Big Data Analytics in the Management of Business

Dorota Jelonek[1,*]

Czestochowa University of Technology, Poland

Abstract. Data, information, knowledge have always played a critical role in business. The amount of various data that can be collected and stored is increasing, therefore companies need new solutions for data processing and analysis. The paper presents considerations on the concept of Big Data. The aim of the paper is to demonstrate that Big Data analytics is an effective support in managing the company. It also indicates the areas and activities where the use of Big Data analytics can bring the greatest benefits to companies.

1 Introduction

Over the past decade, the "Big Data" era has quietly descended on many communities, from governments and e-commerce to health and sports organisations [1].

These data are generated from the Web, online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social media, science data, sensors and mobile phones and their applications.

Within the next decade, the amount of information will increase by 50 times while the number of information technology specialists who keep up with all that data will increase by 1,5 times.

Information overload [2] is one of the most serious problems in the Big Data environment. For information users, searching for what they need from the vast amounts of information accurately is becoming more difficult [3]. However, if companies are able to collect, process and analyse large datasets, then gathered information can be extremely valuable. In the era of exponential growth of business information, the acceleration of data accessibility is becoming vital.

Enterprise database systems, search systems, advanced data, text and Web analytics are becoming important for turning data into actionable knowledge and intelligence. As the data volume is large, the analytics can only be possible if we have highly efficient algorithms and software.

Highly skilled executives in terms of gathering and using knowledge, who also have analytical skills, are nowadays among most competitive factors of the company.

In the turbulently changing environment, making a decision is associated with a high risk, which may hinder the access to relevant information and reliable analysis, delivered when you need them. Advanced analytical tools, supported by innovative ways of processing Big Data become necessary to expand enterprises. Research of Brynjolfsson, Hitt and Kim [4] confirmed that the efficiency is higher in organisations which base their decisions on data and analytics systems. This is facilitated by basing decision-making process on data obtained from analytical systems what is referred to as "data-driven decision-making approach (DDD)" [5].

The purpose of this paper is to demonstrate that Big Data analytics is an effective support in the management of business. The following sections present the essence of Big Data, traditional data analysis vs. Big Data analytics methods as well as the role of Big Data analytics in the management of business. The results of empirical research about priorities for the application of Big Data are also presented.

2 The essence of Big Data

The current explosion of data that is being generated is due to three main reasons [6, 7]:

1. Hundreds of applications such as mobile sensors, social media services, and other related devices are collecting information continuously.
2. Storage capacity has improved so much that collecting data is cheaper than ever, making preferable to buy more storage space rather than deciding what to delete.
3. Machine Learning and information retrieval approaches have reached a significant improvement in the last years, thus enabling the acquisition of a higher degree of knowledge from data.

Figure 1 shows the relations between increasing data variety and complexity and the memory size of databases from megabytes to petabytes. It is shown that ERP systems generate data which companies collect and process in databases or data warehouses. Nevertheless CRM and Web systems have the biggest influence on the rapid increase in the amount of data.

Figure 1 also presents the main sources of data for Big Data.
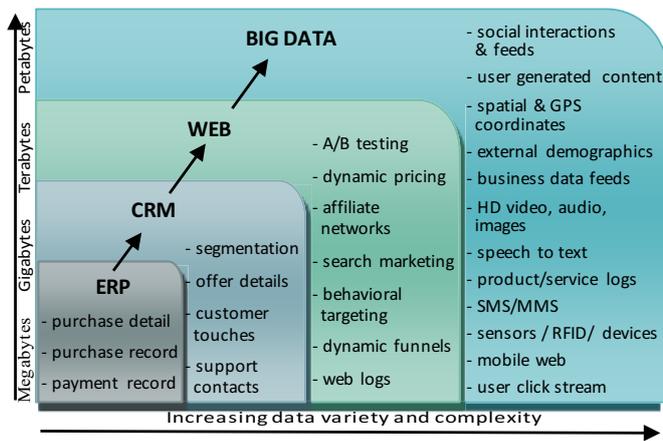
* Corresponding author: jelonek@zim.pcz.pl

Fig. 1. Big Data – variety and complexity [8]

The concept of Big Data is, however, broader than just large data sets. The most important challenge is not to increase data sets but to search for a new approach to the analysis.

Wielki [9] indicates, that Big Data is mainly characterised by:
• type of data - data unstructured,
• size of data - 100 terabytes to petabytes,
• way of information flow - a permanent inflow of data to the organisation (in real time),
• basic analytical method - machine learning,
• primary purpose – a creation of new products.

