Beyond Data Warehousing: What's Next in Business Intelligence?

Matteo Golfarelli
DEIS - University of Bologna
Viale Risorgimento, 2
Bologna, Italy
+39-051-2093542
mgolfarelli@deis.unibo.it

Stefano Rizzi
DEIS - University of Bologna
Viale Risorgimento, 2
Bologna, Italy
+39-0547-338862
srizzi@deis.unibo.it

Iuris Cella
Gruppo Formula S.p.A.
Via Matteotti, 5
Villanova di Castenaso, Bologna, Italy
+39-051-6002111
iuris.cella@formula.it

ABSTRACT

During the last ten years the approach to business management has deeply changed, and companies have understood the importance of enforcing achievement of the goals defined by their strategy through metrics-driven management. The DW process, though supporting bottom-up extraction of information from data, fails in top-down enforcing the company strategy. A new approach to BI, called Business Performance Management (BPM), is emerging from this framework: it includes DW but it also requires a reactive component capable of monitoring the time-critical operational processes to allow tactical and operational decision-makers to tune their actions according to the company strategy. The aim of this paper is to encourage the research community to acknowledge the coming of a second era in BI, to propose a general architecture for BPM, and to lay the premises for investigating the most challenging of the related issues.

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Types of systems – *decision support*

General Terms

Management, Design.

Keywords

Architectures, Business metrics, Business processes.

1. INTRODUCTION

Business Intelligence (BI) can be defined as *the process of turning data into information and then into knowledge*. Knowledge is typically obtained about customer needs, customer decision making processes, the competition, conditions in the industry, and general economic, technological, and cultural trends. BI 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 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. Eventually, the main results obtained on topics such as OLAP, multidimensional modeling, design methodologies, optimization and indexing techniques converged to define the modern architectures of data warehousing (DW) systems, and were absorbed by vendors to form a wide set of on-the-shelf software solutions.

Today, companies and managers are beginning to ask IT vendors for new tools capable of handling the changed business scenario. In fact, during the last ten years the approach to business management has changed from both the technological and the organizational points of view. On the technological side, outsourcing the information system has progressively become the keyword to cut fixed costs: no investments are required and only the actual services are paid. This approach hampers the usage of DWs, since analysis becomes a direct cost. On the organizational side, companies are now 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 [17]. 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 [10]. 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.

Obviously the framework outlined impacts all levels of BI platforms, since it affects the type of data collected, the way information is extracted and distributed as well as its lifetime and freshness. The DW process covers only part of this framework; in fact, it 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. Bridging this gap marks a turning point in the history of BI, that is

DOLAP'04, November 12–13, 2004, Washington, DC, USA.

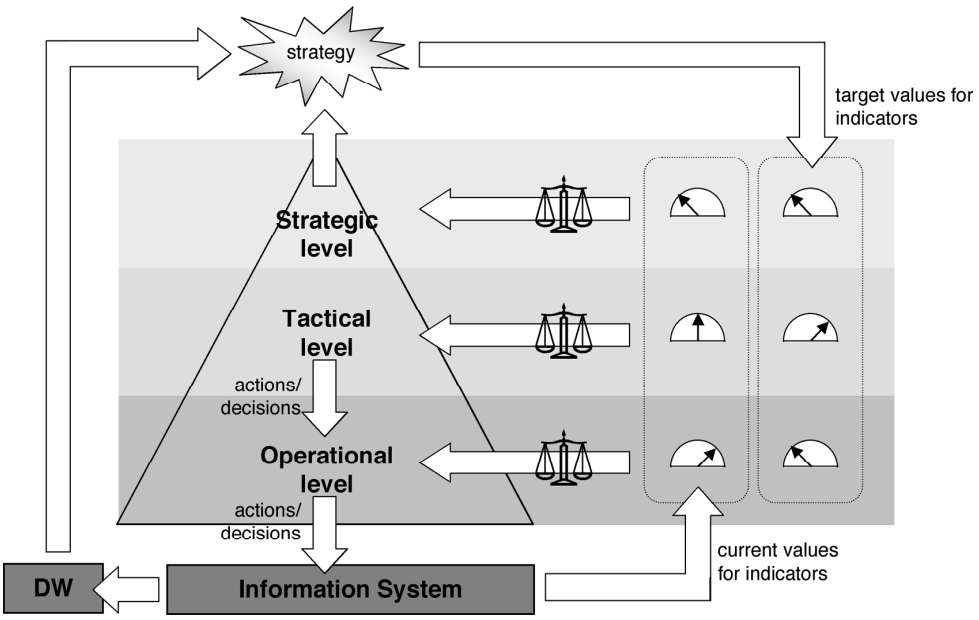

Figure 1. The closed-loop in the BPM approach

no more perceived as a set of techniques for information extraction and processing, but also as an active and concrete approach to business management. Though the convergence between business management and information technology was in progress even before '90s, this is a relevant milestone in their unification.

