Data and Artificial Intelligence Strategy: A Conceptual Enterprise Big Data Cloud Architecture to Enable Market-Oriented Organisations

Caio Moreno[1], Ramón Alberto Carrasco[1], Enrique Herrera-Viedma[2]*

[1] Department of Management and Marketing, Complutense University, Madrid 28223 (Spain)
[2] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Granada 18071 (Spain)

Abstract

Market-Oriented companies are committed to understanding both the needs of their customers, and the capabilities and plans of their competitors through the processes of acquiring and evaluating market information in a systematic and anticipatory manner. On the other hand, most companies in the last years have defined that one of their main strategic objectives for the next years is to become a truly data-driven organisation in the current Big Data context. They are willing to invest heavily in Data and Artificial Intelligence Strategy and build enterprise data platforms that will enable this Market-Oriented vision. In this paper, it is presented an Artificial Intelligence Cloud Architecture capable to help global companies to move from the use of data from descriptive to prescriptive and leveraging existing cloud services to deliver true Market-Oriented in a much shorter time (compared with traditional approaches).

Keywords

I. Introduction

MARKET-ORIENTED (MO) enterprises have as its main objective to perfectly understand the needs of its clients to satisfy them better than the competition and, in this way, to have a competitive advantage over the said competition. In this aspect a key role is played by the data, in such a way that converted into timely knowledge and transmitted to the entire organization to work in coordination, will achieve this common goal [1]-[3].

Precisely, Business Intelligence (BI) is a business philosophy that bases decision-making on data that has been opportunely converted into knowledge process in which, Artificial Intelligence (AI) plays a key role [4][5]. In this way, there have been authors who have defined a formal framework for the effective implementation of Business Intelligence (BI) in a MO company [6].

However, the emergence of Big Data [7] has made these models obsolete, and market-oriented companies do not know how to adapt given the wide variety of products and services that composes the so-called Big Data ecosystems.

The main objective of this work, therefore, is to define a formal framework that allows Market Orientation to be effective in the context of Big Data, including the necessary components.

To achieve this goal, in section II it is introduced market orientation, and formal architectures to achieve it in a BI framework. In section III a review of this proposed formal frameworks is made, adding the appropriate components that allow carrying out Market Orientation in a Big Data context. Finally, in sequence, some conclusions and future work are presented.

II. Market-Oriented Enterprise Strategy Based on Data

The data-driven business strategy has been in existence for several decades. An example of the theoretical framework of strategic planning is showed in Fig. 1, which has three components based, to a greater or lesser extent, on the collection and analysis of data:

1. **Specification of objectives**: Normally, increase the value of the company, present and future, based on doing the same with the value of the clients.

2. **Decision Making (DM)**: Based on the objectives specified through the generation and evaluation of the different strategies.

3. **Monitoring of the results**: Quantifying to what extent the objectives have been reached after the DM.
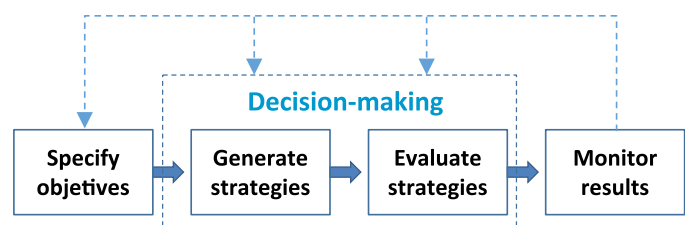

Fig. 1. The formal process of strategic planning (based on [1]).

* Corresponding author.
E-mail address: viedma@decsai.ugr.es

In the eighties and nineties of the twentieth century, companies were in a general situation of lack of differentiation, since they offered products/services (P/S) of very similar quality with differences not perceptible for clients. At this juncture, Market Orientation, proposal, arises as the generation in the entire organization of market intelligence about the current and future needs of customers, the dissemination of information by all departments and the ability to respond to it throughout the organization [2]. In this definition three dimensions are identified:

- **Generation of market intelligence.** It consists of the responsibility of the entire organization to obtain a marketing information system on the present and future needs of customers, as well as distributors, suppliers, lobbyists, competitors and macro environment in general.

- **Dissemination of information.** Among the different areas of the organization in order for them to work in common in the same objective.

- **Response to information.** In such a way that this knowledge obtained and disseminated becomes actions that will result in obtaining competitive advantages in the organization.

Market-Oriented (MO) companies are committed to understanding both the needs of their **customers** and the capabilities and plans of their **competitors**, through the processes of acquiring and evaluating **market information** in a systematic and anticipatory manner [3]. Therefore, a key concept for MO is, undoubtedly, Business Intelligence (BI) that can be defined as [4][5]: the process of converting data into knowledge and this into actions or decisions to create the competitive advantage of the business. In Fig. 2 it is possible to observe a global scheme of the different components that encompasses the concept of BI focused on MO, proposed by Stone and Woodcock [6].

The different components are explained in more detail below:

1. **Data.** It would include the internal data of the organization such as: the contents in the operative processes of the Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) systems; those related to bidirectional communications with customers and society in general commonly supported in collaborative and social CRM; third-party data; qualitative data including those from typical secondary sources; etc.

2. **Data Warehouse (DW).** They are the databases specifically designed for analysis and they would include much of the raw data discussed, through the so-called ETL (Extraction, Transformation and Loading) processes [8]. Data marts are a portion of the DW with a departmental purpose; in our case, the one that interests us is the marketing data mart that would coincide, to a large extent, with the concept widely used in the past of marketing information systems.

