CURRENT TRENDS IN DATA WAREHOUSING METHODS AND TECHNOLOGIES

Vera Ivanova
Assistant Professor, ITEC program, York University, Toronto, Canada

Keywords: Data Warehousing, Business Intelligence, Current Trends, Technologies, System Architecture, Analyzes, Review.

Abstract: Data Warehousing (DW) methods and technologies are in a new stage of their evolution and of their amalgamation with the enterprise businesses, they serve. The main goals of this work are to identify, review and analyze the latest trends in DW. A systematic approach is followed to recognize, define and analyze the most important trends. The approach is based on the trends' corresponding role and value in the business processes and intelligence (BI). For this purpose we start with updated definitions of DW and BI and then consider the generalized Architecture of today's DW. We then "drill down" to analyze the DW problems and trends in their solving for data quality provisions, regulatory compliance, infrastructure consolidation, and standardization, corporate performance optimization and metadata management. This in-depth logical analyzing approach results in comprehensible conclusions to be considered on the important early phases of DW projects, as it is well known that early project decisions carry impacts for the whole DW system life span.

1 INTRODUCTION

Even though DW have been evolving for the last 10 years experts, managers, watchdog agencies, vendors and developers all agree that the DW methods and technologies are in a new stage of evolution (Agosta, 2004), (www.microsoft.com, 2005), (Singh, 2005). The highest priority post-millennium IT project for many corporations was the DW (Glick, 2005). The opinions on the DW new features vary, due to their application and industry specific nature. For example, a large retailer DW in many aspects looks quite different from an University Westermann, 2001), (http://web.mit.edu/warehouse, 2005), yet they may share similar new technology trends, or in some cases identical driving business requirements.

The realities of the DW evolution across many industries changed the way DW are used in terms of operational tasks, publishing and interactive use of broad audiences of employees, suppliers an partners. Many DW implementations brought business improvements with good ROI and still many failed to do so.

DW technologies and methods today deliver far more than key performance indicators to top managers. An update is needed to recognize, capture, review, analyze and define these trends and changes. The value of similar works is in DW high level architectural decisions and in the first steps of large and expensive DW projects considerations, as well as in attempts to improve , refresh the assets, or consolidate old corporate DW that still exist.

2 DATA WAREHOUSING AND BUSINESS INTELLIGENCE

Let us briefly consider the DW and BI updated definitions for the purpose of this work. DW is the concept to consolidate enterprise data from production systems often of heterogeneous sources and platforms and optimizes them for decision making, monitoring, reporting, analysis and interactive publishing. The data are extracted as they are generated from transactional, operational, and other system sources (see p.3), or on scheduled periodical intervals. The main reason to separate the DW data from other data is to optimize the performance of complex queries that typical for the above activities as the execution of similar queries on the source systems would degrade them and is not acceptable or possible. There are many other business and IT reasons for separation, considered further in details in this work. DW is implemented

Ivanova V. (2006).
CURRENT TRENDS IN DATA WAREHOUSING METHODS AND TECHNOLOGIES.
In *Proceedings of the Eighth International Conference on Enterprise Information Systems - DISI*, pages 297-301
DOI: 10.5220/0002498902970301

on its own architecture and often infrastructure to reflect the business specifics and to utilize the available limited resources.

The classic definition of the founder W. Inmon (Inmon, 2002) of DW as "subject oriented, integrated, non volatile and time variant collection of data in support of management decisions" is still valid, but is overloaded and enriched with many new features to accommodate the new technologies and business needs. All new technologies (multimedia, wireless, GPS, RFID, etc.) generate more and more data and several large DW already exceeded 1 Petabyte (1K Terabyte). From the business competition and regulatory compliance perspective DW became a source and environment for many new business processes to model , monitor and optimize, even if they are in fact, operational.

The term BI is accepted for information technologies, which transform business data into meaningful information to support result oriented strategic and sometimes tactical business decisions. This broad term covers a broad field of applications – from a simple reporting query tools to full scale DW. We consider here this last case, where the DW approach is to get data in different formats from different sources (databases, web sites and services, ftp servers etc.) and extract, consolidate, clean and store them in formats, optimal for the business analytical purposes and various processes. Presented with accurate and timely intelligent reports in business terms the enterprise decision makers can develop and optimize processes and strategies. The term BI is also used for advanced data analysis, data mining and corporate performance measurement and management processes. The tools and the applications are industry and department specific. For example, the financial departments use them for budget management, the customer relations department for marketing strategies; and the executives for CPM. Many or all of these tools are part of the DW architecture and infrastructure.

