New Trends in Business Intelligence

Matteo Golfarelli
DEIS
University of Bologna
Via Sacchi 3, Cesena (FC) - 47023 - Italy
Phone: +39-0547-338862 Fax: +39-0547-338890 E-mail: golfare@csr.unibo.it

Abstract - Today, as DWing reached a high level of efficiency, new opportunities in exploiting information coming from the operational databases are requested by users. Consequently, we believe that in the next future the interests of researchers will be more and more oriented to BI applications while DW will play the role of an on the shelf module within a large set of resources placed at the knowledge workers disposal with the aim of exploiting at best the information encrypted in the operational data. The aim of this paper is to sketch our idea about the deve lop prospects for BI and to report in detail one of the most promising evolution in this direction: Business Performance Management.

I. INTRODUCTION

BI, that can be defined as *the process of turning data into information and then into knowledge*, was born within the industrial world in the early 90's, to satisfy the managers' request for efficiently and effectively analyzing the enterprise data in order to better understand the situation of their business and for improving the decision process. In the mid-90's BI became an object of interest for the academic world, and ten years of research managed to transform a bundle of naive techniques into a well-founded approach to information extraction and processing. In particular, most of the researches addressed the specific area of Data Warehouses (DWs) that are large repositories of historical data, organized according to the multidimensional model, that are directly accessed by the final users (i.e. the managers) through user friendly interfaces that allow them to carry out very detailed analyses. The main results obtained on topics such as OLAP [20], multidimensional modeling [18], design methodologies [14] and optimization techniques [25] converged to define the modern architecture of data warehousing systems and were absorbed by vendors to form a wide set of on-the-shelf software solutions.

Today, DWing is mature field from many points of view: the *users* understood the potential of multidimensional analysis and are fully exploiting OLAP capabilities; the *software vendors* made available complete suites of products covering the whole DWing process from ETL extractors, to friendly interfaces, through specialized DBMSs; the *researchers* explored most of the DW aspects at conceptual, logical and physical levels. Becoming aware of this is fundamental for preparing to the next era of decision support systems and the researchers already started to discuss about the next steps in this field. In particular, within a Dagsthul seminar titled "Data Warehousing at the Crossroads" [9], many researchers from all over the World analyzed the state of the art in DWing and more in general in BI and tried to foreseen which is the trend for these areas. Most of the participants agreed that the DWing already reached a high level of efficiency and thus does not provide many research opportunities. On the other hand, DW is at the core of BI and in this broaden area the exploitation of information is still limited and many applications remain to be evaluated or even discovered. Consequently, we believe that in the next future BI will attract the interest of both researchers and users that consider DW as a (crucial) on the shelf module within a large set resources placed at the knowledge workers disposal with the aim of exploiting at best the information encrypted in the operational data.

Understanding which of the several issues to be addressed in BI are the most promising ones is not easy but we believe that the solution can be found only paying attention to the user requests that from the beginning guided the choices in the DW pragmatic field.

The aim of this position paper is to sketch our idea about the develop prospects for BI and to report in detail one of the most promising evolution of the classic DW concepts: *Business Performance Management* (BPM). BPM helps organizations to optimize business performance by encouraging process effectiveness as well as efficient use of financial, human, and material resources. BPM includes DW but it also requires a brand new set of solutions that rely on different technologies and deeply impact on the overall architecture of the BI platform [11]. Due to the requirements, both technological and economical, to be fulfilled to successfully realize BMP, it represents an effective example in evolving from the DW to BI applications.

The rest of the paper is organized as follows: Section II reports some important uncovered issues in the area of DW and BI, Section III proposes the BPM scenario while Section IV describes an architecture supporting BPM and discusses the related research issues.

II. EMERGING RESEARCH ISSUES

In this section we propose some upcoming goals in the area of BI and DW that emerge from the user requirements. We believe that each requirement is strictly related to uncovered research issues that we will discuss separately for DW and BI

