

Big Data: Concept, Potentialities and Vulnerabilities

Fernando Almeida ^{a*}

^a *University of Porto, INESC TEC, Portugal*

Abstract

The evolution of information systems and the growth in the use of the Internet and social networks has caused an explosion in the amount of available data relevant to the activities of the companies. Therefore, the treatment of these available data is vital to support operational, tactical and strategic decisions. This paper aims to present the concept of big data and the main technologies that support the analysis of large data volumes. The potential of big data is explored considering nine sectors of activity, such as financial, retail, healthcare, transports, agriculture, energy, manufacturing, public, and media and entertainment. In addition, the main current opportunities, vulnerabilities and privacy challenges of big data are discussed. It was possible to conclude that despite the potential for using the big data to grow in the previously identified areas, there are still some challenges that need to be considered and mitigated, namely the privacy of information, the existence of qualified human resources to work with Big Data and the promotion of a data-driven organizational culture.

Keywords:

Big Data;
Distributed Computing;
Data Analysis;
Predictive Analytics;
Hadoop.

Article History:

Received: 15 January 2018
Accepted: 23 February 2018

1- Introduction

Information is now increasingly important and a successful differential, as the whirlwind of external events forces organizations to face new situations. The information becomes fundamental for the discovery and introduction of new technologies, as well as for exploring opportunities investment. It has the power to detect new opportunities, sign of the threats and reduces uncertainties during the decision-making process and, consequently, increases its quality. In this sense, the differential of companies and professionals is directly related to the value they give to information, knowledge and how they use it in meeting the demands of the market and in the search for innovative solutions.

The process of decision-making is complex and rational, contemplating factors such as intuition, experiences and knowledge. Business managers are constantly going through situations where they are faced with a number of different paths, and must choose the one that leads the organization to achieve its results. Therefore, information plays a fundamental role in the decision-making process, in order to identify the various alternatives and their consequences.

However, capturing relevant information for the company is a complex and difficult task. Useful data can come from anywhere and there is an increasing number of heterogeneous numbers of devices that capture data from different sources. The compilation and sharing of detailed information is only possible through the use of information and communication technologies (ICT), and this data can come from suppliers, consumers, partners and competitors. To this large volume of data coming from multiple heterogeneous sources we call Big Data, which is the next frontier for business innovation and productivity. For that reason, companies should be aware of the potentialities and vulnerabilities of Big Data and create strategies to handle large volumes of data in order to take advantage of its many potentialities.

2- Concept of Big Data

Although the term "big data" is relatively new, the act of collecting and storing large amounts of information for eventual data analysis is quite old. Companies in diverse sectors of activity, mainly those of bigger dimension and with greater volume of data, have developed solutions of business intelligence (BI) to support business management processes. BI is characterized by the use of a set of methodologies, processes, structures and technologies that transform a large amount of raw data into useful information for making strategic decisions [1]. Table 1 performs a comparative

* **CONTACT:** Almd@fe.up.pt

DOI: <http://dx.doi.org/10.28991/esj-2018-01123>

© This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

analysis between the concept of BI and big data analytics.

Table 1. Traditional Analytics (BI) vs. Big Data Analytics.

Criteria	Traditional Analytics (BI)	Big data Analytics
Focus	Descriptive analytics and diagnosis analytics	Predictive analysis
Data sets	Limited data sets with structured data. Adoption of simple data models	Large scale data sets with more types of data. Adoption of complex data models
Analysis	Looks to what happened, and why?	Provide new insights and forecasts

Big data was initially defined based on the following three Vs [2]:

- **Volume:** organizations collect data from a wide variety of sources, including business transactions, social networks, and information from sensors or data transmitted from machine to machine. In the past, storing such a large amount of information would have been a problem, but the emergence of new technologies, such as Hadoop, have alleviated the burden;
- **Velocity:** data flow at a huge rate and must be handled in a timely manner. RFID tags, sensors, cell phones and smart meters are driven by the need to deal with huge amounts of data in real time;
- **Variety:** data is generated in all types of formats, from structured data, numerical data in relational databases, to unstructured text documents, to e-mail, video, audio, stock quotes, and financial transactions.

However, over the time, six new dimensions have been added to the big data concept [3-7]:

- **Variability:** in addition to the increasing speed and variety of data, data flows can be highly inconsistent with periodic peaks. On a daily basis, seasonal peaks or peak events generated can be very challenging to manage, particularly when we deal with unstructured data;
- **Value:** it should also be recognized that large data must have value. In this sense, it is important to ensure the return on investment and guarantee that the insights generated are based on accurate data that lead to measurable improvements;
- **Veracity:** it is important to ensure that there is a high data quality and document the data provenance in terms of inputs, entities, systems, and processes that influence data of interest;
- **Validity:** valid data is essential for making right and accurate decisions. Therefore, we should avoid to use corrupted data in analytical exploration of data;
- **Visualization:** the presentation of the data must be intuitive and graphically appealing. In fact, as we have more data, the more important is the ability to visualize them, since otherwise it would be very difficult to intuitively identify patterns or correlations between the data;
- **Value:** consideration should be given to the true value of data, in which data value must exceed its cost or management. Therefore, storage and data access may be cost effective.