3. **Insight Generation.** It would include, among others, the typical models of the relational strategy such as valuation, identification, recruitment, retention and client development. In large part, this component is supported by Data Mining or Data Science of which modeling phase is, in turn, supported largely by Artificial Intelligence (AI), and more specifically, by Machine Learning (ML) systems.

4. **Action.** This component is related to business decisions, which is in TD processes, based on the knowledge discovered in the previous component (with visual representation, scorecards, qualitative
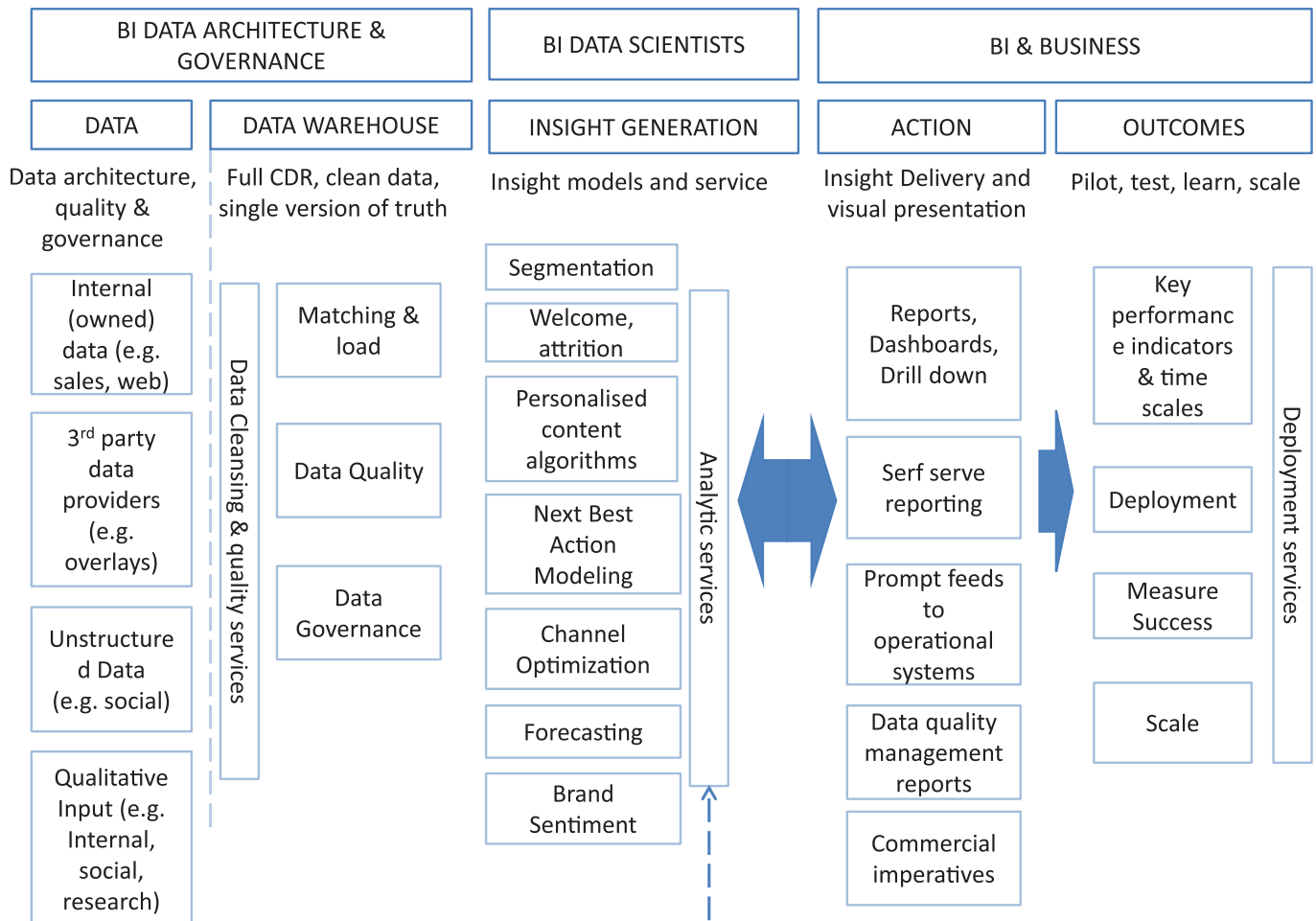

Fig. 2. BI applied to the MO (based on [6]).

reports, etc.). Many of these decisions will end up in the form of specific operations in the corresponding CRM systems.

5. **Outcomes**. In such a way that it is possible to discern to what extent it has been successful in the actions carried out.

III. Definition of a Conceptual Framework for Market Oriented Organizations Based on an Artificial Intelligence Cloud Architecture

Most of the modern organizations have invested, for many years, in a conceptual framework similar to the one proposed by Stone and Woodcock [6], and explained in the previous section, as the way to implement their Market Orientation. However, with the increasing amount of data volume, with a variety of formats (unstructured, semi-structured and structured) that are produced (and therefore they have to be processed) at high velocity, organizations are now obligated to materialize a new Big Data paradigm [9] capable to adapt to these new challenges. Although, there are several proposals for the use of Big Data adapted to certain sectors [10][11], the new components of this MO architecture, that solves these Big Data challenges, have not been fully formalized until now.