The phenomenon called "Big Data" differs as a data source from the previously used sources such as databases or data warehouses. Davenport and others [10] emphasise that organisations which capitalise on big data stand apart from traditional data analysis environments in three key ways:
1. They pay attention to data flows as opposed to stocks.
2. They rely on data scientists and product and process developers rather than data analysts.
3. They are moving analytics away from the IT function and into core business, operational and production functions.

Big Data requires a revolutionary step forward from traditional data analysis, characterised by its four main components: volume, velocity, veracity, variety [11 – 13]. These are shown in Figure 2.
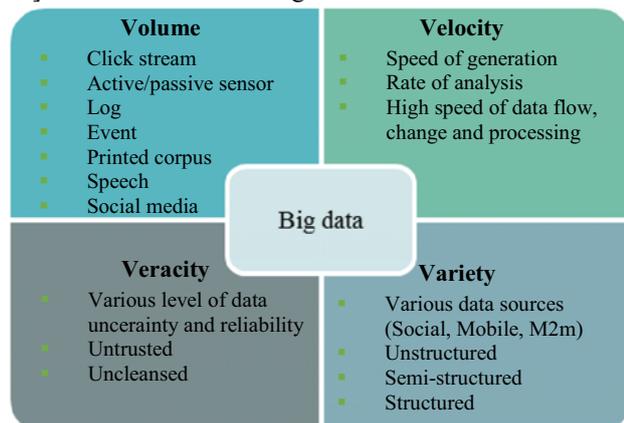

Fig. 2. Four Vs of Big Data.

Each component presented in Figure 2 is shortly expressed below.

Variety makes Big Data really big. Big Data come from a great variety of sources and generally has three types: structured, semi-structured and unstructured. Structured data insert a data warehouse already tagged and easily sorted but unstructured data are random and difficult to analyse. Semi-structured data do not conform to fixed fields but contains tags to separate data elements [14, 15].

Volume or the size of data now is larger than terabytes and petabytes. The grand scale and rise of data outstrip traditional store and analysis techniques [14, 15].

Velocity is required not only for Big Data, but also all processes. For time limited processes, big data should be used as it streams into the organisation in order to maximise its value [14, 15].

Veracity: Big Data are sourced from many different places, as a result you need to test the veracity/quality of the data.

B. Frank states that all of the above-mentioned definitions of the "V" dimensions are secondary to the importance of data value. This parameter was defined by B. Frank as "Uber-V" because it points out the business value that data are for the company. The other dimensions are merely ancillary. B. Frank believes that the complicity, volume and format of data are not important, however the knowledge of how to process them in a way relevant to the enterprise is essential [16].

All these facts are known as the 4V's of Big Data, [17] which lead to the definition given by Steve Todd at Berkeley University: Big Data is when the normal application of current technology does not enable users to obtain timely, cost-effective, and quality answers to data-driven questions.

Other examples of definitions underline the unstructured character of data, i.e. accordingly to Rouse [18] Big Data is a general term used to describe the voluminous amount of unstructured and semi-structured data a company creates - data that would take too much time and cost too much money to load into a relational database for analysis.

Pawełoszek and Wieczorkowski [19] proposed "three-aspect approach", identifying three essential aspects of big data: technological, business and social. The technological aspect in the above classification represents a focus on the methods of big data analysis and information technology used. The business aspect of the described classification focuses on applications of Big Data, especially on they role in decision support. The social aspect of Big Data is associated with social consequences of data processing results. When making decisions on implementing solutions for Big Data analytics all three mentioned aspects should be considered.

3 Traditional data analysis vs. Big Data analytics methods

Big Data are not just data, but also IT infrastructure, analytical systems and employees with high analytical skills.

Big data analytics is the process of examining large data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information.

Big Data analytics may show new relations between data, reveal unseen earlier trends and contribute to the creation of new knowledge, which can then be used to increase the effectiveness and improve the profitability of the company. In the long term, these can compensate for the costs associated with the purchase of specialised software and hiring specialists [20, 21].

There are many differences between the conventional and Big Data analytics. Even though some of them are very fuzzy, Table 1 presents a summary of differences.

Table 1. The differences between Big Data analytics and Traditional analytics [22].

	Big Data analytics	Traditional analytics
Type of data	Unstructured formats	Formatted in rows and columns
Volume of data	100 terabytes to petabytes	Tens of terabytes or less
Flow of data	Constant flow of data	Static pool of data
Analysis methods	Machine learning	Hypothesis-based
Primary purpose	Data-based products	Internal decision support and service

Changes in the analysis of Big Data relate to three main areas [23]:
- the ability to analyse large amounts of data, while not having to use smaller data sets,
- readiness to deal with unstructured data, characterised by low accuracy,
- rising importance of correlations, which tend to look for relations between phenomena rather than their causes.

