How Business Analytics Should Work

Marco Antonio Villegas-García, Fausto Pedro García Márquez, and Diego José Pedregal Tercero

1 A World in Transformation

We are now approaching the so called Information Era. The globalization came to world and the countries connected together, the people's life changed as well as their cultures. As Information and Communication Technologies (ICT) spread all around the world, the digital universe grows astonishingly. It is estimated that the amount of data produced in the world grows at an astounding annual rate of 60 % [1].

The volume of transactions and interchanged data is reaching astronomical scale [2]: IBM estimates that humanity creates 2.4 quintillion bytes (a billion billion) of data everyday. Much of this data is created by digital systems usually linked to internet. International Data Corporation estimates that the digital universe will double in size through 2020 and reach 40 ZB (zetabytes), which means 5,247 GB for every person on earth in 2020. The *digital behavioural universe* is being created from the clickstream and the digital footprints of every person across Earth interacting, participating and consuming this data. At this respect, one of the bigger trends that most drive the behavioural dimension of the digital universe is the mobile computing. Actually, six billion of the world's seven billion people have access to mobile phone, what means that, by far, it is the largest service infrastructure across the world.

Some authors refer to this revolution as a hinge of history [3], highlighting the fact that the humanity is entering in a new scope, a different world, where the game is driven by different rules:

> The combination of massive computing power, massive expansion in data management tools and practices, and exponentialized increases in customer expectations have created a

M.A. Villegas-García (✉) • F.P. García Márquez • D.J. Pedregal Tercero
Ingenium Research Group, University of Castilla-La Mancha, Ciudad Real, Spain
e-mail: marco.villegas@uclm.es; faustopedro.garcia@uclm.es; diego.pedregal@uclm.es

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_5

> world so complex and a customer so demanding that unaugmented human cognition—by this I mean making decisions without the assistance of a robust business analytic tool set—is no longer good enough. What you need to know, whom you need to know, how you come to know and the very abbreviated time window available for making efficacious use of knowledge are transforming.

However, the promise of digital analytics still remains largely unrealized. EMC estimates that the majority of new data is largely untagged, file-based, and unstructured data, little is known about it. Only 3 % of the data being created today is useful for analyses, whereas only 0.05 % of that data is actually being analyzed. Therefore, 99.95 % of useful data available today for analysis is not being analyzed. IDC estimates a 67 % increase in data available for analysis by 2020.

1.1 Some Lasting Stumbling Blocks

In spite of the enormous changes experienced by the society in the last years, a very striking truth is the fact that the way we manage information today is not much different from how we have done for millennia [3]. One of the most ancient form of recorded information are the prehistoric cave drawings. Someone decided to create this record, probably ignoring that he had discover a data storage technique, far more perdurable in time than oral transmission. We realize that for us human, the problem is not storing data, it is in access and meaning. Following the example of the cave drawing, accessing the information means getting there and to stand up in front of it. But once you get there, what do you see? Is the picture a message? Is it perhaps a lyric? Or is it just art? It is evident that the sole data is not enough for us. We need a semantic context in which such bunch of data takes up meaning and becomes valuable information. This enriched information is the required base for building wisdom, which will let us to make better decisions.

The information media evolved with the centuries. The paper became the major support for data recording and communication. Vinyl discs made the miracle of storing audio, whereas electromagnetic tapes did so with video. Now, in the digital era, we are able to carry in the pocket an immense quantity of multimedia content, all with a vulgar USB stick. And the industry continues in this endless run. Definitely we are pretty good at storing data. However, the volume of accessible data far exceeds the human ability to consume it, even more in our days. Yesterday our offices were overflowing with papers, today with emails and digital documents. It is true that Information Management strategies have positively evolved, but they haven't gone very far. In spite of these enormous changes, we have still some of the same problems since the beginning. We could synthesize these with the following questions:

- How to build knowledge up from data?
- How to relate data? How to make it meaningful?

Two separated poles have remain unconnected. By one hand, the fast evolving hardware technology, and by the other the human-minded activity supported by that technology. There are enormous oceans of information to be exploited, but the actual benefits we obtain from them depends on how well connected these two poles are. And here is where Business Analytics come in action. This term refers to a pretty interesting assembly of several fields, like math and statistics sciences, computer science and management sciences, in which the methods, processes and methodologies are continuously enriched by all these knowledge areas. Although this has been a pretty good effort in bringing humans and data-oceans closer, there is still a long way to go.

2 Data Analysis and Synthesis

Another important characteristic of today's society is the *fragmentation*. Our daily experience of life is minced into little pieces, little facets or domains which we have to continually integrate and reconcile: professional and familiar life; business trip plan and restrictions in flights, airlines, budgets. They are labyrinth of decomposition [4]: Organizations are decomposed into regions, divisions, departments, products, and services, not to mention missions, visions, objectives, programs, budgets, and systems; likewise, agendas are decomposed into issues, and strategic issues are decomposed into strengths, weaknesses, threats, and opportunities.