A. Data Warehousing

Today, as several mature implementations of data warehousing systems are fully operational within medium to large contexts, the continuous evolution of the application domains is bringing to the forefront the dynamic aspects related to describing how the information stored in the DW changes over time both from the intentional and from the extentional point of view. So far, research has mainly addressed the second case rather than an evolution of DW schemata, since their handling can be done within the classic star schema by using the so-called *slowly-changing dimensions* [18]. Some commercial systems already allow to track changes in data and to effectively query cubes based on different temporal scenarios. For instance, SAP-BW [24] allows the user to choose which version of the hierarchies to use while querying (e.g., aggregate the sales according to the categories that were true on 1/1/2000).
On the other hand, schema versioning in DWs has only partially been explored [3][5] and no dedicated commercial tools or restructuring methodologies are available to the designer. Thus, both an extension of tools and a support to designers are urgently needed. The DW schema may change in response to the evolving business requirements. New properties and measures may become necessary (e.g., a subcategory property could be added to allow more detailed analysis), while others may become obsolete. Even the set of dimensions characterizing a cube may be required to change. Note that, in comparison with operational databases, temporal issues are more pressing in DWs since queries frequently span long periods of time; thus, it is very common that they are required to cross the boundaries of different versions of data and/or schema. Besides, the criticality of the problem is obviously higher for DWs that have been established for a long time, since unhandled evolutions will determine a stronger gap between the reality and its representation within the database, that will become obsolete and useless very quichly.
A second recurrent problem that is emerging within companies that experienced DW projects is the absence of any standard modeling techniques and design methodology. This lack, typical of poorly engineered systems, is emphasized whenever data marts of the same company but designed by different groups (or external software houses) must coexist or even when different DW are integrated in consequence of mergers and acquisitions. In these cases projects will be slacken by the difficulty in gathering the previous know-out. The absence of standard modeling formalisms is also reflected in the absence of standard meta-data that prevents from easily migrate from different software platform. Differently from the schema evolution problem, design issues have been largely studied and an impressive set of models for describing the main aspects of the DWing process (conceptual [14], logical [29], ETL [30], data and schema quality[6]) is available in the literature. It should be noted that most of the authors agree on the main features of the formalisms but nonetheless the CWM proposal [8] is the only attempt to overcome such heterogeneity, and it is a fact that, till now, it has been very limitedly absorbed by the market.
Finally, the interest of researchers have been recently attracted by applying OLAP on non-conventional/complex data types like geographical and Biological data.

B. Business Intelligence

BI represents the step beyond DWing, it includes classic DW but requires contributions from many other research fields like economy and artificial intelligence. We believe that the two DW features that mainly limit the classic architecture are (1) the batch-update mechanism and (2) the limits in analysis capabilities induced by OLAP. Overcoming these problems requires a broad rethinking of the architecture that leads to new solutions that requires new research issue to be addressed.
In DW systems data update is usually carried out monthly, weekly or even daily when the system is off-line; this is an adequate frequency when data are used for long-medium period analyses and offers the advantage that ETL procedures can be launched on the operational system when no user queries are running. On the other hand daily updates are unacceptable whenever an immediate decision must be kept. This is the case for many applications in the industrial world where the decision-keepers is, for example, the head of an assembling line that is in charge of deciding how many factory workers should be assigned to the different activities according to the information related to the quantity of products already manufactured, the assembling line speed and so on. In that case, streams of data coming from the machineries must be continuously collected and directly loaded in the DW. Obviously, these requirements can hardly fit the current DW architecture and a lot of work must be done in order to allow OLAP on data streams to come true.
OLAP represents, since fifteen years, the perfect trade-off between expressiveness and usability. Its principal asset is the multidimensional model, on which it relies, and that represents a very simple metaphor for group-by query formulation and fruition. On the other hand, OLAP shows very serious limits when more general querying operators are needed. This happens whenever the goal of the analysis is changed for example in order to carry out data mining and what-if analyses; these applications till now had a limited success on the market due the requirement for a strong theoretical background necessary to set up an analysis or a forecast model. While the expertise lack still characterizes the users, the need for more powerful analyses is increased. In fact, the changes in business management imposed by the market makes critical the adoption of scorecard system and more in general the use of complex business models that require complex analyses to be built and tuned. On the other hand, the existence of such models make feasible forecasting using what-if analysis. This trend is confirmed by the interest of several software vendors: for example SAP has recently added to its suite the SEM module [24] that, based on the DW data, supports both inductive and deductive what-if analysis. In an inductive approach the forecasting model is based on

the observations of the behavior of the system in a given period (i.e. it is based on the analysis of the time series obtained from DW data) and no intensional description is derived; on the other hand in a deductive approach the forecasting model is intensionally defined by describing components of the model. As concern the second approach, one of the most promising technique to be considered is based on the system dynamic theory [12] that is particularly useful to model complex feedback systems whose models cannot be derived studying the single components since their interaction cannot be simply obtained as the "sum" of the single component behaviors. System dynamics has been largely adopted in the economic field but have been only recently considered within BI systems [23].

