Implementation of the Big Data concept in organizations – Possibilities, impediments and challenges

Conference Paper · September 2013

CITATIONS
48

READS
9,066

1 author:

Janusz Wielki
Opole University of Technology
87 PUBLICATIONS **284** CITATIONS

SEE PROFILE

Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 985–989

Implementation of the Big Data concept in organizations – possibilities, impediments and challenges

Janusz Wielki
Opole University of Technology
ul. Waryńskiego 4, 45-047 Opole,
Poland
Email: Janusz@Wielki.pl

***Abstract*—This paper is devoted to the analysis of the Big Data phenomenon. It is composed of seven parts. In the first, the growing role of data and information and their rapid increase in the new socio-economical reality, are discussed. Next, the notion of Big Data is defined and the main sources of growth of data are characterized. In the following part of the paper the most significant possibilities linked with Big Data are presented and discussed. The next part is devoted to the characterization of tools, techniques and the most useful data in the context of Big Data initiatives. In the following part of the paper the success factors of Big Data initiatives are analyzed, followed by an analysis of the most important problems and challenges connected with Big Data. In the final part of the paper, the most significant conclusions and suggestions are offered.**

I. Introduction

INCREASING amounts of data are streaming into contemporary organizations as a result of the rapidly growing quantity of data being generated not only by the organizations themselves but also in the organizations' business environments by both their stakeholders and other entities operating there. Thus, it is in this context that such expressions as "a data-centric world" have become more and more common [1].

The above mentioned processes are significant elements of the socio-economical changes taking place worldwide, where the extremely dynamic development of increasingly powerful and pervasive information technology has an important role to play. Advancements in this field have been a significant catalyst for the transformation of the contemporary economy and the emergence of an "interconnected economy". This new type of economy, in the resource dimension, is a knowledge-based economy, where the most meaningful form of capital is intellectual capital [2]. Under these conditions, from an organization's point of view, the ability to collect the right data and information and to transform it effectively into useful knowledge becomes an increasingly important issue.

Recently, the field of information technology has begun to enter into a new era, as a result of the intensification of the progress being made there. It is an era where processing power and data storage have become virtually free, while networks and cloud-based solutions provide users with global access and pervasive services. As a result of these processes, Big Data sets are being generated which have grown exponentially in size [3]. In 2012, about 2.5 exabytes of data were created every day, with this amount doubling about every forty months [4]. Generally, 90% of the global data in existence has been created over the last two years [5].

As a result, the amount of data and information available for organizations for analysis is exploding [4]. This provides organizations with completely new operating possibilities, while simultaneously generating numerous new challenges. In this context the term "Big Data" has emerged and is being used more and more commonly in the business world.

II. Big Data and the Most Important Sources of the Growth of Data

The term Big Data is not universally understood and applied, leading to various approaches to analyzing it. According to the Leadership Council for Information Advantage, this term is not precise "(…) it's a characterization of the never-ending accumulation of all kinds of data, most of it unstructured. It describes data sets that are growing exponentially and that are too large, too raw or too unstructured for analysis using relational database techniques" [6]. On the other hand, NewVantage Partners describes Big Data as "a term used to describe data sets so large, so complex or that require such rapid processing (sometimes called the Volume/Variety/Velocity problem), that they become difficult or impossible to work with using standard database management or analytical tools" [8]. It is important to underline that Big Data not only relates to the storage and consumption of original content but also to the data connected with this consumption [9].

Generally, there have been some significant trends that have caused a considerable increase in data generation [7]. The first trend, the growth in traditional transactional databases is chiefly connected with the fact that organizations are collecting data with greater granularity and frequency. This is due to various reasons such as the increasing level of competition, increasing turbulence in the business environment and the growing expectations of customers. All of these factors require organizations to react rapidly and with maximum flexibility to the changes taking place and then adjust to them. In order to be able to do this, they are forced to conduct more and more detailed analysis concerning marketplaces, competition and the behavior of consumers [7].