Analysis consists primarily in breaking down a complex topic of problem into smaller parts to gain a better understanding of it. We believe that if we have all the data, and we are smart enough, we can solve any problem. In recruiting people, we test their analytical skills. Indeed, regardless their industry sector, many people work as 'analysts'. It has become like an obsession [5]:

> Why is everyone so obsessed with analysis? Analysis is only one style of solving problems. [. . .] We seem to have forgotten all about synthesis, the opposite approach. Take two or more ideas and combine them into a larger new idea. Tackling a problem in this way might lead to entirely new insights, where problems of the "old world" (before the synthesis) do not even occur anymore. Where analysis focuses on working within the boundaries of a certain domain [. . .], synthesis connects various domains.

Managers oversee all this chaos, and they are supposed to integrate the whole confusing mess, most of the times with the sole help of their intelligence and intuition: they make the synthesis on their own. Synthesis is the very essence of managing [4]:

> Putting things together, in the form of coherent strategies, unified organizations, and integrated systems. [. . .] It's not that managers don't need analysis; it's that they need it as an input to synthesis. [. . .] So how can a manager see the big picture amid so many little details?

As companies get large and complex, it is more evident that it is impossible for a single person to watch them conscientiously on detail, even a single department. In some companies, the business operations are carried out by pretty big teams; they work together to accomplish the business goals, keeping certain norms and protocols. Business Analytics give us the possibility of exploiting large volumes of data, extract value from them, and use this value to empower the processes and business decisions at every level in the company. Even though any incorporation of BA is positive to the company's performance, the major benefits come from a wider application of BA throughout the organization [6]. The solutions implemented at this level are very specific to the concrete industry and organization and involve an important amount of time-effort from a considerable number of IT and business experts. A consequence of this is that only big companies can afford these implementation costs, whereas the main population of companies remain far from these possibilities. Even for the BA leaders, at present there is no clear methodology for a comprehensive BA implementation [7]. The big challenge consist on finding a clear path to broadly implement BA without getting stuck (distracted) with "nasty" technical details, but center at the business logic and concepts.

We claim that business intelligence and analytics should evolve to a higher status in which the business users can 'navigate' more fluidly across the processes, data, resources, restrictions, goals... This implies that users should be able to interact directly with the BA systems via *business concepts* like revenue, costs, customer satisfaction, etc. The technical details on how data is collected and analysed, and how they are arranged to built more abstract artefacts, most be hidden for business users. The problem is that even IT and business analysts get frequently trapped into the high complexity of the BA details implementation.

We propose a BA architecture that facilitates both the implementation and exploitation of BA systems. Our idea is strongly based on business modelling techniques, in particular on the BIM proposed by Barone et al. in 2010 [8]. The following section explain the main components of this approach.

3 The Business Analytics Architecture (BAA)

The main objective of Business Intelligence tools is to transform raw data into meaningful and actionable information for business purposes. We understand by 'meaningful and actionable' information that which allow to the readers building a better knowledge about the reality in a given context, and give rise to making wise decisions driving the reality to the desired state.

At this point we can identify two main faces of all BI systems: the 'internal face' deals with the raw data that is going to be transformed into 'meaningful and actionable' information. That data is usually extracted from operational information systems which record everything happening in the company. There are many commercial solutions filling this section, most of them under the names ERP, MRP and CRM.[1] By the other hand, the 'external face' of BA systems is associated with the business purposes we are pursuing. These comprise the designed business strategy, usually stated at the level of enterprise/corporation; and the operational objectives, which are more related with a particular department.

In spite of seem disconnected, these two poles are tightly related. In fact, the enterprise's data is just a representation of the actual business' execution. They absolutely shouldn't be disconnected, as they are nowadays in many business information systems. By far, most of the efforts driven by BIA community could be condensed in this sublime goal: achieve a more fluid an natural connection between the data and the real-world. And this should be done in both directions: (1) from data to real-world, in order to make a fact-based tracking of the company's performance, and be able to drive it to success; and (2) from real-world to data, enabling meaningful data analysis powered by business semantic.

To achieve this goal, the BAA is compose of three main layers (see Fig. 1). The higher one contains the business logic and concepts, using the same terminology that business analysts use everyday. It is usually called 'the semantic layer', because it provides the logic framework that gives the appropriate meaning to all other elements. This is the 'external face' of the BI system, which is supposed to provide an user-friendly interface to embed the business logic into the system. In the next section we will drop a light about how the business users should interact with this layer and the easy-to-use functionalities it should provide.

This semantic layer needs to be connected with low-level data, and this is done through the mapping layer. It contains the conceptual mapping between the data entities and the business entities defined in the semantic layer. Most of the complexity of connecting abstract concepts with concrete facts are embedded in this layer; we follow a simplify version of the model proposed by Rizzolo et al. [9]. This connection is meant to be fully bidirectional, because any relevant change in row data (concrete facts) should be translated and presented in terms of business concepts (synthesis), and vice-versa: the logic depicted by business concepts is used to guide and sharpen the mining of concrete local data (analysis).