The neologism often used to refer to this new picture in BI is *Business Performance Management* [1] (BPM), that can be defined as a set of processes that help organizations 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 [7]. At the moment, the BPM solutions proposed by software vendors mainly couple classical OLAP tools with some specialized ETL and data integration systems [9], [15]. On the other hand we believe that, given the significance of the change, mere rearrangements of the previous solutions will turn out to be ineffective, and that a global rethinking of methodologies, models, and techniques will be required.

The aim of this paper, that was conceived by facing research mentality with industrial experience, is to encourage the research community to acknowledge the coming of a second era in BI, to propose a general architecture for BPM, and to lay the premises for investigating the related research issues. Thus, after describing in Section 2 the BPM scenario by identifying the key concepts from the economical/ organizational and the technical points of view, in Section 3 we sketch an architectural framework for BPM and discuss the main research area that will be involved. Finally, in Section 4 we draw the conclusions and briefly discuss the impact BPM technology is expected to have on enterprises.

2. THE BPM APPROACH

Describing BPM [12] 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 is 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 Figure 1, is based on a closed-loop where:

- the strategy and the corresponding targets on indicators are influenced by the enterprise performance as inferred from the information system;

[1] Synonyms for BPM are *Corporate Performance Management* (CPM) and *Enterprise Performance Management* (EPM). Note that the term BPM is also used as an acronym for Business Process Management, that is the process-based approach to management discussed above [18].

- the actions/decisions taken at the tactical and operational levels are aimed at matching current and target values for indicators;
- the actions/decisions fulfill the company strategy and determine its performance.

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 right 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 [7].
- *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.
- It should be now clear that DW technology covers only partially BPM requirements. With reference to Figure 1, we might say that DW is used by the top management to understand the enterprise and to define the global strategy, while other techniques must be used by tactical and operational decision-makers to “absorb” the strategy and make the best decisions for their tasks.

3. TECHNOLOGICAL CHALLENGES AND RESEARCH ISSUES

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) [6].

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. In some cases, events may trigger actions that create a direct feedback to databases (e.g., automatically re-order an item when out-of-stock) or to business (e.g., tune some production parameters).

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

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 [3], for instance), still:

- most of the cleaning techniques devised so far (e.g. purge/merge problem [2] and duplicate detection [13]) rely on the presence of a materialized integrated level; we expect

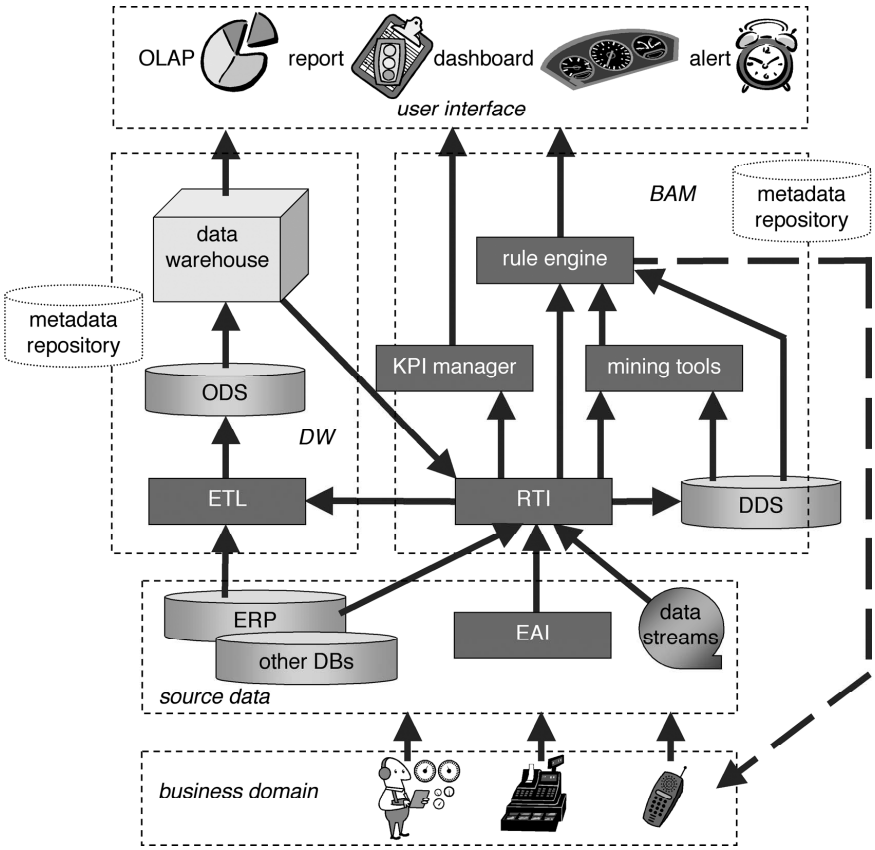

Figure 2. A complete architecture for BPM

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 [11].

