Big Data in industry

T S Latinović¹, D M Preradović², C R Barz³, M T Latinović⁴, P P Petrica⁵ and A Pop-Vadean⁵

¹The University of Banja Luka, Faculty of Mechanical Engineering, Banja Luka, Bosnia and Herzegovina
²The University of Banja Luka, Faculty of Economics, Banja Luka, Bosnia and Herzegovina
³Technical University of Cluj-Napoca, North University Center of Baia Mare, Romania
⁴Agency for Information Society of Republic of Srpska, Bosnia and Herzegovina
⁵The technical University of Cluj Napoca, Department of Mechatronics and Machine Dynamics, Cluj Napoca, Romania

E-mail: drtiho@gmail.com

Abstract: The amount of data at the global level has grown exponentially. Along with this phenomena, we have a need for a new unit of measure like exabyte, zettabyte, and yottabyte as the last unit measures the amount of data. The growth of data gives a situation where the classic systems for the collection, storage, processing, and visualization of data losing the battle with a large amount, speed, and variety of data that is generated continuously. Many of data that is created by the Internet of Things, IoT (cameras, satellites, cars, GPS navigation, etc.). It is our challenge to come up with new technologies and tools for the management and exploitation of these large amounts of data. Big Data is a hot topic in recent years in IT circles. However, Big Data is recognized in the business world, and increasingly in the public administration. This paper proposes an ontology of big data analytics and examines how to enhance business intelligence through big data analytics as a service by presenting a big data analytics services-oriented architecture. This paper also discusses the interrelationship between business intelligence and big data analytics. The proposed approach in this paper might facilitate the research and development of business analytics, big data analytics, and business intelligence as well as intelligent agents.

1. Introduction

In order to clearly define the concept of Big Data (BD), it is first necessary to define the information itself and the difference between data and information. Data are structured facts in a manner suitable for processing for people and computers, and which have no meaning. Data becomes information only when they have a meaning [1].

The amount of data generated in the world continues to grow every day. According to forecasts by Cisco [2] the level of generated data in the world at the end of 2015 will be around 4.7 zettabytes (ZB). Since the beginning of time, by the end of 2003, in total was generated 5 EB data for the whole of humanity, and from 2004 to the end of 2012 what is about 2.7 ZB electronic data. During the period of nine years, it was generated 500 times more data than in all previous periods of civilization [3]. Progression of data can be viewed through the prism of the creators of information. The first data were
bringing employees in the company, then comes the web and web users started to be the ones who created the dominant amount of data. The last group data creators, which still dominates, are machines or devices - IoT (Internet of Things). Data are growing exponentially and they can't be stored before they are analyzed. The rapid growth of data is caused in large part by the devices, as they increasingly generate data, communicate with each other or with humans.

BD is a special concept of digital information which is not suitable for storage, transport and / or analysis. These colossal level of data are so big that goes beyond the technological capabilities for storage, analysis and visualization which we have and know today [4]. The challenge is to create the next generation of specific techniques for the storage, transport and analysis of these data.

The problem of big data is not a particularly new concept. This is a problem they first met physicists at CERN, Switzerland, about 50 years ago. The example of the difficulties in managing and analyzing large data which they had - is the project of discovering the Higgs boson, 2012, because the detectors which has 1.050 million traces of particles zoom protons moving at speeds of light in the accelerator, 40 million time per second, which is a very large amount of data to be analyzed [5].

Traditional data processing systems lose its importance because of growth of volume, speed of arrival the data and unstructured data.

For example, if we have the document size of 100 MB, it cannot send via mail, the image size of 100 GB cannot be seen, and video recording of 100 TB cannot be edited using traditional tools. Tools of a new generation - Big Data Tools, should resolve this as well as many other issues. Areas of special importance to a new generation of tools are a [6]: 1) collection, 2) content, 3) storage, 4) analyze, 5) searching, 6) sharing, 7) transfer and 8) data visualization.

When data becomes BD depends on the capacity of the system of organizations. What is for one organization the BD, for the other it is not? Mir has an opinion that data become BD when they reach a size of 1 TB. A large number of different types of devices becomes smart and transmits data. That kind of data has a big speed, a large amount and variety (an unstructured type of data).

