Techniques, Process, and Enterprise Solutions of Business Intelligence

Li Zeng, Lida Xu, Zhongzhi Shi, Maoguang Wang, and Wenjuan Wu

Abstract—Business Intelligence (BI) has been viewed as sets of powerful tools and approaches to improving business executive decision-making, business operations, and increasing the value of the enterprise. The technology categories of BI mainly encompass Data Warehousing, OLAP, and Data Mining. This article reviews the concept of Business Intelligence and provides a survey, from a comprehensive point of view, on the BI technical framework, process, and enterprise solutions. In addition, the conclusions point out the possible reasons for the difficulties of broad deployment of enterprise BI, and the proposals of constructing a better BI system.

I. INTRODUCTION

As businesses continue to use computer systems for a growing number of functions in today’s competitive, fast-evolving world, most companies face the challenges of processing and analyzing huge amounts of data and turning it into profits. They have large volumes of detailed operational data, but key business analysts and decision makers still cannot get the answers they need to react quickly enough to changing conditions because the data are spread across many departments in the organization or are locked in a sluggish technology environment. In these cases, Business Intelligence (BI) is presented, which are sets of tools, technologies and solutions designed for end users to efficiently extract useful business information from oceans of data.

Nowadays, BI has been viewed as sets of powerful tools and approaches to increasing the value of the enterprise. More and more business sectors have deployed advanced BI solutions to enhance their competitiveness since it is important to the effective and efficient running of the enterprise.

II. BUSINESS INTELLIGENCE TECHNICAL ARCHITECTURE

A. Review Stage

The concept of Business Intelligence was firstly introduced by Gartner Group in 1996, and incipiently referred to the tools and technologies including data warehouses (or data mart), reporting query and analysis. According to Revelli [1], Business Intelligence “is the process of collection, treatment and diffusion of information that has as an objective, the reduction of uncertainty in the making of all strategic decisions”.

Today’s BI evolves from EIS (Executive Information System), DSS (Decision Support System), query and reporting tools and multidimensional analysis which also known as OLAP (On-Line Analytical Processing).

B. Technical Framework

Business Intelligence framework is never delivered by way of a single technology, product or vendor. [2] The very nature of BI encourages business users to have more access to, and control over, the data.

The successful application of BI in an enterprise should consider the following two points. The first is correct, valid, integrated, and in-time data, and another is the means which will transform the data into decision information. However, neither satisfied data nor effective means are easily acquired. BI Technical framework is used to solve these two questions above. The framework consists of Operational Applications Tier, Data Acquisition Tier, Data Warehouse Tier, Platforms and Enterprise BI Suites Tier, Corporate Performance Management Tier. Operational Applications Tier has the system such as Enterprise Resource Planning, Customer Relationship Management, Supply Chain Management, and Legacy. Extraction, Transformation and Loading obviously belong to Data Acquisition Tier. Besides Data Warehouse, Data Warehouse Tier includes Data Marts and Operational Data Store. Data Warehousing, OLAP [3], and Data Mining [4] are three of the most significant technologies in the
Business Intelligence arena. Data Warehouse can be defined as a “very large” repository of historical data pertaining to an organization. OLAP refers to the techniques of performing complex analysis over the information stored in a data warehouse. The complexity of queries required to support OLAP applications makes it difficult to implement using standard relational database technology. Data mining is the process of identifying and interpreting patterns in data to solve a specific business problem. Data mining strategies for BI include classification, estimation, prediction, time series analysis, unsupervised clustering, and association analysis or market basket analysis, e.g., Apriorio [5], Decision Tree Induction, Support Vector Machine [6]–[8], K-Nearest Neighbor, Genetic Algorithms, Rough Set, Fuzzy Set [9], K-means, Case-Based Reasoning, etc.

That is to say, BI is technically a combination of several disciplines and techniques mentioned above instead of an independent, novel or original approach.

III. BUSINESS INTELLIGENCE ENTERPRISE SOLUTIONS AND APPLICATION PLATFORMS

The world-wide leading BI vendors provided advance enterprise platforms, the building blocks for complete, end-to-end BI solutions. Seamless integration between world-class database management, sophisticated analysis services, enterprise servers, and exceptionally flexible, easy-to-use front-end applications make the power of BI accessible to users at all levels of the organization, and realize benefits in some key areas, for example, lower Total Ownership Cost (TOC), increased Return on Investment (ROI), and so on. As a productivity suite, BI platform will be a familiar fixture in business today. By integrating a BI solution with the existing desktop environment, organizations make BI tools available to a larger segment of users, and increase the number of employees who can make informed, BI-based decisions and share their knowledge and best practices with other employees. Providing broader access to BI across the organization ultimately makes the organization more agile and more competitive by empowering all employees to make faster, better decisions based on common understanding of markets, customers, and the data that shapes the business.