Building the Right-Time Integrator for on-the-fly integration is an interesting challenge related to BAM. At the current state of technology, the most promising candidates to this end seem to be the so-called Enterprise Application Integration systems, capable of making different applications talk to each other in real-time on a common message bus through a publish/subscribe mechanism, thus enabling integration of OLTP systems at the application level with no ODS support (see [5], for example). On the other hand, the existing tools still lack in effectiveness and flexibility in profiling data coming from custom, non-standard sources.

An emerging standard that could help in this direction is the *Model Driven Architecture* (MDA) developed by OMG. In this approach, aimed at achieving interoperability in different contexts and at supporting quickly evolving requirements, a *platform independent model* is used to expresses business functionality and behavior independently of the underlying middleware architectures and technologies [14] .

3.1 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 [10], the BI community has only marginally faced the problems related to their modeling and handling [4]. 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 directly signaled by the EAI, 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 [19]).

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 [8].

3.2 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.

Table 1 summarizes the main features of the four interaction paradigms.

3.3 Design

As one might expect, the crucial issues emerging in BPM design are different from those arising in the classical DW context:

- *Right-time*. First of all, the BPM designer should take great care over determining what is the meaning of right-time for the specific business domain. This issue confines all other architectural choices to the background: if some piece of information cannot be delivered at right-time, it is useless in the BPM context and should not be supported.
- *Light architecture*. New pursued strategies will bring to the foreground new functions and behaviors, that will be monitored by new indicators. Thus, requirements for KPIs and business rules will change quickly, and light architectures will be desirable.
- *Process design*. While DW design requires understanding and integrating operational data, in BPM a leading role is played by processes. Hence, BPM design also requires to understand business processes and their relationships in order to find out the relevant indicators and rules, and to determine where the data to compute them can be found. In particular, capturing and modeling the relationships between different indicators has a primary role in this phase to ensure that effective and reliable information is delivered.

4. CONCLUSION

In this paper, we summarized the requirements emerging from modern companies and discussed how they meet into a new architecture, called BPM, that promises to lead business intelligence beyond data warehousing. DW systems led to quantify business information, to make it promptly available and certified. On the other hand, the role of BPM is to quantify the enterprise strategy and targets, in order to decentralize decision making. DW is not enough to this end since its technology is neither suitable for the grain nor for the freshness of the collected information, that should quickly flow throughout the different levels of the company.

We saw that different sophisticated technologies, such as real-time data mining, main-memory databases, and stream processing are involved in BPM. 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 by adding a few lines to describe what kind of enterprise we envision behind BPM technology. The main target of the top-down approach to management is to accomplish goals predetermined by the *whole* management group. This will require a significant horizontal (i.e., inter-divisions) sharing of information, which is currently not largely practiced (or even discouraged!) in several companies. On the other hand, the risk of this approach is that the creativity and initiative of individuals, that represents an invaluable assets of several small/medium companies, will be depressed and left out from the decision process. As to the impact on the enterprise information system and IT infrastructures, we have to consider that deploying a true BPM solution requires a strong system integration at all levels; thus, IT investments should be directed towards open-platform tools based on standard protocols.

Table 1. Interfaces for BPM

Interface	*Structure*	*Freshness*	*Interaction*	*Information*
report	static	short-time	pull	measures/indicators
OLAP	dynamic	short-time	pull	measures
dashboard	static	right-time	pull	indicators
alert	static	right-time	push	event

5. REFERENCES

[1] Baina, K., Tata, S., and Benali, K. A Model for Process Service Interaction. In *Proceedings 1st Conference on Business Process Management* (EindHoven, The Netherlands, 2003).

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

[3] Beneventano, D., *et al.* Information Integration: The MOMIS Project Demonstration. In *Proceedings VLDB Conference* (Cairo, Egypt, 2000).

[4] Buglione, L., and Abran, A. Balanced Scorecards and GQM: what are the differences? In *Proceedings FESMA/AEMES Conference* (Madrid, Spain, 2000).

[5] Charlesworth, I. *EAI Technology Audit - Vitria BusinessWare 4*. Butler Group Report, 2002.

[6] Dresner, H. Business Activity Monitoring: BAM Architecture. *Gartner Symposium ITXPO* (Cannes, France, 2003).

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

[8] Garcia-Molina, H., and Salem, K. Main Memory Database Systems: An Overview. *IEEE Trans. on Knowledge and Data Engineering 4,6* (1992).

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

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

[11] Koudas, N., and Srivastava, D. Data Stream Query Processing: A Tutorial. In *Proceedings VLDB Conference* (Berlin, Germany, 2003).

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

[13] Monge, A., and Elkan, C. The field matching problem: Algorithms and applications. In *Proceedings KDD-96* (Oregon, 1996).

[14] OMG. MDA Guide Version 1.0. http://www.omg.org/mda/

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

[16] Stolorz, P., *et al*. Fast Spatio-Temporal Data Mining of Large Geophysical Datasets. In *Proceedings KDD* (Montreal, Canada, 1995).

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

[18] Weske, M., van der Aalst, W., and Verbeek, H. Advances in Business Process Management. *Data & Knowledge Engineering 50,1* (2004).

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