Big Data and Analytics:
Prospects, Challenges, and the Way Forward

B

Md. Toriqul Islam
https://orcid.org/0000-0003-3004-770X
Transparency International Bangladesh (TIB), Bangladesh

Borhan Uddin Khan
University of Dhaka, Bangladesh

1. INTRODUCTION

Appearing as a buzzword, big data received increasing popularity in the past few decades. The world community has become enthusiastic about big data and advanced big data analytics not solely due to the sheer magnitude of data involved but because of the substantial potential for impacts (McKinsey & Company, 2016). With the discovery of the Internet, the Internet of Things (IoT), other associated emerging technologies, and social networking sites, digital data has exponentially grown in the last few decades. This large volume of data sets requires real-time analysis and processing capacity for getting better insights into diverse businesses in the current information economy. Eventually, this data-driven ecosystem begets the notion of big data.

Big data refers to large-scale, complex, unstructured, semi-structured, or structured datasets that require unique computational techniques for processing in place of traditional ones. Gartner defines big data as ‘high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing, which enable enhanced insight, decision making, and process automation’ (Gartner, 2023). Big data brings both huge opportunities and challenges for numerous industries in the current data-driven economy. According to recent research, the total market value of big data is projected to exceed US $655 billion by 2029, compared to approximately $241 billion in 2021 (Taylor, 2022). Big data research has gained significant importance in the Fourth Industrial Revolution (4IR/Industry 4.0) owing to its substantial role and impact in driving transformative social change.

This chapter provides a succinct overview of big data and analytics, including their usage, advantages, disadvantages, applications, prospects, challenges, general recommendations, future research directions, and concluding remarks. Section two presents the background, focus, and research method. Section three offers a brief conceptualization of big data and analytics. Section four discusses how big data works, while section five outlines the advantages and disadvantages. Section six highlights the applications of big data, and section seven unveils its prospects. Section eight discusses the challenges associated with big data. Section nine proposes the way forward, and section ten outlines future research directions. Finally, section eleven concludes the chapter with a brief summary.

DOI: 10.4018/978-1-6684-7366-5.ch048

2. BACKGROUND, FOCUS AND RESEARCH METHOD

The development trajectory of human civilization has witnessed radical transformation through several industrial revolutions. Throughout history, beginning with the discovery of steam engines, followed by the emergence of fossil fuel-based engines, and the evolution of integrated business models, human society has now entered the information age. In the evolving information age, a significant portion of world affairs relies on computing power and wider applications of Information and Communications Technology (ICT). Presently, approximately 40% of the global population is connected online, over 1.5 billion websites are operational, Google handles an average of more than 40,000 search queries per second, and nearly 650 million Tweets are posted on Twitter (Internet Live Stats, 2023; Bulao, 2023). These statistics undeniably highlight that we are living in a data-driven world that is constantly evolving, offering abundant opportunities but also posing substantial challenges.

Big data has immense potential for various industries by producing precise insights about diverse products and services. The utilization of data-driven technologies holds promising prospects for enhancing productivity, improving trade intelligence, addressing societal challenges, fostering research and innovation, reducing costs, and expediting performance across various fields. Big data opens up a new horizon for innovation, competition, and productivity (Manyika et al., 2011).

In essence, big data appears with a plethora of promises while entailing numerous challenges. Specifically, big data may entail huge ethical, legal, social, practical, and technical challenges. This chapter would focus on the technical, practical, and legal challenges of big data and big data analytics from a regulatory perspective. For data collection and analysis, this chapter adopted doctrinal and qualitative legal research methodologies. The doctrinal legal research method has been adopted due to the nature and context of the current work and the suitability of the doctrinal approach over other research methodologies.

3. BIG DATA AND BIG DATA ANALYTICS

The central focus of this chapter shall be on the usages and applications of big data and analytics, including their prospects, challenge, and way forward. To begin with, it is essential to provide an overview of some conceptual features. Thus, this section aims to conceptualize two fundamental notions: big data and big data analytics.

3.1. Big Data

Although the term 'data' is deeply rooted in the long past, there are differences in opinions regarding the first use of the phrase 'big data'. Some authors argue that the phrase 'big data' has been coined by John R. Mashey in the mid-1990s (Diebold, 2019). Others contended that the concept of 'big data' was introduced in 1997 by NASA scientists to highlight the challenges and limitations of mainframe computers in analyzing and storing large amounts of data (Austin & Kusumoto, 2016; Gangadharan, 2013). Another group of scholars argues that the term first appeared in the field of econometrics and statistics in the early 2000s, referring to the vast amount of potentially relevant public data and the latest advancements in data collection and storage methods (Favaretto et al., 2020).

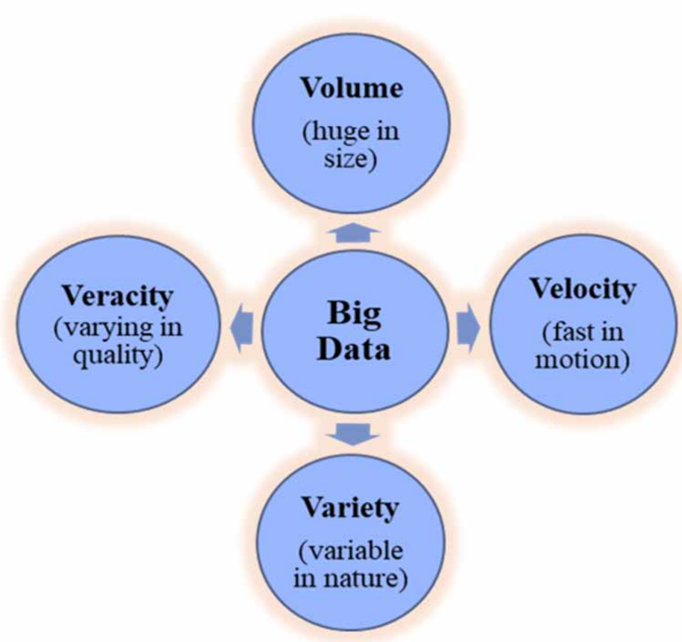
Doug Laney explained big data in 2001 using three characteristic features, e.g., volume, variety, and velocity. To him, big data refers to a dataset that is huge in size, variable in nature (as comprises various types of data, unstructured, semi-structured, or structured, and data of numerous forms or formats, e.g., text, image, video, doc, pdf, graphics used in social media, and wearable devices), and speedy in motion (as contains data of different paces or speeds) but completes processing or analyzing data in or near real-time (Kitchin & McArdle, 2016).

The term 'big data' lacks a precise and universally accepted definition due to its varied interpretations. Some definitions are overly broad, rendering them ambiguous, while others are too concise, which results in a lack of essential details. For instance, some researchers define big data as data with high volume, velocity, and variety that surpasses the processing capabilities of traditional data processing technologies (Yoganingrum, 2022). While this definition is well-articulated, it may be lacking in certain aspects. For instance, it fails to clarify how big data differs from other types of data, and it does not address whether big data processing technologies should be included in the definition of big data or not.

There are several broad definitions of big data, which may lack precision due to the unnecessary widening of its scope by incorporating additional elements. For example, some authors define big data based on at least one of the ten defining characteristics, including big volume, big value, big variety, big velocity, big veracity, big market, big analytics, big infrastructure, big intelligence, and big service (Narongou & Sun, 2022). With this definition, the authors may aim to encompass a wide range of aspects that may fall under the purview of big data. However, this approach may result in a lack of precision, as it may include elements that may not necessarily be relevant to the concept of big data.

Although the attempt to provide a definition using the aforementioned approach may be commendable, it lacks the precision criteria that are typically expected in academic definitions. Definitions are expected to be clear, formal, and exact statements that accurately convey the meaning of a given word, term, phrase, or concept (Dictionary.Com, 2023). Therefore, while the approach may aim to encompass various aspects of big data, it falls short in terms of precision, as it may not provide a concise and exact definition of the concept. Numerous scholars in the field utilize four key structural dimensions of big data, commonly referred to as the '4 Vs', to characterize the phenomenon. These dimensions comprise volume, velocity, variety, and veracity, as depicted in the following figure.

Figure 1. Four Structural 'vs' of Big Data



Big data encompasses a rapidly growing vast amount of digital data characterized by its volume, velocity, variety, and veracity (Weitzenboeck et al., 2022). In essence, big data entails massive datasets that exceed the capacity of conventional computer programs or software to analyze, capture, manage, and store effectively. The four dimensions of volume, velocity, variety, and veracity collectively characterize big data while indicating large data size, high generation speed, diverse data types, and varying data quality levels. It is worthy of note that big data does not refer to the synonym of a 'lot of data' only but includes more than that. The 'lots of data' are indispensably a prerequisite to constitute big data but referring to 'lots of data' with the term 'big data' and stopping thereby is a mistake. It is a method of processing raw data received from various scattered data sources, storing data for analysis, and interpreting it in a unique way to extract meaning. Moreover, big data is not a single technology but includes all technologies working together to acquire, store, and utilize data under the domain of the 'big data' ecosystem (IBM Developer, n.d.).

To put it simply, big data refers to a massive amount of structured, semi-structured, and unstructured data that require unique methods, techniques, or tools for analysis or processing. Big data and small or ordinary data can be distinguished in terms of implications, nature, and processing methods. Compared with small data, big data entails massive implications that contain various sorts of data while requiring unique tools or techniques for data processing or analysis, but not vice versa. Again, to compare with traditional or small or ordinary data, big data contains huge data in volume, variety in nature, and velocity in motion. The technology that processes big data is also regarded as big data (Chen & Kharabsheh, 2019).

