International Journal of Trend in Research and Development, Volume 8(5), ISSN: 2394-9333
www.ijtrd.com

Future of Big Data

[1]Matthew N. O. Sadiku, [2]Olaniyi D. Olaleye, [3]Abayomi Ajayi-Majebi and [4]Sarhan M. Musa,
[1,4]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]Barbara Jordan-Mickey Leland School of Public Affairs, Texas Southern University, Houston, TX, USA
[3]Department of Manufacturing Engineering, Central State University, Wilberforce, OH, USA

Abstract: We live in an era of data. Data is increasingly integrated into every aspect of our lives. It has become is a high-value commodity. The Internet is being flooded with a huge amount of data, which is consolidating the big data concept. Big data is an umbrella term for large volumes of structured and unstructured data that we encounter daily. It may be regarded as a collection of data from traditional and digital sources inside and outside an organization. Big data is a fast-moving field that has taken the business world by storm. The exponential growth of global data is not slowing down anytime soon. The future for big data is exciting. This paper addresses the reasons for the excitement.

Keywords: *Big Data, Data Analytics, Future Of Big Data*

I. INTRODUCTION

Today, the world is powered by data. We swim in a sea of data and data is everywhere. The amount of data generated worldwide is exploding. Big data experts forecast that the amount of generated data will be growing exponentially in the future. There are several reasons for this prediction [1]. First, the number of Internet users working online is growing. Second, billions of connected devices and embedded systems create, collect, and share data across the world. Machines and implanted sensors are generating unimaginable amounts of data. Third, fast data has also spoilt users, making them addicted to real-time and also expect to access data on the go. Consequently, the volume of data is only increasing by the year. Data is going to grow, and grow, without stopping. The amounts of data generated and stored have expanded within a short period of time, leading to what is commonly known as "big data." The major sources of big data are shown in Figure 1 [2].

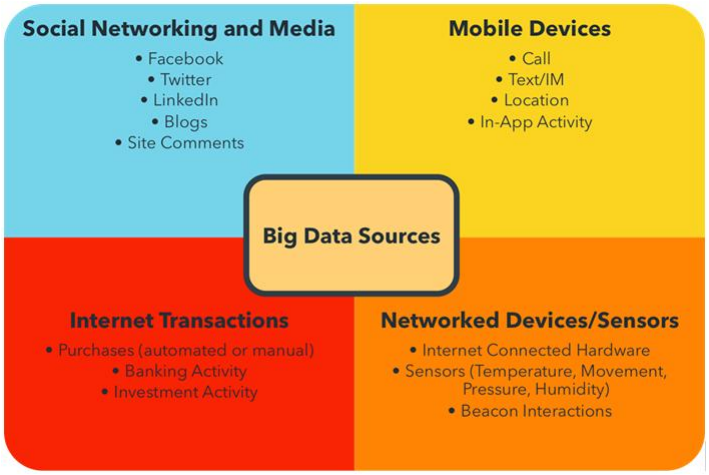

Figure 1: Major sources of big data [2].

Big data refers to data sets too large or complex for traditional data-processing applications. It may be regarded as a large amount of both structured and unstructured data. Structured data is information that is easy to categorize and search, while unstructured data is not as easy to organize [3]. The two typed of data are depicted in Figure 2 [4].

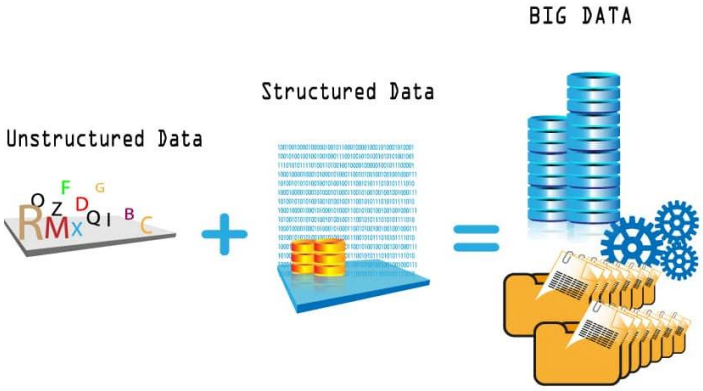

Figure 2: Structured and unstructured data [4].