For better insight into scale BD, we put some examples in Table 1 for the some of the event which resulting is a large amount of data.

| Event               | Description                                                                 |
|---------------------|-----------------------------------------------------------------------------|
| Flight of Air Bus   | 10 TB of data is generated every 30 minutes of flight, during an average year |
|                     | of flight it created about 640 TB of data                                   |
| Smart meters        | In 2009, is about 76 million digital electricity meters, which producing 350 |
|                     | million transactions a year. It is estimated if it goes with current tempo by|
|                     | 2020, the number of these devices amount to 800 million                    |
| Cell phone cameras  | In 2013 in the world was about 5 million phones cameras. Most of them have  |
|                     | GPS. This number increasing from day to day, and also increasing the number  |
|                     | of photos and videos that users create                                      |
| Web users           | At the end of 2014, the number of internet users was nearly 3.08 billion     |
|                     | which contributed to the volume of Internet traffic measured in ZB          |
| Blogs               | Data from 2013 indicate that there are 200 million registered blogs         |
| Email               | About 300 billion emails are sent every day                                 |
| RFID                | The use of RFID tags (chips) is increasing at the global level, 2005 was 1.5 |
|                     | million tags, and for 7 years later, 30 billion tags                        |
| Twitter             | Twitter generates about 12 TB of data per day, with 200 million users and   |
|                     | 230 million tweets per day (97,000 tweets per second)                       |
| NYSE                | New York Stock Exchange daily generates about 1 TB of data                  |

Source: [7], [8]
1.1. Big Data futures
Basic features of the BD can be represented as model 3V: volume (quantity), velocity (speed), variety (diversity).

Large volume or large amounts of data is the first feature of BD and also a challenge, as well an opportunity for large organizations to better understand human behavior and react accordingly, and adjust their resources. An example of Facebook on the illustrative way could contribute to a better understanding of the large volume of data every minute of arriving in the database. Facebook users with their activities and using the services offered by this community to networks, generating 25 TB of data per day (1.04 TB per hour).

The speed of large amounts of incoming data which coming in is closely linked to the capacity of IT systems and ways of exploitation of the BD in real time (Fast Data). This feature BD opens a whole new field of research. Example coming of large amounts of data is a Swiss CERN where during tests every second arriving 40 TB of data.

The diversity of data. Data coming at high speed, on a large scale, but here is a large variety of their forms. The data can be divided into structured (text, numbers, documents, financial data, records, personal data, etc.) and unstructured (photos, text messages, audio and videos, 3D models, simulations, geographical location and so on). Today we have a lot of unstructured data, which makes these data unsuitable for the traditional way of processing data. From the total available data for decision making, the company uses only 10%. The remaining 90% of the company data is not used in decision-making and managers in this area need to search an opportunity for growth and development.

IBM's definition of GDP includes 4V (volume, velocity, variety, veracity). The last term refers to the (UN) reliability of the data or truthfulness. According to IBM, veracity stands out as a major problem, because 3.1 billion dollar is the costs in the United States annually due to incorrect information [9].

If data grow, it grows the need for computer performance. However, data will go up and on one level are starting to exceed the capacity of traditional systems. In order to keep tempo with growing data, BD technology has changed the principle of work. Big data is "shattered" into smaller parts where parallel working multiple computers which are linked in clusters.

2. Big Data resources
With the growth of the data over time organizations falls into a kind of information gap, because the resources at their disposal can process structured (and somewhat semi-structured) data, small-scale (Small Data). The volume of unstructured, available data, is far greater than the volume of structured data. The space between the data that organizations can use (Small Data) and the total available amount of data that organizations do not exploit - represent Big Data Gap or the amount of data that organizations are not taken into account in its business for making decisions.

This data gap is potential for the organization to achieve its competitive advantage. According to research by McKinsey Global Institute's [10] companies that recognize the potential of BD in decision making and with BD tools learn how to get information and knowledge - will achieve a competitive advantage in the market about 20% compared to companies that do not use the BD.

2.1. Examples of exploiting Big Data
Amazon is example a company that has learned how to from the multitude of data (search, clicks per page, status updates on social networks, purchase history, site navigation, place of residence, climate, temperature, etc.) using algorithms, generate knowledge on the basis of which suggests customer items you should buy, or which would have been interesting. Similarly done Facebook also, the company use BD for targeting customers for advertising and marketing campaigns, through well-defined filters.

3. The industries with the greatest potential of Big Data
In every field of human endeavor and work today are present digital data. Except for those in the field of IT, Big Data has become a very important area for the top management of companies and
organizations in many sectors. If we make a comparison of economic and non-economic areas, in some it can be more and in some less exploit the potential of BD.

According to research by McKinsey Global Institute [10], cluster A (which consisting of computer and electronic products, and information sector) had the highest percentage growth in productivity (Y axis) in the reporting period, but the greatest benefit from the exploitation of BD (X axis) (Figure 1, the size of the circles refers to the relative share of GDP). The sector, which had a small increase in productivity, and has a significant potential in the field of BD cluster B (finance, insurance, management).