The bottom layer deals with the integration of multiple data sources in a unified repository. In most business IT infrastructures, the Data-warehouse is devoted to this kind of tasks, which today is becoming more challenging than ever. Some facts causing this are (1) the availability of huge amounts of data (2) the proliferation of new and unstructured data sources like multimedia, hypertext... (3) the very high requirements in terms of availability and response time.

The analytical machinery (algorithms and processes) cross over all the three layers. We must take profit of the analytical power at every level of abstraction. For example, we can use a predictive model to forecast the company sales for the next month, which could be an application in the semantic layer; by the other hand, we

[1] Enterprise Resource Planning, Material Resource Planning and Customer Relationship Management, respectively.

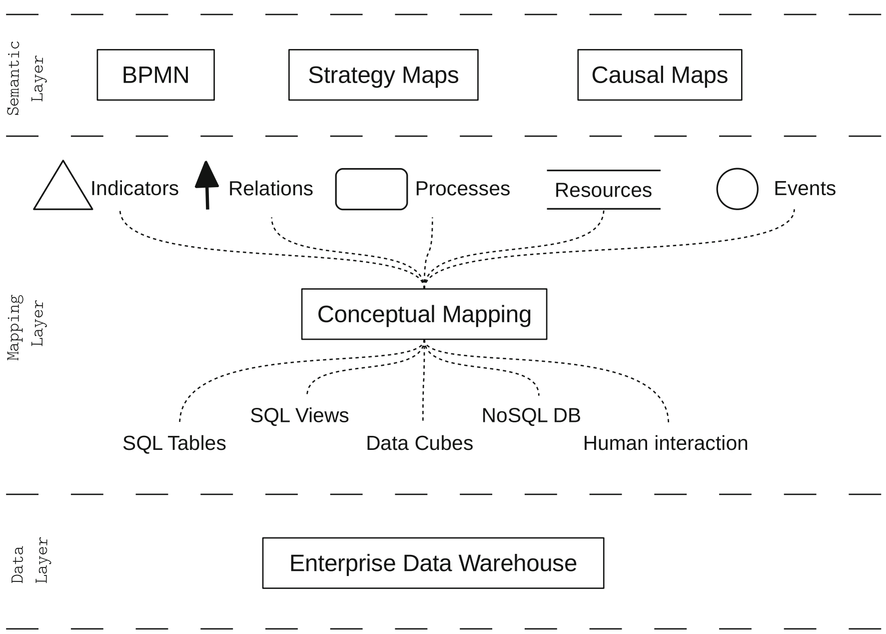

Fig. 1 Business intelligence architecture

could use a clustering model to clean up a dirty data source, before it becomes part of the data layer.

3.1 Semantic Layer: Business Modelling

The semantic layer allows business users to conceptualize their business operations and strategies using concepts that are familiar to them, including: Actor, Directive, Intention, Event, Situation, Indicator, Influence, Process, Resource, and Domain Assumption [7]. These concepts are synthesized from some well known management methodologies like Balanced Scorecard, Strategy Maps and Business Process Management. In this layer we will model the enterprise in different dimensions. The objective is to represent the business knowledge that people use in day to day work.

A *Balanced Scorecard* (BS) [10] is designed to align the work and priorities of employees with the strategic objectives that comprise an organization's defined mission. It allows managers to look at the business from four important perspectives: finance, customers, processes, innovation and learning. While keeping it simple, the BS meets several managerial key points: first, it brings together many seemingly disparate elements of a company's competitive challenges; second, it provides a tool for balancing the strategy across the fundamental business dimensions.

A *Strategy Map* is an illustration of an organization's strategy. It is extremely useful for simplify the translation of strategy into operational terms and to communicate to employees how their jobs relate to the organizations overall objectives. Strategy maps are intended to help organizations focus on their strategies in a comprehensive, yet concise and systematic way [11]. In fact, it works as the integration mechanism, in the sense that, the four BS perspectives, the associated strategic objectives, and the key performance indicators (KPI) are linked together as cause-and-effect relationships [12].

The main objective of *Business Process Management* (BPM) [13] is to align the processes with the business strategy. It essentially pursues the "achievement of the organization's objectives through the improvement, management and control of essential business processes". By the other hand, the Business Process Management Notation (BPMN) [14] is a notation standard for modelling business processes, probably the best known and established in the industry. The primary focus of BPM is in elements and processes, while BS and strategy maps focus on strategy and objectives.