The second trend, the increase of multimedia content, is connected with the rapid increase in the use of multimedia in the various industries of the contemporary economy, such as the health care sector where over 95 % of the clinical data

978-83-60810-53-8/$25.00

generated is in video format. Generally, multimedia data already accounts for over half of Internet backbone traffic and it is predicted that this share will grow to 70% by the end of 2013 [7].

The next trend which has caused a growth in the amount of data being generated is the development of the phenomenon called "The Internet of Things", where the number of physical objects or devices that communicate with each other without any human interference is increasing at a fast pace. They link with each other in a wired or wireless manner, often using IP protocols. As they are equipped with various sensors or actuators they collect and send huge amounts of data [10]. By 2015 the amount of data generated from the 'Internet of Things' will grow exponentially as the number of connected nodes deployed in the world is expected to grow at a rate of over 30% per year [7].

Social media is the next extremely significant source of the increase of data. Facebook users alone generate huge amounts of data. In 2011 the 600 million active users of this social platform spent over 9.3 billion hours a month on the site, with the average Facebook user generating 90 pieces of content (photos, notes, blog posts, links, or news stories) [7]. A year later the number of Facebook users reached 1 billion. Research conducted at the beginning of 2012 showed that if only messages are considered, users receive an average of nearly twelve messages a month, and send nine [11]. In the case of YouTube, every minute 24 hours of video is uploaded, while over the same timeframe Twitter users send 98000 tweets [7], [12]. In addition, smart phones are playing an increasingly important role in social networks. Although the penetration of social networks is increasing for both PCs and smartphones, it is significantly higher for smartphones. If frequent users are considered in the case of PC's it is 11% p.a. while in the case of smartphones it is 28% p.a [7]. This has caused a rapid increase in mobile data traffic which doubled between the third quarter of 2011 and the third quarter of 2012. It is predicted that mobile data traffic will grow twelve fold by 2018 [13].

III. Opportunities and Benefits Connected With Big Data Utilization

The development of the Big Data phenomenon and its associated tools and techniques, is not something which has been separated from the wider processes which have been taking place in organizations over recent years. In fact, it is becoming more and more common in organizations concerned with the field of analytics and has significantly expanded the possibilities available within the scope of business intelligence (BI) tools.

Given their role in providing organizations with numerous possibilities and opportunities in the sphere of analytics, business intelligence systems are well suited for aggregating and analyzing structured data [14]. But there are, however, some types of analyses that BI can not handle. These mainly relate to situations where data sets become increasingly diverse, more granular, real-time and iterative.

Such types of unstructured, high volume, fast-changing data, pose problems when trying to apply traditional approaches based on relational database models. As a result, it has become apparent that there is growing demand for a new class of technologies and analytical methods [6].

There are many diverse benefits arising from the utilization of Big Data, depending on the sector of the economy, as has been confirmed by the results of research conducted by the McKinsey Global Institute. These results show the transformational potential of Big Data in such diverse domains as health care, public sector administration, retail, manufacturing and personal location data [7]. According to the results of a survey conducted in summer 2012 by NewVantage Partners, among C-level executives and function heads from many of America's leading companies, there are seven basic groups of benefits connected with Big Data initiatives. Better, fact-based decision making (22%) and an improved customer experience (22%) are the most important of these benefits, coupled with the overall message that the expectation is to make better decisions faster. The other groups of benefits include: increased sales (15%), new product innovations (11%), reduced risk (11%), more efficient operations (10%) and higher quality products and services (10%) [15].