Thus, in this section, it is proposed a new formal framework for companies that want to adapt to a Big Data MO strategy. For this purpose, the formal framework, presented in Fig. 2, is redefined by adding the necessary components. The new proposed framework can be seen in Fig. 3. The changes introduced are explained below:

A. Data

In this architectural data layer, a fundamental new source for companies, that comes from the called Internet of Things (IoT), also named Internet of Everything, is incorporated. It is a new technology paradigm envisioned as a global network of machines and devices capable of interacting with each other [12]. A practical example of the use of IoT data is the one called smart trains, where the railway industry is exploiting the opportunities to use IoT data. This new data will enable predictive maintenance, smart infrastructure, advanced monitoring of assets, video surveillance systems, energy efficiency, etc. [13].

Special emphasis must be placed on the usefulness of the market data that this new technology will bring for MO companies [14]. Therefore, the value proposition of IoT data in MO companies is to receive real-time data coming directly from connected objects on the internet, capable to provide rich market insights to improve business outcomes.

B. Data Management Solutions for Analytics

Another addition to the architecture is this new layer called Data Management Solutions for Analytics (DMSA) instead of the DW layer. Although the DW component is still valid at present for the functions indicated in section II, it is clear that the large volume of data, which Big Data implies with such data, produced in real-time and with a variety of formats (more and more, among other factors, by the greater implantation of IoT sensors, as seen in the previous point), cannot be included in the conventional DW with the conventional ETL processes discussed.
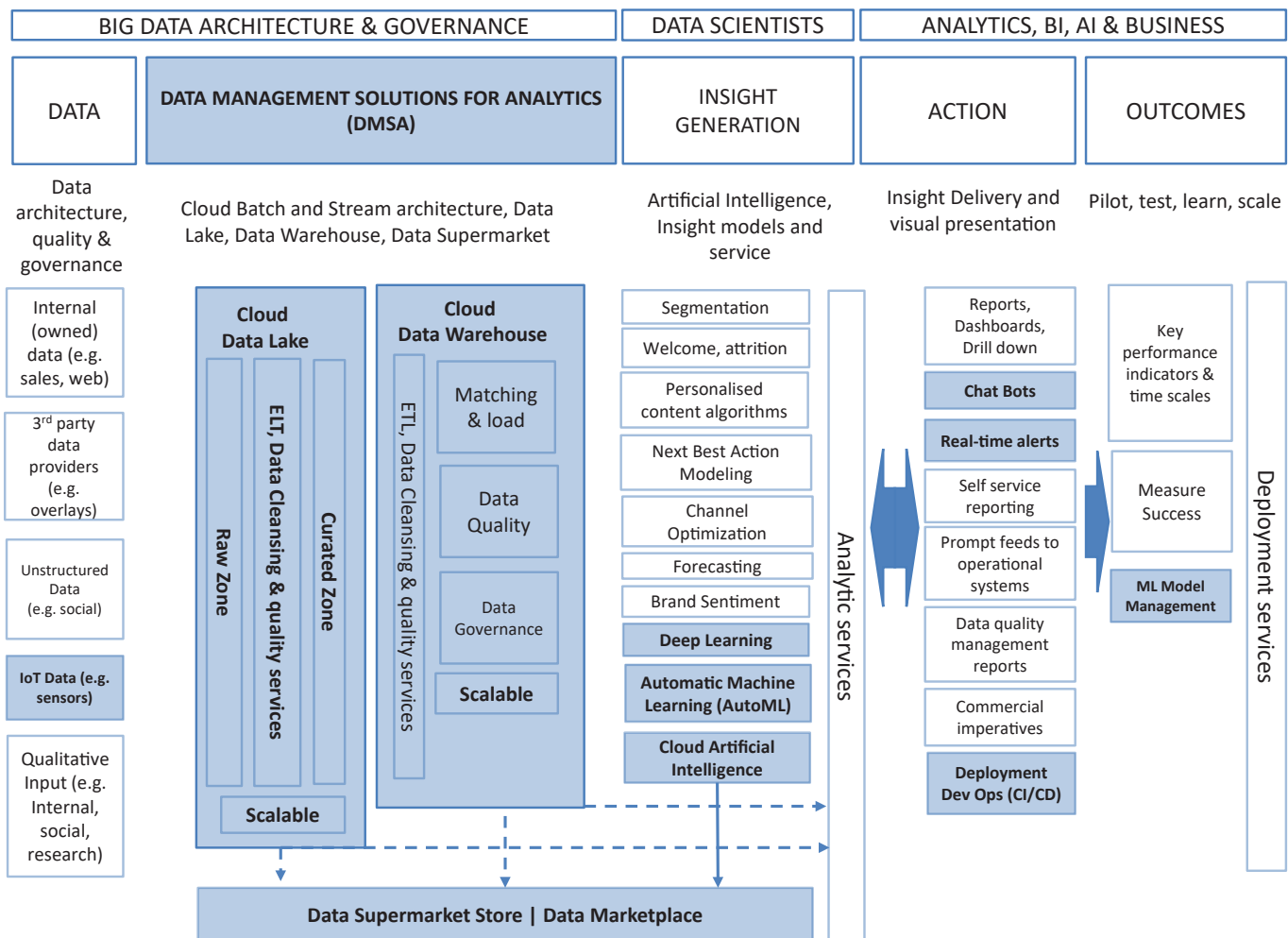

Fig. 3. A conceptual Big Data MO strategy architecture.