While analysing Big Data the focus should be placed on the search for correlations and patterns, which indicate that "something is happening", instead of explaining the reasons "why it is happening". This means that previously used method of hypothesis and seeking arguments to verify them is reversed. The discovery of unexpected correlations can only be a stimulus to formulate hypotheses.

Big Data process continuously incoming data from the environment and from inside the company. Thus, the Big Data analytics is based on data collected in real time, and that is why the results of the analysis are accurate and generated without delay.

Analytic applications are universal and have a wide range of functionality, useful in both large companies and small companies from the SME sector. They allow for integration of complex business processes and quick response to any changes on the operational level, as well as in the business environment. Thanks to them,

companies are able to track on a regular basis the status of each process and rapidly response to events through flexible modification of the processes.

The analysis of Big Data involves analytical methods for traditional data and Big Data, analytical architecture for Big Data, and software used for analysis of Big Data. In the analysis of huge volumes of data multiple analyses are being applied such as: association rule learning aimed at discovery of relationships in databases; A/B testing allowing for comparison of control group with a test group; cluster analysis enabling classification of objects divided into smaller groups; crowdsourcing enabling gathering data generated by communities; data fusion and data integration which data analysis coming from different sources; genetic algorithms based on the process of natural evolution and having its application mainly in optimization; machine learning and natural language processing creating the field of artificial intelligence; neural networks based on the functionality of human's nervous system and finding its application in optimization and pattern recognition; analysis of nodes in networks; predictive modelling and analysis of regression based on mathematical models; spatial analysis and simulations; supervised and unsupervised learning and visualization boards including cloud tags, history and spatial information flows and managerial dashboards [24].

This long list of methods for analysing Big Data is probably not complete as new methods of extracting information and knowledge from the sets of Big Data constantly appear. Enterprises from almost every sector develop the concept of information and data as company's strategic asset [25].

4 Big Data in the management of business

Managers increasingly adopt strategies based on acquiring, processing and using high-quality data for the decision-making (data-driven decision-making approach).

Research carried out by EMC Forum 2013 indicate that:
- 39% of entrepreneurs believe that Big Data provide business success,
- 19% of entrepreneurs are of the opinion that with Big Data they have achieved a competitive advantage,
- 36% of entrepreneurs believe that the introduction of Big Data will increase the safety and security of their data.

The researchers from the Economist Intelligence Unit survey [26], sought the answer to the question "Which of the following business processes do you believe are the most important priorities for the application of big data now, and which will be most important in three years?". The results of their study are presented in Figure 3.
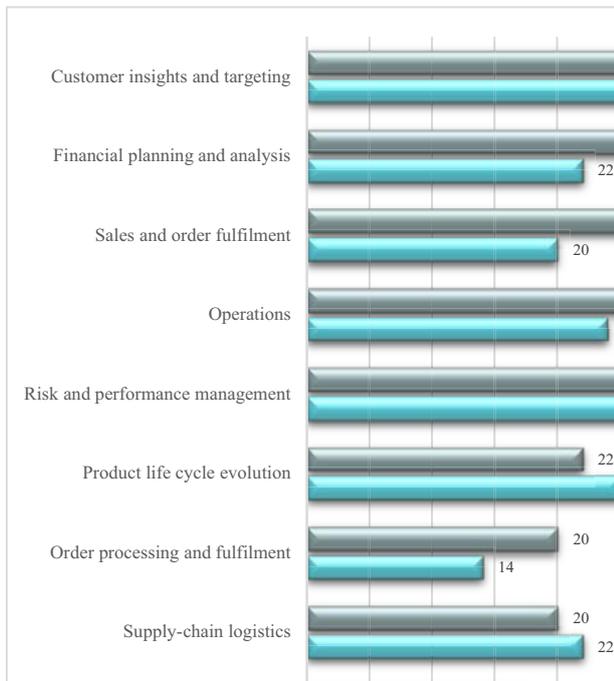
Fig. 2. Priorities for the application of Big Data [26]

Customer processes are currently a priority for the application of big data, according to 42% of the C-level executives. Financial planning with 32% and sales with 29% are the following indications for big data analytics. Other priorities include operations, risk management and performance management as well as the evolution of product life cycle and same others. A long list of priorities suggests broad opportunities for Big Data solutions enterprise-wide. In three years, these priorities will become even harder to segregate. According to respondents a customer insights and targeting will remain the top priority, however it will drop in relative terms as several others emerge.