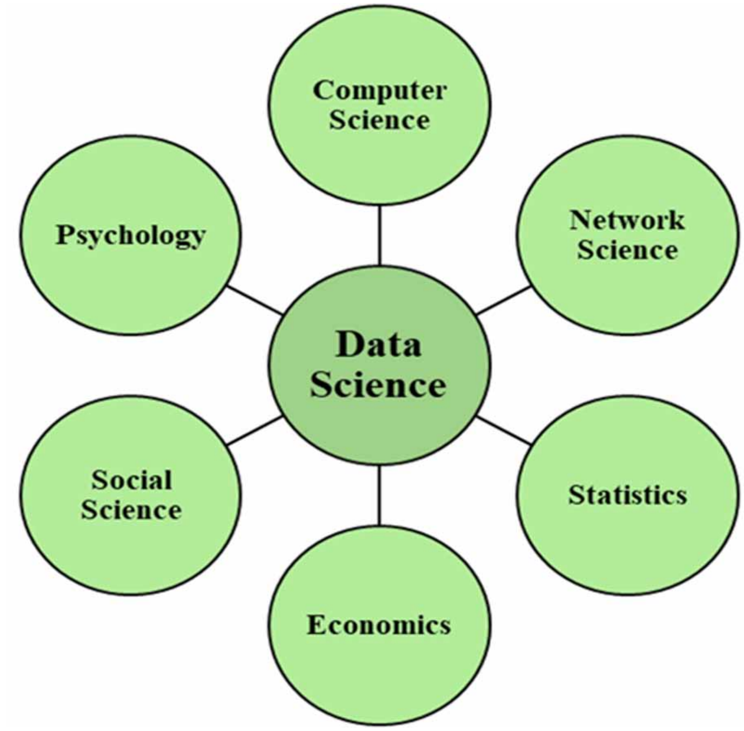
3.2. Big Data Analytics

Data is a highly precious resource in today's data-driven world, often referred to as 'gold' in the global information economy. As such, individuals from all walks of life must have a basic understanding of data science, big data, and big data analytics, which are interconnected phenomena. Data science is a field of study that encompasses various aspects of data usage, including analyzing, compiling, conditioning, extracting, identifying, interpreting, modeling, processing, reporting, visualizing, and presenting data, regardless of its size. It covers a wide range of procedures and techniques used to work with data in a holistic manner.

The concept of 'data' has evolved as a distinct field of study within data science and encompasses a wide range of data processing activities. In parallel, 'big data' is an integral component of data science, representing a subset of data characterized by its massive volume, velocity, and variety. As an umbrella term, data science encompasses various disciplines within its scope and includes computer science, network science, statistics, economics, social science, and psychology (Bhadani & Jothimani, 2016).

Similar to big data, the concept of 'big data analytics' has garnered substantial attention in both industry and academia, as noted by Poola et al. (2017). According to Noel et al. (2016), it entails a continuous process of discovery, while Loshin (2013) posits that it has the potential to generate value through the collection, integration, and analysis of voluminous and unstructured datasets. As a result, big data analytics has become an indispensable tool in various domains of life, as highlighted by Kannan et al. (2016). It involves the use of tools, techniques, technologies, or analytics to uncover hidden patterns, identify correlations, predict market trends and customer preferences, and support decision-making by analyzing large datasets. Big data analytics also involves the use of advanced analytic technologies to analyze large and diverse datasets arising from various sources, which can vary greatly in size. (IBM, 2023).

Figure 2. Integrated Courses of Data Science



In essence, big data analytics is the process of collecting, processing and analyzing large sets of structured and unstructured data to uncover hidden patterns, trends, and insights that can be used to inform business decisions and improve organizational performance. This process involves using advanced technologies such as machine learning, statistical modelling, and data visualization to extract meaningful information from large data sets and make predictions about future outcomes. The goal of big data analytics is to turn vast amounts of data into actionable insights that can drive business growth, improve efficiency, and increase competitiveness. To obtain valuable insights and enable prompt and informed decision-making, big data analytics involves four stages of data processing: data collection, data processing, data cleaning, and data analysis.

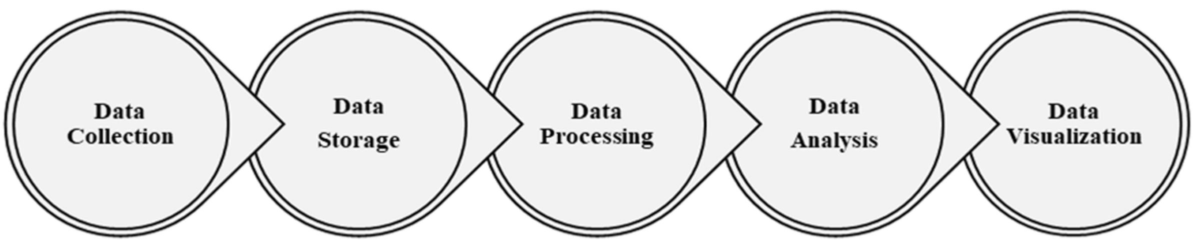
Businesses utilize big data analytics to open up new horizons, search for better prospects, ensure effective operations, and increase profits by harnessing big raw datasets. Put differently, big data analytics enables rapid and informed decision-making, enhances cost-effectiveness and operational efficiency, and also supports the advantages of data-driven marketing (IBM, 2023). The results of the 'Big Data and AI Executive Survey 2021', conducted by NewVantage Partners, indicated that approximately 91.7% of executives in global business and IT sectors planned to increase their investments in AI and big data, while nearly 92.1% acknowledged significant benefits from big data and analytics. However, despite the vast potential, big data analytics pose several challenges, such as data privacy, data security, data storage, fault tolerance, and data quality concerns (Tsai et al., 2015).

4. HOW DOES BIG DATA WORK?

In the global data-driven economy, data emerges as a crucial asset for businesses across industries. By exploiting big data, businesses can uncover valuable insights into areas impacting organizational performance, including market conditions, customer purchasing behaviors, and business processes (Google Cloud, n.d.). The insights received by utilizing big data can facilitate informed and impactful decision-making for businesses. Nonetheless, it is paramount important to learn - how to use big data effectively. In advance of implementing big data in operational settings, organizations must consider its flows across multiple locations, sources, systems, proprietors, and users. According to SAS (n.d.), businesses can gain control over the complex 'big data fabric' by following five key stages: developing a big data strategy, identifying data sources, accessing, managing, and storing data, analyzing data, and making informed, data-driven decisions.

Big data analytics involves a range of techniques including data collection and storage in a data warehouse, data processing for validation and filtering, cleaning, or sanitizing data through formatting and removing errors, conflicts, incompleteness, or redundancies, and finally, analyzing the data by utilizing a diverse range of tools. The workflow of big data can be summarized as integrating, managing, and analyzing big data sets (Oracle, n.d.). Overall, a functional big data flow involves processes that enable the efficient collection, storage, processing, analysis, and visualization of large datasets.

Figure 3. Stages of Data Processing Activities of Big Data



4.1 Data Collection

Big data techniques commence with data collection from diverse sources, including chat records, laboratories, log files, microblog messages, mobile devices, posts on Internet forums, satellites, search entries, sensors, and supercomputers (Khan et al., 2014). The data may also come from internal sources, such as sale-transaction records, or external sources, e.g., posts on social media and reviews of customers. Moreover, data sources may include automobiles, machines, networks, and wearable devices (Krishnan, 2019). Sometimes, specialized techniques are employed in the data collection process to extract raw data from a specific situation. An essential aspect of scientific data management is capturing data transition from scattered raw data to systematically arranged valuable data.

4.2 Data Storage

After the data collection phase, it is crucial to find a secure, scalable, and resilient repository to store data before or even after starting the processing activities. Choosing an appropriate data warehouse is critical, as it depends on specific needs and requirements. The storage solution can be cloud-based, on-premises, or a combination of both, with temporary storage options available for in-transit data. Apart from a traditional data warehouse on-site, there are more flexible and cost-effective options such as cloud solutions, data lakes, data pipelines, and Hadoop (SAS, n.d.). Some commonly used repositories for big data include Hadoop Distributed File System (HDFS), NoSQL databases, object storage, which involves storing data as objects rather than files or blocks, cloud storage, and memory databases, which store data in memory rather than on disk.

4.3 Data Processing

The processing of big data involves the use of various techniques to manage, cleanse, transform, and analyze large and complex data sets. This is critical since it facilitates the extraction of valuable information from extensive data sets. Often, data collected from different sources may be inconsistent, noisy, or repetitive, and thus necessitates processing to enhance data quality for analysis (Bhadani & Jothimani, 2016). Some commonly used techniques for big data processing are distributed computing, which splits data processing into smaller tasks that run simultaneously across multiple nodes; batch processing, which involves processing data at fixed intervals in batches; stream processing, which processes data in real-time as it is generated; and machine learning, which utilizes algorithms to learn patterns and reveal insights from data.

4.4 Data Analysis

The primary objective of big data analysis is to extract meaningful insights and enable businesses to attain their objectives. Various techniques are commonly employed for data analysis, including data mining, which involves utilizing statistical and machine learning techniques to identify relationships and patterns in large data sets; natural language processing, which involves analyzing and comprehending human language, including speech and text; predictive analytics, which utilizes statistical and machine learning techniques to predict future events based on historical data; and statistical analysis, which employs statistical methods to analyze and interpret data. However, data analysis has two fundamental goals: (1) comprehending the relationships among different aspects of data, and (2) devising effective data mining methods to facilitate accurate future predictions or observations of data (Fan & Liu, 2013).

4.5 Data Visualization

Visualization has demonstrated its efficacy as a tool not only for presenting crucial information contained in large amounts of data but also facilitating complex analyses (Xiong, 2013). Utilizing visualization techniques, businesses can comprehend the inner findings of critical data sets and adopt an informed decision based on them. The flexibility of visualization allows for precise control over the placement and visual representation of objects in a plot (Waskom, 2021).

However, the most frequently employed techniques for data visualization include charts and graphs, i.e., visualizing data using bar charts, line charts, scatter plots, and heat maps; interactive dashboards that allow users to explore and interact with data in real-time; geographic visualization, i.e., mapping data onto a geographic map to help identify patterns and trends based on location; network visualization, i.e., mapping relationships between different data points to help identify patterns and trends, and 3D visualization, i.e., creating three-dimensional visualizations of data to help identify patterns and trends using multiple tools, such as medical imaging, architectural design, and geospatial analysis.