More and More Enterprise solutions and platforms for Business Intelligence have been developed. These vendors and platforms include the commercial solutions such as IBM DB2 with Business Intelligence Tools [10], Microsoft SQL Server, NCR Teradata Warehouse [11], Hyperion/Brio, SAS, iData Analyzer, Oracle, Cognos, Business Objects, OLAP4All, and the open source project Pentaho [12], and MSMiner [13]. However, few vendors can claim providing with an enterprise wide implementation of Business Intelligence, with ready access to information for all decision makers through ad-hoc, reporting or other means.

With the absorption of Alphablox's rapid analytic development tools into its IBM DB2 Warehouse, the increasing momentum of WebSphere Information Integrator software, and the recent acquisition of data integration vendor Ascential Software means that IBM is now filling out more parts of the BI stack itself, rather than relying on partners as it has done in the past. The combination of DB2 and WebSphere products now offers a powerful framework for bringing BI, content and processes together. DB2 Intelligent Miner for Modeling delivers DB2 Extenders for modeling operations. DB2 Intelligent Miner for Data provides new business insights and harvests valuable business intelligence from enterprise data. DB2 Intelligent Miner Visualization provides Java visualizers to interact and graphically present the results of associations. IM Visualization provides the following java visualizers to present data modeling results for analysis: Associations Visualizer, Classification Visualizer, Clustering Visualizer, and Regression Visualizer. DB2 Intelligent Miner for Scoring provides scoring technology as database extensions, i.e. DB2 Extenders and Oracle cartridges.

NCR Corp. is also the world-wide leading BI vendors. Its product Teradata Warehouse is a powerful solution that combines parallel database technology and scalable hardware with the world's most experienced data warehousing consultants, and the best tools and applications available in the industry. Today, NCR and Teradata warehouse have occupied the share of BI market more than 50 percent in world wide. What’s more, NCR shapes a set of distinct methodology of data warehousing and practical framework, named as Scalable Data Warehouse. In order to extend the power of Teradata in the area of data mining, NCR has developed Teradata Warehouse Miner data mining software as a set of tools and application interfaces to provide high performance, scalable data mining to Teradata customers. Teradata Warehouse Miner complements traditional data mining software for Teradata customers by addressing the need to handle large volumes of data in a scalable manner.

The latest release of Microsoft's SQL Server 2005 is being expected heavily for its BI capabilities, by customers, partners and competitors alike. Microsoft Business Intelligence Platform reaches significant enhancements in OLAP, data integration, tighter Office integration and new
scorecard capabilities. The ubiquity of Excel and other Office applications on the corporate desktop proves Microsoft with a captive audience for realizing its grandiose scheme of delivering enterprise-scale BI to the masses.

The release of SAS 9 Business Intelligence Platform and Enterprise BI Server has kept a broader range of business users interested in SAS BI and analytic software. The SAS BI Platform provides an extensible Enterprise Intelligence Platform that serves as the foundation for creating and delivering accurate, in-depth intelligence. The SAS BI Platform was developed to support an end-to-end framework for creating and sharing enterprise intelligence, which is called the SAS Intelligence Value Chain. Most organizations have many elements of an intelligence value chain in place: data captured from business processes, data storage and manipulation capabilities, and various query and reporting tools, perhaps from multiple vendors. With the SAS Enterprise Intelligence Platform, users can optimize and extend the value of these existing systems while setting the stage for new levels of Business Intelligence not previously possible.

Oracle’s BI enterprise solutions include Oracle 10g, Oracle Express, Oracle Darwin Data Mining, respectively using for the three technologies of BI, namely, data warehousing, OLAP and data mining. Oracle 10g has enhanced more the capability of developing support than its predecessor Oracle 9i by providing JDeveloper and free BI component Beans. After a couple of years in the wilderness, Oracle makes a welcome appearance and back into Business Intelligence.