II. REVIEW ON BIG DATA

Big data (BD) refers to a collection of data that cannot be captured, managed, and processed by conventional software tools. It is a relatively new technology that can help many industries, including government. The three main sources of big data are machines, people, and companies. Big data can be described with 42 Vs [5]. The first five Vs are volume, velocity, variety, veracity, and value [6].

- *Volume*: This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as data over one petabyte in volume.
- *Velocity*: This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed. The volume of data is only increasing by the year in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- *Variety*: This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- *Veracity*: By this, we mean the truthfulness of data, i.e. weather the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- *Value*: This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

On this basis, small data can be regarded as having low volume, low velocity, low variety, low veracity, and low value. Additional five Vs has been added [7]:

- *Validity:* This refers to the accuracy and correctness of data. It also indicates how up to date it is.
- *Viability:* This identifies the relevancy of data for each use case. Relevancy of data is required to maintain the desired and accurate outcome through analytical and predictive measures.
- *Volatility:* Since data are generated and change at a rapid rate, volatility determines how quickly data change.
- *Vulnerability:* The vulnerability of data is essential because privacy and security are of utmost importance for personal data.
- *Visualization:* Data needs to be presented unambiguously and attractively to the user. Proper visualization of large and complex clinical reports helps in finding valuable insights.

Figure 3 shows the 10V's of big data. In addition, the 10V's above, some suggest the following 5V's: Venue, Variability, Vocabulary, Vagueness, and Validity) [8]. The future of big data will bring more Vs.

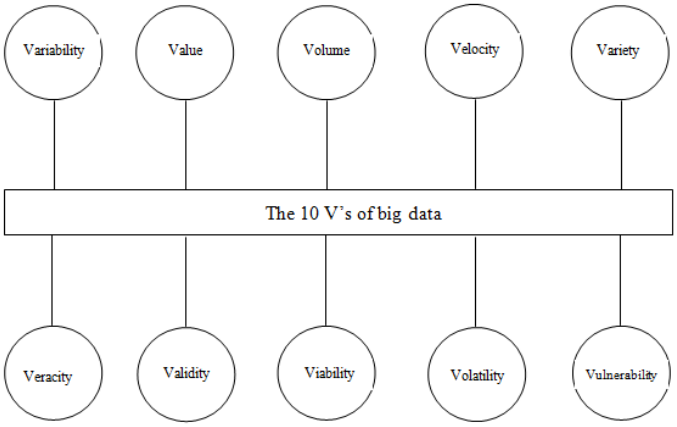

Figure 3: The 10V's of big data.

To thrive in today's complex business environment, businesses must adopt a data-driven culture and leverage analytics platforms to make key decisions that improve productivity. Industries that benefit from big data include the healthcare, financial, e-commerce, education, airline, travel, restaurants, automobile, sports, agriculture, manufacturing, hospitality industries, and government.

III. BIG DATA ANALYTICS

Big data sets can be staggering in size so that its analysis is daunting. Every day, data is growing bigger and bigger, and big data analysis (BDA) has become a requirement for gaining invaluable insights into data such that companies could gain significant profits in the global market. Big data analytics can leverage the gap within structured and unstructured data sources. Once the big data is ready for analysis, we use advanced software programs such as Hadoop, MapReduce, MongoDB, Spark, Cassandra, Apache Storm, and NoSQL databases [9]. Big data analytics refers to how we can extract, validate, translate, and utilize big data as a new currency of information transactions. It is an emerging field that is aimed at creating empirical predictions. Data-driven organizations use analytics to guide decisions at all levels [10].