The characteristics of Big Data can be grouped into five categories [7]: (i) collecting data; (ii) processing data; (iii) integrity data; (iv) visualization data; and (v) worth of data. These nine dimensions are mapped in each of these categories, as indicated in Table 2.

Table 2. Categories of Big Data [7]

Categories	Big Data Characteristics
Collecting data	Veracity, Variety
Processing data	Velocity, Volume
Integrity data	Validity, Variability, Volatility
Visualization data	Visualization
Worth of data	Value

These nine dimensions don't need to be taken together as the perfect definition. There are authors that argue that the combination of "volume + velocity + variety" is sufficient to convey an acceptable notion of big data. From this perspective the additional four dimensions are implicitly present in the above definition.

The process of dealing with large volumes of data is complex, and traditional programming approaches and data organizations models are no longer suitable for working with large volume of data, particularly when most of these data have no structure. Because of these characteristics, handling and processing big data requires specific tools and techniques.

One of the most well-known frameworks is the Hadoop. It is an open source implementation of the Map-reduce programming paradigm, introduced by Google to process and analyze large datasets. The programs that are developed using this paradigm perform parallel processing of data sets and can therefore be run on servers without much effort. The reason for the scalability of this paradigm is the intrinsically distributed nature of the solution's operation. A large task is divided into several small tasks that are then executed in parallel on different machines and then combined to arrive at the solution to the larger task that started it all [8]. Examples of using Hadoop are analyzing user patterns on e-commerce sites and suggesting new products that they can purchase.

The Hadoop framework consists of two main components: storage and processing. The first is the Hadoop Distributed File System (HDFS), which handles data storage between all the machines on which the Hadoop cluster is running [9]. The second, Map-Reduce, manipulates the processing part of the framework [10]. Figure 1 shows the high level architecture of Hadoop.

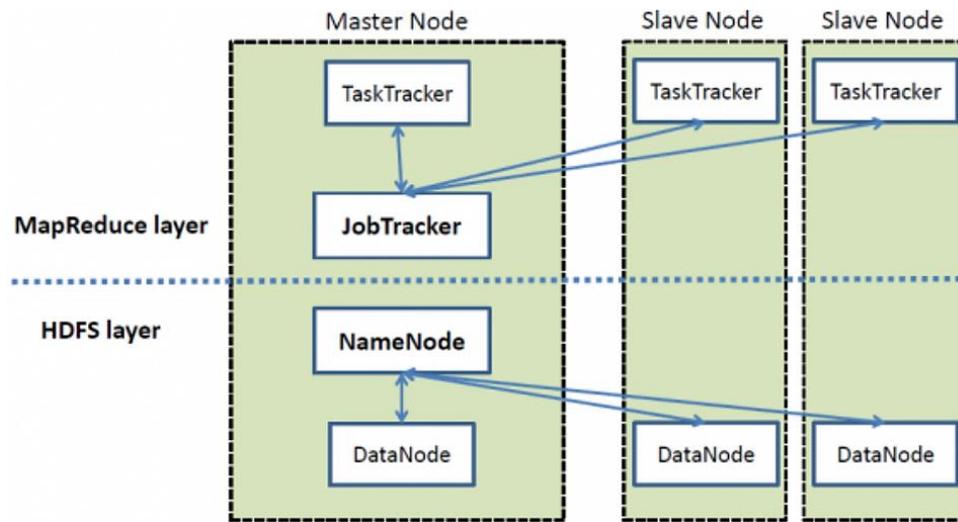


Figure 1. High level architecture of Hadoop [11].

HDFS is a scalable, distributed file system whose design is heavily based on the Google File System (GFS), which is also a distributed file system. Distributed file systems are needed once the data become too large to be stored on only one machine. Because of this, all the complexity and uncertainties from the network environment come into play, which makes network file systems more complex than ordinary file systems. HDFS stores all files in blocks. The default block size is 64Mb. All files on the HDFS have multiple replicas, which help in parallel processing. HDFS clusters have two types of nodes: (i) primary and secondary namenodes; and (ii) datanodes. Below we briefly introduce the mission of these nodes [9]:

- Primary namenodes: manages the file system namespace. It manages all files and directories. Namenodes have the mapping between files and the blocks in which they are stored;
- Secondary namenodes: this node is responsible for checking namenode information. In case of failure, these nodes can be used to restart the system;
- Datanode: stores the data as blocks. Datanodes report to namenodes about the files they have stored so that the namenode is aware and the data can be processed.

Map-Reduce is a programming paradigm in which each task is specified in terms of mapping and reduction functions. Map-Reduce offers a way to process large volumes of data by distributing the processing in many machines so that it is processed in an acceptable time. This distribution implies parallel processing since the same function is applied on all machines, but on different data sets in each of them. The approach adopted by Map-Reduce takes advantage of parallelism to split the data load, rather than split the processing steps. Each component is responsible for completely processing a small set of data, rather than processing all the data at a particular stage of computation.

Map-Reduce maps a dataset into a collection of tuples (key, value) and, after that, reduce all tuples with the same key to produce the final output of the processing. This approach adopts the principles of abstracting all the parallelization complexity of an application using only the Map and Reduce functions. Computing is distributed and controlled by the framework, which uses its distributed file system and messaging exchange protocols to run a Map-Reduce application. The processing is divided into three phases: (i) an initial mapping step, where several mapping tasks are performed; (ii) an intermediate step where the data are collected from the mapping tasks, grouped and made available for the reduction tasks; and (iii) a reduction step where various reduction tasks are performed by grouping the common values and generating the output of the application [10]. The process of how Map-Reduce operates can be visualized in Figure 2.

