Big Data in Engineering: A Primer

Matthew N. O. Sadiku, Guddi K. Suman, and Sarhan M. Musa

Roy G. Perry College of Engineering

Prairie View A&M University

Prairie View, TX 77446

_____

ABSTRACT

With the world witnessing a digital revolution, data is now the fuel that drives forward the 21st century. Big data is a collection of data that the traditional software options cannot handle. Big data is everywhere and the big data revolution is upon us. Big data engineering is a demanding specialization that trains engineers to understand real-time data processing and build massive big data reservoirs. This paper is a primer on the various uses of big data in engineering.

Key Words: Big Data, Data Analytics, Engineering, Big Data Engineers.

_____

1. INTRODUCTION

Our lives revolve around huge data sets. With the advent of various social media platforms and multinational companies, the generation of data has increased drastically. This data is due to Internet, high-speed networks, mobile, IoT, social data, industrial operations, etc. all across the world. There has been a rapid emergence of technology in response to the huge amounts of data generated. Collecting and handling big data is one of the emerging challenges in the 21th century.

As its name implies, big data is a structured, semi-structured, and unstructured data, which is very big, fast, and comes in many forms. Big data is a collection of complex data sets that are so intense in their volumes that traditional data processing software cannot handle them. It refers to the large volume of structured and unstructured data with the potential to be mined for information. Typical examples of big data include [1].

- A million sales transactions by an online retailer
- A million hosted phone calls by a telecommunications provider
- A sensor that produces 50 megabytes of data every two nanoseconds

Big data has become the mainstream technology across various industries. As shown in Figure 1, businesses from all types of industries have greatly benefited by adopting big data solutions [2]. This is true of healthcare, energy, finance, telecommunication, marketing, and sports. To understand the role of big data in engineering, it is expedient to understand what constitutes big data.

2. 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 newer technology that can help many industries. The three main sources of big data are machines, people, and companies. Big data can be described with 42 Vs [3]. The first five Vs are volume, velocity, variety, veracity, and value [4].

- *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 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 have been added [5].

- *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 2 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) [6].

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, airline, travel, restaurants, automobile, sports, agriculture, manufacturing, and hospitality industries.

3. BIG DATA ANALYTICS

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. Once the big data is ready for analysis, we use advanced software programs such as Hadoop, MapReduce, MongoDB, and NoSQL databases [7]. 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 [8].

Big data analytics is capable of processing massive amounts of dirty data and extract the gold information from it. It has the potential to predict performance, upcoming changes, and market trends with unprecedented accuracy. To maximize the value of data and make it useful to end users, engineers need to incorporate analytics into their daily tasks. To handle big data, various frameworks like Hadoop, Spark, Cassandra, and Apache Storm are used.

4. BIG DATA IN ENGINEERING

Big data engineering is a specialization that requires working with big data. It is a field associated with helping organizations capture data from various sources, process, and make it ready for further use. It involves developing, maintaining, testing, and evaluating big data solutions. By using big data engineering skills, it is possible to bring together massive amount of data and process it so that company can make faster decisions. Big data engineering enables streaming analytics for real-time diagnostics and efficient traffic management.

As data keeps increasing and becomes more complex, the role of big data engineering will grow in importance. Opportunities within big data engineering are many. Companies desire to get the maximum value out of their data, and they need data pipeline builders as they migrate to the cloud. For example, the engineering company Raytheon Corp monitors its assembly operations

down to the turn of a screw, as shown in Figure 3 [9]. If a screw in a missile fails to complete its full count of turns, Raytheon will know about it immediately and be able to take corrective action.

5. BIG DATA ENGINEERS

Big data engineers play a major, unique role in enterprises that are facing a data deluge. They are tasked with designing big data solutions and building massive big data reservoirs that can store and process sweeping data streams. They sort through the massive data to find relevant sets useful in predicting behavior [10]. They know how to apply technologies to handle the challenge of dealing with petabyte or even exabytes of data on a daily basis.