The key to big data is analytics. Big data analytics is capable of processing massive amounts of dirty data and extract the gold information from it. It basically learns from data to predict the way individuals will behave in the future. It is on the verge of a major transformation. Industries making investments in big data analytics include education, banking, retail, manufacturing, finance, healthcare, transportation, government, and military. Some of the benefits of big data analytics are shown in Figure 4 [11].

Figure 4: Some benefits of big data analytics [11].

IV. TRENDS IN BIG DATA

Here are some of the predictions of experts on the future of big data [12]:

- Data volumes will continue to grow.
- Ways to analyze data will improve.
- More tools for analysis (without the analyst) will emerge.
- Prescriptive analytics will be built into business analytics software.
- Real-time streaming insights into data will be the hallmarks of data winners.
- Machine learning is a top strategic trend for 2016,
- Big data will face huge challenges around privacy.
- More companies will appoint a chief data officer.
- "Autonomous agents and things" will continue to be a huge trend.
- Big data staffing shortages will expand.
- The big data talent crunch may ease as companies employ new tactics.
- The data-as-a-service business model is on the horizon.
- Algorithm markets will also emerge.
- Cognitive technology will be the new buzzword.
- All companies are data businesses now.
- Businesses using data will see $430 billion in productivity benefits.
- "Fast data" and "actionable data" will replace big data.
- The shortage of skilled staff will persist.
- Do-it-yourself analytics will help more people analyze and forecast than ever before.

V. FUTURE OF BIG DATA IN SOME APPLICATIONS

Different industrial sectors use big data to offer effective business solutions for their customers. Manufacturing, education, healthcare, stock markets, business, and transportation lead in harnessing the power of big data. The future of big data in every industry depends on how information is stored, processed, and applied. We consider how big data is impacting the following industries.

- *Business:* Data and business go hand in hand.. Big data is at the heart of business planning, decision making, and strategies for **customer satisfaction**. Big data solutions allow businesses to consolidate large amounts of information in a short amount of time for easier analysis. This allows financial institutions to gain insight into their markets effectively and efficiently. Big data technology helps businesses to unveil trends and provide actionable insights to businesses. Business analysts leverage data analytics technologies to make more informed business recommendations, predict trends, and help increase profits. Big data also has enormous effect on today's workforce. Financial data underpins everything we do from accounting, auditing, reporting, and sharing insights.
- *Healthcare*: The healthcare industry produces zettabytes of data taken from , mobile health (mHealth), electronic health records (EHRs). medical imaging, and medical devices. Big data in healthcare consists of massive volumes of structured and unstructured data. The future of data analytics for healthcare is in problem-solving. Patient and financial data is not only regulated, but also often spread across many hospitals, administrative offices, and insurance companies [4].
- *Smart Cities:* With the rapid increase in the use of Internet-based technologies in the smart cities, a large amount of data (a.k.a. big data) is generated. This data comes from many sectors of the cities such as e-commerce, energy, transportation., IoT, RFIDs, smart phones, sensor nets, smart household appliances, etc. Smart cities provide the opportunity to connect people and places using innovative technologies [13].

Other sectors that will be greatly impacted by big data in the future include education, tourism, real-estate, tele communication, accounting, and agriculture.

VI. CHALLENGES

Big data is the answer for every company. It is expensive and it requires specialized human resources. There are plenty of technical challenges. Big data will always come with big security challenges. Privacy advocates are worried as more and more data is collected about people. Media and regulators are demonizing Big data and its supposed threat to privacy. Reliable tools to integrate data set in research and healthcare across institutions and countries is a daunting challenge.

CONCLUSION

Big data and data analytics will continue to grow in the coming years. Forward-looking organizations both big and small, are investing in big data and related technologies to make informed decisions, remain competitive, and improve customer experience and satisfaction. Google Cloud, Microsoft Azure, and AWS have all changed the ways they process and store bi data.

Big data is one of the most disruptive technologies in recent time. It is likely to drive improvements across all sectors. With a promising and bright future, big data can be a lucrative career to pursue. As the volume of big data continues to increase, the demand for trained data analysts and data engineers will grow.