Therefore, at this architectural level (the semantic layer) we make use of three different type of models. The strategy map (see Fig. 2) provides a way to depict a strategy that achieves a main goal [7]. A goal is split in several subgoals creating a hierarchy that clearly sets how the subgoals should be accomplished to achieve the main goal. To attain a particular goal, one or more processes most be performed, and a set of indicators are configured to measure the achievement level of every

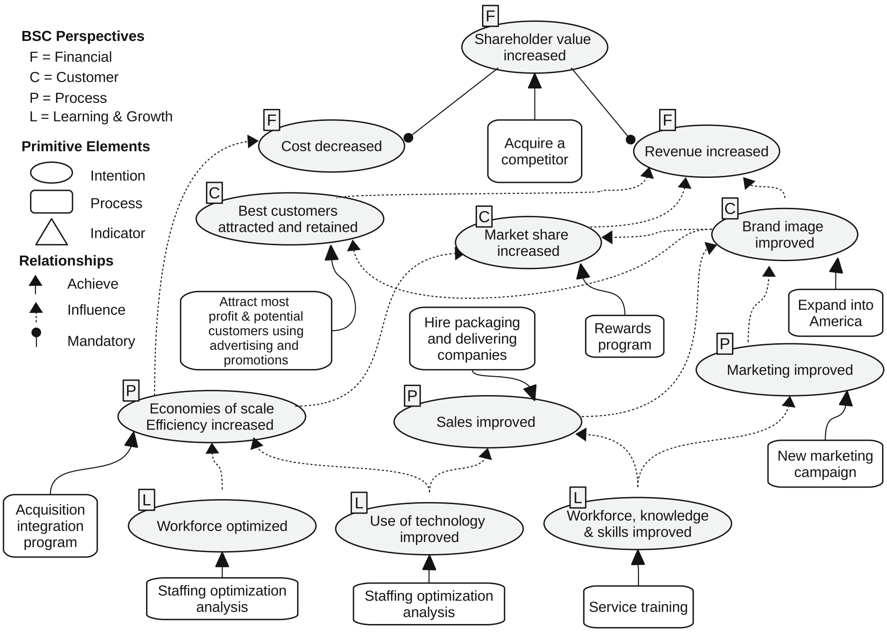

Fig. 2 Strategy map (adapted from [7])

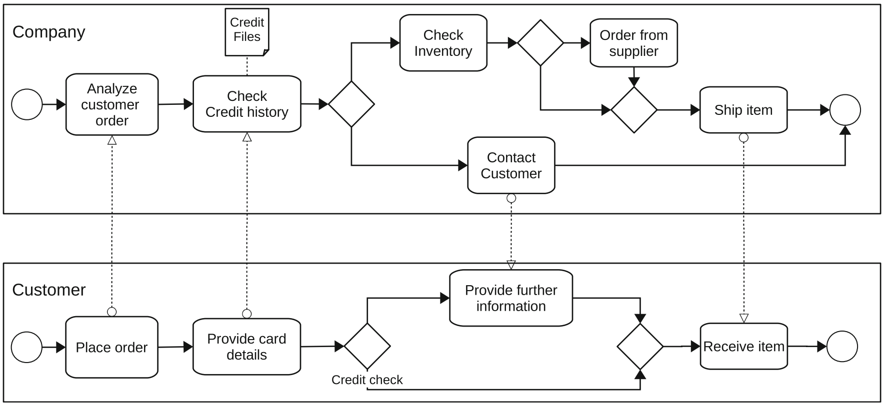

Fig. 3 BPMN model

goal (KPI). Notice that each goal has a label indicating which dimension of BSC corresponds to (financial, customer, process, learning & grow).

Within this modelling technique we regard the strategy, the metrics we are going to use to measure the accomplishment level, the key processes involved in this accomplishment, and the (causal) relation among these elements. Whereas the strategy components are all included in this diagram, not all the business process are depicted on it. For these other processes that we consider worth to monitor but are not included in the strategy map, we use the BPMN (Fig. 3).

The BPMN express the flow of processes, their relationships and interactions from the initial state to the final one. We widely adopt the standard BPMN 2.0 [14]. This type of models contain information about resources and how are they consumed/produced by the processes, as well as detailed information of each element like geographical location, organizational level, starting conditions, processing time. They provide an internal look of the organization, leaving aside the global picture of strategic goals.

In addition to these two modelling techniques, we consider the use of *Causal Maps* as a source of additional although different kind of business information. The causal maps help analysts to express the business intuitions that managers and operational employees hold about they work. These intuitions/ideas could be clearly deviated in any direction, according to the mental model hold by the people and the organizational culture. In spite of all these things, we claim that is worthy to construct such a diagram for many reasons: (1) they serve as a mean to unify and clarify the personal perception of the business; (2) they help to determine and focus on the factors that actually have impact on business; (3) they open the possibility of doing 'automatic' inference (reasoning) about the multiple influence and strengths that play a role in the day to day work (Fig. 4).