5. ADVANTAGES OR DISADVANTAGES OF BIG DATA

The paradigm shift towards big data technology has already commenced and progressed significantly across industries. The transformative impact of big data analytics is evident in the processes of how companies arrange, function, oversee talent, and generate value (Henke, Libarikian, & Wiseman, 2016). The following sub-section aims to explore the advantages of big data, while the next sub-section will briefly analyze the disadvantages thereof.

5.1. Advantages of Big Data

Big data is an approach that can provide useful insights into customer behavior, preferences, and trends, which can enable companies to make informed decisions and customize their products or services to meet the needs of their customers. By analyzing large amounts of data from various sources, big data can help companies optimize their operations, identify areas for improvement, and encourage innovation and product development. Specifically, big data can enhance decision-making, create personalized experiences, promote cutting-edge research and innovation, improve customer experiences, ensure advanced security measures, and increase productivity while reducing costs for businesses.

5.1.1. Enhanced Decision-Making

One of the significant advantages of big data is its ability to provide organizations with valuable insights and information that can inform decision-making processes. With the vast amount of data available, organizations can analyze patterns, trends, and correlations in the data to gain insights into customer behavior, market trends, and operational efficiency. These insights can help organizations make more informed and data-driven decisions, leading to improved business outcomes and competitive advantages (Davenport & Patil, 2012).

5.1.2. Personalized Experiences

Big data allows organizations to capture and analyze vast amounts of data on customer preferences, behaviors, and interactions. This enables organizations to deliver highly personalized and tailored experiences to their customers. For example, e-commerce companies can use big data to analyze browsing behavior and purchase history to provide personalized product recommendations to individual customers. Personalized experiences lead to increased customer satisfaction and loyalty, resulting in improved customer retention and revenue generation (Chen, Chiang, & Storey, 2012).

5.1.3. Cutting-edge Research and Innovation

Big data has extended new possibilities for research and innovation in various fields. Researchers can access and analyze large datasets to identify patterns, trends, and correlations that were previously not possible with smaller datasets. This has led to advancements in fields such as healthcare, where big data analytics has been used to develop predictive models for disease diagnosis and treatment planning (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Big data has also enabled innovation in areas such as smart cities, autonomous vehicles, and renewable energy, leading to significant advancements in technology and societal benefits.

5.1.4. Improved Customer Experiences

Big data can help organizations understand customer needs and preferences better, allowing them to provide more personalized and responsive customer service. For example, call centers can use big data analytics to analyze customer interactions and sentiment analysis to identify and address customer issues more efficiently (Minelli, Chambers, & Dhiraj, 2013). This can lead to improved customer satisfaction, loyalty, and retention, resulting in increased customer lifetime value.

5.1.5. Advanced Security Measures

Big data analytics can be used to detect fraud and manage risks more effectively. By analyzing large amounts of data from multiple sources in real time, organizations can identify anomalies, patterns, and trends that may indicate fraudulent activities. For example, banks can use big data analytics to analyze transaction data to detect potentially fraudulent activities, such as unusual spending patterns or suspicious transactions (Laney, 2001). This can lead to improved fraud detection and prevention, saving organizations significant financial losses.

5.1.6. Increased Productivity with Reduced Costs

Big data analytics can help businesses enhance productivity levels through a methodical analysis of large and complex data. Storing massive amounts of data can be costly, but the advancements in scalable storage systems have enabled organizations to reduce expenses while optimizing operational efficiency (Microsoft, n.d.). As a result, businesses can achieve more profits and increase productivity with reduced costs within their organizational frameworks.

In summary, big data offers numerous advantages to organizations in various industries and sectors, including improved decision-making, enhanced personalization, advanced research and innovation, improved customer service, enhanced fraud detection and risk management, and increased productivity with reduced costs.

5.2. Disadvantages of Big Data

While big data brings massive opportunities for numerous sectors and industries, it entails significant drawbacks too. In the next sub-section, we will explore some major disadvantages of big data and prescribe workable solutions to address those issues.

5.2.1. Data Quality Issues

One of the primary concerns with big data is the issue regarding data quality. The sheer volume and variety of data can make it challenging to ensure the accuracy and comprehensiveness of data. Poor data quality can result in incorrect or biased results, leading to poor decision-making and missed opportunities (Jiang & Kavanagh, 2016). For example, if the data of a particular business is biased toward a specific demographic or geographic region, it may not accurately reflect the wider population. To address this data quality issue, relevant organizations must strictly adhere to the techniques of data governance and robust data quality frameworks. To make sure of this, data has to be collected ethically and legally, and it has to be adequately cleansed and appropriately verified before analysis.

5.2.2. Data Security Concerns

Another significant drawback of big data is the potential privacy and security risks associated with data collection and data storage. The vast amount of personal data or sensitive personal data collected by big data applications can make them a target for hackers or cybercriminals. Besides, collecting a large amount of data without informed consent or proper safeguards can violate established data protection principles damaging the reputation of an organization and its legal standing (Kshetri, 2014).

To address this challenge, organizations must follow the provisions of data protection laws while processing personal data or sensitive personal data of individuals. Specifically, during the processing of personal data or sensitive personal data of a natural person, an organization should comply with key data protection principles, lawful bases for processing, and the provisions of exemptions or derogations allowed in data protection regulations.

5.2.3. Cost and Resource Constraints

The implementation of big data technology can be costly and resource-intensive, requiring significant investments in hardware, software, and expert personnel (Hernández-Rojas et al., 2018). This can pose a challenge for smaller organizations with limited budgets and resources. Moreover, handling a vast quantity of data may require specialized skills and expertise, which could result in increased costs for implementing and maintaining the system. To address cost and resource constraints, organizations must carefully assess their needs and capabilities before investing in big data technology. This includes conducting a cost-benefit analysis to determine the potential return on investment, as well as identifying areas where the use of big data can provide the most significant benefits. Moreover, resorting to cooperative, or joint venture projects may also provide relief for businesses in the areas of cost and resource constraints.

5.2.4. Data Analysis Challenges

Another drawback of big data is the challenge of interpreting and integrating data from multiple sources. The large amount and variety of data can make it difficult to identify meaningful patterns and insights, and integrating data from disparate sources can pose significant technical and logistical challenges (Provost & Fawcett, 2013). To address these challenges, organizations must invest in advanced analytic tools and techniques that can help identify meaningful insights and patterns in large datasets. In addition, the latest advanced data integration tools can help streamline the process of integrating data from multiple sources.

5.2.5. Lack of Skilled and Talented Professionals

One of the biggest challenges with big data is finding skilled professionals who can successfully handle complex analytics and processing activities. Hence, there is a significant skills gap and talent shortage in the field of big data, which makes it difficult for companies to find the right professionals to handle their data-related needs (Thomsen, 2018). This gap is driven by the high demand for professionals with data science skills, combined with a lack of formal education and training programs to help people develop these skills (Brynjolfsson & McAfee, 2012).

As a result, many companies are struggling to find professionals who have the right combination of technical skills and business acumen to make sense of big data. This has led to a highly competitive job market for data scientists, with many companies offering high salaries and benefits to attract and retain talent (Davenport & Patil, 2012). Moreover, companies are often forced to invest significant resources in training and development programs to help their existing employees acquire the required skills to work with big data (Cukier & Mayer-Schönberger, 2013).

5.2.6. Digital Divide

The 'digital divide' refers to unequal access to digital technology and its benefits among different social groups and regions. The emergence of big data technologies has created a unique digital divide between two groups who can effectively access and use big data and who cannot. Big data poses challenges for those who lack the necessary skills, resources, and infrastructure to participate in the data-driven economy. Therefore, policymakers should prioritize investments in digital infrastructure, including high-speed internet, data literacy programs, and data collection and analysis tools. Policymakers should also ensure ethical and transparent usage of big data so that the benefits of data-driven innovation are shared equitably across different social groups and regions (Crawford & Schultz, 2014; Kitchin, 2014; Liu, 2018; van Dijk, 2012).

6. APPLICATIONS OF BIG DATA

Big data applications have become extremely popular in the last two decades, not because of their competence to manage large datasets but rather their capability to extract insights from intricate, heterogeneous, longitudinal, noisy, and voluminous datasets (Khanna et al., 2021). Based on the nature of work, big data technologies can be divided into two specific categories, such as operational big data technology and analytical big data technology (Osadchuk, 2023). Big data technologies can also be classified into four major types, such as data storage technology, data mining technology, data analytics, and data visualization technology (Coursera, 2023).

No single tool can assist businesses in extracting new opportunities by harnessing large datasets but sets of technologies work together for the purposes. The major players that work together to bring about insights from big data include - Hadoop open-source software; machine learning; cloud computing; data mining, management, and storage technology; text mining technology; memory analytics, and predictive analytics (SAS, 2023). Big data brings big changes in numerous other sectors, including construction, manufacturing, retail, transportation, data management and security, aircraft industries, agriculture, and even fraud prevention.

Many organizations employ big data and analytics to process big datasets produced by numerous IoT devices and appliances. Moreover, social networking sites (SNS) also utilize big data and analytics to learn user behavior, preference, satisfaction, and target advertising. The following sub-section analyzes the applications of big data in several sectors, including financial services, healthcare systems, education sectors, public services, and telecommunication industries.

6.1. Financial Services

The financial service sector appears as one of the most crucial areas that utilize the power of big data techniques to process, analyze, and leverage big datasets for practical purposes. Big data offers a revolutionizing solution regarding the functionalities of global stock markets and investors' investment decision-making (CFI Team, 2022). By efficiently nurturing meaningful customer relations and enhancing the capacity of predicting customer preferences, financial institutions can produce unique client-centric goods and services to grab desired commercial prospects. Eventually, big data can contribute to discovering new business models.

With the help of business intelligence tools and big data analytics, financial organizations can handle and track the regulatory processes from maintaining tax files to record keeping of central banks. Along with strong internal control systems and business intelligence tools, big data analytics helps financial institutions to fight against cyberattacks and financial terrorism. By utilizing unique fraud-detecting algorithms and statistical computing, big data analytics help financial institutions to detect financial fraud before its actual occurrence. By employing real-time big data analytics, Alibaba, for example, built a fraud-detection, control, and monitoring system to reduce fraudulent activities within their business (CFI Team, 2022).