This nomenclature has been used by Gartner that defines DMSA as "*a complete software system that supports and manages data in one or many file management systems, most commonly a database or multiple databases*" [15]. Gartner also evaluates the different software products that support the DMSA. This evaluation includes only the main products and many of them are specifically focused on the cloud (Google, Amazon AWS, Microsoft, Alibaba Cloud…) [16]. In effect, the cloud allows the ingestion of this large volume of data.

According to the National Institute of Standards and Technology (NIST) Cloud Computing is [17] "*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*". Cloud providers, according to Iorga & Karmel [18], provides "*economies of scale, cutting-edge technology advancements, and a higher concentration of expertise enabling cloud providers to offer state-of-the-art cloud ecosystems that are resilient, self-regenerating, and secure—far more secure than the environments of consumers who manage their own systems*". Cloud is a key weapon to allow companies to adjust to market demand; the data and the market will grow, and the data architecture will have to grow to adapt. Most of the main cloud providers (Microsoft, Google and Amazon) in the last years invested heavily to create state-of-the-art data and AI services to enable small, medium and large organizations to use the cloud to transform their business into MO companies.

At present, there are two main pieces included in the DMSA: The Data Lake and the DW in which the cloud component will be fundamental. They are explained in more detail below:

A **Data Lake** is a centralized data repository that allows enterprises to collect a larger volume and variety of data and store all structured, semi-structured and unstructured data at any scale without the rigidity and overhead of traditional DW architectures [19]. Very often this data repository runs on a cloud provider in which case it is usually called Cloud Data Lake. Data Lakes have a high degree of flexibility and scalability in such a way that it allows MO companies to ingest all the market data in general, including that of their own customers and competitors. This is possible thanks to the characteristics of these repositories [20][21] that scale as much as possible; plug-in disparate data sources; acquire high-velocity data; store in native format; do not worry about original data schema; run massively parallel SQL queries; and allow advanced algorithms like deep learning that will power real-time decision analytics (as it will be seen in the Insight Generation layer).

In a Data Lake there are two main levels:

- **Raw Zone.** This tier has the capacity of storing data as-is, i.e., raw data, without having to first structure these data (see Fig. 4). Therefore, raw data are data not classified when they are stored. As a result, data preparation, cleansing, and transformation tasks are eliminated. The meaning of this storage philosophy is that [21]: "*Storing data in its rawest form enables us to find answers from the data for which we do not know the questions yet; whereas a traditional data warehouse is optimized for answering questions that we already know*". At a technological level, this layer is a folder stored thanks to distributed file systems such as HDFS (Hadoop File System).

- **Curated Zone.** Where the result of different types of analytics, using Big Data processing engines, is stored (see Fig. 5). Only high-value data are stored as files inside the Curated Zone in a folder (often in HDFS); it means data that passed quality data checks, i.e., data cleansing, data transformation and data enrichment [21]. This data, available in the curated zone, can be used to ad-hoc dashboards and visualizations, real-time analytics, and machine learning to guide better decisions.
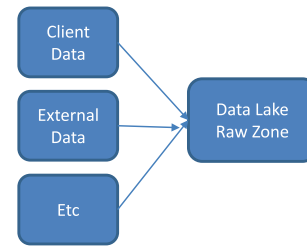

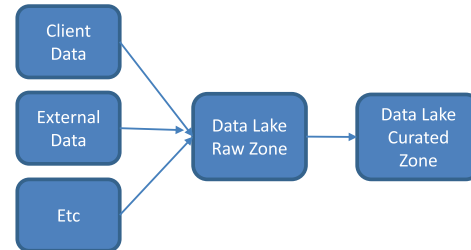Fig. 4. Plug in disparate data sources into the Cloud Data Lake: Raw Zone.

Fig. 5. Plug in disparate data sources into Cloud Data Lake: Raw Zone, clean, transform and enrich and store in the Curated Zone.

About the other main component of the DMSA, i.e., the typical DW, it still plays a fundamental role in the enterprise data architecture of any MO organization and it is a key component of the data landscape. However, in a Big Data environment, it often becomes in a **Cloud Data Warehouse**. The reasons are multiple [22]: to reduce cost, increase security, simplify maintenance, and to make possible unlimited and easy DW growth. As mentioned, data is growing in an exponential form, and MO organisations need to adapt using modern Cloud DW architectures. Parallel database systems [23] have been available for organizations for many years and provided enterprise capabilities to run large SQL queries on a large amount of data for many business users in traditional reporting and analytics systems. This new Cloud DW must: maintain the capability to run large SQL queries in a large amount of data, run in the cloud, use MPP (Massive Parallel Processing) [24] with storage and compute decoupled and allow unlimited growth and high SQL query performance.

Other advantages of using the cloud in this layer are that MO companies can program their cloud infrastructure using automatic scaling capabilities. One of the main benefits is cost reduction. **Automatic scaling** allows services to go to sleep during times of low load and can handle unexpected traffic spikes.

As seen in section II, DWs are loaded by ETL processes. Having expanded the storage philosophy with the Data Lakes explained, it is necessary to also extend the philosophy of these processes incorporating **ELT (**Extract, Load, and Transform) processes where data is extracted from the source, then loaded into a landing area in the Data Lake, transforming it where it sits in the Data Lake and then loading it into the Curated Layer of the Data Lake or DW. When the data is extracted from the source into the landing area (Raw Zone), it is a raw copy, meaning the column "names" is kept the same as in the source database and data is not converted. In most cases, in Big Data projects, it is preferable to use ELT instead of ETL. ELT has the benefit of minimizing the processing on the source, since no transformation is being done, which can be extremely important if the source is a production system where you could be impacting the user experience as opposed to a copy of the source (via replication, database snapshot, etc.). The negative side of this approach is that it may take longer to get the data into the target system data lake or data warehouse, also with the landing area there is an extra step in the process, and more disk space will be needed for the landing area.