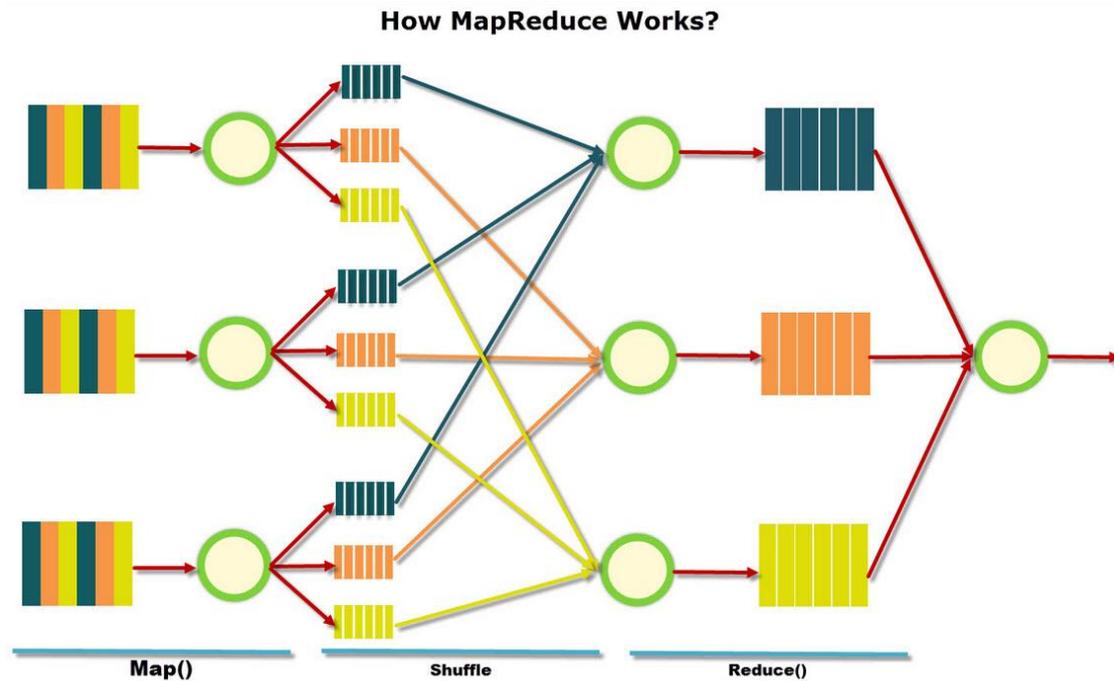


Figure 2. Visualization how Map-Reduce works [12].

3- Potentialities of Big Data

Several studies have reported the use of Big Data in a wide set of different contexts. In this sense, we approach some of these projects by organizing this information by activity sector.

3-1- Financial Services

Financial services are good examples of how technology can help in the processing of large amounts of data. Big data can be used to reduce churn rates, in order to anticipate possible customer exits. It can also be used to tighten customer relationships and customize services. For example, understanding how customers use credit cards and what kind of loans they need, help banks in creating new products that meet the customers' needs.

Sarrocchio et al. [13] advocate that unlike traditional business intelligence systems, new techniques and technologies used with big data allow Chief Financial Officers (CFOs) to gain useful information at a much lower cost. The study identified two groups of benefits: (i) tangible cost reductions; and (ii) intangible benefits. In the first group, we may find significant cost reductions by the alignment of process, harmonized reporting and centralized data sources. In the second group, we have more data-driven decision making, which will allow CFOs to discover new business opportunities.

Other studies complement the previous view, mentioning that financial services present high levels of big data, low levels of knowledge assets, and high levels of competitive intelligence activity [14]. The challenge of dealing with high levels of latency remains with the use of legacy services and critical operations that require a high level of operations' serialization [15].

3-2- Retail Services

Big data is fundamental to the digital transformation of the retail services. Sales depend on the in-depth knowledge of the target audience, which requires a systematic analysis of a large volume of information. The big data can be used to implement multi-channel marketing actions, in which it becomes necessary to integrate all communication channels to fully understanding the behavior of the client. This market awareness can be achieved through greater segmentation of the target audience to understand their consumption habits and preferences. In addition, this information can be used to create loyalty programs.

Shockley & Mercier [16] reported that around 62 percent of US retailers use information (including big data) and analytics in creating competitive advantage for their organizations. They also describe the use of retail services in a multi-channel shopping interaction. For example, a consumer might begin researching a product on a mobile app, purchase it online and pick it up at a store. Coordinating this multi-channel shopping interaction requires the analysis of a large volume of data to identify consumer preferences in each of these different channels of interacting with the retailer. One channel that is expected to have a strong growth potential is e-commerce through mobile devices. At this level,

there are studies that intend to understand emerging mobile checkout scenarios and customer reactions to these scenarios [17].

The potential for applying Big Data to retailing is also proven by Grewal et al. [18] that consider five drivers in the future of retailing, respectively: (i) technology and tools to facilitate decision making; (ii) visual display and merchandise offer decisions; (iii) consumption and engagement; (iv) Big Data collection and usage; and (v) analytics and profitability.

