A REVIEW: IMPLEMENTATION OF OLAP SEMANTIC WEB TECHNOLOGIES FOR BUSINESS ANALYTIC SYSTEM DEVELOPMENT

Miss. Pratiksha P. Dhote¹ and Prof. Arvind S.Kapse²
¹² CSE, P. R Patil College Of Engineering, Amravati

Abstract- Business Intelligence (BI) tools provide fundamental supporting analyzing large volumes of information. Data Warehouses (DW) and Online Analytical Processing (OLAP) tools are used to store and analyze data. Nowadays more and more information is available on the Web in the form of Resource Description Framework (RDF) and BI tools have huge potential of achieving better results by integrating real-time data from web sources into the analysis process. We describe the convergence of some of the most influential technologies in the last few years, namely data warehousing (DW), on-line analytical processing (OLAP), and the Semantic Web (SW). OLAP is used by enterprises to derive important business-critical knowledge from data inside the company. However, the most interesting OLAP queries can no longer be answered on internal data alone, external data must also be discovered (most often on the web), acquired, integrated, and (analytically)queried, resulting in a new type of OLAP, exploratory OLAP. When using external data, an important issue knows the precise semantics of the data. Here, SW technologies come to the rescue, as they allow semantics (ranging from very simple to very complex) to be specified for web-available resources. Next, we goes on to survey the use of SW Technologies for data modeling and data provisioning, including semantic data annotation and semantic-aware extract, transform, and load (ETL) processes. Finally, all the findings are discussed and a number of directions for future research are outlined, including SW support for intelligent MD querying, using SW technologies for providing context to data warehouses, and scalability issues.

Keywords - Business Intelligence, Data warehousing, OLAP, Semantic Web, Reasoning.

I. INTRODUCTION

Business intelligence (BI) is aimed at gathering, transforming and summarizing available data from existing sources to generate analytical information suitable for decision-making tasks. The most widely used approach to BI has been the combination of data warehousing (DW), online analytical processing (OLAP) technologies and the multidimensional (MD) data model. DW/OLAP technologies have been successfully applied for analysis purposes, but always in a well-controlled “closed-world” scenario, where the set of data sources is rather static, and well-structured data is periodically loaded in batch mode applying heavy cleansing transformations. However, the eruption of XML and other richer semi-structured formats like RDF has opened up much more heterogeneous and open scenarios than those of such traditional in house DW applications [1].

The purpose of this technology is to categorize the main requirements of these new OLAP approaches, as well as to show how SW technologies can help to fulfill the new requirements. Then, five criteria related to the different relevant aspects of DW/OLAP systems are defined. By means of these criteria, in current approaches are categorized. Furthermore, the five criteria define a space that allows us to locate Exploratory OLAP use cases and to distinguish them from Traditional OLAP use cases. The name of five criteria is Materialization, Transformations, Freshness, Structuredness, and Extensibility. We firstly find Materialization; This criterion concerns the level of materialization of the integrated data. The next one is Transformations, This criterion concerns the level of transformations applied to the
source data during the integration process. The next criterion is Freshness, which concerns how often the data integration process is performed (i.e., how often the DW is refreshed). The next criterion is Structuredness, which concerns which types of data are found in the data sources or, more specifically, how Structured the least structured type of source data. The next and last criterion is Extensibility. This criterion concerns how Dynamic the set of data sources can be, i.e., how easily new data sources could be brought into the system [2].

1.1 Motivation:

The opportunity and importance of using unstructured and semi-structured data (either textual or not) in the decision making process. Nowadays, Web 2.0 sites and Linked Open Data initiatives are becoming sources of huge amounts of valuable semi-structured data. Currently no one questions the need of adding all this information to the traditional corporate analysis processes. A significant amount of information and thus, knowledge, can be found in “unconventional” data sources like web portals, social media, unstructured or less-structured data stores like product reviews, customer complaints, e-mails, and so on. Enterprises have started to look into such rich information sources to increase their profits and improve their products and services. As an example, populating a business report that shows the effect of a product campaign in a specific time period may require combining information from historical, structured data like product sales and customer data, residing in a DW, with sentiments extracted from Big Data (e.g., tweets) relating to products promoted by the respective campaign [3].

1.2 Objectives:

- To provide interactive analysis of multidimensional data this facilitates effective Data Mining.
- To integrate with OLAP operations to enhance interactive mining of knowledge.
- To represent data & integrate it.

1.3 Scope:
We can integrate the proposed work on distributed structure and unstructured databases for efficient retrieval of review and information required for product analysis and used for the development of business intelligence. Online Analytical Processing (OLAP) is a technology by which Data Warehouse are queried and analyzed. A Data Warehouse does not necessarily need OLAP technology. OLAP enables users to gain insight into data. OLAP uses a multidimensional view of aggregate data to provide quick access to strategic information for analysis purposes. In the future work, we are planning to finish the prototype of the proposed framework and test the solution on large-scale case studies.

