A Study on Big Data Analytics

Mohana H K, Assistant Professor, Seshadripuram First Grade College, Bengaluru, India
Rajashri Padaki, Professor, Seshadripuram First Grade College, Bengaluru, India
Nethra H S, Assistant Professor, SEA College of Science, Commerce & Arts, Bengaluru, India
mohana1.amma@gmail.com

Abstract: A variety of humongous amount of data is being generated at an extremely fast pace in various sectors. Big data analytics is closely associated with statistics, data mining, machine learning, artificial intelligence, data science and systems. Further it is being adopted all throughout the globe to gain numerous benefits from the data being produced. Thus, analysis of huge amount of data popularly known as big data has become extremely inevitable and crucial. Hence, in this paper the attempt has been made to study the importance and role of big data analytics in the current digital era. Further the illustrations of how some companies enormously get benefited by applying big data analytics are presented. Ultimately, the different kinds of big data analytics that are used in various organizations to make accurate and right decisions are also discussed.

Keywords —Big Data Analytics, Data Volume, IoT, Massive Data, Organizations, zettabytes

I. INTRODUCTION

Recent advancement in digital technologies and broadband wireless systems have witnessed tremendous growth in wireless access networks and continues increase in mobile applications and data services. Further an extensive emergence of digital technologies and broadband wireless cellular networks have led to a new generation for information based technologies called industrial Internet of Things (IoT). In the modern world this IoT connects billions of smart devices such as sensors, actuators, and data processors etc. via broadband wireless access networks globally to process their real time data including environmental monitoring, industrial applications, business, human centric pervasive applications and so on. Further the mobile cellular networks have become both generators and carriers of massive data [1-2]. This new generation moreover provides unparalleled integrated computing capabilities to supply manufacturers with better wherewithal to extract value from an increasingly huge amount of data and gain a powerful competitive advantage [3].

Big data might be originated from innumerable Internet of Things (IoT) such as embedded sensors, appliances**,** smart phones, smart home security systems, smart factory equipment, ultra high speed wireless internet, wireless inventory trackers, biometric cyber security scanners etc.. In these IoT, the big data originates from internet transactions, emails, videos, clicks stream logs, mobile apps, social networks, customer databases, documents, medical records, etc.. Thus, the big data analytics is playing a pivotal role to analyze the big data obtained from both public and private sectors. This paper addresses on the benefits and challenges of big data analytics and different techniques that can be used to analyze the big data.

The rest of this paper is organized as follows: Section II outlines the overview of big data, Section III describes the architecture of big data analytics, the components of big data and big data in practice are described in section IV and V, different methods of big data analytics are illustrated in section VI and section VII concludes the paper.

II. OVERVIEW OF BIG DATA

Earlier the big data was limited to only few organizations such as Google, Yahoo, Microsoft and European Organization for Nuclear Research (CERN). However, nowadays due to rapid developments in IoT and digital technologies, the devices such as sensors, computer hardware and the Cloud, the storage and processing speed increasing and the cost is decreasing drastically. Therefore, many sources such as sensor appliances**,** smart phones, smart home security systems, smart factory equipment, ultra high speed wireless internet, wireless inventory trackers, biometric cyber security scanners etc start generating the data and organizations tend to store them for long time due to inexpensive storage and processing capabilities. Once the big data is stored, a number of challenges arise such as processing and analyzing [4-5].

In many perspectives like in retail business, consumer behaviour and preferences can be understood by analyzing the big data. It includes customer movement in the store or online website transactions or product searches. It is also vital to analyze the amount of data generated from various

 DOI : 10.35291/2454-9150.2022.0079

sources and to make real time decisions where money can be saved and operations can be more optimized in both public and private sectors [5]. The process of analyzing data sets to make precise decisions is known as data analytics. Figure 1 depicts the data analytics processes [6].