Big Data are produced in every segment in manufacturing. For enterprise performance, Big Data help to identify the product demands, productivity and performance via diverse business objectives. For production, Big Data make it possible to detect the right facility causing the product flaw [27].

Big Data analysis is used in enterprise management in the following areas and activities [28, 29]
• transformation of key organisational business processes,
• strategic decision support,
• identification of the most cost-effective suppliers in delivering products on-time,
• product development,
• flag machinery and process variances that might be indicators of quality problems,
• analysis of receivables, an anticipation of payments, asset management,
• identification of which marketing promotions and campaigns are most effective in driving customer traffic, engagement, and sales,
• predictions of customer behaviour,
• customer relations management,

• optimization of marketing mixes given marketing goals,
• optimization of sales resource assignments, product mix,
• redefinition of product,
• collaborative filtering,
• supply-demand analysis.

The concept of Big Data is still relatively new, so its implementation in companies often faces barriers. The Big Data + Report indicated that Polish companies, primarily face the following difficulties [30]:
• shortage of qualified specialists - 29%,
• high exploitation costs - 27%,
• unclearly defined purpose and justification for the implementation of Big Data - 22%,
• complicated technology, which does not bring the expected profits - 7%.

Some institutions take steps to eliminate identified barriers. For example, universities educate "data science" and "data analysis" specialists while IT companies try to develop cheaper, more effective and "user-friendly" solutions. Nevertheless, the most difficult barriers to overcome are the mental barriers, lack of knowledge about Big Data and resistance to change.

5 Architecture for Big Data Analysis

Four main components of Big Data: volume, velocity, veracity and variety are the reason why Big Data are difficult to process, and what's more analyse and draw conclusions. For effective use of Big Data companies need new IT architectures, that is a configuration of hardware and software in a way which ensures efficient processing of Big Data. Cloud computing technologies can provide unlimited resources on demand. It could be a solution to the problem of growing volume of data and could allow effective data management [31]. The most common architecture for Big Data is Apache Hadoop.

Many organisations looking to collect, process and analyse Big Data have turned to a newer class of technologies that includes Hadoop and related tools such as YARN, MapReduce, Spark, Hive and Pig as well as NoSQL databases [32].

Hadoop is essentially a distributed data infrastructure: It distributes massive data collections across multiple nodes within a cluster of commodity servers, which means you don't need to buy and maintain expensive custom hardware. It also indexes and keeps track of that data, enabling Big Data processing and analytics far more effectively than was possible previously.

Hadoop quickly emerged as a foundation for Big Data processing tasks, such as scientific analytics, business and sales planning, and processing enormous volumes of sensor data, including from internet of things sensors.

Apache Hadoop is used by large corporations such as Yahoo, Facebook, Amazon, eBay, The New York Times, Chevron and IBM.

6 Conclusions

The practical usefulness of Big Data analytics is evident in many areas of company management, especially in strategic management and stakeholders management. Big Data analytics can lead to more effective marketing, new revenue opportunities, improved operational efficiency, competitive advantages over rival organisations and other business benefits.

As a conclusion, a list of recommendations for companies that wish to implement Big Data solutions is presented below:

1. Verify company's information strategy in terms of Big Data requirements, which include hardware platform, software, application landscape of Big Data analytics and human resources - "data science" and "data analyst" specialists.
2. Adjust the "new" information strategy to the business strategy and the business strategies to new business processes opportunities.
3. Create flexible business models.
4. Build information culture in the organisation. The first question a data-driven organisation asks itself is not "What do we think?" But "What do we known?"
5. Change the decision-making model. Data-driven decisions tend to be better decisions.
6. Collect, process and use data which already exist in the information resources of the company.
7. Identify other data that should be collected.
8. Identify new sources of data.
9. Use Big Data analyses in real time, because with time some data cease to be useful.

In the era of Big Data and new, more advanced analytical capabilities, companies can gain a competitive advantage on the market, by competing on analytics. The analysis is a part of the increasingly exposed current studies on decision-making on the basis of data.