Organizations use Big Data platforms to give them answers to important questions in seconds rather than months. Thus, the key value of Big Data is to accelerate the time-to-answer period, allowing an increase in the pace of decision-making at both the operational and tactical levels [14], [15]. An extremely important new element, in the context of decision-making, connected with the Big Data phenomenon, is the possibility for constant business experimentation to guide decisions and test new products, business models, and customer-oriented innovations. Such an approach even allows, in some cases, for decision making in real-time. There are many examples of companies using this in practice. For example multifunctional teams in Capital One perform over 65,000 tests each year. They experiment with combinations of market segments and new products. The online grocer FreshDirect adjusts, on a daily basis or even more frequently, prices and promotions based on online data feeds. Tesco is another example. This company gathers transaction data on its millions of customers through a loyalty card program and uses it to analyze new business opportunities. For example, it looks at how to create the most effective promotions for specific customer segments and how to inform them about decisions concerning pricing, promotions, and shelf allocation [17]. Walmart is another example. This company created the Big Data platform (The Online Marketing Platform) which is used, among other things, to run many parallel experiments to test new data models [16]. Also such dot-com giants as Amazon, eBay and Google have been using testing in order to drive their performance [17].

According to the McKinsey Global Institute, five key ways in which Big Data creates value for organizations can be distinguished [7]:

- creating transparency by integrating data and making it more easily accessible to all relevant stakeholders,
- enabling experimentation to discover needs, expose variability, and improve performance,

- segmenting populations in order to customize actions,
- replacing or supporting human decision making with automated algorithms,
- innovating new business models, products and services.

Generally, the results of the research conducted in February 2012 among 607 executives from around the world by the Economist Intelligence Unit confirm the value of Big Data utilization by companies. The surveyed executives claim that Big Data initiatives have improved the performance of their organizations over the past three years by around 26%. Simultaneously they expect that such initiatives will improve performance by an average of 41% over the next three years [18]. In addition, it is worth noticing that according to the results of the research of Brynjolfsson et al., firms where decision making is based on data and business analytics have 5-6% higher output and productivity. Decision making based on data and business analytics also impacts on other performance measures such as asset utilization, equity return and market value [18].

As in the case of BI initiatives Big Data systems have been used for two purposes - human decision support and decision automation. According to the results of the above mentioned research conducted by the Economist Intelligence Unit, Big Data is used, on average, for decision support 58% of the time and for decision automation around 29% of the time, based on the level of risk connected with the decision [14].

IV. Tools, Techniques and the Most Useful Data in the Context of Big Data Initiatives

The effective implementation of Big Data initiatives requires an undertaking of appropriate organizational actions, including ensuring organizations are provided with all the necessary resources to enable analysis of the ever-growing data sets to which they have access. In this context, the application of proper techniques and technologies is one of the key issues. In practice, organizations use many various techniques and technologies to aggregate, manipulate, analyze, and visualize Big Data. They come from various fields such as statistics, computer science, applied mathematics, and economics. Some of them have been developed intentionally and some of them have been adapted for this purpose. Examples of techniques utilized for the analysis of Big Data are: A/B testing, data fusion and data integration, data mining, machine learning, predictive modeling, sentiment analysis, spatial analysis, simulation or time series analysis. Examples of technologies used to aggregate, manipulate, manage, and analyze of Big Data are: Big Table, Cassandra, Google File System, Hadoop, Hbase, MapReduce, stream processing, visualization (tag cloud, clustergram, history flow, spatial information flow) [7].

Increasingly, there are a number of new analytical toolkits for the analysis of Big Data. Examples of such solutions are [19]:

- Alterian, TweetReach (network intelligence tools for real-time analysis of the reactions and responses to changes of industry players),
- NM Incite, Social Mention, SocMetrics, Traackr, Tweepi (sentiment analysis tools for estimating the buzz around a product or service, influencer intelligence tools for identifying key influencers and targeting for marketing or insights),
- Attensity, Autonomy (live testing tools for getting direct feedback from users on new products or ideas, data mining tools for text-analytics to estimate market size).