These three diagrams are connected to each other. The strategy plan provides the full map of goals, how they are related, how certain processes will help to achieve them. These 'certain' processes are fully specified in the business process model, in

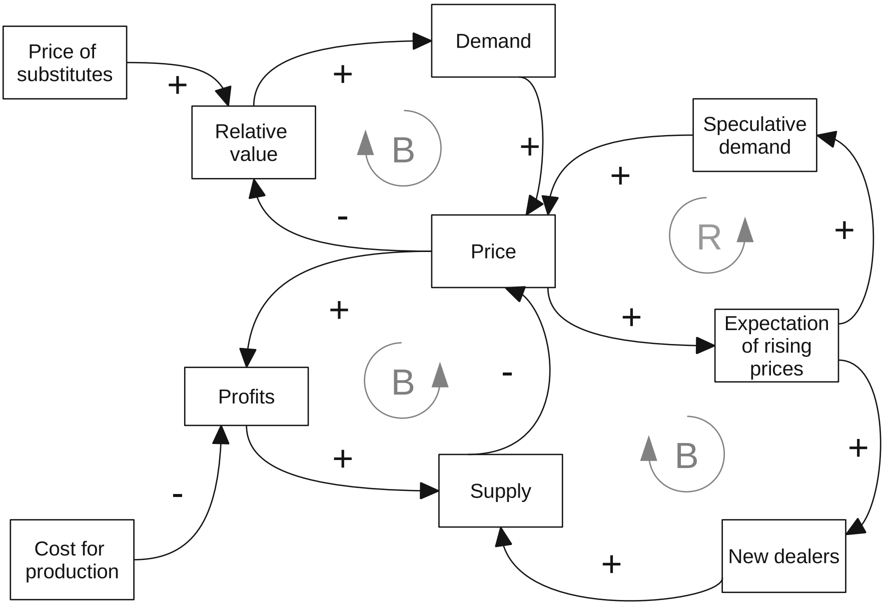

Fig. 4 Causal map

which we set the dynamic that arranges people, processes, resources, events, situations, restrictions... This 'dynamic' defines a set of clear relations or influences among the entities. For all other causal relations, which can not be derived from BPMN, but are contained in the knowledge formed by the people's experience, we use the causal maps.

The idea behind all this is to represent as much business knowledge as we can. It is in this layer where the business concepts acquire their meanings. It provides a semantic context in which the ideas are defined, and constitute the proper environment for direct interaction with the business user. From the users' perspective, this new 'semantic environment' will provide at least two big benefits. First, they will not be bothered with technical stuff, so they can center at business analysis and decision. Second, they are not left alone when analyse data: they are supported, guided and powered by the system, following the logic captured by the diagrams previously introduced.

3.2 *Mapping Layer: Conceptual Mapping*

One of the big challenges that information technology should face nowadays is bridging the gap between the *ideas* or *concepts* that humans use every, and the data elements that populate the enormous digital universe. As the society becomes more and more data-centered, and the data incrementally proliferate every day, the solution of this challenge gets more and more relevance.

Most of the available data is stored in relational format. It usually consists on a set of data tables, which are related by keys-columns following certain rules. The ERM [15] probably constitute the most well established de-facto standard for storing and representing data, even though some new and very promising approaches have emerged in the last years [16].

Many solutions have been proposed to close the representational gap between the storage layer (e.g., ERM) and the conceptual layer, which gives the 'user-semantic' meaning to the data. Some of them include EDM, Hibernate, Doctrine, RedBean, ActiveRecord. These mapping technologies are known as Object-Relational Mappings (ORM).

However, for BI applications these approaches are insufficient due to the huge volume of data involved. For a mid-size food company, it is not rare to handle 100 thousands billing transactions a day. The problems come when the managers and business users want to make sense of datasets that expand several years and interactively explore them, because the underling storage technology cannot timely support such a huge operations. That's why the so called Data-Warehouse technologies has come to existence, as well as the OLAP and DataMarts tools [17]. They intend to solve the problem by pre-aggregating the data into data-cubes, drastically reducing the system response time. In top of this data aggregation layer it is common to find data visualization tools, constituting the 'business intelligence' capability of such a systems. This kind of systems have become very popular in the IT industry, but they suffer some weaknesses. Their implementation usually involve important resources in terms of IT personnel effort and required time; they usually are pretty specific to the concrete industry and company involved; and being constructed following a bottom-up methodology, they can grow in size and complexity without real business necessity. For these reasons, some alternatives to Data-warehousing have been proposed. Some of them primarily focus on hardware optimization, like 'In-memory' technologies [18]. Other approaches to solve the gap between the conceptual and the storage facets of data, are based on conceptual modelling techniques. They help to raise the abstraction level, seamlessly connecting user's concepts with physical data.

The second layer of our Business Intelligence Architecture is the Conceptual Mapping, which is responsible of mapping the business concepts to the raw-data entities. Our approach is partially based on the Conceptual Integration Model (CIM) proposed by Rizzolo et al [9]. They extend MultiDim model [19], which in turn extends Chen's ER model to support multidimensionality (Fig. 5).