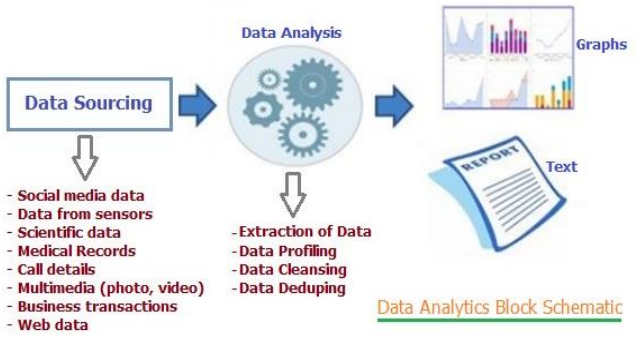

Figure 1: Schematic block diagram of big data analytics

As per [7] in 2022 around 18 billion IoT devices are connected globally and it is predicted to increase around 30.9 billion by 2025 [8]. As per [9] the amount of data consumed in 2021 worldwide is around 74 zettabytes. This enormous collection of data from the IoT units is called as big data. In 2022, the world would produce and consume around 94 zettabytes of data due to increase in the number of Internet of Things connected devices [10]. By 2025, around 175 zettabytes of data would be in the global data sphere [11]. Figure 2 and Table 1 show the data volume consumed and predicted in future [9-11]. Figure 3 show the volume of data produced in various organizations [12].

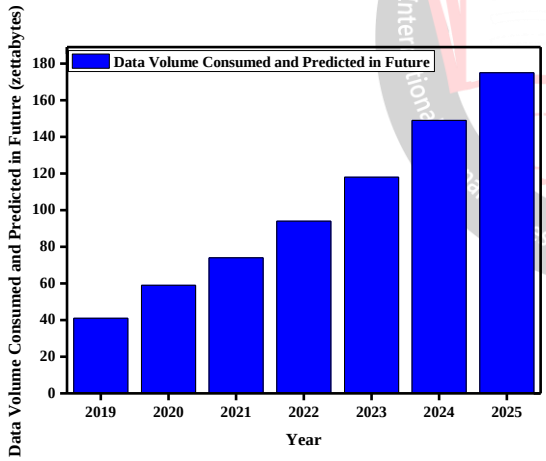

Figure 2: Data Volume Consumed and Predicted in the Future.

Table 1: Data Volume Consumed and Predicted in the Future

Year	Data Volume Consumed and Predicted in Future (zettabytes)
2019	41
2020	59
2021	74
2022	94
2023	118
2024	149
2025	175

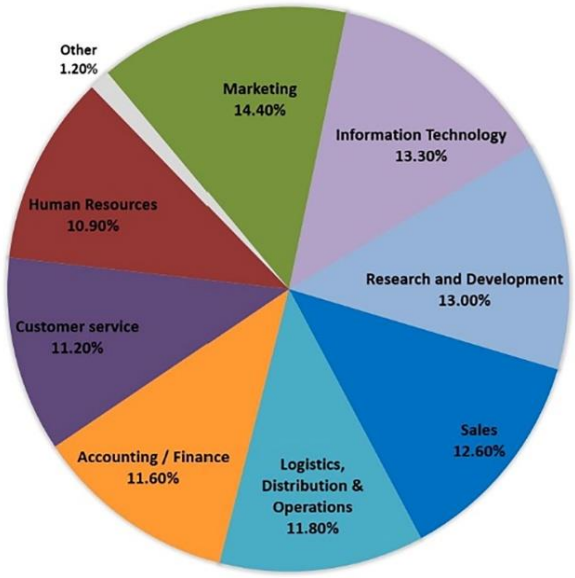

Figure 3: Big data analytics usage in different organizations [12]

III. Architecture of Big Data Analytics

Figure 4 show the architecture of big data designed to handle the ingestion, processing, and analysis of data specifically too large or complex for conventional database systems. This architecture comprises of data sources, data storage, batch processing, real time message ingestion, stream processing, analytical data store, analysis and reporting and orchestration units [13].

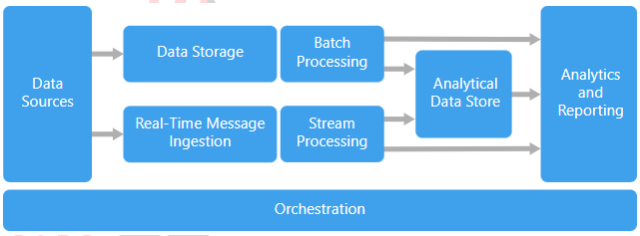

Figure 4: Architecture of big data analytics.