The exponential growth of data and data processing technology brings radical transformation in business operations, processes, and systems. Nonetheless, there are technical and practical challenges for big data applications in the financial sector. The current banking industry is still much slower in adopting innovative technologies. Handling the increasing workloads resulting from big data is unattainable since 92 out of the top 100 banks worldwide still run-on IBM mainframe computers (Mathur, 2022). Arguably, the enormous potential and prospects of big data would surely outweigh all associated risk factors.

6.2. Healthcare Systems

By enhancing the quality of care and reducing error, waste, and cost of services, big data analytics brings enormous potential to healthcare systems (Mehta & Pandit, 2018). By employing big data analytics, the health industry can foresee the outbreaks of potential epidemics, evade curable diseases, and improve the quality of life. Thus, big data applications in healthcare are becoming a frequent phenomenon, and eventually, the investment in big data encompassing healthcare systems is growing considerably over the years. A current study indicates that the global healthcare big data market size may reach US $78.03 billion by 2027 (Emergen Research, 2020).

Big data facilitates the de-identification of health data allowing secondary use of such data (Tenyi et al., 2017). Through the identification of patterns and decoding links, big data can promote autonomous-decision making (Asante-Korang & Jacobs, 2016). It enables healthcare institutions to provide personalized predictions, directed treatment, and cost-effective care by reducing waste and delivering actionable advice to patients while encouraging them to maintain sound health (Raghupathi & Raghupathi, 2014; Bates et al., 2014).

In brief, big data analytics helps the entire healthcare system by minimizing costs, avoiding unnecessary diagnoses, predicting outbreaks of diseases, and designing early preventive measures, while providing patients with evidence-based medicine. Despite having immense value-adding prospects and potential, big data analytics in healthcare witness many challenges as well. The challenges encircling healthcare-big data include, among other things, the absence of fitting IT infrastructure, high costs for analytical tools, data quality, data privacy, data security, data ownership, the multidimensionality of data, and some technical issues (Mehta & Pandit, 2018).

6.3. Education Sectors

Technological advancement may also appear as a catalyst for employing big data analytics in the education sector. Indeed, big data has the potential to make a paradigm shift in the education sector by analyzing and addressing prevailing challenges in the field. By installing sensors in objects, connecting with cloud computing, and employing augmented reality and wearable techniques, big data can produce insightful real-time information about the education system, culture, and atmosphere.

Research in the field of education, humanities, psychology, and social science are broadly classified into two different approaches, e.g., qualitative, or quantitative approaches. These two modes of research are predominantly different in both theory and practice. They are disconnected too, although can be fixed and bridged by a mixed method (Gibson & Ifenthaler, 2017). Big data can help to reduce the gaps between diverse modes of research and advance research and education with the power of real-time analysis and insights. Big data can immensely contribute to designing dynamic programs, modernizing course curriculums, improving grading systems, and predicting suitable career paths for students. Nevertheless, big data is yet a new phenomenon in the education sector. Hence, the promises, prospects, and challenges of big data in the education sector are still unknown to many relevant stakeholders (Daniel, 2017).

6.4. Public Services

The application of big data in various public service sectors holds great potential. To effectively utilize big data analytics, the government can consider a broad range of sectors, including education, finance, healthcare, transportation, energy, industry, environment, agriculture, and foreign trade and investment. Without a comprehensive understanding of these areas and associated issues, the government may struggle to make informed decisions regarding development projects. By leveraging real-time insights, big data can help facilitate effective decision-making in the public sector. By analyzing data from various public bodies, big data can also help the government protect and promote the interests of all sectors. Additionally, big data can support public authorities to take appropriate initiatives whenever required. Therefore, big data opens up a new frontier for the public sector and services (Desouza, 2014).

Previously, it was believed that the rapid progress of big data technology would help public organizations develop sustainable and effective e-government systems. It was also anticipated that in the coming years, all government agencies would face challenges in integrating their diverse data assets, building analytical capabilities, and adopting a data-driven approach to decision-making (Long et al., 2021). That time has now arrived. Currently, big data technology plays a pivotal role in various government services, such as traffic management, weather forecasting, crime prevention, national security, economic forecasting, and predicting global politics, policies, and business trends. Big data can help governments to develop a welfare state system by establishing a free, fair, transparent, effective, and proficient system of governance.

To achieve the aforementioned objective and ensure the fundamental needs, rights, and freedoms of citizens, governments must engage in extensive efforts. A comprehensive analysis of public data can aid governments in implementing various welfare programs in critical areas and providing essential logistical support. By providing authentic and real-time insights, big data analytics can assist governments in dealing with national crises such as poverty, malnutrition, unemployment, drug addiction, natural disasters, cyber threats, and terrorism. Big and public data can foster cooperation, provide real-time problem-solving solutions, encourage greater transparency, and usher in a new era of decision-making and policymaking (Bertot et al., 2014). In summary, big data and analytics present significant opportunities and potential for various government sectors. Nonetheless, processing personal data through different platforms may also jeopardize citizens' right to privacy.

6.5. Telecommunication Industries

Big data analytics has significant implications for all aspects of the global telecommunications industry, including market insights, innovation, and understanding of relevant scientific phenomena. Telecommunications operators can use big data technology to detect illegal IDs and their owners, understand customer preferences, differentiate customer groups, develop policies based on demand, and assess customer engagement rates and areas for improvement. To comprehend the usage and application of big data in telecommunications services, it is crucial to understand the development stages of telecommunications (mobile) networks.

In the 1980s, the first generation of mobile networks (1G) was introduced worldwide, which only provided voice calls and text messages with limited data generation. In the early 90s, the second-generation network (2G) was introduced, which added the capability for pictures, text, multimedia messages (MMS), faxing, and voicemail. For data analysis, the 2G mobile network used general statistical methods developed in the 1970s. The 3G network replaced the 2G network in 2000, focusing on advanced data and customer behavior analysis, along with the addition of graphics, videos, texts, and voice calls. The introduction of the 4G Long Term Evolution (LTE) network in Sweden in 2009 marked a turning point in the application of big data analytics in the mobile ecosystem. The 4G technology enabled real-time analysis of high-quality Internet browsing experience, audio-video calls, texts, multimedia messages, and online games, providing new insights into network performance and users' behaviors (Kastouni & Lahcen, 2020).

To keep up with the latest trends in the telecom industry, meet customer expectations, and ensure service quality, telecom operators must assess customer preferences and choices to upgrade their future client-oriented plans and services. Big data can provide companies with useful insights to offer distinctive products and services (IBM Developer, n.d.a). Understanding customer behavior and preferences enables companies to tailor customer-centered products and services.

7. PROSPECTS OF BIG DATA

In today's digital world, a significant portion of our work is carried out online using digital data. Currently, almost half of the world's population is connected to the Internet (Internet Live Stats, 2023a). As of January 2022, the top 10 countries in terms of internet usage are China, India, the USA, Indonesia, Brazil, Russia, Japan, Nigeria, Mexico, and Germany (Statista Research, 2022). There are about 400 million active websites out of the approximately 2 billion websites available (Chakarov, 2022). Amazon has around 1.4 million servers, Facebook produces about 4 PB of data per day, and on average, an individual generates about 1.7 MB of data per second. Consequently, people worldwide produce approximately 2.5 quintillion bytes of data each day (Wasabi, 2023; Prodanoff, 2022). The question now arises as to whether this vast amount of digital data (big data) has any potential prospects across industries.

In the Industry 4.0 era, data has emerged as a crucial element for driving transformation. Big data analytics can integrate and analyze vast amounts and types of data, providing valuable insights from data warehouses. This helps uncover customer preferences, predict opportunities, and inform successful business strategies. According to Gartner (2023a), better data analysis can reveal innovations that may be overlooked by human analysis. The global big data market is predicted to grow from $271.83 billion to $655.53 billion by 2029 (Fortune Business Insights, 2022). Notably, the top global market leaders in this field are IBM, Dow Jones & Company, Dun & Bradstreet Holdings, Equifax, Moody's Corporation, Salesforce.com, SAP SE, and Verisk Analytics.

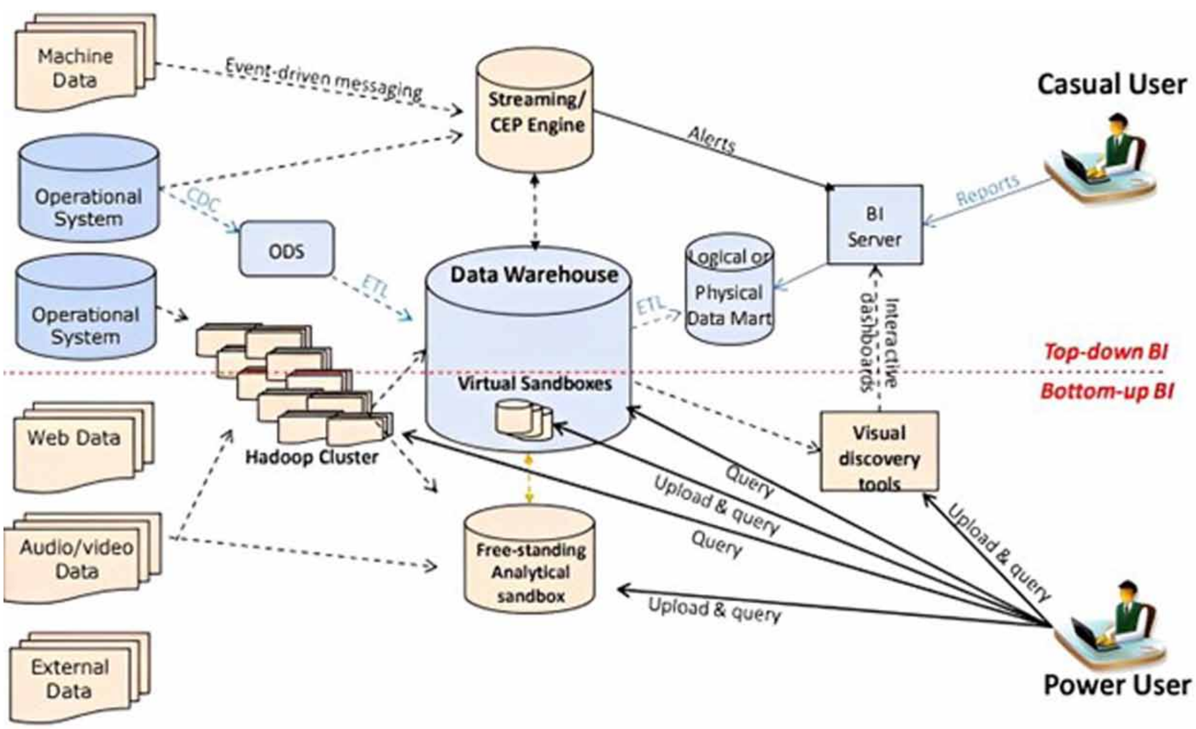
The use of big data has the potential to make significant contributions to various aspects of sustainable development goals. These areas include reducing poverty and hunger, improving health and education, promoting gender equality, providing clean water and sanitation, encouraging affordable and clean energy, economic growth, industry and infrastructure development, reducing inequalities, promoting sustainable cities and communities, responsible consumption and production, climate action, conserving life below water and on land, ensuring peace, justice and strong institutions, and promoting partnerships for the attainment of the sustainable development goals (United Nation, 2023).