MSMiner is a multi-strategy data mining platform for decision support developed by the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences. The goal of research and development is to implement an integrated, extensible decision support tool by employing data warehousing and data mining technologies. The integration of an entire decision-making process on an interactive basis is emphasized in this system. MSMiner Workbench consists of six modules: ETL, Metadata Server, subject management, OLAP analysis, data mining, and dynamic link library interface of algorithms, visual reporting tools. The functions of the system include database access, data modeling, data preprocessing, data mining, and data visualization. In addition to this, emphasis has been put on the extensibility characteristics of the system. As a multi-strategy data mining platform, MSMiner not only provides convenient tools to develop new data mining algorithms, but also includes many build-in algorithms such as SOM and C4.5. In MSMiner, a task scheduling algorithm based on Directed Acyclic Graph (DAG) was developed, which ensures that the data mining task and data preprocessing task can be executed concurrently. ETL is an important subsystem of MSMiner. The main objective of the ETL function module is to transform the operational data from source database to subject data in the data warehouse. The OLAP module in MSMiner is realized by two ways: creating special multidimensional OLAP (MOLAP) and simulating the multidimensional data by using the relational OLAP (ROLAP). MSMiner supports ROLAP which is based on the star schema. The OLAP subsystem in MSMiner support standard OLAP operations, such as slice, dice, roll up, drill down, and pivot, and the results may be displayed in many forms such as cross-tabulation tables, bar charts, pie charts, or other forms of graphical output. MSMiner provides two ways to run a data mining task. One is to execute a task immediately; another is to schedule the task for later execution.

IV. GENERIC BUSINESS INTELLIGENCE PROCESS

In generally, Business Intelligence follows a process that begins with a precise description of the business issue. With the data and mining technique selected, data miners will conduct mining and inspect, and evaluate the results. It’s likely that further iterations of data selection and the application of different mining techniques may be necessary to provide a satisfied solution. Suppose that the mining effort effectively addresses the original business problem, it’s necessary to deploy the results so that it leads to concrete actions taken.

Though people using standard statistical techniques or reporting tools to explore database. As a matter of fact, what people were doing is making a hypothesis about a business issue that they are addressing and then attempting to prove or disprove their hypothesis by looking for data to support or contradict the hypothesis.

Data mining uses an alternative approach that begins with the premise that we do not know what patterns of customer behaviors exist. In this case, we might simply ask the question: what are the relationships between what my customers spend and where they come from? The using of data mining algorithm may tell us about all of the different types of customers and the money they spend. Data mining provides answers without having to ask specific questions. Many business and research fields have been proven to be excellent candidates for data mining, for example, banking, insurance, retail, telecommunications, manufacturing, pharmaceuticals, biotechnology and so on, where significant benefits have also been derived. Well-known approaches are customer profiling and cross-selling in retail, loan delinquency and fraud detection in banking and finance, customer retention in telecoms, and patient profiling and weight rating for diagnosis-related groups in health care. Data mining is about the discovery of pattern and relationship in business data. All of the different applications are using the same data mining concepts and applying them in different ways. That’s not to say that data mining is magic and omnipotent. We still have to understand the overall business process. The process starts with defining the business problem that we want to solve. Then a mining expert can concentrate on the right solution, that is, gather relevant data and discover hidden patterns using mining algorithms. Once the analysis is complete, the new knowledge extracted from the data should be put into action.
A. Locate the business issue

The first step is to identify the business issue that we want to address and then determine how the business issue can be translated into a question, or set of question, that data mining can tackle. As we are formulating the business issue, we need to also think about whether we have access to the right data. It is important to recognize that the data we hold may not contain the information required to answer the question.

B. Create a data model from the business issue

Once the data is being used to support a specific business application, the Metadata and data together form what we call data model that supports the application. It is a complex and hard task to define data models for any application. The challenge is that very often we are not sure at the outset which variables are important and therefore exactly what is required. Mapping the business issue to a data model can therefore be a time-consuming activity. The alternative is to use common data models that have been developed to solve business issues similar to the ones we are trying to address. While these types of models may not initially provide us with all of the information we require, they are usually designed to be extendable to include additional variables. The main advantage of using a common data model is that it provides us a way of quickly seeing how data mining can be used within our business.

C. Data preprocessing phase

Most data preprocessing is the form of data cleaning, which involves dealing with missing information. Ideally, the majority of data preprocessing takes place before data is permanently stored in a structure such as a data warehouse. Common concerns with noisy data, often represented by random error, include incorrect attribute values, duplicate records, and data smoothing. In very large datasets, noise can come in many shapes and forms.

D. Choose suited mining and analysis techniques

It is apparent that when it comes to solving a particular problem we have several techniques to choose from. The question now becomes, how do we know which data mining technique to use? Given a set of data containing attributes and values to be mined together with information about the nature of the data and the problem to be solved, determine an appropriate data mining technique. For a given business issue, the step of selecting mining techniques or algorithms not only includes defining the appropriate technique or mix of technique to use, but also the way in which the techniques must be applied. Data mining techniques can be generally divided into two broad categories: discovery mining and predictive mining. Discovery mining applies to a range of techniques whose primary objective is to find patterns inside the business data without any prior knowledge of what pattern exist. Clustering and sequence analysis are typical examples of discovery mining techniques. Predictive data mining is applied to a range of techniques that find relationships between a specific variable, called the target variable, and the other variables in the business data. Classification and regression are examples of predictive mining techniques.