III. BUSINESS PERFORMANCE MANAGEMENT

In the following sections, we will describe BPM that, according to Gartner group [11], represents one of the most promising applications of BI on the top of DW. The BMP example summarizes all the needs that are causing the change from DW to BI. Besides describing the BPM framework, we will also propose a possible architecture and the related research issues [15].
The increase in the competition on the markets changed in the last ten years the approach to business management. Today companies are more process-oriented than in the past [1]; in fact, in order to reduce the costs and keep pace with the market, they are adopting an end-to-end strategy that involves both customers and suppliers to synchronize all the business activities. At the same time, companies have understood the importance of enforcing achievement of the goals defined by their strategy through metrics-driven management [28]. Thus, the new requirement of managers is to ensure that all processes are effective by continuously measuring their performance through Key Performance Indicators (KPIs) and score cards [17]. Communication and enforcement of the strategy is obtained by sharing goals and measurements at all the company levels, thus promoting the so-called *information democracy*. Translating the company strategy into a detailed set of indicators that are closer to the operational tasks allows employees to better understand the *desiderata* of managers.
As stated before the neologism often used to refer to this new picture in BI is exactly BPM. Describing BPM [21] requires to understand how management is carried out within a process-oriented enterprise where, beside the classical organizational structure, a set of inter-division processes are present. The organizational structure is a hierarchy of divisions, aimed at defining their duties and responsibilities, and is usually organized on three different levels. At the *strategic level*, the global strategy of the enterprise is decided. The *tactical level* is usually composed by multiple divisions, each controlling a set of functions; the decisions taken here are related to the corresponding functions and must comply with the strategy defined at the upper level. Finally, at the *operational level*, the core activities are carried out; the decision power is limited to optimizing the specific production activities in accordance with the main strategy. On the other hand, a *process* identifies a set of logically related tasks performed to accomplish a defined goal. Processes are orthogonal to organizational structure, in fact they usually include tasks carried out by different divisions and require decisions at different levels.
The key point of processes is that the focus is on the global business goals rather than on the single tasks. Of course, employees involved in processes must share the business strategy in order to synchronize their behavior. This result can be achieved by translating the top-level strategy into multiple goals at the lower levels, each defined by a target value for a given indicator; each indicator measures a specific task and should be easily understood by the employer who is in charge. This approach, depicted in Fig. 1, is based on a closed-loop where:

1. the strategy and the corresponding targets on indicators are influenced by the enterprise performance as inferred from the information system;
2. the actions/decisions taken at the tactical and operational levels are aimed at matching current and target values for indicators;
3. the actions/decisions fulfill the company strategy and determine its performance.

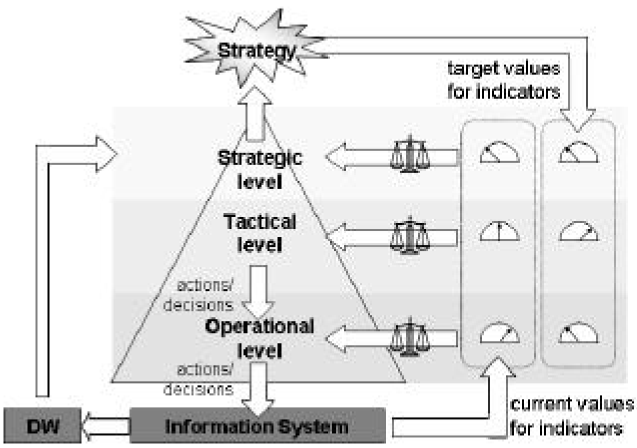

Figure 1. The closed-loop in the BPM approach

Note that, while a business strategy is with no doubt more than a simple set of target values, the attempts made until now to share strategy policies and directives among other levels failed owing to how every single employee perceives the company. At least KPIs allow managers to get results without misunderstandings and personal definitions, while it resulted that implementing behavioral business rules or application code limits the autonomy of the employees with potential loss of flexibility.
The term BPM defines this new approach to management and requires indicators to be constantly fed and made available at the on time, at the proper decision level in the best form. The peculiar features that distinguish BPM from classical DW-based BI are:

