Journal of Mechatronics, Automation and Identification Technology

Vol. 2, No.2. pp. 16 – 20

Big Data Overview, The Vs Characteristics and Applications

Jamal Bzai, Miroslav Bojović

Computer Science Department, Faculty of Electrical Engineering, Belgrade, Serbia
Computer Science Department, Faculty of Electrical Engineering, Belgrade, Serbia
bj145052p@student.etf.rs, mbojovic@etf.rs

***Abstract* — Big data is a collection of massive and complex data sets and data volume that include the huge quantities of data, data management capabilities, social media analytics and real-time data. Big data analytics is the process of examining large 0amounts of data. There exist large amounts of heterogeneous digital data. Big data is about data volume measured in terms of terabytes or Petabytes. After examining of Big data, the data has been launched as Big Data analytics. In this paper, presenting the Vs characteristics of big data and the technique and technology used to handle big data. There are many motivations for the adoption of big data; this data has innovation, the economy, productivity and future growth. Big data analytics has become very popular in the area of marketing, industries, healthcare, and a brief discussion on privacy.**

I. INTRODUCTION

Big Data is something so huge and complex that it is impossible for traditional systems and traditional data-warehousing tools to process and work on them. Data (Big Data) is generated by machines, generated by humans, and also generated by mother nature. With the growth of technologies and services, this large data is produced that can be structured, semi-structured and unstructured from the different sources. Big data can neither be worked upon by using traditional SQL like queries nor can the relational database management system (RDBMS) be used for storage. So that a wide variety of scalable database tools and techniques have evolved. Hadoop, an open source distributed data processing system is one of the prominent and well known solutions. The NoSQL has gained prominence as a non-relational database with the likes of MongoDB, Dynamo DB from Apache[1], [2]

The need of big data comes from the Big Companies like Google and Facebook. For the purpose of analysis of big amount of data which is in unstructured form. Such type of data is very difficult to process that contains the billions records of millions people information that includes the web social media, images, audio and so on. The report is divided in the following sequence: Starting with the introduction, we talk about the characteristics of big data (V's). It is followed with a descriptive note on the various components of Big Data based on Hadoop framework. Apache Hadoop is an open source software framework for storage and large scale processing of data sets on clusters of commodity hardware. Hadoop was developed by Doug Cutting and Mike Cafarella in 2005[1].

II. CHARACTERISTICS

Big data can be described by some of the following characteristics:

A. Volume

The volume presents the most immediate challenge to conventional IT structures. This is the aspect that comes to most people's minds when they think of Big Data. Many companies already have large amounts of archived data in the form of logs, but do not have the capacity to process that data. The benefit gained from the ability to process large amounts of information is the main attraction of big data analytics.

B. Velocity

Velocity refers to the increasing speed at which this data is created, so, the increasing speed at which the data can be processed, stored and analyzed by relational databases. Velocity refers to the speed at which new data is generated and the speed at which data moves around. About social media messages going to viral in seconds In 1999, Wal-Mart's data warehouse stored 1,000 terabytes (1,000,000 gigabytes) of data. In the year 2012, it had access to over 2.5 petabytes (2,500,000 gigabytes) of data. Every minute of every day, we upload hundreds hours of video on Youtube, We send over 200 million emails through Gmail's.

C. Variety

The next aspect of Big Data is its variety. Big Data is not always structured data and it is not always easy to put big data into a relational database. This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts Dealing with a variety of structured and unstructured data greatly increases the complexity of both storing and

analyzing Big Data. 90% of data generated is data is in unstructured form.

D. Veracity

When we are dealing with a high volume, velocity and variety of data, it is not possible that all of the data is going to be 100% correct there will be dirty data. The quality of the data being captured can vary greatly. The data accuracy of analysis depends on the veracity of the source data.

E. Value

Value is the most important aspect in the big data. Though, the potential value of the Big Data is huge. It is all well and good having access to big data but unless we can turn it into value it is become useless. It becomes very costy to implement IT infrastructure systems to store big data, and businesses are going to require a return on investment.

Digitalization of information is the foundation of big data, and the amount of available data is increasing at a rapid pace. In the 1950s, John Hancock's insurance company was thought to have more data than any other organization, with 600 megabytes of data. Throughout history, this number has continued to increase incrementally. The Federal Express had the greatest reported amount of data by the 1970s with 80 gigabytes. By the 1990s, Walmart took the lead with the largest amount of data with 180 terabytes of data, 2250 times the reported amount in the 1970s. To this day, Facebook is reported to have the largest amount of data; approximately 100 petabytes worth[3].