![Figure 1. The potential exploitation of BD by sector in the US](source)

Cluster C (construction, other services, education, arts, and management companies) has seen a decline in productivity, which explains the strong systemic barriers that prevent the growth of productivity. Cluster D, which is represented at the global level had a significant increase in productivity, a cluster E, which more exist at the local level slightly lower productivity growth. Both clusters have widely BD potential. Although all sectors have to overcome the barriers that prevent their exploitation of BD, for some this will be more and for other less demanding. For example, public sector and education will have more difficulties to create a competitive advantage based on BD because they don't have the information and action as organizations that are driven by data. Limited investment and growth in health care have so far led only to a relatively low level of performance of IT investments into health care. Sectors such as manufacturing, sales and service sector can have a relatively small barrier because of the opposite reason.

The report identified the sectors in which the exploitation of BD led to a significant increase of productivity (Table 2). The benefits that would be achieved include the following: increasing the effectiveness and efficiency, production high-quality product/service, or creating large added value to
products and services. For example, entrepreneurial firms can use and analyze the data to create
design products/services that will meet customer needs on higher level indicators and/or in a better
way. Data can be used for example to create value-added products during its use. An example of this is
a smartphone that "learns" the needs and preferences its owner, and holds the supporting applications
and the information that owner used the most commonly, as the user has a higher value than the brand
new device that is not customers oriented.

Table 2 – Estimated benefits in five sectors exploiting of Big Data (2011)

| Sector                  | Estimated annual added value | % Productivity growth per annum |
|-------------------------|------------------------------|--------------------------------|
| Health care (USA)       | $ 300 billion                | 0.7%                           |
| Administration, the public sector (EU) | $ 250 billion | 0.5%                           |
| GPS location (global level) | $ 100 billion               | 700 billion to end-users directly or indirectly |
| Sale (USA)              | 60% growth                   | 0.5 – 1.0%                     |
| Manufacture             | 50% growth                   | 7% reduced costs               |

Source: [10].

Personal location data can contribute to the reduction of transport costs, fuel and time savings. These are ways of generating an additional value based on the use of big data. According to estimates, the report would lead to significant growth of value added in production or service activity in five sectors should the information used in making decisions. The largest effects would be felt in the health care system (in the US), which would benefit from the exploitation of BD got $ 300 billion of added value per year for their services, which would raise the quality of customer service. In Europe would most benefit the users get the service administration of the public sector, as much as 250 billion dollars a year additional value, with an increase in productivity of 0.7% per year (Table 2).

3.1. Big Data potential in industry of Bosnia and Herzegovina

In this section will be referred to several examples of potential exploitation BD in Bosnia and Herzegovina.

3.1.1. Tax Administration. The fiscal processes in BiH were completed in 2011. The introduction of fiscal cash registers in the Tax Administration, through the network, every day, arriving large amounts of data. The analysis of these data might show a different pattern that showed tax evasion, but also to get information about the structure of demand recorded on the basis of which to be able to adjust certain economic policies and take appropriate corrective measures to stimulate the economy. The volume of unused data is significant because the Tax Administration uses data on total daily traffic on the basis of which we compare the financial statements of organizations. In this way only can perform traffic cash control while data on the structure and quantity of sold products/services remain unanalyzed.

3.1.2. Telecoms in BiH. Data from the end of 2011 show that in BiH were 4.12 million mobile phones. Telecoms are an obvious example of large amounts of data that are constantly generated. According to available data, of the telecom environment, Croatia Vipnet in 2013 began with the introduction of the BD technology (we can't find the data for BiH).

3.1.3. The network of highways. In BiH so far have been built 124 km of highways. Although it is not much, in the future may represent a potential for application of BD tools (data are generated at the paying tolls places, traffic counters, the speed of driving, etc.). These data could be used for the purposes of managing traffic jams, "bottlenecks", but also during accidents.

Hydroelectric power plants in Bosnia and Herzegovina are the source of large amounts of data that are generated daily and monitored. Based on the analysis of data on water levels, but the amount of
energy produced, it could better manage production at the commercial level and qualify for electricity abroad, the current market, when capacities exceed the needs of consumers, but also to reveal the place of electricity losses, and its illegal use.

4. Conclusion

Big Data technologies in recent years have full expansion. A large number of scientific papers, reports and projects confirm the necessity to adopt BD tools to the daily operations. While this area is very popular in the world, in the countries of the region it is still in its infancy (Croatia officially has one project Big Data in Telecom, while in Serbia gathers online community through the blog hadoop-srbija.rs). In Bosnia and Herzegovina, there are still no projects to introduce Big Data, either forums or blogs on this topic. Also, Big Data is not mentioned in the curriculum in higher education at the University of Banja Luka.