In addition, a very important element of Big Data initiatives is properly trained people. In this context, a specific type of worker is indicated, known as data scientists, who are properly trained to work with Big Data. In practice, it means that they should be people who know how to discover the answers to an organization's key questions from huge collections of unstructured data. These people should be a hybrid of analyst, data hacker, communicator and trusted advisor [20]. In addition to analytical abilities and substantial and creative IT skills, they should be close to the products and processes inside the organization [21]. As the acquisition of in-depth domain knowledge from data scientists typically takes years [14], most organizations build platforms to close the gap between the people who make decisions and data scientists, such as that created by Walmart - the Social Genome Platform. It facilitates cooperation among buyers, merchandisers, product managers and other people who have worked in retail for years and data scientists [22].

In addition to proper techniques, tools and people, the basic resource required for Big Data initiatives is appropriate data. As was mentioned earlier, a lot of data from various sources is currently flowing into contemporary organizations but not all Big Data sets are equally valuable,. Business activity data such as sales, purchases, costs etc. is definitely the most important source of data. Office documentation is the second key source of data, closely followed by social media. In certain sectors such as healthcare, pharmaceutical, and biotechnology, data sets from social media are more important that those from office documentation. The other important types of data sets include: point-of-sale data, Website clickstream data, RFID/logistics data, geospatial data, telecommunications data, telemetry data [18].

V. Success Factors of Big Data Initiatives

Through an analysis of implemented Big Data initiatives, various success factors can be determined, each with their own set of recommendations.Marchand and Peppard have identified five important guidelines for the success of a Big Data project. They include [24]:

1. Placing people at the heart of the Big Data initiative.
2. Emphasizing information utilization as the way to unlock value from information technology.

3. Equipping IT project teams with cognitive and behavioral scientists.
4. Focusing on learning.
5. Worrying more about solving business problems than about deploying technology.

Based on their experiences gained from cooperation with companies from data rich industries, Barton and Court, on the other hand, came to the conclusion that full exploitation of data and analytics requires three capabilities [25]:

1. Choosing the right data. In this context two aspects are important: creative sourcing of internal and external data and upgrading IT architecture and infrastructure for easy data merging.
2. Emphasizing information utilization as the way to unlock value from information technology. In this context two aspects are important: focusing on the biggest drivers of performance and building models that balance complexity with ease of use.
3. Equipping IT project teams with cognitive and behavioral scientists. In this context two aspects are important: creating simple, understandable tools for people on the front lines and updating business processes and developing capabilities to enable tools utilization.

According to Barth et al., organizations that benefit from Big Data base their activities on three fundamental issues [21]:

1. Paying attention to data flow as opposed to stocks.
2. Relying on data scientists and product and process developers rather than data analysts.
3. Moving analytics away from the IT function, into core business, with operational and production functions.

VI. The most significant obstacles and challenges connected with Big Data eferences

As with other IT-related initiatives, Big Data also has its own set of problems and challenges. The Economic Intelligence Unit research mentioned earlier indicates some of the impediments to the effective utilization of Big Data for decision-making [14]. "Organizational silos" were the most significant barrier (55,7%), which result from the fact that data connected with particular organizational functions (i.e. sales, distribution etc.) are collected in "function silos" rather than pooled for the benefit of the entire company. The second (50,6%), although no less important, issue is the lack of appropriately skilled people (data scientists) prepared to analyze data. The third aspect (43,7%) is the excessively long time it takes organizations to analyze huge data sets. As was mentioned earlier, organizations expect to be able to analyze and act on data in real time. The fourth barrier (41,7%) is the difficulties concerned with the analysis of ever increasing amounts of unstructured data. Finally, the inability of senior management to view Big Data in a sufficiently strategic way (34,9%) is the fifth key impediment [14].

McAfee and Brynjolfsson indicate five management challenges which prevent organizations from reaping the full benefits of Big Data utilization. They are: leadership, talent management, technology, decision making, company culture.