3-3-Healthcare Services

Digital Health is the convergence of digital revolutions with the health field, genetic advances, and the speed of technological improvements. Digital Health is very important for the health industry, as well as for society in general, and it is only possible to reach them with the use of the big data. Digital Health requires the use of information systems to collect, aggregate and work structured and unstructured data linked to the health sector, generating clinical information that strengthens precision medicine. The big data can be used to prevent epidemics by monitoring a population on social networks. In conjunction with the use of statistical analysis, it can help in the early visualization of the possibility of an outbreak of an epidemic, giving health institutions time to adjust to sudden increases in demand for care and medication. On the other hand, it can also be used in telemedicine, through the use of information technologies and large data analysis in the provision of clinical information of patients.

Wu et al. [19] studied the adoption of wearable device together with big data technology. They provide practical guidance to wearable device manufacturers on optimizing competition strategies and offer insights to social planners on potential policy-making to promote better healthcare services. Dimitrov [20] suggests the use of medical devices and mobile apps in telemedicine solutions via medical Internet of Things (mIoT).

Despite the existence of several studies that report on the successful use of Big Data in Healthcare Services, some obstacles persist. Kruse et al. [21] report issues of data structure, security, data standardization, storage and transfers, and managerial skills such as data governance. Actually, Big Data security and privacy is still the obstacle most widely considered in scientific studies, in which several user's accessing mechanisms and strategies are proposed [22, 23].

3-4-Transport Services

Transport infrastructures are increasingly involved in technology. Travel and transportation companies must consider an increasing amount of information to make strategic decisions. These data come from multiple sources such as transactional data, asset data, supply chain data, call center logs, competitor pricing, weather data, NFC or RFID. Therefore, in order to deliver "smart data", more elaborate and sophisticated algorithms are needed. These algorithms can be used in the optimization process of routes, resolution of traffic issues, road prevention, among others.

Kochhar [24] states that big data is already transforming both the plan phase and operations phase of the rail transportation. He suggests that big data can be used in the following phases: (i) planning and demand modeling; (ii) predictive maintenance; (iii) event response; and (iv) personalized services. For its part, Kanniyappan and McQueen [25] predict what can be the future truck of 2025. They suggest that it will communicate with other vehicles and connect to growing sources of online information as big data balloons on the road. Special cameras, wireless Lan, and multiple radar system can be used to watch the road, the side of the road, and cars and trucks behind the vehicle. Additionally, computerized controls of the vehicle will make it more fuel efficient.

3-5-Agriculture Sector

Precision farming has gained prominence in several countries, such as USA, Brazil or Australia. The use of big data technologies generates a new agricultural revolution, guaranteeing an increase in productivity and production without the need to increase the area. The decision on the type of seed to plant, how to plant, how to fertilize, how to use defensive, harvest time, storage, transportation, and marketing decisions that need to be made by the farmer are already being made by technology. Wolfert et al. [26] state that big data is expected to have a large impact on smart farming and in the whole supply chain. It will also cause major shifts in roles and power relations among traditional and non-traditional players.

Bronson and Knezevic [27] reported the use of Integrated Field Systems (IFS) on the agri-food sector in the North American food system. The IFS can be used to collect information about soil conditions, weed varieties and weather. It can also be used to identify weeds and map weed pressures in a digital mapping tool, which will help framers identifying chemical needs and therefore areas of possible investment in research and development. This information will help framers to minimize risk and streamline decision-making.

3-6-Energy Sector

The potentialities of using big data analytics for companies in the energy sector are expressive. It can be used in the improvement of marketing campaigns, in operational fields and to explore new business opportunities. The advantages offered by the adoption of big data are not restricted to very competitive environments, but also in more regulated

markets. Big data applications can be present along the entire value chain, from the generation to the commercialization of services in real time, in the process of back-end and front-end optimizations, in the management of clients through the analysis of the profile of consumption, and in the distribution energy process to ensure a more balanced between supply and demand data.

Porter [28] describes the adoption of big data using different scenarios in the energy sector. For example, generators use market data to optimize their dispatch, ramping flexible assets up and down in response to real time supply and demand forecasts. This can be done by collecting information regarding plant operational performance, availability and technical parameters. Other example is the adoption of smart meters to provide both users and suppliers with granular consumption data. This can be done by integrating a wide range of IoT devices and sensors.

In addition to these scenarios, two areas emerge as presenting a high potential for Big Data usage. One of them is the adoption of smart grids, which allow optimize the process of generation, distribution and consumption of electricity through the introduction of information and communication devices and technologies [29]; another area is energy efficiency through creation of energy efficient buildings in the process of managing temperature and solar exposure [30].

3-7- Manufacturing Sector

In recent years manufacturers have been able to reduce waste and variability in their production processes by introducing new management practices in the fight against waste. In addition to cost savings, these practices provided significant improvements in the final product quality. Big data can be used across multiple domains. One of them is the integration of the supply chain to achieve better condition in the supply of raw material and to launch real-time orders according to the needs of customers. Furthermore, it can be used in the analysis of historical process data to identify patterns and relationships between various internal processes.