- *Users:* the users of BPM systems are still decision-makers, but at the tactical and operational levels. These users have limited view of the company strategy, and only have to deal with the subset of indicators related to their specific tasks.
- *Delivery time*: Decisions at the lower levels must be faster then the strategic ones, thus the freshness of information must be set accordingly. BPM systems are not supposed to operate in real-time, but rather in *right-time*, meaning that it is crucial for information to be fresh enough to be useful for decision making [11].
- *Information coarseness and lifetime*: information circulated in BPM systems is usually more detailed than in DW systems, since it concerns single events related to specific tasks. Besides, lifetime of information required by BPM is limited, since users are interested in the current performance of their tasks. Such characteristic leads to considering data streams as potential sources. The state of an automated assembling line or the performance of the stock exchange may be definitely part of the input for a BPM system. Finally, the high dynamicity of information encourages to resort to rule engines and mining techniques for identifying outliers and remarkable business situations.
- *User interface*: tactical and operational decision-makers will not probably have time and skills to run OLAP sessions, hence, information will be mainly accessed in the form of reports and dashboards carrying the relevant indicators, as well as through automated alerts activated by business rules.

IV. BUSINESS ACTIVITY MONITORING

Obviously the framework outlined so far is only partially covered by the DW process that essentially helps managers to understand their companies by supporting bottom-up extraction of information from data, thus lacking in enforcing the company strategy in a top-down fashion. At the moment, the BPM solutions proposed by software vendors mainly couple classical OLAP tools with some specialized ETL and data integration systems [16], [26].

An architectural sketch for a complete BPM solution is proposed in Figure 2. The left side of the figure shows the classical DW architecture: an ETL tool extracts data from the operational data sources and cleans/transforms/integrates them into an *Operational Data Store* (ODS); data are then loaded from the ODS into the DW, accessed by reporting and OLAP tools. On the right side of the picture, the architecture is completed by a reactive data flow, more suited for monitoring the time-critical operational processes. The technology implementing this flow is often called *Business Activity Monitoring* (BAM) [10].

The main components introduced by BAM are:

- a *Right-Time Integrator* (RTI) that integrates at right-time data from operational databases, from the DW, from Enterprise Application Integration (EAI) systems, and from real-time data streams;
- a *Dynamic Data Store* (DDS), that is a repository capable of storing short-term data for fast retrieving, to support rule inference and mining;
- a *KPI manager* that computes all the indicators necessary at the different levels to feed dashboards and reports;
- a set of *mining tools*, capable of extracting relevant patterns out of the data streams;
- a *rule engine* that continuously monitors the events filtered by the RTI or detected by the mining tools to deliver timely alerts to the users.

DW and BAM together implement the closed loop on which BPM is based:

1. The strategic management analyzes the medium- and long-period trends through OLAP tools and is enabled to quantify the effectiveness of the strategy pursued in the short period by KPIs and dashboards.
2. Tactical and operational decision-makers, in turn, use other KPIs and dashboards to direct and tune their actions and decisions according to the company strategy.
3. Finally, alerts allow the unexpected events occurring at all levels to be monitored and reactively managed.

With reference to the architecture outlined in Figure 2, in the following subsections we discuss the research and technological issues we consider more relevant.

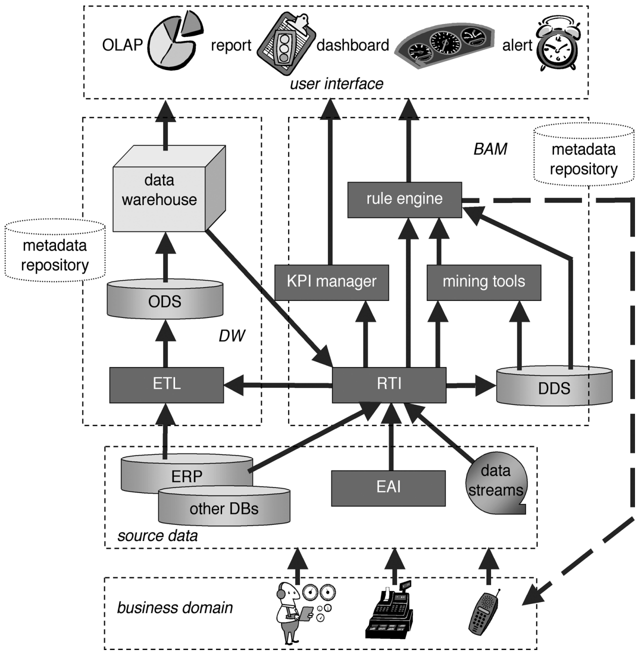

Figure 2. A complete architecture for BPM

A. Data latency

Data latency is the interval between the time an event occurs and the time it is perceived by the user. BAM emphasizes the need for reducing data latency by providing a tool capable of right-time filtering/ cleaning/ transforming/ integrating the relevant data coming from OLTP/OLAP databases as well as from data streams. In practice, in most cases this requires to abandon the ODS approach typically pursued in DW systems and to adopt on-the-fly techniques, which raises serious problems in terms of data quality and integration. In fact, while on-the-fly integration by query rewriting on heterogeneous sources has been widely investigated and in some cases implemented in research prototypes (see [4], for instance), still:

- most of the cleaning techniques devised so far (e.g. purge/merge problem [2] and duplicate detection [22]) rely on the presence of a materialized integrated level; we expect that, in its absence, some of these techniques can be modified to be re-implemented on proper data structures in main memory while others cannot be applied at all.
- manipulating data stream still presents many technical challenges: complex queries over the data are performed in an offline fashion, and real-time queries are typically restricted to simple filters [19].

B. Informative power

The informative power of a BPM system is mainly related to the types of rules and indicators supported.

As concerns business rules, we believe that the ECA paradigm (Event-Condition-Action) will provide the best trade-off between effectiveness and simplicity for the industrial context. In fact, though more powerful solutions exist (one might be even tempted to implement a dedicated expert system), providing and managing very complex business rules would probably discourage most enterprise users.

As to indicators, while different approaches have been devised in the business economics field and are widely spread and appreciated in the industrial context [17], the BI community has only marginally faced the problems related to their modeling and handling [7]. An interesting issue on this subject is related to the need for defining a consistent set of indicators, which requires techniques for simulating how indicators are related and affect each other. Some works in this direction have been carried out in the fields of budgeting and what-if analysis: while the first assumes a tree-based hierarchy between indicators, the second does not consider any predefined relationship between indicators, thus requiring the effects of correlations to be manually defined. In the BPM context, indicators are defined at different level of detail and are related to each other according to a graph, induced by the constraints on the structure of both the organization and the processes. A further research issue is related to the definition of the KPI target values, that should be based on the historical data stored in the DW also considering the forecasts made by managers. Also the tuning of these values requires a complex set of simulations.

The events monitored by the rule engine should not be restricted to those obtained from the ETL process, they might also be associated to relevant patterns more deeply hidden in the input data streams. In order to let such patterns emerge, BAM could take advantage of mining tools, particularly those oriented to time-series analysis. Though most techniques devised over the years for this purpose are made inapplicable by the right-time constraint, there is some on-going research on real-time data mining and mining applied to streams (see, for instance, the work on high-performance time series mining in [31]).

Though indicators and rules usually describe short-term information, they may achieve higher flexibility by relying on some history of data: for instance, a notable event may occur when the sensor readings are over the threshold for 50% of the time during the last minute. Thus, the problem of storing data for fast retrieving arises; for this reason the BPM architecture includes the DDS component. Simple buffering techniques will not be appropriate in this context, since data will be accessed in different ways by several services concurrently running on the architecture (e.g. by the KPI manager, the rule engine, the mining tools). Indeed, it seems that the most promising technology to deal with this issue is that of *main-memory databases* or *real-time databases*, that guarantee appropriate performances and high reliability [13].

C. Interface

As sketched in Figure 2, interaction with the user for a BPM architecture will be organized around different paradigms, seamlessly merged into a common interface. The classical paradigms of DW systems, namely reporting and OLAP, will still be present, though static reports will be integrated with KPIs to give users a full picture of the trend of their business in the short- and medium-time. Even dashboards will include KPIs, but there the information latency will be shorter in order to allow users to monitor the progress of their tasks at right-time. Finally, alerts will be quickly delivered to enable users to timely react to the relevant events.

IV. CONCLUSION

In this position paper, we summarized which are, in our opinion, the most promising development of DW and BI in the medium long term as emerging from the requirements of modern companies. From that picture it appears clear that users ask for more reactive architectures and more powerful querying tools. An emerging application that represents an effective example of such combination is BMP that involves different sophisticated technologies, such as real-time data mining, main-memory databases, and stream processing. Most of this fields are not mature enough in terms of commercial products, but all of them are object of a lively research activity, which promises that the most relevant issues will be solved soon.

We close the paper claiming that the need for a new generation of BI tools is strongly felt but many research issues must be further explored before creating a comprehensive solution.