Figure 4. Big Data for Sustainable Development1

Big data allows institutions to become data-driven by gathering information from various sources such as devices, logs, sensors, social networking sites, and transactional activities, and analyzing them as needed. Additionally, big data enables data-driven entities to make faster and better decisions, reduce costs, and adopt enhanced data-driven marketing strategies. IBM has identified several benefits that big data can provide to businesses, such as improved decision-making, optimized cost management, enhanced operational efficiency, and improved data-driven strategies for market entry and expansion (IBM Developer, n.d.b). Other advantages of big data analytics include real-time analysis, quick insights, better investment plans, targeted marketing, reduction of revenue loss, data monetization, improved client experience, innovation of products and services, increased customer satisfaction, and greater profitability.

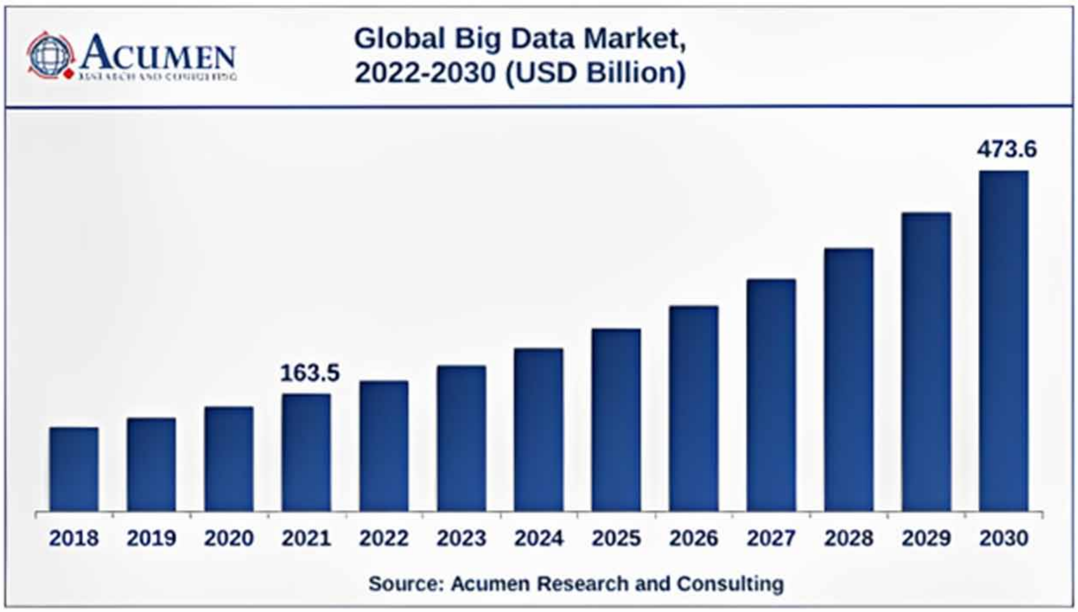
Big data can expose people's confidential behavioral patterns and reveal their preferences (Abbas, 2011). This helps fill the gaps between what people express and their actual activities and interactions with others in a given context. Such insights are vital for public and private entities in decision-making processes that range from law enforcement and social welfare to national security. Big data analytics enables its users to obtain a comprehensive view of people's behavioral patterns, choices, preferences, and trends in socioeconomic and cultural contexts. In the absence of big data analytics, people's perception of the data warehouse is too limited, whereas proper utilization of big data may unlock unprecedented opportunities, revealing untapped horizons for the wider community, as indicated in the following figure.

Figure 5. The Next-Generation Big Data Analytics Architecture[2]



In the past, companies used to accumulate data, perform analysis using analytics, and store it for future use. By utilizing big data, companies can now collect and analyze data simultaneously, which allows them to make quick and informed decisions in real-time. It enables businesses to operate faster and with greater agility and gives them a competitive advantage they did not have before (SAS, 2023). The benefits of big data can be summarized by using five key terms: speed, efficiency, real-time insights, reduced costs, and increased profits. As demonstrated in the figure below, the global market size of big data is increasing rapidly and is expected to grow even larger in the coming years (Acumen Research and Consulting, 2023).

Figure 6. Global Big Data Market from 2022-2030



8. CHALLENGES OF BIG DATA

Big data technology offers enormous potential for businesses, including cost reduction, improved decision-making, and enhanced products and services. However, big data also presents a wide range of challenges. This section focuses on technical, practical, and legal challenges associated with big data technology.

8.1 Technical and Practical Challenges

Big data presents several technical challenges concerning data, data processing, and data management. Firstly, the data itself poses a challenge as it has various dimensions such as volume, variety, velocity, veracity, volatility, quality, discovery, and dogmatism. Secondly, data processing challenges involve data capture, integration, transformation, selection of an appropriate analysis model, and how the results of data analysis are produced. Lastly, data management issues are affected by data security, privacy, governance, and ethical considerations (Nasser & Tariq, 2015).

Undoubtedly, big data brings promises for discovering insights into people's behavioral patterns and variances, which are not available in small-scale data. Nonetheless, the nature of this high volume and veracity of data causes some unique computing and statistic challenges, including incidental indigeneity, measurement errors, noise accumulation, scalability, storage bottleneck, and spurious correlation

(Fan et al., 2014). The large scale of data accumulated from different sources, times, and technologies causes the issue of heterogeneity, experimental divergences, and statistical biases. This high volume and veracity of data create high computational costs and algorithmic instability.

A big data user faces numerous other challenges while analyzing, capturing, searching, sharing, storing, and visualizing data. The challenges of big data also include incompleteness, inconsistency, scalability, security, and timeliness issues regarding data. Due to the veracity of data, big data poses huge challenges in terms of the efficiency of representation and access to data by diverse users (Khan et al., 2014). The threats are not straightforward but rather complex, necessitating unique computational and statistical solutions with more adaptable and comprehensive procedures.

Big data faces numerous practical challenges due to its large size, high speed, and variability that traditional data processing technologies cannot efficiently handle. As a result, modern data processing solutions are required to overcome these challenges. However, only a limited number of tools are presently available to handle big data (Khan et al., 2014). Even popular tools such as Hadoop, HBase, and Cassandra sometimes appear insufficient in addressing the challenges of real-time data searching, sharing, storage, and visualization.

More specifically, tools such as Hadoop and MapReduce were limited in their ability to process queries and maintain a low-level architecture for data management and processing. Meanwhile, SAS, R, and Matlab were unsuitable for handling large-scale datasets. Although GraphLab developed its infrastructure for calculating graph-based algorithms in machine learning, it could not handle big data effectively. Therefore, suitable tools for effectively managing and processing big data were lacking (Khan et al., 2014).

A fault-tolerant tool is essential to obtain accurate insights into customers' behavior and preferences and the operational performance of businesses. However, designing and assembling such systems can be challenging and costly due to their complex algorithms (Katal et al., 2013). Although fault-tolerant systems aim to reduce errors, their implementation often incurs additional costs. Despite the desire of businesses to process vast amounts of data for predictive decision-making, ensuring data quality and relevance remains a challenge. While data may reveal valuable insights into customer behavior and patterns, there is no guarantee of accuracy and authenticity.

One of the major concerns of big data is the high infrastructure costs of big data (Tole, 2013). The immense processing capability and steady and intricate network designs require specialists' expertise. If not chosen the open-source software, the beneficiary has to pay huge for both the hardware and software systems of big data. Due to maintenance and support of the system, open-source resources also require experts' expertise which costs a lot. Even businesses always need an outside maintenance team for the proper functioning and operation of big data technology, which leads big data to appear as an expensive venture.

The advancement and diffusion of big data technology and its advantages are not evenly distributed, leading to digital disparities between developed and developing nations (Hilbert, 2016). Due to insufficient infrastructure, scarcity of human and financial resources, and differences in organizational settings, the use of big data technology is unequally distributed among developed and underdeveloped nations. This creates a new level of the digital divide.

8.2 Legal Challenges

Big data technology has significant implications for the current legal framework surrounding privacy and personal data protection. Despite the enormous benefits of big data, the data processing activities of many large corporations pose significant privacy risks. In today's world, people use various apps for a range of purposes, including browsing the internet, learning, conducting business, socializing, banking, shopping, and making reservations. These apps, by default, collect valuable personal data, such as user names, profile pictures, email addresses, phone numbers, and location data. App builders can then use this information to create user profiles, leading to excessive surveillance at all times and places.

Big data and big data analytics have the potential to pose threats to individuals, institutions, and their reputations (Krasnow Waterman & Bruening, 2014). Businesses often employ big data to enhance productivity, reduce costs, and increase profits. However, this insight may require continuous monitoring of employees' actions and performances based on industry benchmarks, which can violate human dignity. As a result, the use of big data technologies and their applications can lead to serious privacy threats for many individuals and institutions. Big data has the potential to uncover unique trends, practices, and business-client relationships by analyzing large datasets from various sources.

