

BIG DATA: THE UBIQUITOUS DATA THAT ORGANIZATIONS NEED TO LEVERAGE NOWADAYS

Alexandra G. *PRICOP*¹
Gheorghe *ORZAN*²

Abstract: Information is modeling each aspect of everyday life: the personal environment, the cultural, religious, social, political and economic environments. The innovations in technology provide a wide distribution of ideas, thoughts, decisions, behaviors, facts and events, which enlarge the traditional data field, overpassing the processing capacity of standard analysis devices and the use of conventional analysis models and defining the 'Big Data' concept. This paper provides an understanding of Big Data from an economic standpoint and proposes five areas to focus on when leveraging Big Data: transparency, test environment, real-time decision-making, management augmentation and prediction of the future trends.

Keywords: information, big data, analytics, organizational asset, competitive advantage

JEL: M21, M31

Introduction

During the last decades, the technological and informational revolution has driven the fast growth of the business environment. In this new society, a high percent of the population owns a smartphone and connects to the Internet, gaining access to all the online information available. The advantages are not only seen by consumers, but by the organizations and governments as well. By conducting analyzes, the vast amount of available information is being compressed into complex sets of data.

In the new context, data is more than a financial and accounting tool. Its value is reflected through the capacity of driving real-time decisions, which would bring impactful changes to the way business is being done. The fast development of mobile networks, cloud computing and new technologies have driven the expansion of large data pools, described through the notion of "Big Data" (The Aspen Institute, 2010, p.1).

But this new technology raises new questions and challenges. Organizations need to understand what defines Big Data, what potential impact could it have on their business and which are the methods of leveraging Big Data to gain business value. A major challenge is integrating Big Data into the organizational culture, flows and processes, as well as finding strategic insights as the result of analyzing Big Data.

Big Data definition

We are generating vast and varied flows of information, that are subject to the different analysis techniques, which form the body of the Big Data technology (Dumbill, 2013). Facebook, in just over two short years, has quintupled in size to a network that touches more than 500 million

¹ PhD Candidate, The Bucharest University of Economic Studies, Doctoral School of Marketing, 6 Romana Square, Bucharest, Romania, alexandraa.pricop@gmail.com.

² Professor PhD, The Bucharest University of Economic Studies, Doctoral School of Marketing, 6 Romana Square, Bucharest, Romania, orzang@ase.ro.

users. More than 4 billion people around the world now use cell phones and for 450 million of those people the Web is a fully mobile experience (Bughin et al., 2010).

The term “Big Data” is vague and variable. In the business environment there are multiple perspectives of looking at Big Data, in correlation with the area of activity and the informational flows. Many scholars have tried to define “Big Data”: Provost and Fawcett (2013) think of Big Data as “data-sets that are too large for traditional data-processing systems and that therefore require new technologies”; from the point of view of Cooper and Mell (2012), “Big Data is where the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing”.

As Boyd and Crawford (2012) observes, the term “Big Data” is poorly defined, being used in the recent past to refer to data sets large enough to require new computational machines and technology, but what once required such technologies can now be analyzed on an ordinary desktop, with standard software. The main problem is not its size. “There is little doubt that the quantities of data now available are often quite large, but that is not the defining characteristic of this new data ecosystem. In fact, some of the data encompassed by Big Data are not nearly as large as earlier data sets that were not considered Big Data (e.g. census data)” (Boyd, Crawford, 2012).

Big Data is in strong connection with the available processing and analysis technologies. For that reason, “what is defined as big data today may not be the same tomorrow. For instance, with the on-going advancements in the storage technologies, the data that is deemed as big data today might be captured” (Singh et al., 2016, p.5).

Big Data is being defined in terms of 3 V’s: “high-volume, high-velocity and high-veracity information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making” (Gandomi, Haider, 2015), while IBM researchers define Big Data in terms of four dimensions: volume, velocity, variety and veracity (Power, 2014).

The volume characteristic is the first one to think of when talking about Big Data and it refers to the enormous quantity of data in analogical and digital form, that has hundreds or even thousands terabytes in size when saved on a stocking device.

Big Data’s high velocity deals with the pace at which data is being generated and transmitted from the sources to the receivers. Data flow is enormous and continuous.