Big data engineers have a spectrum of responsibilities. They may be responsible for the following tasks [11]:

- Build highly scalable data management systems from the design phase to completion
- Design and develop data pipelines
- Design top-tier algorithms, predictive models, and prototypes
- Create data set processes to be used for data modeling, mining, and production
- Develop custom analytics apps and other kinds of software
- Ensure that data systems meet specific requirements
- Oversee disaster recovery preparations
- Research improvements to data quality, reliability, and efficiency
- Look for data acquisition opportunities as well as new uses for existing data and tools
- Turn messy, chaotic information into clean, accurate, and actionable data
- Develop, construct, test, and maintain frameworks like large-scale data processing systems and databases
- Collaborate with other teams such as data analysts, data scientists, data architects, modelers, and IT specialists
- Have sound oral and written communication skills

Figure 4 summarizes the roles of a big data engineer [12]. Nearly all fields need big data engineers, as they provide organizations with analyzes and insights that are useful in business, finance, government, healthcare, engineering, science, telecommunications, manufacturing, and other industries. Many enterprises in numerous sectors are actively collecting big data and are facing a shortage of the necessary expertise.

To become a big data engineer requires showing an interest in computer science, software engineering, computer engineering, applied mathematics or related field. Big data engineer should hold at least a bachelor's degree in one of those fields. They must also have considerable knowledge of Java and high-level programming languages such as C/C++, Python, R, SQL and Scala. To advance their career, they should pursue vendor-specific certifications such as Oracle, Microsoft, IBM, Google, and Cloudera.

6. APPLICATIONS OF BIG DATA IN ENGINEERING

Big data engineering finds application across the spectrum as data analytics in any industry where huge data challenges have emerged. These industries include healthcare, architecture, agriculture, finance, marketing, engineering, automotive industry, and construction. Big data has changed many professions. Its use is emerging and transforming all branches of engineering including chemical engineering, software engineering, electrical engineering, mechanical engineering, and construction engineering. The following are typical engineering applications of big data.

- *Software engineering:* This engineering discipline may be regarded as the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software. It is the application of engineering to software. Software engineering practices can be used to improve the production of big data systems. Big data helps to improve software engineering processes, which can in turn develop better big data projects [13].

- *Electrical Engineering:* The electrical engineering field is one of the fields that have been forced to evolve in the big data world. The value of engineering data largely depends on the insight that can be gained from it. One should develop the data skills needed to thrive in the field. Electrical engineer is responsible for inventing and developing new electrical systems and for solving problems relating to the electrical field. Big data is a new technology which is helping electrical engineers [14].

- *Chemical Engineering*: Like other industries, chemical companies are using big data to innovate and improve operations. Chemical engineers, whether in industry or in academia, are being asked to manipulate, transform, and analyze complex data sets. Big data tools are already having a great impact in chemical engineering. Big data analytics are now being used by leading chemical companies to recognize patterns, to enhance manufacturing processes, improve marketing, and support innovation [15, 16].

- *Mechanical Engineering:* A large portion of mechanical engineering focuses on domains such as product design and development, manufacturing, material, and energy, which are likely to benefit from big data. Mechanical engineers may improve reliability of the existing manufacturing processes, predict emerging equipment failures, design next generation of manufacturing equipment, and invent new technologies based on the historical data [17].

- *Construction Engineering*: Big data is impacting how we design and construct buildings. Big data in construction offers a way to improve existing processes and make informed decisions about future projects. Big data is changing construction industry in the following ways [18]: (1) improved health and safety, (2) better cost control and prediction, (3) enhanced operational efficiencies, (4) informed asset management.

- *Manufacturing Engineering:* Data has long been the lifeblood of manufacturing. The big data revolution is rapidly changing the manufacturing industry. Manufacturing data is key to conducting data-driven manufacturing. Most manufacturing data is standardized which is supported by various industrial vendors and associations. Big data in manufacturing is generated from the production and process, from machines such as pumps, motors, compressors, meters, sensors, controllers or conveyers. The big data in manufacturing does not mean anything without analysis. Six key drivers of big data applications in manufacturing have been identified as system integration, data, prediction, sustainability, resource sharing, and hardware [19].