Businesses often attempt to reuse data to achieve maximum output from their data processing activities. However, reusing data can have privacy implications. Even when data processing entities use de-identification techniques, such as encryption, anonymization, key-coding, pseudonymization, and data sharding, computer scientists can still link the anonymized data to specific individuals (Desouza, 2014). Moreover, due to delays in the law-making process, policymakers may not adopt the final piece of legal regulation on time to address the challenges posed by big data.

Big data and analytics shake the foundation of the EU's General Data Protection Regulation (GDPR) in several ways (GDPR, 2016). It prioritizes data protection during the data collection stage but pays little attention to post-analysis protections. As a result, it fails to address the unforeseen privacy risks that may arise after data collection and processing for analysis. Although the primary aim of data protection laws is to ensure the protection of privacy and personal data of individuals, they do not establish clear criteria for assessing and evaluating data. This legal vacuum leaves predictions regarding employee performance, economic cash flow, and people's lifespans largely unregulated (Wachter, 2019).

Simply put, the legal issues related to big data involve data privacy, security, ownership, governance, and the accuracy of algorithms used to process and analyze big data for decision-making. Notably, the existing principles of data protection, e.g., data minimization and purpose specification, are incompatible with the operation and applications of big data and analytics.

9. WAY FORWARD

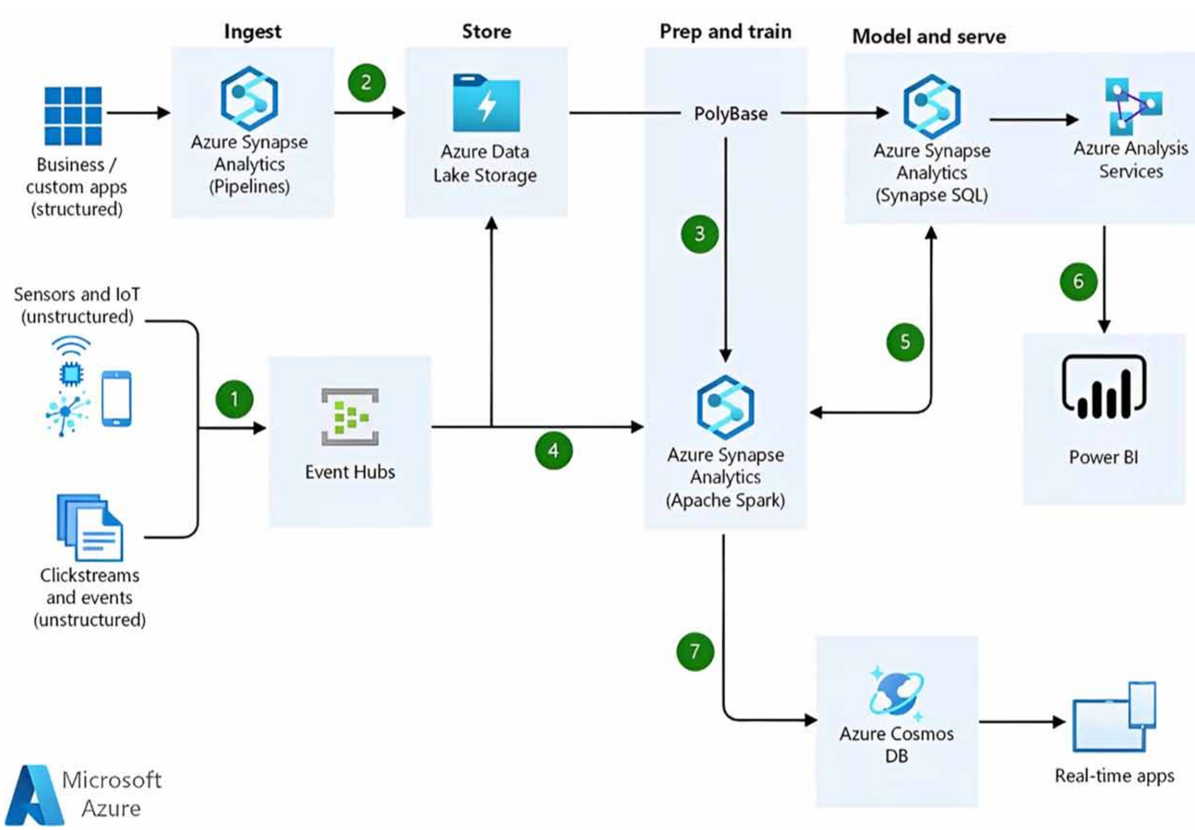
In order to ensure the smooth operation of various professional, formal, official, or institutional activities, insights derived from relevant datasets have become essential over the centuries. To gain better insights into specific schemes, projects, or businesses, the processing of relevant big data is indispensable. However, humans are notoriously inefficient at data processing, as they spend too much time on it (Satell, 2013). Spending a significant amount of time processing only a few documents and potentially mixing them up does not effectively utilize people's time.

With the innovation of computers, the Internet, IoT, and new business models, data generation has become so voluminous that ordinary data processing tools and techniques have become inefficient in dealing with those huge datasets. Big data and analytics have emerged as a solution to address these challenges. By freeing humans from the task of data processing, big data enables individuals to focus on the primary goal of the project (Satell, 2013). Big data can be a key enabler for achieving competitive advantage, boosting productivity, promoting innovation, and enhancing customer satisfaction. Nonetheless, big data also poses some technical, practical, and legal challenges. Fortunately, there are ways to overcome these challenges and move forward.

Earlier, it was contended that the prevailing data processing technology cannot achieve desired performance and requires sophisticated data processing software. It was also claimed that only a few tools can address the existing big data challenges. It can be argued that no system is inherently good or bad, but rather presents opportunities and challenges. Presently, several successful integrated initiatives are offering comprehensive solutions to the past big data challenges. For example, IBM's big data infrastructure integrates various data processing techniques, including content processing, database processing, stream processing tools, and business data systems, resulting in improved performance, services, and decision-making (Zhuge, 2016). By utilizing NoSQL technology, Google handles vast amounts of data ranging from Gigabyte to Yottabyte generated from hundreds of its services and products such as Google Apps, Browsers, Docs, Drive, Gmail, Maps, YouTube, and Translator (Neto, J. A. R. (Zezinho), 2020; Hewage et al., 2018).

Social media giants such as Facebook, Twitter, and LinkedIn also follow in Google's footsteps by successfully employing cutting-edge tools and techniques for their data processing activities (Neto, J. A. R. (Zezinho), 2020; Hewage et al., 2018). Microsoft's SQL Server 2019 data clusters provide scalable HDFS, Spark, and SQL Server clusters that enable users to read, write, and process big data by utilizing Spark or Transact-SQL libraries. Furthermore, Microsoft's existing big data infrastructure, as shared in the bellow figure, allows users to analyze and integrate both relational and non-relational data in high volume (Assaf MSFT et al., 2022).

Figure 7. Real-Time Analytics on Big Data Architecture[3]



Despite significant advancements in big data and analytics research, numerous issues remain unresolved using conventional tools and techniques, which may take years to address. According to McKinsey & Company, the emergence of Lambda or Kappa architectures is expected to bring about a revolutionary transformation in the data generation, processing, analysis, and visualization process by 2025 (McKinsey, 2022). Consequently, human-assisted data processing will be replaced by big data and analytics, enabling work to be completed in a shorter timeframe, rather than taking months or even years.

Although the use of big data and analytics poses legal challenges, progress is being made in addressing these issues. Based on legal frameworks and case law related to privacy and data protection, it can be argued that privacy and data protection laws generally apply to big data and analytics (van der Sloot & van Schendel, 2016). For example, in *United States v Jones*, the court restricted the collection of large-scale locational data by police officers (United States v. Jones, 2012). Similarly, in *ACLU v Clapper*, the Second Circuit Court of Appeals declared the collection of massive phone call metadata illegal as it was not covered by section 215 of the Patriot Act, 2001 (ACLU v. Clapper, 2013). However, in another case, the Foreign Intelligence Surveillance Court allowed the collection of metadata for lawful purposes.

It is important to note that data is inherently dynamic and follows a circular process (van der Sloot & van Schendel, 2016). Typically, data is interconnected, combined, anonymized, de-anonymized, and merged with other data to create profiles, and may be pseudonymized for statistical or profiling purposes. Data protection laws cover personal, sensitive, statistical, private, anonymous, or metadata that is naturally static data. Despite limitations in legal regulations, some countries, including England, Estonia, France, and Japan, are exploring the possibility of implementing new laws to regulate big data and analytics.

If data protection law applies to big data, the issue of compliance may arise. In such cases, organizations are expected to adhere to data protection laws, rules, and regulations. This can be achieved by conducting regular data protection impact assessments and ensuring that data protection is incorporated into the design and default settings of new products and services. Additionally, transparency regarding data collection, storage, usage, and disclosure must be maintained by every institution. Organizations are also responsible for providing individuals with clear and easily comprehensible information regarding the use of their data.

Large data processing entities should adhere to the key data protection principles. These principles require that personal data be processed in a lawful, fair, and transparent manner (lawfulness, fairness, and transparency principle) and only for specified, explicit, and legitimate purposes (purpose limitation principle). Additionally, personal data should be adequate, relevant, and limited to the original purposes (data minimization principle), accurate and up to date (accuracy principle), not stored for longer than necessary for the original purposes (storage limitation principle), protected from unauthorized or unlawful processing and any accidental loss, destruction, or damage (integrity and confidentiality and security principle), and that the controller is responsible and accountable for complying with these principles (accountability principle) (GDPR, 2016).

In order to establish data governance, every organization must have a transparent data protection policy for processing personal data and periodically review and update this policy as needed. Additionally, every data processing entity must ensure that the algorithms and tools used to process big data are unbiased and accurate. To safeguard personal data from accidental loss, unauthorized access, or breaches, an organization should adopt appropriate legal and security measures, including the adoption of an encryption system, restricting data access, and conducting regular audits.

To overcome legal challenges related to big data, organizations should invest in employee education and training. This should cover the fundamentals of data protection laws, regulations, and best practices to ensure compliance with data protection laws. In addition to businesses and public bodies, private individuals have also a responsibility to adhere to existing data protection laws and practices.