When considering leadership, having more or better data does not guarantee success. The leaders still have to have a vision of the organization's development, set clear goals, understand the market, etc. Big Data changes the way organizations make many of their decisions. Talent management is connected with the necessity of providing the organization with the right people (such as data scientists) who are prepared to work with huge sets of data. The next challenge relates to the problem of assuring the data scientists have the proper tools to handle the Big Data. Although the technology alone is not enough to succeed in Big Data initiatives, it is a necessary part of it. The next challenge is connected with the problem of ensuring there are mechanisms in place to guarantee that the information and the relevant decision-makers are in the same location. It is important to make sure that the people who understand the problems are able to use the right data and to work with people who have the necessary problem solving skills.

The final challenge is connected with changes related to organizational culture. The key issue in this context is to make decisions as data-driven as possible, instead of basing them on hunches and instinct [4]. It is worth mentioning that the significance of such cultural transformation is also mentioned in other research e.g. that concerning sectoral Big Data projects [23].

In addition, the numerous challenges connected with data and legal rights should be noted. They relate to such issues as copyright, database rights, confidentiality, trade marks, contract law, competition law [1]. There is another important challenge also connected with legal aspects. It relates to the transparency in data collection practices. A further important risk is around the utilization of Big Data to increase the automation of decision-making. There is one more important danger which is underlined in the context of Big Data. It is connected with the fact that Big Data might not be providing the whole picture for a particular situation. There are several reasons for this i.e. biases in data collection, exclusions or gaps in data signals or the constant need for context in conclusions [26].

At the same time, existing pre-Big Data challenges and threats are still developing, such as the problem of securing collected data and information. These issues chiefly relate to the problem of how to protect competitively sensitive data and data that should be kept private by organizations (e.g. various types of consumer data) [7]. As a result, the problems connected with the broadly defined security of the IT infrastructure of organizations and protection against various attacks becomes an even more important issue than previously [27]. The increasing dependence, as a result of the Big Data phenomenon, of organizations on the efficient and reliable functioning of their IT infrastructure, means that securing it has become even more important.

VII. Conclusions

The rapidly growing amount of data which organizations have at their disposal and the opportunities connected with its practical utilization are increasingly changing the processes relating to making decisions at various organizational

levels. Thus, Big Data offers huge potential to positively impact on the functioning of organizations generally and gives them a competitive advantage. Companies are now trying to utilize to an even greater degree the opportunities and chances that are emerging.

But if initiatives aimed at the practical usage of Big Data sets are to be successful at giving an organization a competitive advantage and be of value, it is not enough to just collect and own the appropriate data sets. In fact, this is only the starting point of every Big Data initiative. Further essential elements are suitable analytical models, tools, skilled people, and organizational capabilities. Lack of all of these necessary components can lead to a situation whereby instead of expected benefits there is only disappointment and a belief that Big Data initiatives are only the next wave in a long line of management fads.

Generally, although Big Data solutions have a huge potential for both commercial organizations and governments, there is uncertainty concerning the speed with which they can be utilized in a secure and useful way [3].

REFERENCES

[1] Kemp Little LLP, "Big Data – Legal Rights and Obligations", http://www.kemplittle.com/Publications/WhitePapers/Big%20Data%20-%20Legal%20Rights%20and%20Obligations%202013.pdf, January 2013.

[2] D. Tapscot, A. Williams, *Radical Openness*. New York: TED Books (Kindle Edition), 2013.

[3] National Intelligence Council, "Global Trends 2030", http://globaltrends2030.files.wordpress.com/2012/11/global-trends-2030-november2012.pdf, December 2012.

[4] E. Brynjolfsson, A. McAfee, (2012), "Big data: The management revolution", *Harvard Business Review*, pp. 60-68, October 2012.

[5] A. Lampitt, "Hadoop: Analysis at massive scale", in *InfoWorld*, http://resources.idgenterprise.com/original/AST-0084522_IW_Big_Data_ rerun_1_all_sm.pdf, pp. 8-12, Winter 2013.

[6] LCIA, "Big Data: Big Opportunities to Create Business Value", http://poland.emc.com/microsites/cio/articles/big-data-big-opportunities/ LCIA-BigData-Opportunities-Value.pdf, 2011.