Agrawal [31] reported that one of the areas that will benefit more from the growth of big data market is the manufacturing industry, with revenues projected to reach \$39 billion by 2019. Delgado [32] emphasizes the role of big data in improving quality and safety. He refers that companies can install computerized sensors on assembly lines to allow that generated data could be used to enable the improvement of the quality and safety of the manufacturing process.

Smart manufacturing appears as a new paradigm in the development of the manufacturing sector, which intends to take advantage of information technologies to increase the degree of autonomy and flexibility of the production process. Associated to this concept arises the Industry 4.0 concept that intends to change the paradigm of the productive process, in which we have data-driven decisions. To this end, and in order to concretize the Industry 4.0 revolution, the use of Big Data becomes essential [33-35].

3-8- Media and Entertainment Sector

Media and entertainment companies seek to increase the value of their media assets by combining creativity with technological advancements. This offers significant opportunities for broadcasters, publishers, advertising agencies, owners of digital platforms, and content providers capable of understanding the impact of this change. Nowadays, many audiences are becoming one's audience, forcing the media sector to become more data-driven. Digital and media groups need to understand changes in consumption patterns, including what programs and content are viewed through digital and traditional channels, as well as other content sources.

Lippell [36] considers that media companies are an early adopter of big data technologies, because it enables them to drive digital transformation, exploiting new sources of data from both inside and outside the organization. Almost every media associated business with large volumes of data can leverage big data, namely [37]:

- Video publishers – independent or private video creators who publish content in different types of format;
- Media owners – businesses that sell content through retail or mass content distribution mediums;
- Gaming companies – video game makers that can log gamer reaction and customize the experience of the game according to the experience of players' use;
- TV channels – television channels that can tailor content and advertising according to the audience at different hours of the day.

3-9- Public Sector

Decisions by public managers involve difficult daily budget choices, program prioritization, natural disaster prevention and epidemics, as well as infrastructure investments. Organizing all this operational and strategic action plans based solely on intuition invariably generates administrative collapse. Big data for public management can help public

managers in areas such as combat against corruption, strengthening the implementation of smart cities and monitoring the level of population satisfaction.

The benefits of big data in the public sector can be grouped into three major areas [38]: (i) improvements in efficiency; (ii) improvements in effectiveness; and (iii) learning from the performance of public services. Lauchlan [39] describes the adoption of big data to deliver social benefits. The Home Office Child Abuse Image Database in the UK has benefited from the use of big data in the investigation of child abuse crimes and child protection. Using the image's unique identifiers and metadata, local protection authorities can check devices they've seized from suspects against the material on the database much more quickly. Previously, a process containing 10000 images would take around 3 days to review, can be reviewed only in an hour, which originates a cheaper, less labor-intensive and more efficient.

4- Vulnerabilities of Big Data

In addition to the many potentialities offered by the big data, there is also a set of vulnerabilities that need to be known and understood. Three types of vulnerabilities associated to big data management and processes were identified: (i) security; (ii) privacy; and (iii) absence of standards.

Big data users must be concerned about the potential vulnerability of their data, and take a proactive approach to implement security mechanisms. The importance of the data for the organization, the legal requirements to process and store data, and whether the data storage is totally secure against unauthorized access should be questioned. According to Moreno et al. [40] the major security flaws that can be found in big data are related to the lack of authentication mechanisms, such as user id and password, and the lack of secure channels for access cloud databases.

The big data computing process is distributed across multiple servers or nodes. A set of data running in the Hadoop is composed of several nodes, sometimes in the order of hundreds or thousands, and a large volume of information is constantly migrated between each of these nodes. This highly dynamic environment of distributed computing, and the enormous amounts of elastic data involved, makes big data security a major challenge. Many traditional security approaches, such as firewalls and intrusion detection systems (IDS), are not designed to protect distributed file systems and handle this huge amount of data.

Hadoop currently supports security mechanisms to mitigate these risks, including the use of Kerberos, the adoption of firewalls, and the implementation of basic HDFS permissions [41]. However, Kerberos is not a mandatory requirement for a Hadoop cluster, which makes it possible to run entire groups of data without implementing any security. In addition, Kerberos is difficult to install and configure in the cluster, and the integration of Active Directory and Lightweight Directory Access Protocol (LDAP) is also a challenge. Therefore, other security practices have emerged, such as the implementation of security controls closer to the data, and the implementation of cryptographic models in all data, both statically and in transit.

Big Data, social media, and predictive algorithms open up a new field of privacy discussion. All technology is neutral. However, what we do about it can have a positive or negative impact on society. Companies can use the vast amount of data that they have on their customers for a promotion that significantly improves their customers' experience, or it can be used unlawfully to violate their privacy. Big Data unquestionably increases the risk of privacy breaches in various domains, particularly in the Internet and social networks contexts. Data extracted from social networks such as Facebook, LinkedIn or Twitter allow companies to obtain very diverse information about their users. For example, it would be possible to know patterns of travel, where they eat lunch every day, entertainment options, etc.

The question of privacy is associated with the concept of what will be ethically acceptable or not. In addition to the need for a set of more explicit and comprehensive laws on this subject, it is essential that companies, governments and public entities adopt a code of conduct. This code of conduct should unambiguously identify the purpose of the data collection, its scope and the limits for its collection. Pentland from MIT advocates the existence of a "New Deal" of data, in which personal data are treated as a good, where people would have their rights assured over their own data [42]. That is, regardless of who collects the data about the user, they will belong to the user and he/she can access to the data whenever he/she wants.