References

1. M. Beyer, *Gartner says solving 'big data' challenge involves more than just managing volumes of data.* (2011) http://www.gartner.com/it/page.jsp?id=1731916

2. D. Jelonek, *The problem of information overload in the information society (in Polish)*, Zeszyty Naukowe Ekonomiczne Problemy Usług Uniwersytet Szczeciński, T.1, **650** (2011)

3. J. Dong, Y. Qin, X.Y. Sun, L.M. Du, *Research on Improved Collaborative Filtering Recommendation Algorithm on MapReduce.* In MATEC Web of Conferences, **63**, p. 04018, EDP Sciences (2016)

4. E. Brynjolfsson, L.M. Hitt, H.H. Kim, *Strength in numbers: How does data-driven decisionmaking affect firm performance?* (2011)

5. F. Provost, T. Fawcett, *Data science and its relationship to big data and data-driven decision making*, Big Data, **1(1)** (2013)

6. T. Kraska, *Finding the needle in the big data systems haystack.* IEEE Internet Comput **17** (2013)

7. A. Fernández, S. del Río, V. López, A. Bawakid, M.J. del Jesus, J.M. Benítez, F. Herrera, *Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **4(5)** (2014)

8. J. Bloem, M. Van Doorn, S. Duivestein & E. van Ommeren, *Creating clarity with big data*. Sogeti VINT, (2012).

9. J. Wielki, *Analysis of the possibilities of using big data in e-business* (in Polish), Prace Naukowe/ Uniwersytet Ekonomiczny w Katowicach (2014)

10. T. Davenport, P. Barth, R. Bean, *How 'Big Data' is Different,* MIT Sloan Management Review, **54(1)** (2012)

11. Big Data. What it is and why it matters, http://www.sas.com/en_us/insights/big-data/what-is-big-data.html (10.01.2017)

12. A. McAfee, E. Brynjolfsson, T.H. Davenport, D.J. Patil, D. Barton, *Big data. The management revolution.* Harvard Bus Rev, *90(10)* (2012)

13. IBM, *Big Data at the Speed of Business*, (2014) [Online]. Available: http://www.01.ibm.com/software/data/bigdata/. (12.01.2017).

14. S. Sagiroglu, D. Sinanc, *Big data: A review. In Collaboration Technologies and Systems (CTS)*, 2013 International Conference on IEEE, (2013)

15. C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P.C. Zikopoulos, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data,* Mc Graw-Hill Companies, 978-0-07-179053-6, (2012)

16. http://www.forbes.com/sites/teradata/2014/11/19/defining-big-data-in-two-words-who-cares (2017.01.22).

17. R. Gupta, H. Gupta, M. Mohania, *Cloud computing and big data analytics: what is new from databases perspective?,* In: 1st International Conference on Big Data Analytics (BDA), New Delhi, India, (2012)

18. M. Rouse, "Big data", http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data, 2011. (2017.02.20)

19. I. Pawełoszek, J. Wieczorkowski, *Big data as a business opportunity: an Educational Perspective*, In: Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on. IEEE, (2015)

20. S. LaValle, *Big data, analytics and the path from insights to value,* MIT Sloan Management Review, (2011) http://sloanreview.mit.edu/article/big-data-analytics-and-the-pathfrom-insights-to-value/ (20.02.2017)

21. J. Liebowitz, *Big data and business analytics*, Boca Raton FL [etc.]: CRC Press: Taylor & Francis Group. (2013)

22. T. Davenport, *Big Data @ Work: Dispelling the Myths*, Uncovering the Opportunities, Harvard Business School Press, Boston (2014)

23. V. Mayer-Schonberger, K. Cukier, *Big Data. Revolution which will change the way we think, work and live* (in Polish), MT Biznes, Warszawa (2014)

24. A. Chluski, L. Ziora, *The role of big data solutions in the management of organizations. Review of selected practical examples*, Procedia Computer Science **65** (2015)

25. F. Provost, T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media, Inc. (2013)

26. Economist Intelligence Unit survey, September 2014 https://www.eiuperspectives.economist.com/sites/default/files/Whosbigonbigdata.pdf (2017.01.20)

27. B. Qing, J. Wang, J. Cheng, *Research on Ontology Modeling of Steel Manufacturing Process Based on Big Data Analysis*, MATEC Web of Conferences, **45**, EDP Sciences (2016)

28. B. Schmarzo, *Big Data. Understanding how data powers big businesses*, Wiley (2013)

29. T. Davenport, J. Dyché, *Big data in big companies*, International Institute for Analytics (2013)

30. Computerworld Polska (2014). Raport Big data +, Systemy analityki wielkich zbiorów danych w polskich organizacjach. Computerworld Polska.

31. D. Jelonek, C. Stępniak, T. Turek, L. Ziora, *Identification of mental barriers in the implementation of cloud computing in the SMEs in Poland,* In Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on IEEE (2014)

32. http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics (2017.02.10)