provide very fast answers to complex queries. This allows the user to maintain his flow of thought, his attention not being distracted by slow response time [16].

- *Multi-granular*: Decision-support typically requires having access to aggregated information (the global view) as well as more detailed information, typically following several levels of granularity of information. Such levels of granularity apply to the phenomena being analyzed as well as their time and space dimensions. Combining the thematic dimensions with the spatial and time dimensions allows one to analyze the evolution of phenomenon, to discover trends, to unveil correlations, etc.

3. The Conceptual Framework

In the previous section, we have shown that OLAP-SOLAP techniques are easy, rapid and multi-granular. These features have attracted us to integrate these

techniques with e-business field in order to analyze the hybrid data behind e-business websites. In our work we propose a conceptual framework in which

we combine these technologies. This conceptual framework is presented in Figure (1) and will be detailed in this section.

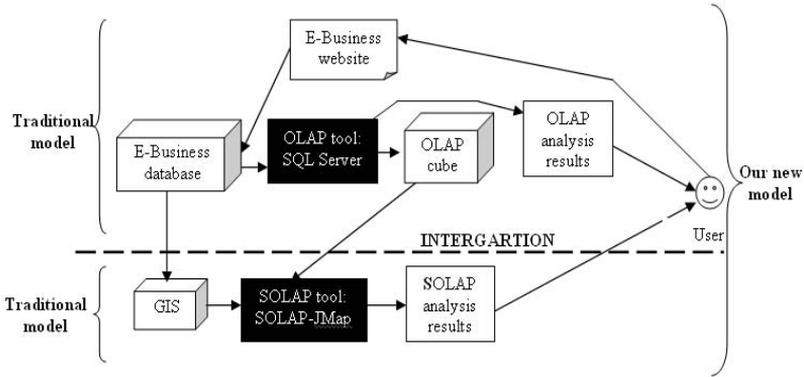


Fig. 1: The proposed conceptual framework.

The components of this framework and the relationships between them are detailed in the following points:

- The e-business website: This website can be related to any field of the e-business domain: e-commerce, e-government, or e-banking.
- The e-business website database: This is the database behind the website database. This database contains the data to be analyzed. This data can be hybrid, which means that it can contain non-spatial data (textual or numerical data) or spatial (geographic data) or both.
- The GIS (Geographic Information System): The spatial data of the E-business website’s database can be stored in a Geographic Information System (GIS) which contains the geographic aspect of this spatial data (exp. Addresses, street names, regions, countries, etc.).
- The non-spatial data of the E-business database are analyzed using an OLAP tool such as the OLAP tool of SQL-Server. This

tool will create an OLAP cube which can be easily browsed by the users and the OLAP analysis results are presented to them.

- The spatial (geographic) data stored in the GIS with the OLAP cube are passed to the SOLAP tool in order to be analyzed and SOLAP analysis results are presented to the users. As example of SOLAP tool we can site the SOLAP-JMap one developed recently by a Canadian company called Kheops technologies inc. (<http://www.kheops-tech.com/>).
- The users of the framework: The user receives the INTEGRATED analysis results from both the OLAP and SOLAP tools in formats which are user-friendly and easy to interpret (table, graphs, maps, etc.). As example of decisions, the user can suggest some enhancement of the website form and content, can propose new business strategies related to the domain of the application, etc. These users can be novice users since OLAP/SOLAP analysis

results are straightforward and can be easily interpreted.

4. The Importance of the Framework

The proposed conceptual framework bridges the E-business and the OLAP-SOLAP worlds. This framework can be considered as new because of the following contributions:

- It can be considered as the only framework which INTEGRATES the OLAP-SOLAP concepts with e-business. It is important to mention that the OLAP (only) and e-business have been integrated in several previous works. The SOLAP has never been used in e-business since it is a recent concept;
- It is a generic framework and can be applied to any fields of the e-business domains such as e-commerce, e-government, e-learning, e-banking, etc. In our future work we plan to apply this framework to a job-seekers website in the UAE. The implementation of the framework is straightforward because it needs only the data related to the application field to be analyzed;
- When implemented to a specific field (e-commerce, e-government, etc.), the analysis results can be easily interpreted by non-expert users;
- The analysis results, which are the outputs of the application of the framework, can be used to make decisions related to the field of the application.

5. Conclusions and Suggestions for Future Research

Several analysis techniques have been used to analyze e-business websites' databases for decision making purposes.