References

[1] K. Baina, S. Tata, and K. Benali. A Model for Process Service Interaction. In *Proc. 1st Conf. on Business Process Management*. The Netherlands, 2003.

[2] R. Baxter, P. Christen, and T. Churches. A Comparison of Fast Blocking Methods for Record Linkage. In *Proc. Workshop on Data Cleaning, Record Linkage and Object Consolidation*. USA, 2003.

[3] B. Bebel *et al.*. Creation and management of versions in multiversion data warehouse. In *Proc. SAC*. Cyprus, 2004.

[4] B. Beneventano *et al.* Information Integration: The MOMIS Project Demonstration. In *Proc. VLDB*. Egypt, 2000.

[5] M. Blaschka. *FIESTA - A framework for schema evolution in multidimensional databases*. PhD thesis, Technische Universitat Munchen. Germany, 2000.

[6] M. Bouzeghoub and Z. Kedad. A quality-based framework for physical data warehouse design. *In Proc. DMDW*. Sweden. 2000

[7] L. Buglione, and A. Abran. Balanced Scorecards and GQM: what are the differences? In *Proc. FESMA/AEMES Conf.*. Spain, 2000.

[8] OMG. Common Warehouse Metamodel Specification. http://www.omg.org/technology/documents/formal/cwm.htm, 2004.

[9] J. Hammer, M. Schneider and T. Sellis . *Data Warehousing at the Crossroads*. Dagsthul Seminar 04321. Germany, 2004.

[10] H. Dresner. Business Activity Monitoring: BAM Architecture. *Gartner Symposium ITXPO*. France, 2003.

[11] H. Dresner and F. Buytendjik. Business Intelligence and Data Warehousing Scenario: Key Trends and Evolving Markets. *Gartner Symposium ITXPO*. France, 2003.

[12] J. Forrester. *Industrial Dynamics*. MIT Press, 1961.

[13] H. Garcia-Molina and K. Salem. Main Memory Database Systems: An Overview. *IEEE TKDE* 4:6, 1992.

[14] M. Golfarelli D. Maio and S. Rizzi. The Dimensional Fact Model: A Conceptual Model for Data Warehouses. *Int. J. Cooperative Inf. Syst.* 7:(2-3), 1998.

[15] M. Golfarelli S. Rizzi and I. Cella. Beyond data warehousing: What's next in business intelligence? In *Proc. DOLAP 2004*, USA, 2004.

[16] Istante Software. Istante: Product Overview. http://www.istantesoftware.com

[17] R.S. Kaplan and D.P. Norton. The Balanced Scorecard – Measures that Drive Performance. *Harvard Business Review* 70:1, 1992.

[18] R. Kimball *et al.*. *The data warehouse lifecycle toolkit*. John Wiley & Sons, 1998.

[19] N. Koudas and D. Srivastava. Data Stream Query Processing: A Tutorial. In *Proc. VLDB*. Germany, 2003.

[20] A. Maniatis *et al.*: Advanced visualization for OLAP. DOLAP. USA, 2003.

[21] W. Martin and R. Nußdorfer. *"Pulse Check": Operational, Tactical and Strategic BPM.* CSA Consulting GmbH White Paper, 2003.

[22] A. Monge and C. Elkan. The field matching problem: Algorithms and applications. In *Proc. KDD-96*. USA, 1996.

[23] PowerSim Software. http://www.powersim.com/

[24] SAP. *Business Planning And Simulations (SEM-BPS).* 2003.

[25] Y. Sismanis et al. Dwarf: shrinking the PetaCube. In *Proc. SIGMOD Conf.*. USA, 2002.

[26] D. Sonnen and H. Morris. *BusinessFactor: Event-Driven Business Performance Manager*. TIBCO White Paper, 2004.

[27] P. Stolorz *et al.* Fast Spatio-Temporal Data Mining of Large Geophysical Datasets. In *Proc. KDD*. Canada, 1995.

[28] K.E Sveiby. *The New Organizational Wealth: Managing and Measuring Knowledge Based Assets*. Berret Koehler Publishers, San Francisco, CA, 1997.

[29] D. Theodoratos and T. Sellis. Designing Data Warehouses. *Int. J. DKE* 31:3, 1999.

[30] P. Vassiliadis et al. A Framework for the Design of ETL Scenarios. In *Proc.* CAiSE. Austria, 2003.

[31] Y. Zhu. *High Performance Data Mining in Time Series: Techniques and Case Studies*. PhD Thesis, New York University, 2004.