10. FUTURE RESEARCH DIRECTIONS

The primary objective of this chapter was to present a concise overview of big data and analytics, encompassing their utilization, merits, demerits, applications, prospects, challenges, and general recommendations, through the lens of legal and regulatory perspectives. The study was carried out by employing qualitative and doctrinal legal research methodologies, which were appropriate for the nature and context of the research. It is pertinent to emphasize that there are no universally applicable policy measures for challenges related to big data, and resolving these challenges may require a combination of scientific, technical, and legal tools and techniques. Hence, forthcoming research on big data could explore diverse viewpoints beyond the legal perspective, such as techno-scientific and legal studies. This may be particularly valuable due to the intricate and ever-evolving nature of this field of investigation.

11. CONCLUSION

Big data refers to vast and complex datasets generated from a variety of sources, such as social media, online businesses, cell phones and apps, search queries, click streams, emails, images, health data, scientific research data, videos, sensors, logs, and audio (Sagiroglu & Sinanc, 2013). These datasets are characterized by their size, motion, variety, and fluctuating quality, making them challenging to process using traditional data processing methods. Analytics involves the examination, interpretation, filtering, transformation, and modeling of these massive datasets to extract valuable insights, knowledge, and information.

Big data and analytics have the potential to provide significant value to businesses, giving them a competitive advantage. Therefore, numerous public and private entities use big data and analytics in a variety of industries, including finance, healthcare, education, public services, telecommunications, social media, IoT devices, and fraud detection systems. With advancements in machine learning (ML) and artificial intelligence (AI), businesses can now obtain more precise and accurate analyses of products, services, and customer's behavioral patterns, trends, preferences, and choices. Thus, big data and analytics have grown in significance and offered tremendous opportunities in the past two decades.

On the other hand, big data also brings many challenges, including technical, practical, societal, and legal ones. Moreover, big data may pose problems related to data quality, privacy, security, integration, and ownership. Additionally, the increasing demand for big data analytics has led to a shortage of skilled professionals in this area, resulting in a rise in the complexity of associated fields.

To move forward, governments should develop robust legal frameworks to ensure compliance by individuals and organizations to minimize the associated risks. However, legal regulations have inherent limitations, including delays in implementation and a lack of dynamism compared to the pace of advanced technology. Furthermore, data protection principles, such as purpose limitation, data minimization, and storage limitation, may not always apply to big data and analytics. Hence, policymakers may consider enacting specific regulations for these fields. Public-private investments in education, training, and awareness campaigns are also necessary to address skill shortages and promote the responsible use of big data and analytics. Moreover, private organizations should also pay due attention when dealing with big data and analytics to ensure the rights and interests of all stakeholders involved.

REFERENCES

Abbas, R. (2011). The social implications of location-based services: An observational study of users. *Journal of Location Based Services*, *5*(3-4), 156–181. doi:10.1080/17489725.2011.637970

Acumen Research and Consulting. (2023). Retrieved from https://www.acumenresearchandconsulting.com/big-data-market

Asante-Korang, A., & Jacobs, J. P. (2016). Big data and pediatric cardiovascular disease in the era of transparency in healthcare. *Cardiology in the Young*, *26*(8), 1597–1602. doi:10.1017/S1047951116001736 PMID:28148322

Assaf, W., Roth, J., Schoenack, N., West, R., Buck, A., & Coelho, D. (2022). Big data options on the Microsoft SQL Server platform. *Microsoft.* Retrieved from https://learn.microsoft.com/en-us/sql/big-data-cluster/big-data-options?view=sql-server-ver16

Austin, C., & Kusumoto, F. (2016). The application of big data in medicine: Current implications and future directions. *Journal of Interventional Cardiac Electrophysiology*, *47*(1), 51–59. doi:10.100710840-016-0104-y PMID:26814841

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, *33*(7), 1123–1131. doi:10.1377/hlthaff.2014.0041 PMID:25006137

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, *19*(1-2), 5–16. doi:10.3233/IP-140328

Bhadani, A. K., & Jothimani, D. (2016). Big data: challenges, opportunities, and realities. In M. K. Singh & D. G. Kumar (Eds.), *Effective big data management and opportunities for implementation* (pp. 1–24). IGI Global. doi:10.4018/978-1-5225-0182-4.ch001

Brynjolfsson, E., & McAfee, A. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–68. PMID:23074865

Bulao, J. (2023). How much data is created every day in 2023? *Tech Jury*. Retrieved from https://techjury.net/blog/how-much-data-is-created-every-day/#gref

Business Wire. (2021). Retrieved from https://www.businesswire.com/news/home/20210104005022/en/NewVantage-Partners-Releases-2021-Big-Data-and-AI-Executive-Survey

Cano, I., Tenyi, A., Vela, E., Miralles, F., & Roca, J. (2017). Perspectives on big data applications of health information. *Current Opinion in Systems Biology*, *3*, 36–42. doi:10.1016/j.coisb.2017.04.012

Castrounis, A. (2018). Data science and big data, explained. *KDnuggets*. Retrieved from https://www.kdnuggets.com/2016/11/big-data-data-science-explained.html

CFI Team. (2022). *Big data in finance*. Corporate Finance Institute. Retrieved from https://corporatefinanceinstitute.com/resources/data-science/big-data-in-finance/

Chakarov, R. (2022). How many websites are there? How many are active in 2022? *Web Tribunal.* Retrieved from https://webtribunal.net/blog/how-many-websites/#gref

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly*, *36*(4), 1165–1188. doi:10.2307/41703503

Chen, S. C. I., & Kharabsheh, R. A. (2019). The link between innovation and prosperity: How to manage knowledge for the individual's and society's benefit from big data governance? In *Big data governance and perspectives in knowledge management* (pp. 200–217). IGI Global. doi:10.4018/978-1-5225-7077-6.ch009

Coursera. (2023). *4 types of big data technologies (+ management tools).* Retrieved from https://www.coursera.org/articles/big-data-technologies

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review. Boston College. Law School*, *55*(1), 93–128.

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data. *Foreign Affairs*, *92*(3), 28–40.

Daniel, B. K. (2017). Big data in higher education: the big picture. In *Big data and learning analytics in higher education* (pp. 19–28). Springer. doi:10.1007/978-3-319-06520-5_3

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, *90*(10), 70–76. PMID:23074866

Desouza, K. (2014). *Realizing the promise of big data*. IBM Center for the Business of Government.

Dictionary.Com. (2023). *Definition*. Retrieved from https://www.dictionary.com/browse/definition#:~:text=%2F%

Diebold, F. X. (2019). On the origin (s) and development of "big data": The phenomenon, the term, and the discipline. *Big Data & Society*, *6*(1), 1–16. https://www.sas.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

Eckerson, W. (2011). Big data analytics: Profiling the use of analytical platforms in user organizations. *BeyeNetwork*. Retrieved from http://docs.media.bitpipe.com/io_10x/io_103043/item_486870/Big%20Data%20AnalyticsMarkLogic.pdf

Emergen Research. (2020). Big data in healthcare market size worth USD 78.03 billion by 2027. *Emergen Research.* Retrieved from https://www.globenewswire.com/news-release/2020/12/17/2146728/0/en/Big-Data-in-Healthcare-Market-Size-Worth-USD-78-03-Billion-by-2027-Emergen-Research.html

Exabyte. (n.d.). *Wasabi.* Retrieved from https://wasabi.com/glossary/exabyte-definition/#

Expert Panel, Forbes Technology Council. (2023). Big data, big changes: 15 industries and functions big data can revolutionize. *Forbes.* Retrieved from https://www.forbes.com/sites/forbestechcouncil/2019/07/25/big-data-big-changes-15-industries-and-functions-big-data-can-revolutionize/?sh=7ea968c3a35b

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, *1*(2), 293–314. doi:10.1093/nsr/nwt032 PMID:25419469

Fan, J., & Liu, H. (2013). Statistical analysis of big data on pharmacogenomics. *Advanced Drug Delivery Reviews*, *65*(7), 987–1000. doi:10.1016/j.addr.2013.04.008 PMID:23602905

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of big data? Researchers' understanding of the phenomenon of the decade. *PLoS One*, *15*(2), e0228987. doi:10.1371/journal.pone.0228987 PMID:32097430

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27–34. doi:10.1145/240455.240464

Fortune Business Insights. (2022). *The global big data analytics market is projected to grow from $271.83 billion in 2022 to $655.53 billion by 2029*. Retrieved from https://www.fortunebusinessinsights.com/big-data-analytics-market-106179

Gangadharan, S. P. (2013). How can big data be used for social good? *Guardian*. Retrieved from https://www.theguardian.com/sustainable-business/how-can-big-data-social-good

Gartner. (2023). *Gartner glossary: Big data*. Retrieved from https://www.gartner.com/en/information-technology/glossary/big-data

Gartner. (2023a). *Gartner glossary: Big data*. Retrieved from https://www.sas.com/fi_fi/customers/boragen.html

General Data Protection Regulation. (2016, April 27). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119/1-88. art 5.