The information gap that occurs due to the inability of processing unstructured data is recognized as a potential for organizations that bridging this gap can achieve about 20% of competitive advantage, compared to organizations that do not use the BD tools.

Economic areas that have the greatest potential for exploitation of the BD in order to raise profitability, increase efficiency and reduce costs are, in the first place, in the field of the information sector, followed by the financial sector, insurance, management, and retail.

However, one should not forget to mention the risks faced by organizations implementing BD projects. These risks can have a source of support to vendors (BD tools do not have the classic support for users), lack of knowledge in the organization, the staff employment for the BD, the interpretation of the results, the risk of diversion, unnecessarily high costs and so on. For all these risks organizations must learn to create internal policies to manage all identified risks and have reduced them to an acceptable level.

The paper provides the guidelines for the potential use of Big Data in Bosnia and Herzegovina. The first area of use is the tax area - here the application of BD tools and data analysis could identify tax evasion, certain market anomalies in supply and demand for bringing corrective or other economic measures, etc. Telecoms in BiH are next identified users of BD technology to enhance their services and form packets based on user needs, but also to prevent the outflow of users, or their transfer to other operators. Except those, there is a network of highways and hydroelectric power plants.

Overall, the most of the authors emphasize the information gap between the used data to make decisions and the total available data, of which the vast majority unstructured. Organizations ignore the colossal amounts of data when making decisions. BD gives businesses and organizations a deep insight into the data on business partners, customers, weather, pandemics, business forecasts, traffic and transport and other areas. We can get more knowledge if we connect all the data, which will give us new aspects of life. New knowledge in the field of medicine, the development of disease and pandemics, transport, traffic, military, security, trade, in situations where it is necessary to predict: Business, weather conditions, a pandemic.

The Future of Big Data can be seen, among others, and through the prism of Fast and Extreme Data. The world is slowly starting to shift focus from Big to Fast and the possibility of exploiting the data generated in real time and on the basis of their response time paramount. Companies can no longer link the banner to do something better than others, because everyone already have a high quality, but they are quicker than others. The need for speed is imperative as it increases profitability and customer satisfaction. The primary objective is the satisfaction of the client's needs when he wants, where he wants and how he wants, otherwise, it will be "hijacked" by the competition. Gartner in its report states: "In a consumer society, the winner will be decided by company's ability that knows your client in order to react faster and better than the competition".

In addition, the future of big data will bring more V: value - the value of the obtained knowledge from Big Data use; virtual - a tendency to create virtual models and simulations for BD. We believed that in the near future for BD will be possible apply intelligent information systems (their use in transport, in order to increase safety, is already well advanced) that could make decisions for the same
act. A particular problem in the business world, but also in public administration, represents a large amount of stored text documents, PDF or other formats. BD could go in the direction of developing a document management system for BD, which would meet the needs of users and the possibilities of new applications.

Further research could be related to the technical requirements, infrastructure and BD architecture, exploring the possibilities of introducing new technologies for exploitation, exploration of new methods of access to the Fast Data, but also and the Extreme Data.

References
[1] Aleksić-Marić V 2008 Business informatics Second Edition, Banja Luka, Faculty of Economics, Banja Luka, Bosnia and Herzegovina
[2] Cisco 2013 Cisco Global Cloud Index: Forecast and Methodology 2013–2018 White Paper. Available at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html [Last access 7 2015]
[3] Intel 2013 Learn Big Data and Make a Big Impact, Available at http://www.intel.co.uk/content/www/uk/en/big-data/learn-about-big-data.html [Last access 7 2015]
[4] Mir H 2013 Tutorials: Zero To Pro Training, Available at http://zerotoprotraining.com/index.php?mode=video&id=327 [Last access 7 2015]
[5] Smith T 2013 YouTube, Available at https://www.youtube.com/watch?v=j-0cUmUyb-Y [Last access 7 2015]
[6] Mir H 2013 The Value Amazon, Facebook, and Google are Extracting From the Big Data, Available at http://zerotoprotraining.com/ index.php?mode=video&id=330 [Last access 22 7 2015]
[7] Mir H 2014 Big Data Risks, Available at http://zerotoprotraining.com/ index.php?mode=video&id=1181 [Last access 26 7 2015]
[8] Stats I W 2015 Stats: Internet World Stats, Available at http://www.internetworldstats.com/stats.htm [Last access 19 7 2015]
[9] IBM 2015 Big Data Hub, Infographics & Animations, The Four V’s of Big Data, Available at http://www.ibmbigdatahub.com/infographic/four-vs-big-data
[10] McKinsey Global Institute 2011 Big data: The next frontier for innovation, competition, and productivity, s.l.: s.n.