Resilience or Resistance? Outreach of Big Data in the Digital Age

Bibhu Dash, Pawankumar Sharma, Swati Swayamsiddha

Abstract— **Big data has become a ubiquitous part of modern business due to the development of technology and influence of social media. In dynamic business environments, managing this much big data is neither very seductive nor particularly simple. These data are consumer-driven and consumer-focused. That makes these data special. As data storage prices decrease as a result of cloud growth, data volumes are growing across the board for the company. Every firm is putting a greater emphasis on data-driven decision making, which forces them to concentrate on information extraction and data management. The article explores whether the explosion of big data and its applications is advantageous for the digital age, or on the contrary, it serves as a distraction from more pressing issues and confuses us more than it aids. We also aim on the intellectual underpinnings of big data and how it affects management and organizational performance.**

***Index Terms*—Big data, Data Science, Data-driven decision making, 3C Factors, Organizational Performance.**

I. INTRODUCTION

Digitalization is at its pinnacle around the world, and it has an impact on how we store and manage big data for greater outcomes [1]. The definition of big data is often ambiguous and not widely accepted. The phrase is often used in conjunction with data mining and business analytics to find links and associations in vast volumes of transaction data. massive data in the context of data science typically refers to structural or non-structural datasets that have gotten so massive that working with them using conventional database management systems and analytical techniques is difficult. They are datasets that are too large to be captured, stored, managed, and processed within a reasonable amount of time by often employed software tools and storage systems [2].

As reported in Statistica, big data sizes range from a few terabytes (TB) to many zettabytes (ZB) of data in a single dataset, and they are constantly growing (see Figure 1) [3]. As a result, gathering, storing, searching, sharing, analyzing, and visualizing massive data can be challenging. Enterprises are now examining vast amounts of extremely detailed data to find trends and information previously thought to be impossible to obtain. Examples of big data are web traffic/log data, GPS data, social network data, Satellite data, Sensor or Smart device data etc.

The bottom line is that an enterprise's performance, profitability, and security are all compromised by having too much data because it creates too much noise. We are more likely to face management, security, and privacy challenges the more data we have. Additionally, our high signal-to-noise ratio causes business decisions and IT operations to be complicated. In fact, negative signals might cause teams to become so preoccupied that they fail to notice important signals. That indicates the ideal amount of data that organizations should have according to the 3C factor analysis: correct data, complete data, and concise data [4]. The detail analysis has been conducted later in this paper.

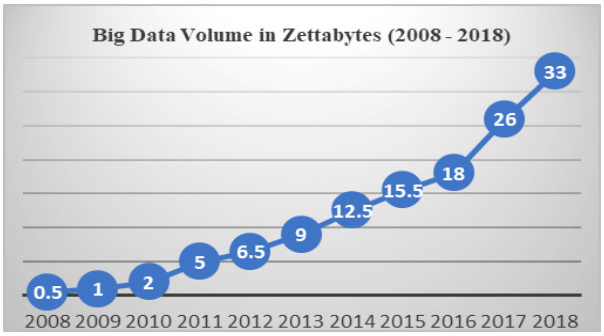

Figure 1. Big data growth worldwide from 2008 - 2018

II. 5V'S OF BIG DATA

Big data is characterized by the five V's, or volume, variety, value, veracity and velocity (See Figure 2). The amount and scope of the data are measured by its volume. Velocity is the pace of change or the frequency of creation of data. The diverse formats and types of data, as well as the various uses and methods of data analysis, are all examples of variety.

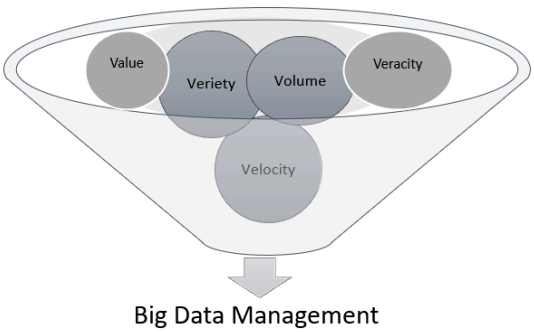

Figure 2. 5V's of Big Data

A. Volume

In a data-driven organization, volume alludes to the quantity of data produced by websites, portals, and online applications. Those insights' quality and quantity determine how valuable they are. The precise method by which an organization obtains value from big data depends on its operational needs and procedures [4, 5]. Value is generally

influenced by the other Vs.