3.3 Data Layer: Data-Warehouses

Data are proliferating in volume and format, they are present throughout the organizations. Many data storage solutions have appeared in the last years, so

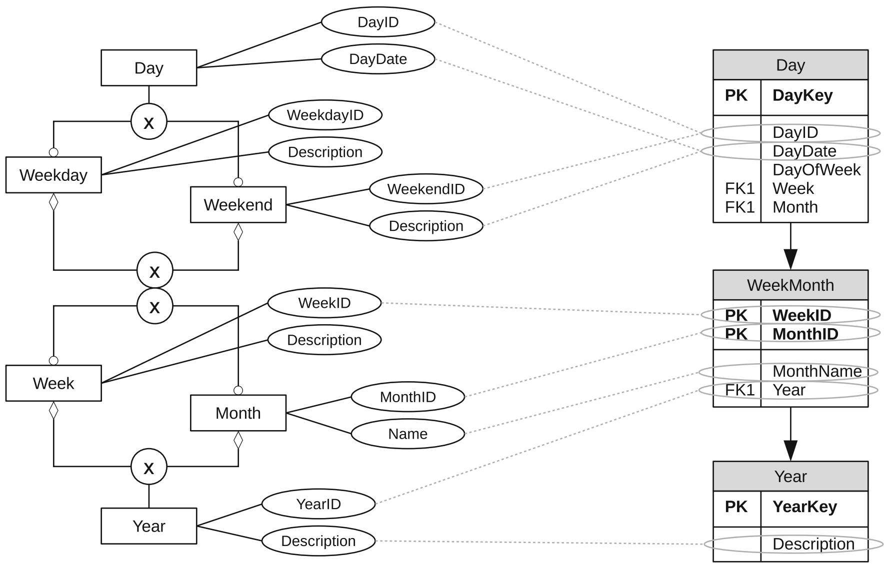

Fig. 5 Conceptual mapping excerpt

companies usually combine many of them. This layer is responsible for unifying all enterprise data sources available, in such a way it is possible to connect them with the Conceptual Mapping layer. We widely adopt ETL industry standard to this purpose, in addition to data warehousing technologies.

Data warehouses typically contain data from multiple organizational silos. This data is often more integrated, better understood, and cleansed more thoroughly. However, for building predictive analytic models they hide some drawbacks, as the problem of eliminating critical outlier data within the process of data cleansing. Despite this, data warehouses can be very useful for the construction of predictive analytic models, if built correctly. Doing so, you will regard at least the following tips [20]: (1) Data warehouses are not as space-constrained as operational databases, and mostly are used for historical analysis. As such, there is less pressure to delete unused records. So maintain as much data as you can. (2) The ability to store more data makes it more practical to store a new version of the record every time it is updated in the source system. It will avoid leaks from the future. (3) Many data warehouses are used to produce reports and analysis at a summary level. Taking wide profit of predictive analytic models imply storing transactional data as well as the roll-up and summary data. A well designed and implemented data warehouse is a great source of data that leverage the power of predictive analytics (Fig. 6).

Trying to gather in a single system all the information of a company, could become an arduous task, and sometimes impractical due to the volume of data involved. For that reason, some organizations have developed what has become to called data marts: data is extracted from operational databases of from an enterprise data warehouse and organized to focus on a particular solution area. Owned by a

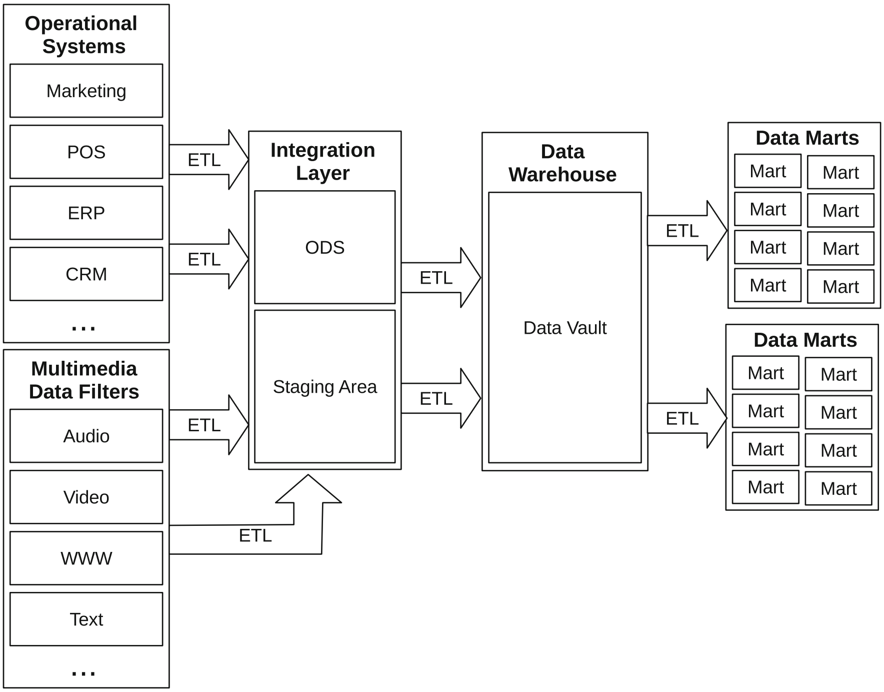

Fig. 6 Basic components of a data warehouse

single department of business area, data marts allow a particular group of user more control and flexibility when it comes to the data they need.