The most famous technique which has been used is DM. This technique has shown its efficiency but it is intended for expert people who can interpret the DM results. In our conceptual framework, we combine e-business websites' database with OLAP and SOLAP techniques in order to analyze non-spatial and spatial data. In the future, we plan to perform a practical implementation of the proposed conceptual framework using a real e-business website database. Thus:

- We plan to test our model by collecting data from an e-business website for Job seekers in the UAE. The data to be collected is hybrid as it contains non-spatial data such as demographic information (ex. gender, age, marital status, etc.), professional information (ex. education level, work experience, etc.), and spatial data which refers to each one of the Emirates of the UAE;
- We plan to develop the GIS database of the UAE. This GIS contains only the high level which is the level of the Emirates in the UAE;
- Once we have collected enough data, we plan to conduct an OLAP analysis and connect the GIS with the OLAP in order to perform the SOLAP analysis

References

- [1] Alba, J., Lynch J., et al. (1997). Interactive home shopping, consumer, retailer and manufacturer incentives to participate in electronic marketplaces, *Journal of Marketing*, 61(3) 38-54.
- [2] Bédard Y., Rivest S., et al. (2005). Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, Idea Group Publishing.
- [3] Berendt B., Hotho A., et al. (2002). Toward semantic web mining. In *proceed-*

- ings of the First International Semantic Web Conference. Sardinia, Italy.
- [4] Codd E.F. & Salley C.T. (1993). Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate. *Hyperion white paper*, 20.
- [5] Ferreira A.C., Campos M.L., et al. (2001). An Architecture for Spatial and Dimensional Analysis Integration. In the proceedings of the SCI 2001. XIV: 392-395.
- [6] Fidalgo R.N., Times V.C., et al. (2004). GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas. In the proceedings of DaWaK 2004, Lecture Notes in Computer Sciences 3181 26-37.
- [7] Kohavi R., Mason L., et al. (2004). Lessons and challenges from Mining Retail E-Commerce Data. *Machine Learning Journal*, 57(1-2) 83-113.
- [8] Kohavi R. and Perkh R. (2003). Ten supplementary analyses to improve E-Commerce web sites. In the proceedings of the Fifth WEBKDD workshop: 29-36.
- [9] Kohavi R. & Provost F. (2001). Applications of datamining to electronic commerce. *Journal of Data and Knowledge Discovery*, 5(1-2).
- [10] Kouba Z., Matousek K., et al. (2000). On Data Warehouse and GIS integration. In the proceedings of DEXA 2000, Lecture Notes in Computer Sciences 1873: 604-613.
- [11] Lau T.W, Hui C.L, et al. (2008). An implication of E-Commerce in textile and clothing industries: A relational online analytical processing (OLAP) approach. *International Journal of Information Technology and Management*, 7(3) 231-243.
- [12] Levy M. & Weitz B. (2001). *Retailing management*. New York, NY, McGraw-Hill.
- [13] Liu H. and Keseli V. (2007). Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users' future request. *Journal of Data and Knowledge Engineering*, 61(2) 304-330.
- [14] Lo E., Ben K., et al. (2008). OLAP on sequence data. In the proceedings of the ACM SIGMOD International Conference on Management of Data, Vancouver, BC, Canada.
- [15] Mokhtarian P.L. (2004). A Conceptual analysis of the transportation impacts of B2C e-Commerce." *Transportation Journal*, 131, 257-284
- [16] Rivest S., Bédard Y., et al. (2001). Toward Better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP). *Geomatica*, 55(4) 539-555.
- [17] Salam I., El Mohajjer M., et al. (2008). Development of SOLAP patrimony management application system: Fez medina as a case study. *International Journal of Computer Science and Applications Vol. 5(No. 3a): 57-66*.
- [18] Scotch M. and Parmanto B. (2005). SO-VAT: Spatial OLAP Visualization and Analysis Tool. In *Proceedings of the 38th Hawaii International Conference on System Sciences*, 142.2.
- [19] Shekhar S., Lu C.T, et al. (2001). "Map Cube: A Visualization Tool for Spatial Data Warehouses." *Geographic Data Mining and Knowledge Discovery*, 74-109.
- [20] Silva J., V. Times, et al. (2005). Providing Geographic-Multidimensional Decision Support over the Web. In *APWeb 2005: 7th Asia-Pacific Web Conference*, Lecture Notes in Computer Sciences 3399: 477-488.
- [21] Srivastava J., Cooleyz R., et al. (2000). Web usage mining: Discovery and Applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, 1(2) 12-23.
- [22] Stefanovic N., Han J., et al. (2000). Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. *IEEE Transactions on Knowledge Discovery and Data Engineering*, 12(6) 938-958.
- [23] Tchounikine A., Miquel M., et al. (2005). Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hy-percubes. *Revue des nouvelles technologies de l'information*, 21-33.
- [24] Thomsen E., Spofford G., et al. (1999). *Microsoft OLAP Solutions*, John Wiley and Sons eds.
- [25] Yougworth P. (1995). OLAP Spells Success for Users and Developers. *Data Based Advisor*, 38-49.