II. LITERATURE REVIEW

2.1 Background History

OLAP is not a new concept and has persisted through the decades. As a matter of fact, the origin of OLAP technology can be traced way back in 1962. It was not until 1993 that the term OLAP was coined in the codd white paper authored by the highly esteemed database researcher Ted codd, who also established the 12 rules for an OLAP product. Like many other applications, it has undergone several stages of evolution whose patterns of progress are relatively intricate to follow through [4].

It was Kenneth Iverson who first introduced the base foundation of OLAP through his book “A Programming Language”, which defined a mathematical language with processing operators and multidimensional variables. The APL was regarded as the first multidimensional language and its implementation as a computer programming language happened during the late 1960’s by IBM. Iverson created brief notations by employing Greek symbols as operators. During this period, high resolution GUIs had not yet surfaced and, as APL uses Greek symbols, it requires support of special hardware like special keyboards, screens and printers [5]. On top of this, since early APL programs were interpreted as opposed to being compiled, it tends to inefficiently exhaust more machine resources and is known for consuming too much RAM space, to name only a few of its drawbacks. Maintenance of APL-based mainframe products is very costly and most programmers encounter difficulties in programming multidimensional applications using arrays in other languages [6].

Eventually, there was a decline in the market significance of APL, but it still survives to a limited degree. Although it was not deemed a modern OLAP tool, several of its ideas can be seen living through some of the modern day multidimensional applications.
1). The term OLAP was created as a slight modification of the traditional database term OLTP (Online Transaction Processing).
2). Databases configured for OLAP employ a multidimensional data model, allowing for complex analytical and ad-hoc queries with a rapid execution time.
3). They borrow aspects of navigational databases and hierarchical databases that are speedier than their relational kind [6].

2.2 Existing System

Most of the existing work for bringing together OLAP and SW technologies focus on the scheme design issues, but our analysis work still unresolved The common characteristics of existing approach is that these can be deal with unstructured data only

A first challenge involves the automatic derivation of such mapping but researches says that related fields like schema mapping and data exchange has provided very useful results that have not been exploited yet by research on DW Design Existing technique LAP / SW [7].

2.3 Limitations

The previous system doesn’t provides a common data model for all data of interest regardless of the data’s source.
Prior to loading data into the data warehouse, it’s consistencies are not identified and resolved. This complicated reporting and analysis.

Information in the data warehouse is under the control of data warehouse users so that, even if the source system data is cleared over time, the information in the warehouse can’t be stored safely for extended periods of time [8].

III. PROBLEM FORMULATION

3.1 Aim
To implement OLAP Semantic Web Technologies For Business Analytic System Development.

3.2 Problem Definition
Most of the existing work for bringing together OLAP and SW technologies focus on the scheme design issues, but our analysis work still unresolved The common characteristics of existing approach is that these can be deal with unstructured data only .To solve these problems we are going to use OLAP technology [9].

IV. PROPOSED SYSTEM

We propose a series of criteria that aims to capture the main aspect of emerging DW/OLAP system and how these components are evolving to cover the new requirements posted by the new scenario.

The resulting categorization schema aims to identify the commonalities and differences of emerging approaches in term of the changes they proposed with respect to the traditional components of DW/OLAP system. We Proposed technique OLAP / DW.

4.1 Methodology
Nowadays, a new trend of OLAP work has emerged, which applies SW technologies to mainly address data integration issues and the automation of data processing. OLAP technology is aimed at gathering, transforming and summarizing available data from existing sources to generate analytical information suitable for decision-making tasks. Traditionally, OLAP has been associated with data warehouses (DW), following the three layered structure shown the data.

Sources layer, which consists of all the potential data of any nature (e.g., relational, object-oriented, semi-structured, and textual) that can help to fulfill the analysis goals. The integration layer, which transforms and cleanses the data gathered from the sources, as well as stores them in an appropriate format for the subsequent analysis, and The analysis layer, which contains a number of tools for extracting information and knowledge from the integrated data and presenting it to the analysts.

As it is clear from this description, the integration model of Traditional OLAP systems (DW/OLAP) is based on a global schema, which is seen as a view over the underlying data source schemas. In this integration model, query answering is simple. The external data sources are assumed to be known in advance as are the user needs guiding the design of the global schema. This works well when the sources and requirements are indeed known in advance, but encounter problems when this does not occur. For those cases, more flexible integration models are needed. In particular, the integration of external data schemas in terms of a global schema has been studied .From the global schema, local schemas can be derived; i.e., the local schemas are seen as views of the unified general global schema. The resulting integration model is thus highly extensible, at the expense of considerably more complicated query answering. Therefore, in this integration model the reasoning power of SW technologies is especially needed. DFD is as follows:
V. CONCLUSION

In this paper, we surveyed techniques to implement OLAP Semantic Web Technologies for Business Analytic System Development. When using external data, an important issue knows the precise semantics of the data. Here, SW technologies come to the rescue, as they allow semantics (ranging from very simple to very complex) to be specified for web-available resources. Next, we go on to survey the use of SW Technologies for data modeling and data provisioning, including semantic data annotation and semantic-aware extract, transform, and load (ETL) processes. Finally, all the findings are discussed and a number of directions for future research are outlined, including SW support for intelligent MD querying, using SW technologies for providing context to data warehouses, and scalability issues.

REFERENCES