Gibson, D. C., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In *Big data and learning analytics in higher education* (pp. 29–42). Springer. doi:10.1007/978-3-319-06520-5_4

Google Cloud. (n.d.). *What is big data?* Retrieved from https://cloud.google.com/learn/what-is-big-data#section-2

Henke, N., Libarikian, A., & Wiseman, B. (2016). *Straight talk about big data.* McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/straight-talk-about-big-data#

Hernandez-Rojas, G., Maldonado-Guzmán, G., Hernández-Rodríguez, G., & Cortez-Murillo, E. (2018). Big data for smart cities: A systematic literature review. *IEEE Access*, *6*, 32292-32311. doi:10.1109/ACCESS.2018.2848354

Hewage, T. N., Halgamuge, M. N., Syed, A., & Ekici, G. (2018). Big data techniques of Google, Amazon, Facebook and Twitter. *Journal of Communication*, *13*(2), 94–100. doi:10.12720/jcm.13.2.94-100

Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, *34*(1), 135–174. doi:10.1111/dpr.12142

IBM Developer. (n.d.a). *Analytics: real-world use of big data in telecommunications*. Retrieved from https://www.ibm.com/thought-leadership/institute-business-value/report/big-data-telecommunications

IBM Developer. (n.d.b). *Big data analytics*. Retrieved from https://www.ibm.com/analytics/big-data-analytics

IBM Developer. (n.d.). *What is big data? More than volume, velocity, and variety*. Retrieved from https://developer.ibm.com/blogs/what-is-big-data-more-than-volume-velocity-and-variety/

Internet Live Stats. (2023). *Google search statistics*. Retrieved from https://www.internetlivestats.com/google-search-statistics/

Internet Live Stats. (2023a). *Internet users*. Retrieved from https://www.internetlivestats.com/internet-users/

Jiang, L., & Kavanagh, D. (2016). Big data analytics in supply chain management: A state-of-the-art literature review. *International Journal of Production Research*, *54*(23), 7323–7345. doi:10.1080/00207543.2016.1224782

Kannan, S., Karuppusamy, S., Nedunchezhian, A., Venkateshan, P., Wang, P., Bojja, N., & Kejariwal, A. (2016). Big data analytics for social media. *Big Data*, 63–94. doi:10.1016/B978-0-12-805394-2.00003-9

Kannan, S., Karuppusamy, S., Nedunchezhian, A., Venkateshan, P., Wang, P., Bojja, N., & Kejariwal, A. (2018). Big data analytics for social media. In *Big data analytics for social media* (pp. 31–49). Elsevier. doi:10.1016/B978-0-12-815856-4.00003-7

Kastouni, M. Z., & Lahcen, A. A. (2020). Big data analytics in telecommunications: Governance, architecture and use cases. *Journal of King Saud University-Computer and Information Sciences*, *32*(6), 6706–6712.

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.

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., & Gani, A. (2014). Big data: Survey, technologies, opportunities, and challenges. *TheScientificWorldJournal*, *2014*, 1–18. Advance online publication. doi:10.1155/2014/712826 PMID:25136682

Khanna, A., Gupta, D., & Dey, N. (Eds.). (2021). *Applications of big data in healthcare: Theory and practice*. Academic Press.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. doi:10.1177/2053951714528481

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1). doi:10.1177/2053951716631130

Krasnow Waterman, K., & Bruening, P. J. (2014). Big data analytics: Risks and responsibilities. *International Data Privacy Law*, *4*(2), 89–95. doi:10.1093/idpl/ipu002

Krishnan, K. (2019). Big data introduction. In *Building big data applications*. Academic Press.

Kshetri, N. (2014). Can blockchain strengthen the Internet of Things? *IT Professional*, *16*(2), 68–72. doi:10.1109/MITP.2014.35

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. *META Group Research Note*, *6*(70), 1-25.

Liu, Y. (2018). How big data analytics is changing the face of e-commerce. *Journal of Business Research*, *86*, 69–75.

Long, C. K., Agrawal, R., Trung, H. Q., & Van Pham, H. (2021). A big data framework for e-government in industry 4.0. *Open Computer Science*, *11*(1), 461–479. doi:10.1515/comp-2020-0191

Loshin, D. (2013). Developing a strategy for integrating big data analytics into the enterprise. *Big Data Analytics*, 29-37. . doi:10.1016/B978-0-12-417319-4.00004-1

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Mathur, V. (2022). Big data in banking industry: Benefits, uses and challenges. *Analyticstep.* Retrieved from https://www.analyticssteps.com/blogs/big-data-banking-industry-benefits-uses-and-challenges

McKinsey & Company. (2016). *Getting big impact from big data*. Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/getting-big-impact-from-big-data

McKinsey & Company. (2022). *The data-driven enterprise of 2025*. Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025

Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*, *114*, 57–65. doi:10.1016/j.ijmedinf.2018.03.013 PMID:29673604

Microsoft. (n.d.). What is big data analytics? *Azure*. https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-big-data-analytics/

Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big data, big analytics: Emerging business intelligence and analytic trends for today's businesses*. John Wiley & Sons. doi:10.1002/9781118562260

Narongou, D., & Sun, Z. (2022). Applying intelligent big data analytics in a smart airport business: Value, adoption, and challenges. In *Handbook of research on foundations and applications of intelligent business analytics* (pp. 216–237). IGI Global. doi:10.4018/978-1-7998-9016-4.ch010

Nasser, T., & Tariq, R. S. (2015). Big data challenges. *Journal of Computer Engineering & Information Technology*, *4*(3), 1–3. doi:10.4172/2324-9307.1000126

Neto, J. A. R. (2020). Google–big data success case. *Medium.com*. Retrieved from https://medium.com/xnewdata/google-big-data-success-case-949e1b99e678

Noel, S., Harley, E., Tam, K. H., Limiero, M., & Share, M. (2016). CyGraph: graph-based analytics and visualization for cybersecurity. In Handbook of Statistics (Vol. 35, pp. 117-167). Elsevier.

Oracle. (n.d.). *What is big data?* Retrieved from https://www.oracle.com/big-data/what-is-big-data/#how

Osadchuk, S. (2023). Top big data technologies in 2023: How they can benefit your business. *DOIT Software*. Retrieved from https://doit.software/blog/big-data-technologies#screen6.

Poola, D., Salehi, M. A., Ramamohanarao, K., & Buyya, R. (2017). A taxonomy and survey of fault-tolerant workflow management systems in cloud and distributed computing environments. *Software Architecture for Big Data and the Cloud*, 285-320.

ProdanoffJ. (2022). *Tech Jury*. Retrieved from https://webtribunal.net/blog/how-much-data-is-created-every-day/#gref

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59. doi:10.1089/big.2013.1508 PMID:27447038

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, *2*(1), 1–10. doi:10.1186/2047-2501-2-3 PMID:25825667

Sagiroglu, S., & Sinanc, D. (2013). Big data: a review. In 2013 international conference on collaboration technologies and systems (CTS) (pp. 42-47). IEEE. doi:10.1109/CTS.2013.6567202

SAS. (2023). *Big Data Analytics - What it is and why it matters*. Retrieved from https://www.sas.com/en_us/insights/analytics/big-data-analytics.html

SAS. (n.d.). *What is big data?* Retrieved from https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#technical

Satellh, G. (2013). Yes, big data can solve real world problems. *Forbes*. Retrieved from https://www.forbes.com/sites/gregsatell/2013/12/03/yes-big-data-can-solve-real-world-problems/?sh=7704437e8896

Statista Research. (2022). *Countries with the largest digital populations in the world as of January 2022*. Retrieved from https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/

Taylor, P. (2022, October 19). Big data - Statistics & Facts. *Statista*. Retrieved from https://www.statista.com/topics/1464/big-data/#topicHeader__wrapper

Thomsen, M. (2018). The advantages and disadvantages of big data in healthcare. *On-Line Journal of Nursing Informatics*, *22*(1), 1–13.

Tole, A. A. (2013). Big data challenges. *Database Systems Journal*, *4*(3), 31–40.

Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*, *2*(1), 1–32. doi:10.118640537-015-0030-3 PMID:26191487

United States v. Jones, 565 U.S. 400, 132 S. Ct. 945, 181 L. Ed. 2d 911 (2012).

van der Sloot, B., & van Schendel, S. (2016). Ten questions for future regulation of big data: A comparative and empirical legal study. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, *7*, 110.

van Dijk, J. (2012). The network society: Social aspects of new media. *Sage (Atlanta, Ga.)*.

Wachter, S. (2019). Data protection in the age of big data. *Nature Electronics*, *2*(1), 6–7. doi:10.103841928-018-0193-y

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. doi:10.21105/joss.03021

Weitzenboeck, E. M., Lison, P., Cyndecka, M., & Langford, M. (2022). The GDPR and unstructured data: Is anonymization possible? *International Data Privacy Law*, *12*(3), 184–206. doi:10.1093/idpl/ipac008

Xiong, J. (2013). Design and simulation of PID controller for ball and beam system based on LabVIEW. In *2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering* (pp. 291-294). IEEE. 10.1109/ICIII.2013.6562707

Yoganingrum, A., Rachmawati, R., & Koharudin, K. (2022). Past, present, and future of artificial intelligence in library services. In *Handbook of research on emerging trends and technologies in librarianship* (pp. 91–114). IGI Global. doi:10.4018/978-1-7998-9094-2.ch007

Zhuge, H. (2016). Introduction. *Multi-dimensional summarization in cyber-physical society*, 1-19. doi:10.1016/B978-0-12-803455-2.00001-9

ADDITIONAL READING

Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: A survey. *Future Generation Computer Systems*, *56*, 684–700. doi:10.1016/j.future.2015.09.021

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209. doi:10.100711036-013-0489-0

Dong, X., Li, G., Liu, Y., & Chen, H. (2016). Deep learning for event-driven stock prediction. In *Proceedings of the 2016 SIAM international conference on data mining* (pp. 104-112). Society for Industrial and Applied Mathematics.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, *1*(2), 293–314. doi:10.1093/nsr/nwt032 PMID:25419469

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future, 2007*(2012), 1-16.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In *Proceedings of the 2013 international conference on Advances in computing, communications and informatics* (pp. 225-232). IEEE. 10.1109/IC3.2013.6612229

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T., & Lapis, G. (2011). *Understanding big data: analytics for enterprise class Hadoop and streaming data*. McGraw-Hill Osborne Media.

ENDNOTES

[1] The authors prepare the figure based on the contents of the United Nations Global Pulse program. For details, see United Nations. (2023). Big Data for Sustainable Development. Retrieved from un.org/en/global-issues/big-data-for-sustainable-development.

[2] The authors collected this figure from a research report produced by Wayne Eckerson in 2011. For details, see Eckerson, W. (2011). Big data analytics: Profiling the use of analytical platforms in user organizations. BeyeNetwork. Retrieved from http://docs.media.bitpipe.com/io_10x/io_103043/item_486870/Big%20Data%20AnalyticsMarkLogic.pdf.

[3] The authors collected this figure from Microsoft Documentation. For details, see Microsoft. (2023). Real-time analytics on big data architecture. Retrieved from https://learn.microsoft.com/en-us/azure/architecture/solution-ideas/articles/real-time-analytics?source=recommendations.