Variety characterizes Big Data from structural point of view, as well as of the type of the source that generates it. Data is being stored from sources like databases, emails, photos, videos, voice records, internet transactions, social media, etc.

The veracity attribute refers to the accuracy of the data: the biases, noise and irregularities in the structure of the data. One of the main challenges that are being encountered is cleaning the data while deciding which data is meaningful and which not.

Given that the phenomenon is widely spread across the economical and governmental sectors, there is a constant preoccupation in trying to define Big Data through the addition of several characteristics: validity, value, variability, venue, vocabulary, vagueness. As Moorthy (2015) explains, “such characterization with V’s is both fortunate and unfortunate. The challenges in Big Data deployment are:

- Volume: Quantum of data generated, stored and used is explosive now. Several organizations provide different estimates and forecasts.
- Variety: For a marketing manager, data can now be generated through multiple channels. Apart from the conventional data sources such as market research, readership surveys, television rating points, household panels, online click streams, etc., there are now the social media, such as Facebook and Twitter, call centers, chats, voice data, video data, smartphones, etc.

- **Velocity:** Real-time data is accessible in many cases, such as mobile telephony, barcode scan downs, click streams, online transactions and blogs. The data generated from all such sources can be accumulated with the speed at which they are generated.
- **Veracity:** Authenticity of the data increases with automation of data capture. With multiple sources of data, it would be possible to triangulate the results for authenticity.
- **Validity:** The terms veracity and validity are often confusing. Perhaps the term validity should be understood as in the market research methodology that the data should be representing the concept that is expected to represent.
- **Value:** Return on Investment and business value are being emphasized more than value for multiple stakeholders.
- **Variability:** Variance in the data is often treated as the information content in the data. With a large temporal and spatial data, there can be considerable difference in the data at different sub-set levels.
- **Venue:** Multiple data platforms, data bases, data warehouses, format heterogeneity, data generated for different purposes and public and private data sources.
- **Vocabulary:** New concepts, definitions, theories and technical terms are now emerging; they were not necessarily required in the earlier context.
- **Vagueness:** It relates to the confusion about the meaning and overall developments around Big Data. Though it is not necessarily characteristic of the Big Data deployment, it reflects the current context. This may change and more clarity is likely to emerge in the future”.

These challenges could be solved by investigating the source and by finding new techniques and methods of collection, storage and analysis (Fig. no. 1)

The volume challenge could be overpassed through high capacity software technologies, which could allow a high storage rate. In term of velocity, it should be paid attention to the accuracy of the data sources, as well as to the improved storage technology.

The variety challenge is in linkage with the sources of the data, as well as with the structure of the data and the methods through which the data is being merged together.

On the other hand, the veracity rates could be improved by gathering data from complex and differentiated sources.

Validity refers to the data quality, governance and data management, driven by the accuracy of data sources and the extent to which data represents the concept that is being under investigation, but also by the structure of the data.

The value characteristic of Big Data is perhaps the most important one from the business standpoint, reflecting the potential to bring an impactful change in the organizational flows, being objectified in a higher return of investment. In order to solve the challenges imposed, the focus should be set on the data structure and the analysis technologies and techniques used to get the most impactful insights out of Big Data.

The variability of Big Data is being generated in relation with the temporal context. Big Data congregates dynamic, spatiotemporal, monthly, bimonthly, seasonal, yearly data and other non-static data behaviors that characterize the source items, being linked with the data source and the methods in which data is being structured.

Venue refers to the process of collecting and structuring the multiple sourced-data, while the data models, semantics, ontologies, taxonomies and other content- and context-based metadata that describe the data's structure, syntax, content, and provenance (Borne, 2014), describe the vocabulary characteristic.

The Vagueness characteristic refers to confusion over the meaning of data, being correlated with the syntax and driven by analysis technologies and techniques.