Finally, one point of vulnerability lies in the absence of widely accepted standards in the big data domain. Standards are used to enable, to promote, to measure and govern the use of technology across a wide spectrum of communities. Standardization is essential in any mass-deployed system, considering that it increases independent use and comparative evaluations of the technology [43].

Standard communication protocols are also needed to build the new generations of smart appliances. Currently, there are a large number of communication protocols in various application domains of the big data, which creates interoperability and large-scale customization issues. The lack of standardization inhibits innovation and creates barriers to the adoption of big data by users, since it becomes necessary to know a great number of different protocols. Therefore,

there is a need for greater collaboration between industry and government entities that simultaneously promote standardization and competition in the market.

5- Big Data: Opportunities and Privacy Challenges

The ability to process and analyze a large volume of data was only available to a minority of large companies until few years ago. However, with technological advances, the emergence of open source frameworks and tools has enabled data to be exploited by more organizations effectively and efficiently. It is also important to recognize that the popularization of electronic devices and the advent of IoT have exponentially increased the volume of data generating new market opportunities. Research has been carried out on the proposed models that allow IoT to encrypt data to enforce data privacy [44-45].

Data emerge as a new driver of productivity, work, and innovation. However, most of the business models in the various sectors of activity mentioned in the previous section do not fully exploit the potential value of these data. Therefore, Big Data present a high potential still unexplored to get better business results.

The fear of investing in technology still in the embryonic stage and the need of changing organizational work processes are considered major issues that still delay the entry of many companies into Big Data [46-47]. It should also be mentioned that when we talk about Big Data, we are also addressing the topic of Cloud Computing, with many managers claiming the fear of sharing important information in an emergent field, and consequently considered by many as insecure [48]. In addition to this challenge, there are also some fears of companies on adopting a technology with limited availability. In fact, the use of the Internet can cause delays and unavailability of services, since the architecture of the Internet was not designed to offer Quality of Service (QoS) [49].

Privacy at Big Data is a growing business concern. Several models have been proposed and developed for privacy protection at different stages of the Big Data life cycle, such as data generation, data storage, and data processing [50]. Despite this, there are still few companies that have a Chief Privacy Officer, enough Big Data skills, and concrete measures to audit the processes [51]. All technology is neutral by itself, so its impact depends on how we use it.

Nowadays new tools for manipulating and exploiting Big Data appear, but these technological advances must also be aligned with the organizational culture of companies. It is essential to create a culture of data-based decision-making, the culture of transforming information into knowledge, which implies the creation of a data-driven culture in the enterprise.

6- Conclusion

The ultimate goal of big data techniques is to be able to identify useful and usable information in a timely fashion. In order to produce actionable analytics there is a need to analyze lots of disparate data and the ability to discover, access, store and retrieved lots of data. The solutions that enable big data analytics processes are systems that involve technical aspects, both at the time of collection, processing and presentation of unstructured obtained data. A central point of the today's big data evolution is the Hadoop framework, which offers an efficient file system and an ecosystem of solutions for storing and analyzing large data sets.

Big data can help companies in the process of taking operational, tactical and strategic decisions. There is a large panoply of potentialities in the use of big data in various sectors of activity, such as financial, retail, healthcare, transport, agriculture, energy, manufacturing, media and entertainment, and public sector. However, despite the many potentialities offered by the big data, there are also vulnerabilities related to data security, privacy, and standardization processes.

7- References

- [1] R. Sharda, D. Delen, and E. Turban, *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. Pearson Education, 2017. ISBN: 978-0134633282.
- [2] O. Ylojoki, and J. Porras, "Perspectives to Definition of Big Data: A Mapping Study and Discussion", *Journal of Innovation Management*, vol. 4, no. 1, pp. 69-91, 2016. <http://hdl.handle.net/10216/83250>.
- [3] R. Kune, P. Konugurthi, A. Agarwal, R. Chillarige, and R. Buyya, "The Anatomy of Big Data Computing", *Journal Software – Practice & Experience*, vol. 46, no. 1, pp. 79-105, 2016. doi: 10.1002/spe.2374.
- [4] I. Lee, "Big Data: Dimensions, Evolution, Impacts, and Challenges", *Business Horizons*, vol. 60, no. 3, pp. 293-303, 2017. doi: 10.1016/j.bushor.2017.01.004.
- [5] C. Yang., Q. Huang, Z. Li, K. Liu, and F. Hu, "Big Data and cloud computing: innovation opportunities and challenges", *International Journal of Digital Earth*, vol. 10, no. 1, pp. 13-53, 2017. doi: 10.1080/17538947.2016.1239771.
- [6] U. Sivarajah, M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods", *Journal*

of Business Research, vol. 70, pp. 263-286, 2017. doi: 10.1016/j.jbusres.2016.08.001.