3 COMMON DW ARCHITECTURE AND INFRASTRUCTURE HIGH LEVEL TRENDS

The amalgamation and realization of BI in DW is a strong trend, altering in many ways the architecture and infrastructure of the DW design and implementation decisions. Fig.1 shows the general system architecture of today's medium and large DW. Though in appears basically as it did years ago, it is now loaded with solutions to support business models and processes tracking, management and optimizing, shown inside the middle DW layer. The other new trend here is the broadly represented advanced user interfaces for publishing, data delivery and self-services, running usually from corporate portals. They now often encompass the whole enterprise, giving access to employees, customer, suppliers and partners. Thus DW essentially evolved from a decision support tool for top managers to an IT infrastructure necessity in many aspects vital for the enterprise (Gorla, 2003), (Golfarelli, 2004), (Inmon, 2002) and certainly used by a much broader audience.

The main high level attributes of this architecture are power, flexibility and scalability. The *power* comes from the consolidated DB ("knowledge is power"). For example with an easy navigation from the corporate portal any (or designated) employees can browse the CPM dashboards. In companies with thousands of employees worldwide this improves the team efforts and environments. Next trend is the enriched business analytical support, and ad-hoc, or scheduled reporting services, available to anyone from a thin browser client, or through a familiar spreadsheet interface. The required by the business *flexibility* is provided by the consolidated metadata solution. Business rules change, processes improve, and targets are re-defined frequently "on the go". If properly designed and propagated, the metadata repository will effectively accommodate these changes and enforce them to downstream applications and services. The architectural *scalability* is a factor on the ETL, DBMS and server levels, shown in Fig.1. A DW configuration with one DB server with one central metadata repository, an application server to run all the tools and a Web server for the user interface may be adequate for medium scale businesses. For large global businesses a distributed DW Architecture utilizes several *Data Marts* with separate metadata repositories and separate, or common ETL. The idea here is to "tune" the individual metadata in every Data Mart to the business requirements, defined by business subjects, departments, or regions in order to achieve adequate reporting and data mining performance.

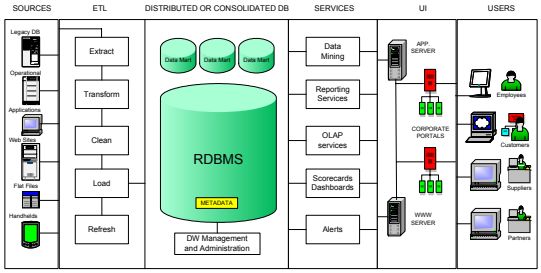

Figure 1: General DW system architecture.

The best Architectural and Infrastructure strategy should match the present and forecasted business requirements with the available project resources (Chadhuri , 2004), (Imhoff , 2003). The known approaches with clusters, federated DBMS, table partitioning, load balancing Web farms etc. are all applicable here, depending on the project budget.

4 CURRENT BI RELATED TRENDS IN DW

4.1 Providing Data Quality in DW

Perhaps the most important trends in DW are the well founded provisions for data quality. The rule that "business is as good as its corresponding data" has strong practical implications on DW. The DW Institute estimates (www.knightsbridge.com, 2005) the business losses due to poor data quality about $600 bln. per year in USA.

The *sources* of poor data quality in DW are numerous: data input errors, loading and consolidation errors, conversion errors, DB integrity errors, reporting exception errors etc. In business processes such poor data can result in errors like wrong customer order generation, incorrect invoices, duplicate payments etc. The effects are low efficiency employment and unhappy customers, i.e. losses and low profit. On the executive level low quality data can result in non adequate tactical and strategic decision making and in inaccurate performance estimations.

The data quality can be generally defined as *compliance* to the particular set of IT and business *requirements*. The best practices here are to designate a person, or a team, responsible for the data quality management and to specify the requirements, usually in several improving cycles (Leahy, 2004). Reviews show that many businesses still do not have clearly defined, enforced and managed data quality requirements. This is the driving force behind the fast market growing for data quality products and services (Geiger , 2005).

In some cases the DW quality requirements are implemented as validation and consolidation rules on the DW system ETL level, as well as on the reporting and data mining levels. However the most flexible and scalable way is the practice to include and manage the data quality requirements on the DB level as part of the metadata. There are many benefits with this approach – end to end system coverage, central management,, scheduled/on demand cleansing and easy adding of new data quality procedures, which all combined give cost effectiveness.