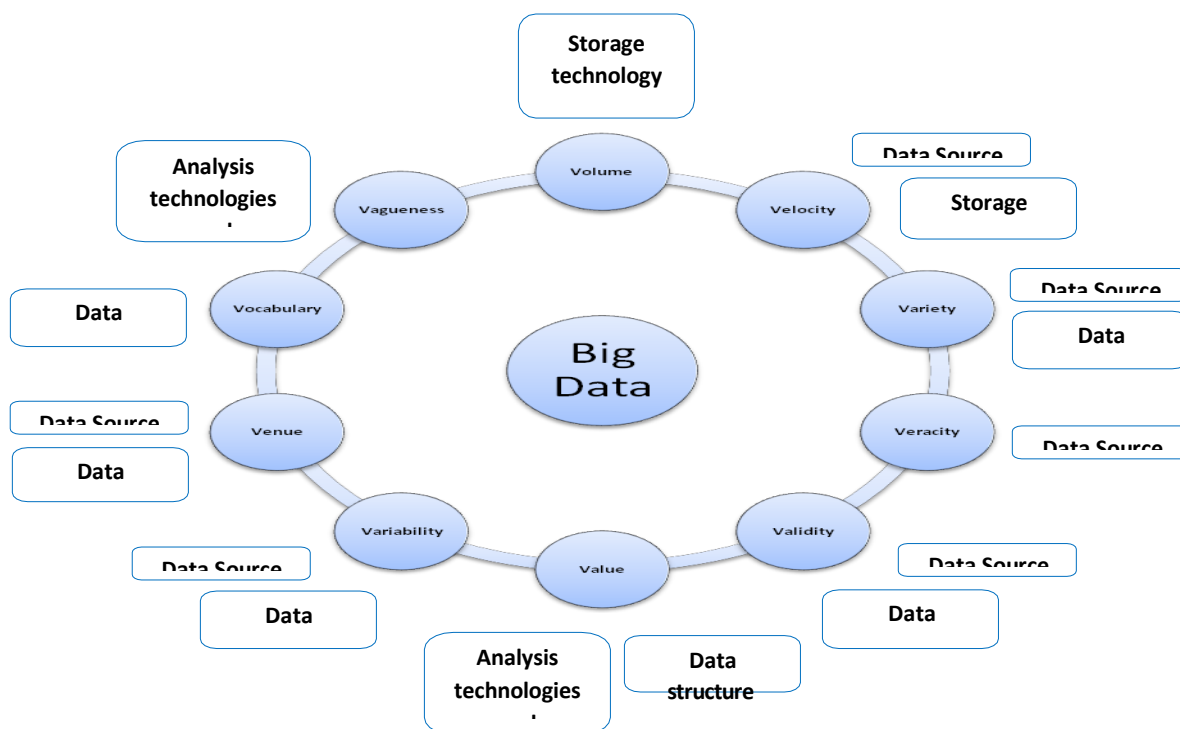


Figure 1. **Big Data characteristics and their link to data components and data manipulation techniques**

Leveraging Big Data

The fast integration of the new technological wave in everyday life is giving birth to new business models, replacing the traditional ones.

“What if we could analyze every transaction, capture insight from every customer interaction and didn’t have to wait for months to get data from the field? What if...? Data are flooding in at rates never seen before – doubling every 18 months – as a result of greater access to customer data from public, proprietary and purchased sources, as well as new information gathered from Web communities and newly deployed smart assets” (Bughin et al., 2010).

The degree at which information is expanding across the economic sectors is in a high development, becoming a real challenge for the companies that have built their competitive advantage by analyzing internal data. One of the biggest challenges is that mountains of data, collected over time are being stocked in achieve systems that are being hard to manage on request (Brown et al., 2011).

The technological innovations from the last decade, as well as their high adoption rate have significantly changed the business flows. Therefore, the top management faces the need to find new

strategic methods of integrating these advances into the organizational culture, processes and structure (Bughin et al., 2010).

The Big Data technology can help companies make better business decisions by providing real-time insights. Therefore, it has the potential to bring a radical change in the research activities, the innovation processes and overall, in marketing. But, as Bughin et al. (2010) notices, in order to leverage the potential of Big Data, companies need to develop new capabilities, as well as organizational and cultural changes.

“Data, big, medium or small, has no value in and of itself. The value of data is unlocked through context and presentation” (Power, 2013). For this to happen, it is necessary to focus on developing specialists in the field of Big Data analytics.