In our modern age, the amount of data being generated is tremendous. In 2010, the International Data Corporation reported that the amount of data stored was approximately 1 zetabyte. By 2011, the International Data Corporation reported the amount of data has increased to almost 2 zetabytes[3].

The amount of data being stored is so much more in our modern age because we are following a trend of digitalizing everything. There are vast amounts of emails, social media updates, and geolocation tags; everything is made available online in the form of data. This type of information is what is being mined to process and make use for big data analytics. Being able to understand just how great the amount of data available this date is challenging, but it is reported that there are "nearly as many bits of information in the digital universe as stars in our physical universe"[3].

III. TECHNIQUES AND TECHNOLOGY

Since big data is not only large, but also varied and fast-growing many technologies and analytical techniques are needed in order to attempt extracting relevant information. For processing the large amount of data, the big data requires exceptional technologies. This techniques and technologies have been introduced for manipulating, visualizing and analyzing of big data.

So, to handle big data there are many solutions are available, but the Hadoop technology is one of the most widely used technologies[4], [5].

IV. USES FOR BIG DATA

There are many motivations for the adoption of big data. Data has a tremendous potential to drive innovation, the economy, productivity and future growth. An outstanding example of the benefits of big data is Google Flu Trend; a service that predicts and locates flu outbreaks. This service uses aggregate search queries for processing. This service has the potential to reduce the impact of influenza[6].

Many corporations are taking advantage of big data to boost their Return on Investment (ROI). As IBM stated, "Plenty of customers are seeing tangible ROI using IBM's big data and analytics platform"[7]. Within the healthcare industry there has been a 20% decrease in patient mortality by analyzing streaming patient data. The telecommunications experienced a 92% decrease in processing time by analyzing networking and call data. Further, the utilities industry witnessed a 99% improvement in accuracy 99% in placing power generation resources by analyzing 2.8 petabytes of untapped data. These are just some of the industries making use of big data analytics. When organizations can make use of the full potential of big data analytics rather than just a segment, they gain a truly powerful tool to boost their Return on Investment[7].

A. Marketing

Big data marketing is about driving up value by engaging customers more effectively. This can be done in many forms. e.g., a coffee shop could be making great use of big data marketing by combining a potential customer's geolocation and search history. This way they could engage customers who they know both like coffee and are in the area. This is called target marketing and is a very effective way of segmenting the market into a subset of who is most likely to respond to the advertisement[8].

This form of marketing sounds simple at first, but big data marketing is rather complex. Technological advancements are making this form of marketing easier, allowing modern organizations to fully take advantage of the digital information made available. The old ways of non-segmented marketing focused at consumers are over. In our modern age, organizations must engage consumers, act on customer feedback and deliver services that are truly personalized towards the consumer in a timely manner. In order to meet these demands, marketing teams need to promote their services online through websites, mobile apps, and social media pages. This allows organizations to allocate their spending from less productive marketing strategies to much more productive elements of their marketing mix[8]. The modernization of marketing is full of potential. Marketing teams must take advantage of new technology, and revaluate current processes to fully leverage their full grasp of the marketing industry. This is essential as, "Companies that integrate people, processes, and technology will deliver a ROMI that is 50% higher than those that don't"[8].

Behavioral profiling is one of the ways marketing teams can segment their marketing toward a target market. With the latest technology, a combination of

complex algorithms and large amounts of data enables the creation of customer profiles. Organizations in a vast array of industries are taking advantage of behavioral profiles, to completely understand and connect with potential customers. Combining data from social media, blogs, surveys, search history, sales data, geolocation and many other sources can create detailed profiles on customers and allow marketing teams to segment their advertisements to those who are most likely to respond positively. This allows organizations to more fully understand the customer; their interests, what they are doing and even as far as what they are thinking, from mining texts or social media updates. This is a new age of marketing[9].

These marketing teams that can identify their customer's behavioral profile in such complexity, they can send very personal and truly targeted messages to the customer, creating a personal relationship, even if they have millions of other customers. e.g., a user who has been searching the internet for toddler toys, who also has provided details of his location, can be effectively delivered advertisements for toy stores near his current location[9].

With behavioral profiling, there are certain aspects that need to be addressed.

There must be an understanding of the users' online and offline behavior: How do they interact online and/or in person? What does this person purchase? What is the quantity in which the customer purchases and how often? These are all important topics which help marketing teams understand what exactly their customers want and allow them to reach out on a personal level and create a relationship[9].