The conclusion for this trend is that the DW data quality assurance is a critical enterprise IT component.

4.2 Regulatory Compliance and DW

Regulatory compliance brings new requirements to both DW and BI in terms of certified financial results, privacy protection, risk declaration, disaster recovery, environmental protection etc. Recent examples are legislation like the Sarbanes-Oxlley (SOX) Act in USA, the Health Information Privacy Act (HIPA) in Canada, the Basel II banking accord requirements in Switzerland and alike. Demanding data policies and special new business processes are required to achieve compliance (Geiger, 2005), (www.knightsbridge.com, 2005). These processes and policies affect the DW as mandatory new additions and improvements are needed in the DW system and data change control, data quality assurance, accelerated warning financial reports and non-compliance monitoring by designated officer, or a team. A part of these processes is the data quality assurance, as considered in p. 4.1.

Whether the DW requirement changes are significant, or not, when a certain business, as a result of its activities, is affected by one, or several legislative regulations, they can not be ignored and the compliance should be provided by the required processes.

4.3 DW Infrastructure Consolidation and Standardization

Many companies today are undertaking infrastructure consolidation and standardization in order to reduce and effectively manage the IT cost (Glick , 2005). This is an ongoing trend after year 2000. This trend fully affects the DW systems, as by their nature they grow, expand and increase in cost

in all terms of hardware spending/amortization, software purchase and licensing and managing/support cost. The consolidation should be very well planned and leveraged by such goals and provisions for system and data growth, timely query and reporting responses, simplified and low operational cost management and support. These should be combined IT *and* business efforts.
On the operating system level the new virtual servers technologies (Glick, 2005) allow accommodation of large heterogeneous DW with servers, running Windows, Unix, or Linux in a very flexible consolidated hardware environment. The administration, managing, support and versioning costs are greatly reduced. On the DB level, it pays to reconsider the consolidation of old and legacy data marts and ETL tools with centralized approach and streamlined new ETL processing tools.
The business trend is that calls for consolidation are coming not only as a result from acquisitions and mergers. It also comes from the understanding that the latest technologies as federated DBMS servers, web farms, network storage devices etc. allow alignment of the BI processes with DW architecture in a very cost effective manner with excellent ROI. If properly designed and executed the consolidation and standardization bring significant savings, up to billions of dollars in many cases (Glick, 2005), (Violino, 2005).

4.4 Corporate Performance Optimization using DW

As a consolidated foundation of the enterprise business data, the DW can be well designed for loading, reporting and data mining. That is why it is the ideal "natural" source and platform for corporate performance measurement (CPM), management and optimization, which makes this important usage of DW a well recognized trend (www.intelligentbusiness.biz,2005). The generalized *functional processes* of CPM, based on DW, are:

- monitor financial and non financial results on demand;
- - link strategies to day to day operational activities;
- - configure and setup early warnings about problems;
- - achieve accord with regulatory compliance;
- - broaden decision making by modelling new business scenarios.

The above functions comprise the BI essentials in using DW for CPM. The emerging trend here is to use DW for *both* strategic and operational decisions, affecting back the DW performance requirements. This allows the CPM activities to cover the optimizing of all departments like customer relations, financial operations, risk management, sales, planning and supply chains.
The strategic and tactical *models* , used in CPM evolved together with DW and BI. The most widely recognized and developed to advanced stage are: The Balanced Scorecard , Six Sigma and Malcolm - Baldridge (Solomon, 2003) and others. The implementation of similar models goes to the business bottom line and core values and properly applied sometimes brings dramatic improvements with proven savings. This is a huge area and alone may be a subject for a study. All DBMS vendors include balanced scorecard and dashboard tools in their DW solutions: Oracle in 10G, Microsoft in SQL Server 2005, IBM in DB2 etc. Also, many integration software vendors offer universal add-on CPM frameworks (Violino, 2005), which could be rapidly implemented – from days to weeks – on top of existing DW, for example (Business Objects, 2005), (Amateo Data Warehouse, 2005) , and many others. These tools get their data feed directly from the DW, ensuring validated, consisted and timely CPM.
Therefore in the case of newly projects, the CPM will most likely come as a part of the rollout, while in case of a relatively older implementations it will be an added-on framework.