C. Insight Generation

As described in section II, this layer is supported largely by AI and ML. The new proposed architecture adds two new key elements for MO companies to increase the monetary value of the current data asset. On one hand, some components that will imply a change in the AI model making process conception that will allow increasing their power and the industrialization of them: Cloud AI and AutoML (Automated machine learning). On the other hand, there is a fundamental element that is the incorporation of the new AI models that are revolutionizing the business world, with special mention to Deep Learning algorithms due to the repercussion that it will have in the MO companies. These commented elements are explained below in more detail:

- **Cloud AI.** As seen, AI plays a fundamental role in the Insight Generation layer. Thus, Cloud providers invested a considerable amount of money and time to build research teams to create state-of-art AI services and more specifically Cloud ML services [25]. Some of these important providers are [26]: Google Cloud Machine Learning Platform, Microsoft Cognitive Services, Amazon Machine Learning; and IBM Watson Analytics. Having said that, it is proposed that MO companies select from the available Cloud AI services offers the services that will enable them to differentiate from their competitors instead of building their own (including the typical models presented in Fig. 2: customer segmentation, channel optimization, brand sentiment, etc.). This approach will reduce the cost of hiring highly skilled, expensive and scarcest experts, also reducing the time to market of new insights generation solutions and spending less time/money in AI research. Most of these AI services will not even require minimum AI skills and will be available in the Cloud for software developers to integrate using APIs (Application Programming Interfaces). It should be noted that the suppliers of these new Cloud AI services also have the advantage of being able to train the ML algorithms with much more data that the company itself could have. This implies that, in many cases, they will be much more efficient models. A real example could be for a MO company to use cloud text analytics services to better understand what customers are writing about companies' products and services in the social media or to automatic review customer's comments in a website or in a contact center.

- **Automatic Machine Learning (AutoML)** is a software that can run in the cloud or in a local computer that enables developers with limited ML expertise to train high-quality models. With the shortage of specialized professional [27], the use of AutoML tools [28][29] in MO companies can increase the capacity of the insight generation team.

- **Deep Learning.** They are an evolution of artificial neural networks composed of multiple processing layers to learn representations of data with multiple levels of abstraction [30]. In MO organizations Deep Learning can play an important role in improving existing ML models to forecast time series data (sales), monitor the brand sentiment, understanding customer behavior using text analytics/ NLP (Natural Language Processing), detect customers using face recognition techniques, etc.[31][32].

Most of MO companies in the last decades used the Insight Generation process to increase company value, reduce costs, understand better their customers and competitors, and provide relevant insights for the business. Our purpose in this new framework is to add the **Data Supermarket**, a place to commercialize the data products generated by the insight generation process to other consumers exchanging them for a monetary value [33]. A typical insight generation process is developed by a multidisciplinary data scientist team, where raw data is converted into insights and data products. Our proposed contribution is to define the Data Supermarket as a key element in the Big Data monetisation strategy where data products created by this Big Data

Scientist team, will be shipped as data products in a form of services or products to internal or external consumers. Smart Steps from Telefonica is a successful example of building data products and providing them to external companies. Smart Steps is an insight solution that uses anonymized and aggregated mobile network data to provide useful insights [33]. Therefore, the principal goal of the Data Supermarket is to generate new revenue to the company and enable the organization to increase penetration in new markets, not explored before. Organizing the data into one single repository and creating a data product catalog is also a benefit for the end business users. Data will be democratized to the business users with a proper data definition [34].

In Fig. 6 the lifecycle to build data products and sell in the Data Supermarket is showed. As mentioned, the Data Supermarket concept is based on a normal supermarket. In daily life people can buy multiple products in a single place, they are all available in the supermarket. The same concept can be transposed to data. Using this same idea, it is proposed that data will come from different data suppliers (source systems), ingested/moved to the Raw Zone (landing zone) where the raw data is organized in a catalog (raw area). The next step is to transform the raw data into different data products according to the internal and external business user's needs. Like the experience of going to a food supermarket, where all the products are gathered and arranged in an organized manner.
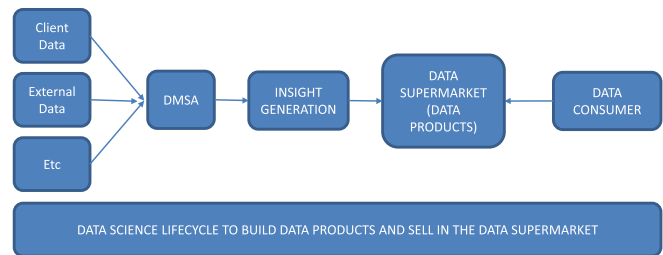

Fig. 6. Data science lifecycle to build data products and sell in the data supermarket.

The benefit of a data supermarket is the synergy of having best practices applied to activities like data collection, data transformation, data storage and data consumption. All the complexity of dealing with data is hidden from the end business user and presented in the Data Supermarket in a better user experience. The presentation or Data Supermarket store will handle the important tasks of data access, data security and data commercialization (free or paid) for internal or external users. The data supermarket store is the broker responsible for selling data products and for allowing market-oriented companies to profit from the data products available in the company data asset. The future data supermarket/data marketplace could become an important key asset of the modern MO organization.