Data Sources: The real time data sources such as IoT devices, static files produced by applications such as web server log files and application data stores such as relational databases are the prime data sources.

Data Storage: Data for batch processing operations is typically stored in a distributed file store that can hold huge volumes of large files in various formats and this kind of storage is called a data lake.

Batch Processing: As the data sets are too large, the data files are processed using long running batch jobs to filter, aggregate and then the data is prepared for analysis. These jobs involve reading source files, processing them and writing the output to new files.

Real Time Message Ingestion: For real time sources the architecture includes stream processing unit to capture and store real time messages. This can be a simple data storage unit where incoming messages are dropped into a folder for processing. However, many solutions need a

 DOI : 10.35291/2454-9150.2022.0079

message ingestion store to act as a buffer for messages, and to support scale out processing, reliable delivery and other message queuing semantics.

STREAM PROCESSING: Once the real time messages are captured the data may be processed by filters and aggregated and later the data is prepared for analysis. The processed stream of data is then fed to an output sink.

ANALYTICAL DATA STORE: Many big data solutions prepare data for analysis and then serve the processed data in a structured format that can be queried using analytical tools.

ANALYSIS AND REPORTING: The goal of big data solutions is to provide insights into the data through analysis and reporting. To empower users to analyze the data the architecture may include a data modeling layer. Analysis and reporting can also take the form of interactive data exploration by data scientists or data analysts.

ORCHESTRATION: Most big data solutions consist of repeated data processing operations that transform source data move data between multiple sources and sinks load the processed data into an analytical data store or push the results straight to a report or dashboard.

IV. COMPONENTS OF BIG DATA

The components of big data comprises of volume, variety, value, velocity and veracity as shown in Figure 5 [12].

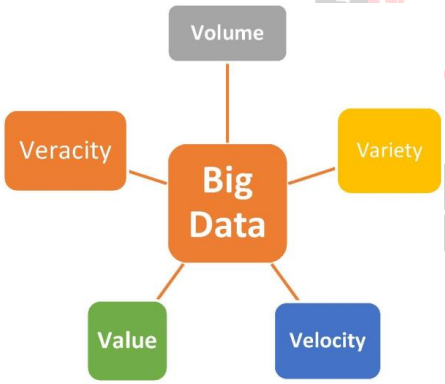

Figure 5: Big Data Components [12].

VOLUME: Volume represents amount of data that is being generated from various sources viz. computer systems, networks, social media, mobile phones, etc., either in structured like database or unstructured like human generated text/ speech or semi-structured like HTML [14]. Enterprises must implement modern business intelligence tools to effectively capture, store and process such unprecedented amount of data in real time [15].

VELOCITY: The rate at which the data is being generated, collected and analyzed is called velocity. The data is being continuously generated through multiple sources such as computer systems, networks, social media, mobile phones etc.. Nowadays in data driven business environment the pace at which data is being generated can be described as torrential and unprecedented. This data must be captured in real time to make accurate and right decisions. Most of the instances a limited amount of data available in real time yields better results compared to a large volume of data that needs a long time to capture and analyze [15].

VARIETY: The data is being generated from the different sources either in structured, unstructured or semi-structured formats. The relational databases may generate the data in structured form, the unstructured format includes audio, video, image files etc. and semi structured files are in the form of HTML files. However 80% of globally generated data is in unstructured format [15].

VALUE: The data is being generated in large volumes and simply collecting it is of no use. Instead data from which business insights are gained add value to the company for decision making process [15].

VERACITY: The veracity of big data represents the assurance of quality or credibility of the collected data. Since big data is vast and involves so many data sources, there is the possibility that not all collected data will be of good quality or accurate. Hence, when processing the big data sets it is essential to check the validity of the data before proceeding for processing [15].