- [7] S. Owais, and N. Hussein, “Extract Five Categories CPIVW from the 9V’s Characteristics of the Big Data”, International Journal of Advanced Computer Science and Applications, vol. 7, no. 3, pp. 254-258, 2016. http://thesai.org/Downloads/Volume7No3/Paper_37-Extract_Five_Categories_CPIVW.pdf.
- [8] H. Bhosale, and D. Gadekar, “A Review Paper on Big Data and Hadoop”, International Journal of Scientific and Research Publications, vol. 4, no. 10, pp. 1-7, 2014. <http://www.ijsrp.org/research-paper-1014/ijsrp-p34125.pdf>.
- [9] B. Sahare, A. Naik, and K. Patel, “Study of Hadoop”, International Journal of Computer Science Trends and Technology (IJCTST), vol. 2, no. 6, pp. 40-43, 2014. <http://www.ijcstjournal.org/volume-2/issue-6/IJCTST-V2I6P9.pdf>.
- [10] V. Pellakuri, and R. Rao, “Hadoop Mapreduce Framework in Big Data Analytics”, International Journal of Computer Trends and Technology (IJCTT), vol. 8, no. 3, pp. 115-119, 2014. doi: 10.14445/22312803/IJCTT-V8P121.
- [11] TutorialsPoint, Hadoop Tutorial, available at <https://www.tutorialspoint.com/hadoop/>.
- [12] P. Dave, Big Data – Buzz Words: What is MapReduce, 2013, available at <https://blog.sqlauthority.com/2013/10/09/big-data-buzz-words-what-is-mapreduce-day-7-of-21/>.
- [13] F. Sarrocco, V. Morabito, and G. Meyer, Exploring the Next Generation Financial Services: The Big Data Revolution, 2016, available at https://www.accenture.com/t20170314T051509__w__/_nl-en/_acnmedia/PDF-20/Accenture-Next-Generation-Financial.pdf.
- [14] S. Erickson, and H. Rothberg, “Intangible dynamics in financial services”, Journal of Service Theory and Practice, vol. 26, no. 5, pp. 642-656, 2016. doi: 10.1108/JSTP-04-2015-0093.
- [15] X. Tian, R. Han, L. Wang, G. Lu, and J. Zhan, “Latency critical big data computing in finance”, The Journal of Finance and Data Science, vol. 1, no. 1, pp. 33-41, 2015. doi: 10.1016/j.jfds.2015.07.002.
- [16] R. Shockley, and K. Mercier, Analytics: The Real-World Use of Big Data in Retail, 2017, available at <https://www-935.ibm.com/services/us/gbs/thoughtleadership/big-data-retail/>.
- [17] J. Aloysius, H. Hoehle, and V. Venkatesh, “Exploiting big data for customer and retailer benefits”, International Journal of Operations & Production Management, vol. 36, no. 4, pp. 467-486, 2016. doi: 10.1108/IJOPM-03-2015-0147.
- [18] D. Grewal, A. Roggeveen, and J. Nordfalt, “The Future of Retailing”, Journal of Retailing, vol. 93, no. 1, pp. 1-6, 2017. doi: 10.1016/j.jretai.2016.12.008.
- [19] J. Wu, H. Li, S. Cheng, and Z. Lin, “The Promising Future of Healthcare Services: When Big Data Analytics Meets Wearable Technology”, Information & Management, vol. 53, no. 8, pp. 1020-1033, 2016. doi: 10.1016/j.im.2016.07.003.
- [20] D. Dimitrov, “Medical Internet of Things and Big Data in Healthcare”, Healthcare Informatics Research, vol. 22, no.3, pp. 156-163, 2016. doi: 10.4258/hir.2016.22.3.156.
- [21] C. Kruse, R. Goswamy, Y. Raval, and S. Marawi, “Challenges and Opportunities of Big Data in Health Care: A Systematic Review”, JMIR Medical Informatics, vol. 4, no. 4, 2016. doi: 10.2196/medinform.5359.
- [22] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, “Big Data security and privacy in healthcare: a review”, Procedia Computer Science, vol. 113, pp. 73-80, 2017. doi: 10.1016/j.procs.2017.08.292.
- [23] K. Abouelmehdi, A. Beni-Hssane, and H. Khaloufi, “Big healthcare data: preserving security and privacy”, Journal of Big Data, vol. 5, no. 1, pp. 1-18, 2018. doi: 10.1186/s40537-017-0110-7.
- [24] D. Kochhar, Big Data in Public Transportation, 2016, available at <https://br.hortonworks.com/blog/big-data-public-transportation/>.
- [25] R. Kanniyappan, and B. McQueen, What’s the Big Deal about Big Data in Transportation?, 2014, available at <http://www.fdot.gov/planning/statistics/symposium/2014/BigDataTransport.pdf>.
- [26] S. Wolfert, L. Ge, C. Verdouw, and M. Bogaardt, “Big Data in Smart Farming”, Agricultural Systems, vol. 153, pp. 69-80, 2017. doi: 10.1016/j.agry.2017.01.023.
- [27] K. Bronson, and I. Knezevic, “Big Data in Food and Agriculture”, Big Data & Society, vol. 3, no. 1, pp. 1-5, 2016. doi: 10.1177/2053951716648174.
- [28] K. Porter, Big Data in Energy: Big Opportunities and Big Risks, 2017, available at <http://watt-logic.com/2017/03/29/big-data-in-energy/>.
- [29] H. Dakl, A. El Hannani, A. Aqqal, A. Haidine, and A. Dahbi, “Big Data management in smart grid: concepts, requirements and implementation”, Journal of Big Data, vol. 4, no. 13, pp. 1-19, 2017. doi: 10.1186/s40537-017-0070-y.