Other applications include civil engineering, highway engineering, and transportation engineering.

7. BENEFITS AND CHALLENGES

The incorporation of big data in engineering has already yielded multiple benefits and helped engineering industries gain a competitive advantage. Big data is being leveraged to increase margins, optimize production, enable innovations, and produce highest returns. It is commonly accepted that big data is a source of competitive advantage. Big data has the ability to transform and change the game for businesses of all sizes and types. Mining big data can offer benefits in design, such as design improvement, saving design time and costs, fast response to market, matching customers' needs, and gaining customer satisfaction. Big data should be measured by the impact it brings into every sphere of our lives.

However, countless challenges come with leveraging big data in engineering. Big data is inherently complicated at every turn. Implementing big data systems is complex, with numerous dynamic components. Any of these components can fail and finding the root cause of the failure can be challenging. The rise of big data also raises fundamental challenges in privacy, security, and data ownership. In spite of these challenges, companies that fail to adopt the new norm of data-driven operations could lose traction in the market and become irrelevant in the future.

8. CONCLUSION

The growth in big data has been phenomenal. Big data as a research discipline is still evolving. As more enterprises begin to appreciate the powerful insights and value of big data, adoption will increase. It is up to each company to figure out how it can best turn insights into impactful decisions.

Big data engineer is a professional who is eager to join a fast-growing field that aligns with their interest for computer science, statistics, or computer engineering. Big data engineers are coveted and highly in demand; their salary reflects this. More information on big data in engineering can be found in the books in [20-22] and related journal: *Journal of Big Data.*

REFERENCES

[1] N. Campana, "What does a big data engineer do?" July 2020,
https://www.freelancermap.com/blog/what-does-big-data-engineer-do/

[2] "Real-time applications of big data that drive various industries,"
https://elysiumpro.in/applications-big-data-drive-industries/

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

[4] 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.

[5] 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.

[6] 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

[7] 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.

[8] 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.

[9] L. Wang and C. A. Alexander, "Big data in design and manufacturing engineering," *American Journal of Engineering and Applied Sciences*, vol. 8, nol. 2, 2015, pp. 223-232.

[10] "5 Must-have data engineering skills to land big data engineer job in 2019," December 2019, https://analyticsindiamag.com/5-must-have-data-engineering-skills-to-land-big-data-engineer-job-in-2019/

[11] "How to become a big data engineer," https://online.maryville.edu/online-masters-degrees/business-data-analytics/careers/big-data-engineer/

[12] "Big data engineer: Role, responsibilities, and job description," https://www.altexsoft.com/blog/big-data-engineer/

[13] T. Arndt, "Big data and software engineering: Prospects for mutual enrichment," *Iran Journal of Computer Science*, December 2017.

[14] S. Parker, "Navigating your career in electrical engineering in the big data era," https://www.smartdatacollective.com/navigating-career-in-electrical-engineering-in-big-data-era/

[15] M. N. O. Sadiku, S. M. Musa, and O. M. Musa, "Big data in chemical engineering," *International Journal of Advances in Scientific Research and Engineering,* vol. 5, no. 6, June 2019, pp. 100-102.

[16] M. N. O. Sadiku, S. M. Musa, and O. S. Musa, "Big data in the chemical industry," *International Journal of Advances in Scientific Research and Engineering*, vol. 3, no. 10, Nov. 2017, pp. 20-23.

[17] "Big data in mechanical engineering," April 2015, https://www.asme.org/topics-resources/me-today/big-data-in-mechanical-engineering

[18] "How big data is changing engineering & construction," March 20, 2020, https://blog.allplan.com/en/how-big-data-is-changing-engineering-and-construction

[19] M. N. O. Sadiku, J. T. Ashaolu, A. Ajayi-Majebi, S. M. Musa, "Big data in manufacturing," *International Journal of Scientific Advances*, vol. 2, no. 1, Jan.-Feb. 2021, pp. 63-66.