Power (2013) identifies “three major types of analyses that could be conducted with these new data sources and data manipulation technologies:

- Retrospective data analyses – using historical data and quantitative tools to understand patterns and results to make inferences about the future. This is the area of business intelligence.
- Predictive data analyses – using simulation models to generate scenarios based on historical data to understand the future. Predictive means ‘looking forward’ and making known in advance.
- Prescriptive data analyses – using planned, quantitative analyses of real-time data that may trigger events”.

From McGuire et al. (2012) point of view, there are 5 ways of leveraging Big Data to gain the competitive advantage:

1. “Big Data can unlock significant value by making information transparent. There is still a significant amount of information that is not yet captured in digital form, e.g., data that are on paper, or not made easily accessible and searchable through networks. We found that up to 25 percent of the effort in some knowledge worker workgroups consists of searching for data and then transferring them to another (sometimes virtual) location. This effort represents a significant source of inefficiency.
2. As organizations create and store more transactional data in digital form, they can collect more accurate and detailed performance information on everything from product inventories to sick days and therefore expose variability and boost performance. In fact, some leading companies are using their ability to collect and analyze big data to conduct controlled experiments to make better management decisions.
3. Big Data allows ever-narrower segmentation of customers and therefore much more precisely tailored products or services.
4. Sophisticated analytics can substantially improve decision-making, minimize risks, and unearth valuable insights that would otherwise remain hidden.
5. Big Data can be used to develop the next generation of products and services. For instance, manufacturers are using data obtained from sensors embedded in products to create innovative after-sales service offerings such as proactive maintenance to avoid failures in new products”.

Conclusions

With the passage of time, a society driven by data has been taking shape (Power, 2013). In this context, the business implications are not being well understood. The new data sources could be used as management and marketing tools, helping the companies understand the trends, consumer’s needs and opinions. Overall, it is being expected that in the near future the impact of high-volume, high-variety and high-velocity data based decisions will grow in the near future (Power, 2014). Big

Data is becoming more and more an important organizational asset that has the potential of improving the business market positioning.

In this article, we first tried to define Big Data from the business perspective, after which we presented some of the ways to leverage Big Data to gain business value.

References

1. Borne, K., 2014. Top 10 Big Data Challenges - A Serious Look at 10 Big Data V's, [online] available at <https://www.mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs> [Accessed at 22 September 2016].
2. Boyd, D., Crawford, K., 2012. Critical Questions for Big Data, *Information, Communication & Society* 15.5.
3. Brown, B., Chui, M., Manyika, J., 2011. Are You Ready For The Era Of 'Big Data'?, McKinsey & Company, [online] available at <http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data> [Accessed at 22 September 2016].
4. Bughin, J., Chui, M., Manyika, J., 2010. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, McKinsey & Company, [online] available at <http://www.mckinsey.com/industries/high-tech/our-insights/clouds-big-data-and-smart-assets-ten-tech-enabled-business-trends-to-watch> [Accessed at 7 June 2016].
5. Cooper, M., Mell, P., 2012. Tackling big data, Federal Computer Security Managers' Forum, [online] available at <http://csrc.nist.gov/groups/SMA/forum/documents/> [Accessed at 7 June 2016].
6. Dumbill, E., 2013. Making sense of Big Data, Big Data (I), Mary Ann Liebert Inc.;
7. Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35(2).
8. McGuire, T., Manyika, J., Chui, M., 2012. Why Big Data is the new competitive advantage, [online] available at <http://iveybusinessjournal.com/publication/why-big-data-is-the-new-competitive-advantage/> [Accessed at 7 June 2016].
9. Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K. and Ghosh, P., 2015. Big Data: Prospects and Challenges, *Vikalpa, The Journal of Decision Makers*, 40(1).
10. Power, D., 2013. *Decision Support, Analytics, and Business Intelligence*, New York, NY: Business, Expert Press.
11. Power, D., 2014. Using 'Big Data' for analytics and decision support, *Journal of Decision Systems*, 23(2).
12. Provost, F., Fawcett, T., 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making, *Big Data*, 1(1).
13. Singh, M. K., Dileep, K. G., 2016. *Effective Big Data Management And Opportunities For Implementation*, United States of America, Information Science Reference.
14. The Aspen Institute, 2010. *The promise and peril of Big Data*, United States of America, The Aspen Institute.