D. Action

The MO scheme requires to be continuously adapting to changes in this market precisely (behavior of customers, competitors, etc.). It is for this reason that this Action layer should facilitate this continuous adaptation. In this layer, some components were added with the aim of automating many of their tasks and therefore drastically reduce the staff that performs them. This staff will be largely replaced by automated agents to perform some specific tasks with more velocity and accuracy. Thus, it is proposed the addition of new components, real-time alerts and chatbots, which are described below:

- **Real-time alerts system.** A MO company needs to be able to notify in real time, to their decision makers, the essential information for decision-making. The real-time notification system will communicate with the decision maker through the best notification option available and desired by the user. The most common

notification mechanisms/systems are SMS, Phone Calls, e-mail, WhatsApp, iOS/Android mobile notification, desktop alert, mass notification, led wall boards, drones, autonomous cars and robots. These systems provide only the essential information that needs to be monitored in real time; and in case that certain rules happen the required actions can be taken consequently.

- **Chatbots.** Answering to clients is another important task that MO companies need to accomplish, and there is a technological capability that can help organizations with this requirement: the chatbots. A chatbot is a software aimed at simulating the conversation of a human being [35][36], i.e. that can interact with humans by text or voice, responding queries using sophisticated natural language processing and speech recognition techniques, they can also retrieve historical information from the DMSA layer and respond personalized questions like what is my current bank balance, what is my national insurance number, how many days do I still have available for annual leave, etc. The insurance company Norwich Union, an Aviva company, is a real example of using chatbots as an automated customer service representative. These virtual agents (chatbots) were designed to help with general queries regarding products [37].

- **DevOps (Development Operations)** [38]-[40]**.** In order to enable MO companies to better serve their customers and compete more effectively in the market, it is proposed the usage of DevOps, which is "*a combination of cultural philosophies, practices, and tools that increase an organization's ability to deliver applications and services at high velocity*"[41]. This component is especially important for the development of technical services of companies since it is a methodology that allows its automation. In this way, the knowledge obtained from the corresponding layer can be implemented more quickly.

E. Outcomes

Outcomes layer includes the continuous motorization of the Key Performance Indicators (KPI) in order to check the effectiveness of the actions. It should be noted that many of these business actions are proposed or executed directly by ML models, so it is necessary to control these automatic decision makers. Thus, the knowledge discovered by these AI techniques is not universal, and its results will be degraded as the behavior of the market changes (consumers, competitors, etc.). Therefore, it is considered essential to incorporate constant monitoring and management of these ML models:

- **ML Model Management** [42]. It is a set of techniques and software that are used to manage ML models. It automatizes the ML lifecycle [43] in MO organizations. It helps organizations to measure the results of the ML models and guarantee the historical of all the ML iterative process. In this way, when it is identified that a ML model is no longer useful, it will be automatically replaced by others.

IV. Concluding Remarks and Future Work

The market is constantly evolving; thus, MO organization has a huge advantage compared to non-market-oriented companies [44]. This difference has been increased in those MO companies that have effectively used the data for decision making, incorporating AI techniques to increase sales, reduce costs, improve customer satisfaction, develop new value propositions, etc.

Because of the emergence of Big Data and AI, every company will have to invest in a new data and AI strategy, learn how to leverage existing internal data, buy new data from other companies and extend new capabilities to become data-driven companies [45]. This journey is mandatory in the digital era that we all live, there is

no way a company will survive without shifting to a data culture, the market is evolving rapidly, and new products will disrupt companies and change entire industries.

In the context that there are many disoriented companies on how to adapt their classic conceptual architectures and take advantage of the great potential of this immense volume of data, this work proposes a conceptual Big Data Science Cloud Architecture, that aims to set the foundational building blocks for a modern MO organisation, capable to support Data and AI use cases and provide the flexibility necessary to add or remove capabilities according to market needs.

Companies that adopt the proposed formal framework will bring agility and many possibilities to get new marketing insights and to innovate and create new P/S individually adapted to the needs of each of its customers [46].

As an additional advantage, companies that know how to make their own decisions based on data in the context of Big Data, can consider new lines of business based on the commercialization of data and knowledge. For this reason, it has been incorporated into our formal framework the Data Supermarket component dedicated to selling Data Products that are powerful weapons that will drive innovation and additional revenue for organizations.

As future work, we plan to focus on topics such as: IoT, AutoML, Chatbots, Robotic Process Automation (RPA), Cloud, Data, AI Strategy and Data Supermarket applied to MO companies in different industries and measure the results achieved in short, medium and long term. We also plan to include automatic TD [47]-[49] applied to MO companies in the formal framework proposed.

References

[1] J. S. Armstrong, "The value of formal planning for strategic decisions: Review of empirical research", *Strategic management journal*, vol. 3, no. 3, pp. 197-211, 1982.

[2] A. K. Kohli, B. J. Jaworski, "Market orientation: the construct, research propositions, and managerial implications", *Journal of marketing*, vol. 54, no. 2, pp. 1-18, 1990.

[3] J. C. Narver, S. F. Slater, "The effect of a market orientation on business profitability", *Journal of Marketing*, vol. 54, no. 4, pp. 20-35, 1990.