B. *Velocity*

The term velocity describes the rate at which data is generated, and it has significantly increased along with both internet speeds, technology and user numbers. In essence, velocity is the frequency of data generation or data delivery. Streaming data, which is gathered in real-time from websites (social media), is the cutting edge of big data [5].

C. *Variety*

All the structured, semi-structured and unstructured data that could be produced by either people or machines is referred to as variety in big data. Whatever information a spreadsheet might hold is considered structured data. It is simple to catalog and to calculate summary statistics for. Text messages, tweets, images, videos, emails, voicemails, handwritten notes, ECG readings, and audio recordings are examples of unstructured data. Unstructured data typically doesn't fit into common data models, making it a poor fit for relational databases in general. Semi-structured data is data that has related information, such metadata, but has not been arranged into a specific repository [5, 6].

D. *Veracity*

Veracity reflects of the accuracy and superiority of the data. The gathered information can be incomplete, erroneous, or unable to offer any useful, insightful information. Veracity, in general, refers to the degree of confidence in the data that has been gathered [5]. Data can occasionally become disorganized and challenging to use. If the data is incomplete, a big volume of data may produce more confusion than insights. For instance, in the medical industry, if information about the medications a patient is taking is lacking, the patient's life may be in danger. Big data's fourth character reflects the data's quality and any hidden insights in it.

E. *Value*

The fifth and final V in the big data acronym is value. This refers to the advantages that big data can provide, and it has a direct impact on what organizations can do with the information they collect [6]. The ability to extract value from big data is required since the value of big data is heavily dependent on the insights that can be gleaned from it.

III. Pros and Cons of Too much Big Dada

Too much Big Data in the digital age have both advantages and disadvantages. Too much data can mislead and sometimes helpful when it comes data analysis and information extraction. Some of the pros and cons are discussed in detail below.

A. *Pros of having too much Big Data*

- *In-Depth Understanding*: Large amounts of data can provide in-depth understanding of numerous elements of business, society, and science. It improves understanding of trends, patterns, and correlations (see Figure 3) [7].
- *Better Decision-Making*: Big Data can be utilized to help people make better educated, data-driven decisions. This is especially useful in fields like as banking, healthcare, and marketing.
- *Improved Personalization*: Using massive volumes of data, businesses may adapt their products and services to individual interests, so improving the customer experience.
- *Productivity and Efficiency*: Big Data may be utilized to optimize processes, cut costs, and increase overall efficiency. Predictive maintenance, for example, can reduce equipment downtime in the manufacturing industry.
- *Predictive Analytics*: Predictive modeling is used to foresee future trends, demands, and concerns.
- *Scientific Research*: Researchers can study large datasets to acquire a better knowledge of numerous phenomena ranging from climate change to genetics.

Figure 3. Business Advantages of Big Data

B. *Cons of having Too Much Big Data*

- *Data Overburden*: Managing and analyzing massive amounts of data can be exhausting, resulting in information overload, IT operation cost increase and difficulty deriving valuable insights.
- *Privacy problems*: Collecting and storing large amounts of personal data presents serious privacy problems. Individuals and organizations can suffer serious consequences because of data breaches and misuse [7].
- *Data Security*: The more data you have, the more secure it must be. Protecting sensitive data from hackers and breaches is becoming more difficult.
- *Costs*: Big Data storage, processing, and management can be costly. To handle it efficiently, organizations require a strong infrastructure as well as trained staff.
- *Data Value and Accuracy*: Not all data is valuable or correct. Cleaning and assuring data quality takes effort, and poor data quality can result in inaccurate insights.
- *Regulatory Compliance*: Many regions have enacted data protection legislation (for example, the GDPR in Europe) that set rigorous rules on how data is gathered, stored, and utilized. Compliance with these regulations might be difficult.
- *Bias and Fairness*: There is a possibility of hidden biases in huge datasets, which can lead to unfair or

discriminating conclusions, particularly in machine learning applications.

- *Intricacy*: Big Data needs knowledge and infrastructure, making it difficult for smaller firms and organizations to realize its full potential.