B. Pervasive Computing

Pervasive computing is becoming a key form of gathering data for big data analytics. Pervasive computing, also known as ubiquitous computing, is a growing trend associated with inserting technology into everyday life. Pervasive means that it is everywhere, hence the implementation into our everyday life. These gadgets are completely connected and always available in real-time[10]. This form of computing is vast; it is a modernized version of the traditional desktop environment. This includes smart watches, smart phones, and even smart homes that communicate with appliances within the residency. Day-to-day equipment is being equipped with microprocessors which truly connect our devices and lives.

Pervasive computing makes daily activities easier through a wide range of technologies including Wi-Fi compatibility, voice recognition, Bluetooth, networking, etc. These technologies simplify our lives, which is why they are becoming a growing trend. With the addition of such pervasive technologies collecting mass amounts of data on our daily lives, there are security and privacy risks. To defeat this issue, trust models should be put in place to make these devices more secure. There are also other drawbacks to this form of computing; slow communication, expensive and limited bandwidth. These all cause a security risk to pervasive computing due to the system vulnerabilities they can create[10].

C. Internet of Things (IoT)

The IoT is becoming the primary grounds for data mining for big data analytics. The concept of IoT involves embedding sensory and network capable devices into our everyday lives. IoT has been growing in popularity with the adoption of wearable devices such as smart watches that track heartbeats, smart home technology, and users publishing their lives on social media via Smartphones. These devices are all communicating together and storing information in the cloud. The immense amount of data being collected and processed by big data analytics from the IoT introduces privacy as a major concern[11].

Furthermore, due to the nature of such technology, there are also security concerns. These systems are very resource-constrained, resulting in a large amount of attention in cryptography and security engineering. These technologies usually work in uncontrolled environments where they are vulnerable to malicious misuse. In order to alleviate these issues special attention needs to be focused on the following areas as stated by Antonio Skarmeta and M. Victoria Moreno[12]:

- Design of lightweight security protocols and cryptographic algorithms.
- Lightweight and efficient implementations of security protocols and crypto- graphic algorithms.
- Secure implementations in hardware and/or software.

D. Smart Cities

Smart Cities are cities that use technology to improve processes, reduce environ- mental footprint and improve citizen's lives through the use of technology. Technological advances in information and communication technologies are enticing the movement into smart cities. For example, a city could analyze the traffic during peak hours with big data analytics and design traffic lights and new road ways to take these elements into consideration.

Advances in information and communication technologies are creating a movement in the environment in which we live; it is creating intelligent systems known as Smart Cities. These intelligent environments capture data using sensors in the boundaries of the city. From processing the data captured by these sensors, the city is able to adapt their behavior to better fit the citizens in the community[13].

V. BIG DATA ANALYTICS

The benefits of using big data analytics are significant. Many organizations are adopting the use of this new technology and are experiencing substantial results. As reported by IBM[14]:

- Sixty-three percent of respondents reported that the use of information including big data and analytics—is creating a competitive advantage for their organizations
- When compared to companies that rely on traditional analytics alone, organizations that implemented big data and analytics pilot projects or deployments are 15% more likely to report a significant advantage from their information assets

and analytics

Companies that use big data analytics outperform in their industry. This data allows companies to identify their most profitable customers and enables them to reach out to them on a personal level, enabling a better customer experience. This technology also allows organizations to improve the speed of processes, better understand customer needs, and improve innovation. By properly understanding customer needs, companies know exactly what type of innovations will be successful; where they should invest in with their R&D department[14].

The implementation of big data analytics also helps manage risk. Without the speed of big data analytics, companies often cannot find the proper information quick enough, which can often lead to unnecessary risk from poor decisions.

Companies need a proactive way of finding valuable information when they need it; making the use of data analytics a valuable tool. Big data analytics can help companies acquire and retain customers. With the use of this technology, organizations are able to create customer profiles. They understand how their customers interact, they purchasing history, their interests, quantity of purchases and much more. With this information they are able to attract the right customers through target marketing and then attain their business by providing the proper experience that the customer requires.

Companies can improve operations and lower their costs with data analytics. They can do this by drastically improving process efficiency and through the ability to make quality decisions faster. Through improving operations, organizations are better prepared to identify and pinpoint fraudulent activity. This can be accomplished through processing data and identifying trends[14].

Using a powerful analytics-based software platform, MoneyGram International can now better understand their users. These insights helped them prevent more than 37.7 million US dollars in fraudulent transactions, reduce customer fraud complaints by 72 percent, and promptly address stringent regulatory requirements[14].