Another less common extension of data warehouses consist of connecting them with unstructured sources of data, like the Web, written documents, and any other media like images, audio and video. To do so, it is necessary to configure and specify how the raw data is transformed into regular data, what is the valid output range and what to do if something is wrong. This work is done by separated modules that we call Multimedia Data Filters (MDF). Basically an MDF receive a source media like a video and return a corresponding bunch of data. They focus on a particular type of source media and are powered by analytics, and need to be specify separately due to the involved complexity. Some example are sentiment analysis, web mining, video analysis and image decomposition.

4 Analytical Foundations

As we have already mentioned, at every level in the process of abstraction we take profit of analytical algorithms. They are usually gathered under the names of Pattern Recognition and Machine Learning algorithm, but is not unusual hear about Artificial Intelligence, Data Mining and Knowledge Discovery. What are all they really about?

Pattern Recognition deals with the problem of (automatically) finding and characterising patterns or regularities in data [21]. By patterns we understand any relations, regularities or structure inherent in a source of data. By detecting patterns, we can expect to build a system that is able to make predictions on new data extracted from the same source. If it works well, we say that the system has acquired *generalization power* by learning something from the data.

This approach is commonly called the *learning methodology*, because the process is focused on extracting patterns from the sample data that lead us to make generalizations about the population data [22]. In this sense, it is a *data driven* approach, in contrast with *theory driven* approach. However, it is extremely useful to tackle complex problems in which an exact formulation is not possible, for example, recognising a face in a photo or genes in a DNA sequence.

Consider a dataset containing thousands of observations of pea plants, in the same format of Gregor Mendel's observations. It is obvious that the characters (color and size, for example) of certain pea plant generation could be predicted by using the Mendel's laws. Therefore, the dataset contains an amount of redundancy, that is, information that could be reconstructed from other parts of the data. In such cases we say that the dataset is *redundant*.

This characteristic has an special importance for us, because the redundancy in the data leads us to formulate relations expressing such behaviours. If the relation is accurate and holds for all observations in the data, we refer to it as an *exact relation*. This is the case, for example, of the Laws of Inheritance: Mendel found that some patterns surprisingly held for all his experiments. For that reason, we say that this part of the data is also *predictable*: we can reconstruct it from the rest of the data, as well as predicting future data, like the color and size of new plants by using the current plants data.

Finding exact relations is not, by far, the general case for someone who analyses data. Certainly, the common case is finding patterns that hold with a certain probability. We call them *statistical relations*. Examples of such relations are: forecasting the total sales of a company for the next month, or inferring the credit score [23] of a new client in a bank by analysing his information.

The *science* of pattern analysis has considerably evolved from its early formulations. In the 1960s efficient algorithms for detecting linear relations were introduced. This is the case of the Perceptron algorithm [24], formulated in 1957. In the mid 1980s a set of new algorithms started to appear, making possible for the first time to detect nonlinear patterns. This group includes the backpropagation algorithm for multilayer neural networks and decision tree learning algorithms.

The emergence of the new pattern analysis approach known as kernel-based methods in mid 1990s, changed the field of pattern analysis towards a new and exciting perspective: the new approach enabled researchers to analyse nonlinear relations with the efficiency of linear algorithms via the use of kernel matrices. Kernel-based methods first appeared in the form of support vector machines (SVM), a classification algorithm that quickly gained great popularity in the community for its efficiency and robustness. Nowadays we have a variate and versatile toolbox composed by the algorithms developed by the scientific community during the short live of this research area.

5 Discussion and Related Work

Companies have traditionally adopted business intelligence solutions to support business decision making on a consistent daily basis, bringing data from disparate sources into a common data infrastructure or warehouse for reporting, analysis and creating analytic applications. By the other hand, emerging data discovery tools have focused instead on providing data-savvy user with free-form, more tactical analysis on single data sets. Definitely both are valuable in maximizing the value of data to an organization, but they do not use up the worth enclosed in the data. Moving forward, organizations will maximize value from BI and analytics by integrating those models and tools into a broader architecture that ultimately unify them. This approach will enable the "single source of truth" throughout the organization, as well as a more seamlessly and productive use of data by the staff.

This scenario provides several advantages in front of most current BA implementations:

- The final business users can interactively reason with the BA system by using business concepts, focusing on business issues.
- The business analysts can focus on business rules and logic. They can apply analytics' power directly into business reasoning, because the internal data-gears are transparent for them.
- The IT analysts can integrate multiple sources of data, and map them to 'business entities' to feed the whole architecture.