IV. What's to be Done?

What can be done when we have too much Big Data? Many firms are now concentrating on creating a new technology layer that ingests all of their data and transforms it into pertinent information using a set of data processing algorithms that truly moves the needle for the business. Data-generating and data-gathering monitoring and log management solutions look to be able to help, but historically they haven't. Any monitoring or log management tool faces the challenge of keeping data-source independence. How can an input-only tool maintain performance across all inputs without "picking pieces"?

In simple terms, the problem, not the solution, is the data generators. They examine your applications and infrastructure, producing a ton of fresh data. You should think about a special new system that takes in all of this extremely varied data and then filters, normalizes, enriches, and correlates it. Such a framework is impartial. It uses all available data points without bias and is more concerned with creating business insights than it is with creating data.

Bringing down silos and fusing together various data sources will help you generate actionable insights. Those who accomplish this will be able to shorten outages, prevent downtime more proactively, and enhance the performance and availability of their services.

V. Innovations in Big Data Analytics

It requires special handling with the aid of new data science methods, tools and technologies to extract meaning from that as data volume, velocity, and variety increase. The need for quicker and more effective methods of processing this data has developed rapidly along with technological development and the rise in data volumes, as previously noted. Big data is no longer sufficient on its own to enable timely and effective decision-making [8].

Big Data cannot be simply analyzed using conventional data management and analysis methods and infrastructures, as we have already explained. As a result, new big data analytics-specific tools and techniques as well as the necessary architectures for storing and handling such data are needed (see Figure 4). Big data's growth has an impact on every aspect of data processing, including the data's generation, collecting, analysis, and presentation, as well as the conclusions that are ultimately drawn from it.

As the smart city concepts are evolving across the globe, the data volume and data management procedures are becoming complex and tedious to work with [11]. Future growth in the volume of big data will put more pressure on the quality and integrity of data and create resistance to the process of extracting insights.

Figure 4. Tools and Technologies for Big Data Processing

Big data analytics' primary activities can be seen in four distinct contexts:

1) Big data distribution and warehousing
2) Bigdata storing solution.
3) computational platforms for big data
4) Big data evaluation, analysis, and visualization

In big data-driven businesses, a framework like this can be used for knowledge discovery and well-informed decision-making [10].

VI. Big Data Privacy Risks

Big data security and user privacy have emerged as key concerns for the digital age. Monitoring and studying user behavior is becoming more and more unsustainable, and current methods that track, monitor, or fingerprint users will become more and more privacy vulnerable [12, 13]. The difficulty for the majority of organizations is that the definition of what data can be deemed personal is a fluid and ever-evolving concept. It is no longer limited to only email addresses or unmistakable personal identifiers, but also to items that can be profiled by combining them with other datasets. Regulators are increasingly targeting the collection and use of IP addresses, a common result of digital advertising operations.

Data management is very important and critical step. Any mismanagement may affect to company's brand value and financials. Any company that doesn't filter and manage its collected, purchased, and borrowed datasets runs the danger of incurring fines, reputational harm, and other negative effects as regulators expand their privacy requirements. The most recent significant example is the 1.2 billion Euro fine levied against Facebook.

VII. Bigdata and Digital Carbon Foot Prrint

The dangers of too much data incoming and handling means impact to processing cost and misinformation [12]. Though cloud is changing the way we store, manage and process data, the reality is that cloud is exemplified with thousands of data centers around the world with 24*7 availability. There are more than few millions of these data centers on the earth, and even one of them can consume as much electricity as a medium-sized town. Notably, they are the main source of carbon emissions in the world of IT.

Too much processed data is neither beneficial for

environment nor organizations. The expense of managing human and non-human resources is too high to gauge the true goal of data cultivation because the generation, storing, and processing of big data are all resource-constrained endeavors. Organizations frequently struggle to understand the underlying value of data and its benefits because only 5% of genuine big data has beneficial effects for company [14].

VIII. Solutions to Manage Too Much Big Data

The need of the hour is for solutions to control the production, storage, and processing of big data. Prioritizing quality data above huge data in a short amount of time is the answer. Signal and noise are separated by quality and time. To comprehend corporate demands and necessities, stronger data governance and audit standards are necessary (see Figure 5) [15, 16].