E. Display, interpret, and evaluate the results

The purpose of interpretation and evaluation is to determine whether a learning model is acceptable and robust that can be applied to problems outside the test environment. If acceptable results are achieved, it is at the stage where acquired knowledge is translated into terms understandable by users. The results from performing data mining can provide a wealth of information that can be difficult to interpret. This step often requires assistance from a business expert who can translate the mining results back into the business context. Therefore, it is important that the results are presented in such a way that they are relatively easy to interpret. We need tools that enable us to visualize the results and to provide the necessary statistical information. Remarkably, IBM Intelligent Miner Visualization, a BI component mentioned before, is good for this.

F. Respond to the mining and analysis results

Deploying the results of the data mining into business is possibly the most important of all. We create mathematical representations of the data and called it models. They contain the rules and patterns found by the mining algorithm. These models provide us with a deeper insight of our business, and can be deployed in or used by other business processes. A number of possible actions may result from successful application of the knowledge discovery process, e.g. relocation of retail items for purchase or placement of selected items on sale together; the mailing of promotional information to a bias sampling of a customer population; incorporation of a developed learner model as a front-end system designed to detect fraudulent credit card usage. To apply what has been learned or mined is the ultimate goal of BI. It is at this point where we see our return on investment.

V. CONSTRUCT A BETTER BI SYSTEM

Business Intelligence initiatives of most organizations lack the maturity and breadth of deployment which need to meet business demands. Gartner group refers to this as the Business Intelligence gap. [14] Though it is considered that some trends are dramatically driving the market need for better BI tools, e.g. daily rising data volumes, geographically dispersed users, and complex existing tools. However, there are still several factors for the difficulties of broad distribution of enterprise Business Intelligence.

Most Business Intelligence applications have to take a long time to install, build and deploy. The average implementation time for some larger BI solutions is between about three to six months. Unfortunately, this timeframe is almost as long as the initial implementations for the transactional systems that BI is expected to improve. Requirements and budgets often change after a long installation and implementation cycle. Many Business Intelligence applications are still difficult to use. A majority of the effort of any BI project is focused on implementation, for the professional thinking, but adequate
user training is often overlooked. As a result, end user acceptance is the important reason that hampers BI to be deployed broadly. In most cases, BI has actually increased the workload while BI was originally conceived as a way to relieve workload through intuitive reporting and analysis. Most BI applications require significant end user training, and are not intuitive enough to allow end users to ever attempt to create their own analysis. A BI application that was intended to reduce costs and reduce the workload whereas actually increases costs and workloads. This eventually limits wider deployment of Business Intelligence throughout the enterprise. Further more, the cost and benefit also come into question. After a rather lengthy and costly implementation, demands have changed. Once the applications cannot demonstrate a return on investment, or few benefits are realized, the end users are about to disenchant with Business Intelligence.

Business Intelligence delivers greater returns as the number of people who use it, and the frequency with which they use it increases. True “Business-wide” intelligence would definitely ensure that users who needed access and analysis of information to support a business process or a decision would have a powerful yet intuitive solution or platform at their fingertips. Some Proposals which define this type of Business Intelligence solution or platform are:

The BI solution should be broadly distributed to all users who need access to information. The more people using a technology, the more valuable the technology becomes. The BI platform is willing to be easier to use and understand, and allow the user to evaluate alternatives, draw conclusions and make decisions. The tools and techniques to access and analyze the information must be powerful yet easy to learn and to use. To be truly effective, a Business Intelligence solution should help users draw conclusions or support a business decision. This can only happen if the information is easy to understand, is timely and is relevant to the user.

Only limited by security or authorization reasons, user should have access to more resources necessary to perform his or her job. For true insight and effectiveness, understanding of data across boundaries helps the user make the business more productive. Analysis tools of Business Intelligence Platform should be powerful yet simple for users to learn, to deploy and to maintain. These solutions need to be flexible and adaptable to changes in the on-demand business environment. Broader deployment of BI throughout the enterprises will only occur if users can learn an application, deploy it and manage it effectively.

VI. CONCLUSION

Though the work is part of a set of related research piece, it still can be explicitly concluded that a good Business Intelligence system should provide not only the strong capabilities and the tools to analyze what happened, but more importantly to help a user understand what is going to happen. Business Intelligence is about creating answers to analyze new questions. Suppose the criteria are achieved, Business Intelligence can become a ubiquitous tool that every person uses every day, much like the familiar desktop applications Microsoft Word and Excel.

REFERENCES