[20] S. S. Roy et al. (eds.), *Big Data in Engineering Applications*. Springer 2018.

[21] M. Atzmueller, S. Oussena, and T. Roth-Berghofer (eds.), Enterprise *Big Data Engineering, Analytics, and Management.* Hershey, PA: IGI Global, 2016.

[22] F. Calisir, E. Cevikcan, and H. C. Akdag (eds.), *Industrial Engineering in the Big Data Era.* Springer 2019.

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.

Guddi K. Suman is currently working towards a PhD in Electrical and Computer Engineering at Prairie View A&M University, Prairie View, TX. Her areas of research interest include space radiation effects on electronics, nanomaterials, characterization of nanomaterials and semiconductors, thin film nanofabrication, nanosensors, experimentation and computer simulation to explore new materials and Artificial intelligence/Data science in Material science. She worked as an intern at Pacific Northwest National Laboratory, Richland, WA, in 2021. She is a member of IEEE.

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.

.

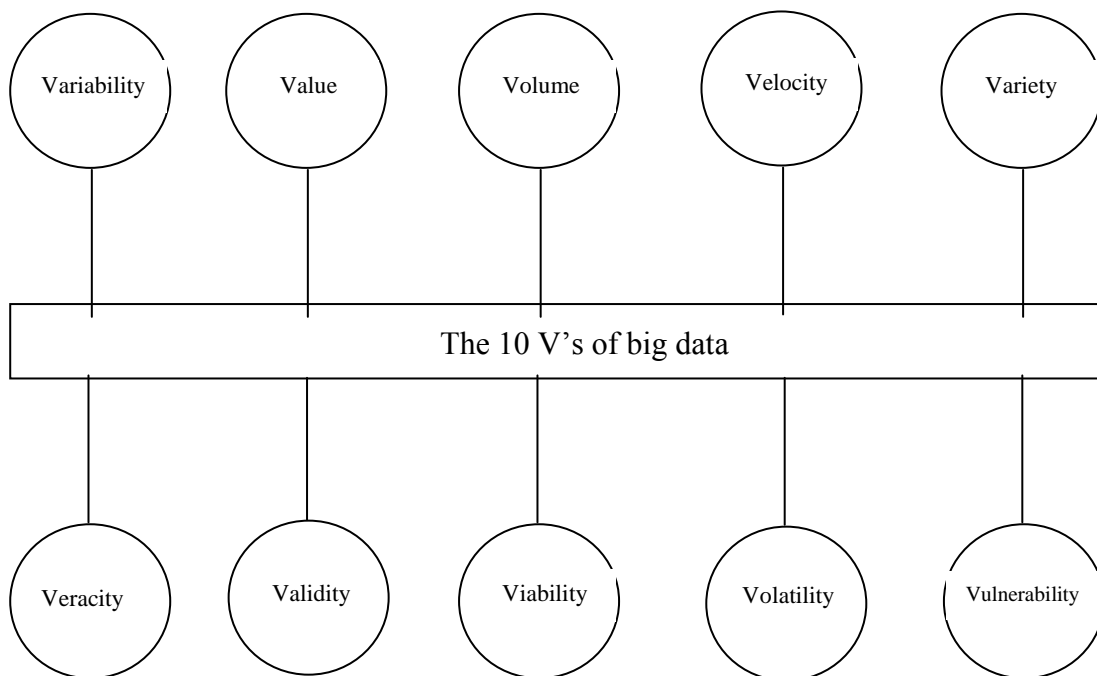**Figure 1 Real-time application of big data [2].**

Figure 2 The 10V's of big data.

Figure 3 Big data at a missile plant [9].

Figure 4 The role of a big data engineer [12].