Organizations or individuals should consider the following points:

1) Smart Data: Data must be contextualized. Labels are necessary to help it make sense. Data points that lack context or meaning are referred to as naked numbers. Digital advertising is rife with naked numbers, which degrade the environment. Frankly, a lot of naming standards and labeling systems are disorganized [15]. Successful businesses follow tight naming guidelines, and they often encourage their IT partners to do the same.
2) Clean Data: A guarantee that the dataset you are optimizing against is totally free of signals based on bot activity is required given that half of the digital ecosystem is powered by bots and progressively infused with AI.
3) Purposeful Data: Data must serve a purpose in addition to being accurate and comprehensive. A dataset's purpose is mostly determined by who is paying for it at the most fundamental level. However, purposeful measures are the best kind of policies [16].

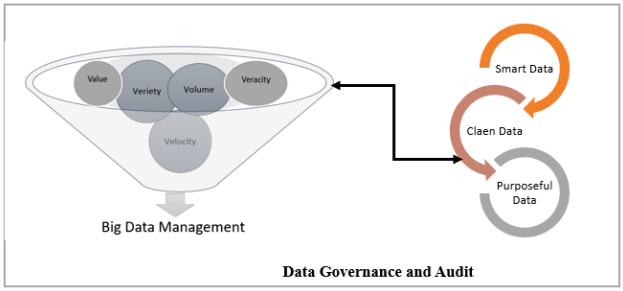

Figure 5. Solutions for Too Much Big Data handling

IX. Conclusion

Data is the new oil, and it is crucial for all kinds of contemporary businesses. Data offers us the advantage to better understand our consumers, processes, and products. Even though it offers us the resiliency to thrive and get insights, having too much data makes it difficult to identify the precise issue that needs to be addressed. Therefore, a path-breaking solution for businesses to understand their big data demands and how to manage those effectively in the future is being created by a suitable quantity of data combined with the appropriate rules and governance standards. In addition to data security as its primary focus, the upcoming study will concentrate more on data governance and data storing strategies.

Acknowledgment

We would like to thank our University of the Cumberland's professor, Dr. Archie Addo and my colleague Meraj Ansari for their valuable time to review and give us feedback to complete this paper.

References

[1] Dash, B., & Gatharia, J. (2015). Impact of Digital Transformation on Organizational Behaviors.
[2] Coveney, P. V., Dougherty, E. R., & Highfield, R. R. (2016). Big data need big theory too. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2080), 20160153.
[3] Simon, P. (2013). *Too big to ignore: the business case for big data* (Vol. 72). John Wiley & Sons.
[4] Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In 2013 Sixth international conference on contemporary computing (IC3) (pp. 404-409). IEEE.
[5] Andrejevic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. Routledge.
[6] George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of management Journal*, *57*(2), 321-326.
[7] Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... & Zwitter, A. (2019). Will democracy survive big data and artificial intelligence?. *Towards digital enlightenment: Essays on the dark and light sides of the digital revolution*, 73-98.
[8] Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. MIT press.
[9] Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, *93*, 67-78.
[10] Dash, B., Sharma, P., & Ansari, M. F. (2018). A Data-Driven AI Framework to Improve Urban Mobility and Traffic Congestion in Smart Cities.
[11] Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 2053951716631130.
[12] Karakaya, Z. (2017, October). Software engineering issues in big data application development. In *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 851-855). IEEE
[13] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, *30*(4), 431-448.
[14] Janssen, M., Van Der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of business research*, *70*, 338-345.
[15] Rawat, D. B., Doku, R., & Garuba, M. (2019). Cybersecurity in big data era: From securing big data to data-driven security. *IEEE Transactions on Services Computing*, *14*(6), 2055-2072.
[16] Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big data governance frameworks. *Procedia computer science*, *141*, 271-277.

Bibhu Dash is a Ph.D. student in School of Computer and Information Sciences at University of the Cumberlands, KY. His research interests are in Big data, data security and AI. He has more than 12 years of IT experience in database design, data security and big data management.

Pawankumar Sharma is a Ph.D. student in School of Computer and Information Sciences at University of the Cumberlands, KY. His research interests are in Cloud, Cyber security and Customer analytics.

Dr. Swati Swayamsiddha is an Assistant Professor at School of Electronics at KIIT University, Odisha India. She has more than 10 years of academia experience and her research interests are 5G, Cognitive IoT, and Big data mining.