[7] McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity", http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf, May 2011.

[8] NewVantage Partners, "Big Data Executive Survey: Themes & Trends", http://newvantage.com/wp-content/uploads/2012/12/NVP-Big-Data-Survey-Themes-Trends.pdf, 2012.

[9] J. Gantz, D. Reinsel, "Extracting Value from Chaos", http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf, June 2011.

[10] M. Chui, M. Löffler, R. Roberts, "The Internet of Things", *McKinsey Quarterly*, https://www.mckinseyquarterly.com/article_print.aspx?L2=4&L3= 116&ar=2538, March 2010.

[11] K. Hampton, L. Goulet, C. Marlow, L. Rainie, "Why most Facebook users get more than they give", *Pew Internet & American Life Project*, http://pewinternet.org/~/media//Files/Reports/2012/PIP_ Facebook%20users_2.3.12.pdf, February 3, 2012.

[12] HP, "Big security for big data", http://www.hpenterprisesecurity.com/collateral/whitepaper/BigSecurityforBigData0213.pdf, December 2012.

[13] Ericsson, "Ericsson Mobility Report", http://hugin.info/1061/R/1659597/537300.pdf, November 2012.

[14] Capgemini, "The Deciding Factor: Big Data & Decision Making", http://www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/?d=6C800B16-E3AB-BC55-00F4-5411F5DC6A8C, February 2012.

[15] NewVantage Partners, "Big Data Executive Survey: Creating a Big Data Environment to Accelerate Business Value", http://newvantage.com/wp-content/uploads/2012/12/NVP-Big-Data-Survey-Accelerate-Business-Value.pdf, 2012.

[16] Walmartlabs, "Big Data Platform and Demand Generation", http://www.walmartlabs.com/platform/, 2013.

[17] J. Bughin, M. Chui, J. Manyika, "Clouds, big data, and smart assets", *McKinsey Quarterly*, https://www.mckinseyquarterly.com/article_print.aspx?L2=20&L3= 75&ar=2647, August 2010.

[18] E. Brynjolfsson L. Hitt, H. Kim, "Strength in Numbers", http://ssrn.com/abstract=1819486, December 12, 2011.

[19] M. Harrysson, E. Metayer, H. Sarrazin, "How 'social intelligence' can guide decisions", *McKinsey Quarterly*, https://www.mckinseyquarterly.com/article_print.aspx?L2=21&L3=37&ar=3031, November 2012.

[20] T. Davenport T., D. Patil, "Data scientist", *Harvard Business Review*, pp. 70-76, October 2012.

[21] P. Barth, R. Bean, T. Davenport, "How Big Data is Different", *Sloan Management Review*, Fall 2012.

[22] R. Ferguson, "It's All About the Platform: What Walmart and Google Have in Common", *Sloan Management Review*, http://sloanreview.mit.edu/article/its-all-about-the-platform-what-walmart-and-google-have-in-common/, December 5,.2012.

[23] P. Groves, B. Kayyali, D. Knott, S. van Kuiken, "The_big_data_revolution_in_healthcare", *McKinsey Quarterly*, http://www.mckinsey.com/ insights/health_systems/~/media/7764A72F70184C8EA88D805092D72D58.ashx, January 2013.

[24] D. Marchand, J. Peppard, "Why IT fumbles analytics", *Harvard Business Review*, pp. 104-113, January-February 2013.

[25] D. Barton, D. Court, "Making advanced analytics work for you", *Harvard Business Review*, pp. 78-83, October 2012.

[26] R. Ferguson, "Competitive Advantage with Data? Maybe … Maybe Not", *Sloan Management Review*, http://sloanreview.mit.edu/article/competitive-advantage-with-data-maybe-maybe-not/?utm_source=facebook&utm_medium=social&utm_campaig, March 26, 2013.

[27] J. Wielki, *Modele wpływu przestrzeni elektronicznej na organizacje gospodarcze*, Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego, 2012.