Nauman Sheikh [25] presents a nice guidebook on how to plan, design, and build analytics solutions to solve business problems. The approach is rather broad, dropping a light on every aspect of implementing analytics on organizations. For Sheikh, *analytics solution* means the process of collecting data, learning from it, anticipating scenarios, automating decisions and monitoring. One the big objectives of this work is the *simplification* of the full implementation of an analytics solution, as well as demystifying some topics associated to business analytics.

A very promising initiative in the industry of analytical software is the development of the Predictive Model Markup Language (PMML) [26] which provides a way for applications to define statistical and data mining models and to share models between PMML compliant applications. PMML provides applications a vendor-independent method of defining models so that proprietary issues and incompatibilities are no longer a barrier to the exchange of models between applications. It allows users to develop models within one vendor's application, and use other vendors' applications to visualize, analyze, evaluate or otherwise use the models [27].

Some efforts have been devoted to extend and profit the power of business modelling. Yu et al. [28] propose an approach that incorporate the intentional dimension of motivations, rationales and goals. While most Enterprise Architecture frameworks define components for data (what), function (how), network (where), people (who), time (when), the motivational aspect (why) has not received much attention, and is seldom supported by modelling. Nevertheless, understanding the motivation and intention of stakeholders is critical for architectural decisions and actions. Modelling the intentions will help to make them *transparent*, making it possible a *systematic analysis* of design implications, the exploration of possible strategies in a *rational way*, and the justification of activities by *tracing them back*.

References

1. Cukier K (2010) Data, data everywhere: a special report on managing information. Economist Newspaper
2. Phillips J (2014) Building a digital analytics organization
3. May T (2009) The new know: innovation powered by analytics, vol 23. Wiley, Hoboken, NJ
4. Mintzberg H (2009) Managing. Berrett-Koehler, San Francisco, CA
5. Buytendijk F (2010) Dealing with dilemmas: where business analytics fall short. Wiley, Hoboken, NJ
6. Davenport TH, Harris JG (2007) Competing on analytics: the new science of winning. Harvard Business Press
7. Barone D, Yu E, Won J, Jiang L, Mylopoulos J (2010) Enterprise modeling for business intelligence. In: The practice of enterprise modeling. Springer, Berlin, pp 31–45
8. Barone D, Mylopoulos J, Jiang L, Amyot D (2010) The business intelligence model: strategic modelling. Technical report, University of Toronto (April 2010)
9. Rizzolo F, Kiringa I, Pottinger R, Wong K (2010) The conceptual integration modeling framework. pp 1–18
10. Kaplan RS, Norton DP, Horv6th P (1996) The balanced scorecard, vol 6. Harvard Business School Press, Boston
11. Lawson R, Desroches D, Hatch T (2008) Scorecard best practices: design, implementation, and evaluation. Wiley, Hoboken, NJ
12. Kaplan RS, Norton DP (2004) Strategy maps: converting intangible assets into tangible outcomes. Harvard Business Press
13. Jeston J, Nelis J (2008) Business process management: practical guidelines to successful implementations. Routledge, London
14. Decker G, Dijkman R, Dumas M, García-Bañuelos L (2010) The business process modeling notation. In: Modern business process automation. Springer, pp 347–368.
15. Chen PPS (1976) The entity-relationship model—toward a unified view of data. ACM Trans Database Syst 1(1):9–36
16. Han J, Haihong E, Le G, Du J (2011) Survey on nosql database. In: Pervasive computing and applications (ICPCA), 2011 6th international conference on. IEEE, pp 363–366
17. Lenzerini M, Vassiliou Y, Vassiliadis P, Jarke M (2003) Fundamentals of data warehouses. Springer, Berlin
18. Plattner H (2009) A common database approach for oltp and olap using an in-memory column database. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, pp 1–2
19. Malinowski E, Zimányi E (2008) Advanced data warehouse design: from conventional to spatial and temporal applications. Springer, Berlin

20. Taylor J (2011) Decision management systems: a practical guide to using business rules and predictive analytics. Pearson Education, Upper Saddle River, NJ
21. Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis
22. Villegas-García MA (2013) An investigation into new kernels for categorical variables. Master Thesis, 2013. http://upcommons.upc.edu/handle/2099.1/17172
23. Huang C-L, Chen M-C, Wang C-J (2007) Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications 33(4):847–856
24. Rosenblatt F (1957) The perceptron—a perceiving and recognizing automaton. Technical Report 85-460-1
25. Sheikh N (2013) Implementing analytics: A blueprint for design, development, and adoption. Newnes
26. Data Mining Group. Predictive model markup language. See www.dmg.org
27. Guazzelli A, Lin W-C, Jena T (2012) PMML in action: unleashing the power of open standards for data mining and predictive analytics. CreateSpace
28. Yu E, Strohmaier M, Deng X (2006) Exploring intentional modeling and analysis for enterprise architecture. In: Enterprise Distributed Object Computing Conference Workshops, 2006. EDOCW'06. 10th IEEE International. IEEE, pp 32–32.