We are in the midst of a revolution that is transforming the way we do business, marketing, accounting, research, education, construction, transportation, defense, fight wars, etc. Big data technologies will attain major heights in the years to come. More information about the future of big data can be found in the books in [14-17] and the following related journals:

- *Journal of Big Data*
- *Big Data Research*
- *Big Data & Society*
- *Government Information Quarterly*

References

[1] S. Khvoynitskaya, "The future of big data: 5 predictions from experts for 2020-2025," January 2020, https://www.itransition.com/blog/the-future-of-big-data

[2] K. Sinha, "Big data: The future has arrived," March 2016, http://www.dbtalks.com/article/big-data-future-has-come/

[3] "The future of big data in business: using data analytics to provide insight," https://online.maryville.edu/blog/future-big-data/

[4] J. bailey, "How big data impacts healthcare" https://www.chetu.com/blogs/healthcare/big-data-impact-in-healthcare.php

[5] "The 42 V's of big data and data science," https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

[6] M. N. O. Sadiku, M. Tembely, and S. M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research*, vol. 3, no. 2, 2016, pp. 106-108.

[7] P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in N. Bouchemal (ed.), *Intelligent Systems for Healthcare Management and Delivery*. IGI Global, chapter 4, 2019, pp. 72-100.

[8] J. Moorthy et al., "Big data: Prospects and challenges," *The Journal for Decision Makers*, vol. 40, no. 1, 2015, pp. 74–96. https://www.grandviewresearch.com/industry-analysis/industrial-wireless-sensor-networks-iwsn-market

[9] M. N. O. Sadiku, J. Foreman, and S. M. Musa, "Big data analytics: A primer," *International Journal of Technologies and Management Research*, vol. 5, no. 9, September 2018, pp. 44-49.

[10] C. M. M. Kotteti, M. N. O. Sadiku, and S. M. Musa, "Big data analytics," *Invention Journal of Research Technology in Engineering & Management*, vol. 2, no. 10, Oct. 2018, pp. 2455-3689.

[11] S. O'Brien, "What is big data and how will it affect the future of businesses," May 2020, https://www.ringcentral.co.uk/gb/en/blog/big-data-future-of-business/

[12] B. Marr, "17 Predictions about the future of big data everyone should read," March 2016, https://www.forbes.com/sites/bernardmarr/2016/03/15/

17-predictions-about-the-future-of-big-data-everyone-should-read/?sh=52dbc8051a32

[13] Z. Khan et al., "Towards cloud based big data analytics for smart future cities," *Journal of Cloud Computing*, vol. 4, no. 2, 2015.

[14] E. Dumbill, Planning for Big Data: A CIO's Handbook to the Changing Data Landscape. O'Reilly, 2012.

[15] S. Klous and N. Wielaard, *We are Big Data: The Future of the Information Society*. Springer, 2016.

[16] V. Mayer-Schönberger and K. Cukier, *Learning With Big Data: The Future of Education*. New York: Houghton Mifflin Harcourt Publishing Co., 2014.

[17] B .Williamson, *Big Data in Education: The Digital Future of Learning, Policy and Practice*. SAGE Publications, 2017.

ABOUT THE AUTHORS

Matthew N.O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a fellow of IEEE.

Olaniyi D. Olaleye is a project management professional. He is a Ph.D. candidate in Urban Planning and Environmental Policy at Texas Southern University, with emphasis on urbanization, climate change, renewable energy studies, and infrastructural sustainability.

Abayomi Ajayi-Majebi is a professor in the Department of Manufacturing Engineering at Central State University in Wilberforce, Ohio. In 2015 he was honored by the White House as a Champion of Change for his significant contributions to the engineering education of minority students. He is a senior member of both the Society of Manufacturing Engineers and the American Society for Quality.

Sarhan M. Musa is a professor in the Department of Electrical and Computer Engineering at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow. His areas of research interest include computational electromagnetics and computer networks.