4.5 Metadata Management in DW

Considered the most important component of the DW, the metadata essential role as definitive information about data covers the entire contents of DB, ETL, the reporting data mining, in other words all data in DW and many, if not all processes. That is why the DW metadata requires an careful management. The functional groups of the metadata could be conditionally divided as administrative, business and operational. The *administrative* includes: the definitions of setting and using the DW itself, such as the DW schema, the sources DB schemas, ETL rules, prepared queries and reports, user profiles, access roles and control, data quality procedures, multilingual support etc. The *business* metadata includes the business rules and terms, definitions and processes. The *operational* metadata includes the structure of the DW monitoring, including logs, audit trails, prepared archiving packages of scripts, replication etc.
The *management* of the metadata encircles such functions of sharing the metadata for design, setting, using and operating across internal and external

applications, outside of DW; cleaning, normalizing, restructuring, updates and publishing of catalogs. All DW big vendors Oracle, Microsoft, IBM, Terradata, Sybase et al. invest heavily in metadata management and offer their integrated repository systems. There is also a market for infrastructure integration vendors like Informatica, CA, SAP, Platinum, Prism et al.

4.6 Other New trends in DW

The following trends in DW will be considered in details in the second part of this work:

- disaster recovering processes and practices;
- - new hardware technologies and devices in DW;
- - new data mining methods and technologies in DW;
- - efficient DW and metadata design;
- - outsourcing and DW as a paid for services;
- - open source low price DW solutions.

5 CONCLUSION

We can now summarize the most important DW trends, considered in this work:

-The DW methods and technologies track, support and optimize more and more business *processes;* than before;

-The amalgamation of BI in DW affects the requirements to DW architecture design and implementation, bringing more operational requirements;

-DW is evolving from a decision support tool for top managers to an IT infrastructure necessity for the whole enterprise;

- Additional research is needed to analyze the influence of new technologies in hardware, data mining, outsourcing and open source movement on DW.

REFERENCES

Agosta L. A Nov.2004. Time of Grouth for Data Warehousing., DM Review

Overview of Data Warehousing in SQL Server 2005, *www.microsoft.com.*

Singh C.,Sept.2005. Enterprise Datawarehousing is next big thing, Economic Times.

Gorla N. 2003. Features to consider in a Data Warehousing Systems. Comm. of the ACM, Nov. 2003, vol 46, No11.

Chadhuri S., Dayal U., 2004. An Overview of data warehousing and OLAP Technology.

Geiger J., October 2005. Intelligent Solutions: You can be sake if it is been certified. *DM review magazine.*

Golfarelli M. e.a., Nov. 2004. Beyond Data Warehousing: What's next in Business Intelligence? *DOLAP'04*.

Data Quality management. Executive Brief, ACC Services Ltd., *www.acl.com*

Sen A., Sinha A., 2005. A comparison of Data warehousing Methodologies. *Comm. of the ACM,* March, V.48. No3

Demarest G., April 2005. Zen and the Art of Information, *Oracle Magazine.*

Oracle Business Intelligence Warehouse Builder 10G, *www.oracle.com.*

Knightbridge Solutions, 2005. Top ten trends in BI and Data Warehousing for 2005, *http://www.knightsbridge.com*

Overby S. , Nov.2003 41 Business Units, 9 CIOs, 1 Standard, *CIO Magazine.*

Inmon W. A history of database evolution, *www.b-eye-network.com*

Riccardi E. , July 2005 Balanced Scorecard and its information systems: The performance Data Warehouse. *SASE.*

Leahy T. , April 2004. Better Data, Better Performance Measurement, *Business Finance*.

Glick B. , October 2005. Measuring IT value methodologies, *Computing Business.*

Violino B., March 2005. Frameworks boost business efficiency. *Optimize Mag*.

Kamram N., Process Warehouse – The missing link in Business Performance Management, *www.datawarehouse.com*

Solomon M., Nov. 2003. Linking BUsiness INtelligence to the balanced scorecard., TDWI lessons, v.16

White C., 2005. Corporate Performance Optimization Guide, *www.intelligentbusiness.biz.*

Inmon W., 2002. *Building the Data Warehouse*, John Willey & Sons.

Inmon W. e.a., 2001. *Data Warehousing for E-Business*, John Willey & Sons.

Imhoff C e.a., 2003. *Mastering Datawarehouse Design*, John Willey & Sons.

Westermann P., 2001 Data Warehousing: Using the Wal-Mart Model, Academic Press.

MIT's Data Warehouse, *http://web.mit.edu/warehouse*, 2005.

Business Objects Data Warehouse, 2005, *http://www.businessobjects.com*.

Amateo Data Warehouse. 2005, *http://www.amateo.com*.