- [30] N. Koseleva, and G. Ropaite, "Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges", *Procedia Engineering*, vol. 172, pp. 544-549, 2017. doi: 10.1016/j.proeng.2017.02.064.
- [31] V. Agrawal, *The Impact of Big Data and Analytics on Manufacturing*, 2016, available at <https://tech.co/big-data-analytics-manufacturing-2016-12>.
- [32] R. Delgado, *Big Data's Transformation of the Manufacturing Industry*, 2017, available at <http://data-informed.com/big-datas-transformation-of-the-manufacturing-industry/>.
- [33] K. Nagorny, Lima-Monteiro, P., J. Barata, A. Colombo, "Big Data Analysis in Smart Manufacturing: A Review", *International Journal of Communications, Network and System Sciences*, vol. 10, no. 3, pp. 31-58, 2017. doi: 10.4236/ijcns.2017.103003.
- [34] F. Tao, Q. Qi, A., Liu, and A. Kusiak, "Data-driven smart manufacturing", *Journal of Manufacturing Systems*, 2018. doi: 10.1016/j.jmsy.2018.01.006.
- [35] K. Witkowski, "Internet of Things, Big Data, Industry 4.0 – Innovative Solutions in Logistics and Supply Chains Management", *Procedia Engineering*, vol. 182, pp. 763-769, 2017. doi: 10.1016/j.proeng.2017.03.197.
- [36] H. Lippell, "Big Data in the Media and Entertainment Sectors", in J. Cavanillas, E. Curry, and W. Wahlster (eds.) *New Horizons for a Data-Driven Economy*, Springer, 2016. ISBN: 978-3-319-21569-3.
- [37] S. Philips, *5 Ways Big Data Plays as a Major Role in the Media and Entertainment Industry*, 2017, available at <http://analyticsindiamag.com/5-ways-big-data-plays-major-role-media-entertainment-industry/>.
- [38] R. Munné, "Big Data in the Public Sector", in J. Cavanillas, E. Curry, and W. Wahlster (eds.) *New Horizons for a Data-Driven Economy*, Springer, 2016. doi: 10.1007/978-3-319-21569-3.
- [39] S. Lauchlan, *Government's Big Data Dilemma – Building Public Trust During a Data Science Skills Crisis*, 2017, available at <http://diginomica.com/2017/02/22/governments-big-data-dilemma-building-public-trust-data-science-skills-crisis/>.
- [40] J. Moreno, M. Serrano, and E. Fernández-Medina, "Main Issues in Big Data Security", *Future Internet*, vol. 8, no. 44, pp. 1-16, 2016. doi: 10.3390/fi8030044.
- [41] D. Hu, D. Chen, Y. Zhang, and S. Pei, "Research on Hadoop Identity Authentication Based on Improved Kerberos Protocol", *International Journal of Security and its Applications*, vol. 9, no. 11, pp. 429-438, 2015. doi: 10.14257/ijjsia.2015.9.11.39.
- [42] A. Pentland, *With Big Data Comes Big Responsibility*, 2014, available at <https://hbr.org/2014/11/with-big-data-comes-big-responsibility>.
- [43] A. Lambshead, "The Importance of Standards in a Time of Innovation", *SMPTE Motion Imaging Journal*, vol. 123, no. 8, pp. 7-7, 2014. doi: 10.5594/j18480.
- [44] C. Tsai, "Big data analytics: a survey", *Journal of Big Data*, vol. 2, no. 1, pp. 1-32, 2015. doi: 10.1186/s40537-015-0030-3.
- [45] C. Maple, "Security and privacy in the internet of things", *Journal of Cyber Policy*, vol. 2, no. 2, pp. 155-184, 2017. doi: 10.1080/23738871.2017.1366536.
- [46] V. Brock, and H. Khan, "Big data analytics: does organizational factor matters impact technology acceptance?", *Journal of Big Data*, vol. 4, no. 21, pp. 1-28, 2017. doi: 10.1186/s40537-017-0081-8.
- [47] M. Bendre, and V. Thool, "Analytics, challenges and applications in big data environment: a survey", *Journal of Management Analytics*, vol. 3, no. 3, pp. 206-239, 2016. doi: 10.1080/23270012.2016.1186578.
- [48] B. Balachandran, and S. Prasad, "Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence", *Procedia Computer Science*, vol. 112, pp. 1112-1122, 2017. doi: 10.1016/j.procs.2017.08.138.
- [49] S. Singh, I. Chana, and M. Singh, "The Journey of QoS-Aware Autonomic Cloud Computing", *IT Professional*, vol. 19, no. 2, pp. 42-49, 2017. doi: 10.1109/MITP.2017.26.
- [50] P. Jain, M. Gyanchandani, and N. Khare, "Big data privacy: a technological perspective and review", *Journal of Big Data*, vol. 3, no. 25, pp. 1-25, 2016. doi: 10.1186/s40537-016-0059-y.
- [51] A. Auld, *The big problem in big data: a lack of skills*, 2017, available at <https://www.computing.co.uk/ctg/opinion/3013853/the-big-problem-in-big-data-a-lack-of-skills-skills>.