[4] S. Negash, P. Gray, "Business intelligence", Handbook on decision support systems, pp. 175-193, Springer, Berlin, Heidelberg, 2008.

[5] DWI: The Data Warehouse Institute, URL (consulted 2019): https://tdwi.org/Home.aspx.

[6] M. D. Stone, N. D. Woodcock, "Interactive, direct and digital marketing: A future that depends on better use of business intelligence", *Journal of research in interactive marketing*, vol. 8, no. 1, pp. 4-17, 2014.

[7] A. G. Picciano, "Big data and learning analytics in blended learning environments: Benefits and concerns". *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, no. 7, pp. 35-43, 2014.

[8] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, B. Becker, "The data warehouse lifecycle toolkit". John Wiley & Sons, 2008.

[9] A. Gandomi, M. Haider, "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management*, vol. 35, no. 2, pp. 137-144, 2015.

[10] A. Baldominos, F. de Rada, Y. Saez, Yago, "Big Data Analytics Solution for Intelligent Healthcare Management", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no 7, pp. 13-20, 2018.

[11] D. J. Bodas-Sagi, J. M. Labeaga, "Big Data and Health Economics: Opportunities, Challenges and Risks", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no 7, pp. 47-52, 2018.

[12] L. Atzori, A. Iera, G. Morabito. "The Internet of Things: A survey", *Computer Networks*, vol. 54, no. 15, pp. 2787-2805, 2010.

[13] P. Fraga-Lamas, T. Fernandez-Carames, L. Castedo, "Towards the Internet of smart trains: A review on industrial IoT-connected railways", *Sensors*, vol. 17, no. 6, pp. 1457, 2017.

[14] B. Nguyen, L. Simkin, L, "The Internet of Things (IoT) and marketing: the state of play, future trends and the implications for marketing", *Journal of marketing management*, vol. 33, pp. 1-6, 2017.

[15] Gartner: Data Management Solutions for Analytics, URL (consulted 2019): https://www.gartner.com/reviews/customers-choice/data-warehouse-solutions

[16] Gartner: Magic Quadrant for Data Management Solutions for Analytics, URL (consulted 2019): https://www.gartner.com/doc/3898487/magic-quadrant-data-management-solutions

[17] P. Mell, T. Grance, "The NIST definition of cloud computing", *NIST Special Publication*, pp. 800-145, 2011.

[18] M. Iorga, A. Karmel, "Managing risk in a cloud ecosystem", *IEEE Cloud Computing*, vol. 2, no. 6, pp. 51-57, 2015.

[19] J. Dixon, "Pentaho Hadoop and Data Lakes." URL (consulted 2019): http://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes

[20] D. E. O'Leary, "Embedding AI and crowdsourcing in the big data lake. *IEEE Intelligent Systems*, vol. 29, no. 5, pp. 70-73, 2014.

[21] P. Pasupuleti, B. S. Purra, "Data Lake development with big data". *Packt Publishing Ltd*, 2015.

[22] A. LaPlante, B. Sharma, "Architecting Data Lakes", O'Reilly Media, 2014.

[23] D. J. DeWitt, J. Gray, "Parallel database systems: The future of high-performance database processing". University of Wisconsin-Madison Department of Computer Sciences Technical Report, 1992.

[24] W. H. Inmon, D. Linstedt. "A Brief History of Big Data in Data Architecture: a Primer for the Data Scientist", Morgan Kaufmann, Boston, 2015.

[25] R. Arulmurugan, K. R. Sabarmathi, H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques", *Cluster Computing*, pp. 1-11, 2017.

[26] A. L'heureux, K. Grolinger, H. F. Elyamany, M. A. Capretz, "Machine learning with big data: Challenges and approaches". *IEEE Access*, vol. 5, pp. 7776-7797, 2017.

[27] T. H. Davenport, D. J. Patil, "Data Scientist", *Harvard business review*, vol. 90, no. 5, pp. 70-76, 2012.

[28] F. Hutter, Balázs Kégl, R. Caruana, I. Guyon, H. Larochelle, et al.. "Automatic Machine Learning (AutoML)". ICML 2015 Workshop on Resource-Efficient Machine Learning, 32nd International Conference on Machine Learning, Lille, France, Jul 2015.

[29] R. S. Olson, J. H. Moore, "TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning". Proceedings of the Workshop on Automatic Machine Learning, in PMLR, pp. 66-74, 2016.

[30] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, vol. 521 no. 7553, pp. 436-444, 2015.

[31] Z. C. Lipton, "The mythos of model interpretability", *ArXiv e-prints*, pp. 1-3, 2016.

[32] W. Samek, T. Wiegand, K. R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models", *arXiv preprint*, 2017.

[33] Telefonica. "Smart Steps". URL (consulted 2019): https://www.business-solutions.telefonica.com/en/products/big-data/business-insights/smart-steps/

[34] B. Bengfort, J. Kim, "Data Analytics with Hadoop: An Introduction for Data Scientists." O'Reilly Media, Inc, 2016.

[35] B. A. Shawar, E. Atwell, "A comparison between Alice and Elizabeth chatbot systems". University of Leeds, School of Computing research report, 2002.

[36] J. Weizenbaum, "ELIZA---a computer program for the study of natural language communication between man and machine", *Communications of the ACM*, vol. 9, no. 1, pp. 36-45, 1966.

[37] M. C. Jenkins, R. Churchill, S. Cox, D. Smith, "Analysis of user interaction with service oriented chatbot systems", in International Conference on Human-Computer Interaction, pp. 76-83, 2007.