With big data analytics, it makes revamping products much more efficient. This analysis helps organizations understand how customers perceive their products. Simply mining text on social media about the related product provides endless customer opinions. With big data, you can test thousands of new variations of the product based on customer needs for costs, lead times, and performance[14].

Big data can also keep organization's sensitive data safe. Companies can now map their data landscape allowing them to analyze threats before they occur and then mitigate them. This allows them to detect sensitive information that is not properly protected. This could potentially save companies millions through preventing sensitive data leaks. This can also save companies their integrity to the public.

The use of big data is also making our cities smarter. This technology mitigates the risks of fast expansion. For example, Oslo, a city in Norway, managed to reduce the street light energy consumption by approximately 62%[15]. Police in Memphis began to implement the use of big data predictive technology; resulting in 30% reduction in serious crime. "The city of Portland, Oregon, used technology to optimize the timing of its traffic signals and was able to eliminate more than 157,000 metric tons of CO_2 emissions in just 6years—the equivalent of taking 30,000 passenger vehicles off the roads for an entire year"[15]. These are just some of the many examples of cities that are making improvements for both the environment and their citizens with the use of big data technology.

VIII. CONCLUSIONS

Big data provides an opportunity for "big analysis" leading to "big opportunities" to advance the quality of life, or to solve the mysteries of the world. We are in the development area of big data.

Big data is a wildly growing topic which has been proven to be a beneficial tool. The implications of being able to collect and analyse such vast amounts of information are phenomenal in a range of applications including marketing, healthcare, telecommunications, and more. Businesses and governments are beginning to adopt big data solutions to address a wide range of issues.

This paper has provided an introduction to big data. We provided an overview of big data concepts, motivations for using big data and applications, also solutions have been applied to solve important problems in real world.

REFERENCES

[1] Apache Software Foundation. Official website http://hadoop.apache.org/

[2] University of Texas at Austin School of Information Big Data Analytics Dylan Maltby 1616 Guadeloupe, Austin, TX 78701 512-471-3821.

[3] K. Bottles, E. Begoli, B. Worley. Understanding the pros and cons of big data analytics. (2014). Retrieved from http://go.galegroup.com.uproxy.library.dcuoit.ca/ps/i.do?p=AONE&u=ko_acd_uoo&id=GALElA377410232&v=2.1&it=r&userGroup=ko_acd_uoo&authCount=1

[4] Bakshi, K, (2012)," Considerations for big data: Architecture and approach"

[5] Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.

[6] O. Tene, J. Polenetsky. Data, privacy in the age of big data. (2012, February 2). Retrieved from

Stanford Law Review: https://www.stanfordlawreview.org/online/privacy-paradox/big-data

[7] IBM. Big data and analytics. (2015, June). Retrieved from IBM: http://www.ibm.com/bigdata/

ca/en/big-data-and-analytics/operations-management/industries/index.html

[8] L. Arthur, in Big Data Marketing, ed. by L. Arthur (John Wiley & Sons, 2013). Retrieved Aug 2015, from http://library.books24x7.com.uproxy.library.dcuoit.ca/assetviewer.aspx?bookid=58128&chunkid=174813851&rowid=235

[9] M. V. Rijmenam, Why 360-degrees customer profiles created with big data are nothing new. (2015). Retrieved from DATAFLOQ: https://datafloq.com/read/360-degrees-customer-profiles-created-big-data-not/109

[10] Techopedia. (n.d.). Pervasive computing. Retrieved from Techopedia: http://www.techopedia.com/definition/667/pervasive-computing

[11] Datamation. Why big data and the internet of things are a perfect match. (2015). Retrieved from Datamation: http://www.datamation.com/applications/why-big-data-and-the-internet-ofthings-are-a-perfect-match.html

[12] A. S. Moreno. Internet of Things Security, Privacy and Trust Considerations. (2014)

[13] K. B. Ahmed, M. Bouhorma, M. Ahmed, Age of big data and smart cities: privacy trade-off. (2014, October). Retrieved from arXiv: http://arxiv.org/ftp/arxiv/papers/1411/1411.0087.pdf

[14] IBM. Better business outcomes with IBMBig Data & Analytics. (2014, January). Retrieved from IBM: http://www.ibmbigdatahub.com/sites/default/files/whitepapers_reports_file/59898_Better%20Business%20Outcomes_White%20Paper_Final_NIW03048-USEN-00_Final_Jan21_14.pdf

[15] Data Science Series. (n.d.). Examples of what you can accomplish with big data. Retrieved from Data Science Series: http://datascienceseries.com/stories/ten-practical-big-data-benefits