[38] M. Walls, "Building a DevOps culture", O'Reilly Media, Inc, 2013.

[39] E. Mueller, "What's a DevOp? The Agile Admin". URL (consulted 2019): http://theagileadmin.com/2010/10/08/whats-a-devop/

[40] D. Samovskiy, "The rise of DevOps"; URL (consulted 2019): http://www.somic.org/2010/03/02/the-rise-of-devops/

[41] Amazon AWS. What is DevOps. URL (consulted 2019): https://aws.amazon.com/devops/what-is-devops/

[42] M. Vartak, H. Subramanyam, W. E. Lee, S. Viswanathan, S. Husnoo, S. Madden, M. Zaharia, "Model DB: a system for machine learning model management." In Proceedings of the Workshop on Human-In-the-Loop Data Analytics, pp. 14, ACM, 2016.

[43] M. Zaharia, A. Chen, A. Davidson, A. Ghodsi, S. A. Hong, A. Konwinski, et al, "Accelerating the Machine Learning Lifecycle with MLflow" *Data Engineering*, vol. 39, 2018.

[44] V. Kumar, E. Jones, R. Venkatesan, R. P. Leone, "Is market orientation a source of sustainable competitive advantage or simply the cost of competing?", *Journal of marketing*, vol. 75, no. 1, pp. 16-30, 2011.

[45] M. Ye, G. Li, "Internet big data and capital markets: a literature review", *Financial Innovation*, vol. 3, no.6, pp. 1-18, 2017.

[46] R. A. Carrasco, M. F. Blasco, J. García-Madariaga, E. Herrera-Viedma, "A Fuzzy Linguistic RFM Model Applied to Campaign Management", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 4, pp. 21-27, 2019.

[47] J.A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, E. Herrera-Viedma. "Carrying out consensual Group Decision Making processes under social networks using sentiment analysis over comparative expressions", *Knowledge-Based Systems*, vol. 165, pp. 335-345, 2019.

[48] F.J. Cabrerizo, M.R. Ureña, W. Pedrycz, E. Herrera-Viedma. "Building consensus in group decision making with an allocation of information granularity", *Fuzzy Sets and Systems*, vol. 255, pp.115-127, 2014.

[49] Y. Liu, Y. Dong, H. Liang, F. Chiclana, E. Herrera-Viedma. "Multiple attribute strategic weight manipulation with minimum cost in a group decision making context with interval attribute weights information", *IEEE Trans. Syst. Man and Cyb: Systems*, Accepted, 2019.

Caio Moreno

Caio Moreno is a Big Data Architect, Data Scientist and Data and AI Solution Architect with 18 years of professional experience, working for companies like AOL (America Online), Pan American Health Organization, IT4biz (CEO and Founder), Hitachi Vantara, Pentaho, and many companies and governments in multiple countries and continents. He is currently a part-time Ph.D. Student at Complutense University of Madrid and Advanced Analytics Manager at Avanade London Office (An Accenture and Microsoft owned company). He holds a master's degree in Business Intelligence and Data Mining at the Complutense University of Madrid, a specialization in Data Science at University Camilo José Cela (UCJC)/U-TAD, an MBA at Getulio Vargas Foundation (FGV-SP) and a Specialization in Software Development/Java at The Federal University of Technology – Paraná (UTFPR). His main field of research is focused on Data Science, Big Data, IoT, Cloud and AutoML (Automated Machine Learning) applied to Marketing and in the financial sector. He has been a professor at The State University of Western Paraná (UNIOESTE) – Brazil (2010-2012) and EOI (Escuela de Organización Industrial) – Spain (2016-2017). Also, a public speaker in many international events in Brazil, Europe, Africa and USA. Fluent in English, Spanish and Portuguese. He is married and has 2 daughters.

Ramón Alberto Carrasco

Ramón Alberto Carrasco has worked in the financial sector for over 24 years and has extensive experience acting as a leader of several departments related to Analytical Marketing and Business Intelligence. He completed his PhD in data mining at the University of Granada. His main field of research is focused on Data Science applied to Marketing (Data Mining, Multicriteria Decision-Making, Sentiment Analysis, Recommender Systems, etc.) on which he has written several textbooks as well as around 60 research papers published in international journals and conferences specialized on these topics. He is currently involved in various (European, national, regional and local) Research, Development and Innovation (R&D&I) Projects. Dr Carrasco is a member of the SCI2S and SECABA Research Groups and is also serving as reviewer for several international journals and conferences. He has been an professor at the University of Granada, Granada, Spain (2004-2013) and currently at the Complutense University, Madrid, Spain (from 2013).

Enrique Herrera-Viedma

Enrique Herrera-Viedma is Professor in Computer Science and A.I in University of Granada and currently the new Vice-President for Research and Knowlegde Transfer. His current research interests include group decision making, consensus models, linguistic modeling, and aggregation of information, information retrieval, bibliometric, digital libraries, web quality evaluation, recommender systems, and social media. In these topics he has published more than 250 papers in ISI journals and coordinated more than 22 research projects. Dr. Herrera-Viedma is Vice-President of Publications of the IEEE SMC Society and an Associate Editor of international journals such as the IEEE Trans. On Syst. Man, and Cyb.: Systems, Knowledge Based Systems, Soft Computing, Fuzzy Optimization and Decision Making, Applied Soft Computing, Journal of Intelligent and Fuzzy Systems, and Information Sciences.