V. BIG DATA IN PRACTICE

The big data can carry unlimited amount of information in single platform. It helps the organizations to make faster decision to meet several benefits to become smarter. The benefits of big data analytics are given below:

SMART AND FAST EFFECTIVE DECISION MAKING: The big data analytics helps the police departments to anticipate and identify the criminal activity before it occurs. With IoT data police can analyze crime patterns and trends. By applying predictive analytics and machine learning to vast sets of data, police departments can more easily forecast where and when violent crime will break out and ensure that they have the resources in place to prevent it [16]. Further accounting, auditing and financial advising firms use big data to recognize risks and frauds during audits [17].

IMPROVING HEALTHCARE AND PUBLIC HEALTH: The computing technique of big data analytics enables to decode entire DNA strings in minutes and will allow us to find new cures in order to for better understand and predict disease patterns [18].

OPTIMIZATION OF BUSINESS OPERATIONS BY ANALYZING CUSTOMER BEHAVIOUR: The big data has an enormous amount of digital information that businesses use to make data driven decisions which in turn improve business related outcomes. The benefits may include more effective marketing, new revenue opportunities, customer

 DOI : 10.35291/2454-9150.2022.0079

personalization and improved operational efficiency [19]. Further big data reveal patterns, trends and associations as they relate to human interactions and behaviors. Based on particulars instead of perception or past personal experience the evidence based big data information can be used to computerize processes, gain insight into aim audiences and improve performance using gladly accessible feedback. For example, Amazon is the recognized leader that uses big data to collect customer names, addresses, search histories, purchases and wish lists to improve customer services and to produce advertising algorithms [17].

COST REDUCTION: The big data processing uses Hadoop technique in which data will be located close to the processing node to minimize the communication overhead. Further, Hadoop stores the big data in a distributed fashion so that data is processed in parallel resulting in cost reduction [20].

OPTIMIZING MACHINE AND DEVICE PERFORMANCE: The big data analytics help machines and devices become smarter and more autonomous. For example, big data tools operate Google's self driving cars. The Toyota Prius is fitted with cameras, GPS and powerful computers and sensors to safely drive on the road without the intervention of human beings [18].

IMPROVING AND OPTIMISING CITIES AND COUNTRIES: The big data is used to improve many aspects of our cities and countries. For example, it allows cities to optimise traffic flows based on real time traffic information as well as social media and weather data [18].

VI. BIG DATA ANALYTICS METHODS

The big data analytics examines huge and various kinds of data from different sources to uncover hidden patterns, correlations and other insights. It helps the organizations to facilitate their growth and development by making better decisions. Various methods to analyze the big data are described below [21]:

DESCRIPTIVE ANALYSIS: This approach uses data aggregation and data mining techniques to provide insight into the past and explains what is happening now based on the incoming data. It describes or summarizes the raw data and interprets it. Google analytics tool is the best example for descriptive analysis. A business gets result from the web server through the tool which helps understand what actually happened in the past and validate if a promotion campaign was successful or not based on the basic parameters like page views. So it is an important source to determine the next action [21].

PREDICTIVE ANALYSIS: This approach uses statistical models and forecast techniques to predict the outcomes through sensors and other machine generated data. Companies can sense when the malfunction can occur and pre-emptively order parts and make repairs to avoid downtime and losses. For example southwest airlines analyses the sensor data on their planes in order to identify patterns that indicate a potential malfunction, thus allowing the airlines to the necessary repairs before the schedule without interrupting or putting the passengers in danger [21].

PRESCRIPTIVE ANALYSIS: This approach uses optimization and simulation algorithms to advise on the possible outcomes and prescribes the possible actions and guide them towards a solution. It uses business techniques, algorithms and machine learning techniques. Google's self driving car is a perfect example of prescriptive analytics. It analyses the environment and decides the direction based on data [21].

DIAGNOSTIC ANALYSIS: This approach takes a deeper route to find out the root cause of the events. It is helpful in mining the factors that contributed to a particular outcome. For example, for a social media marketing campaign, diagnostic tool can be used to assess the number of posts, mentions, followers, fans, page views, reviews, pins etc. and analyse the failure and success rate of a campaign at a fundamental level [21].

VII. CONCLUSION

Nowadays the advancement in digital technologies and broadband wireless networks have led to a new generation called IoT. The smart devices such as sensors, smart phones, smart home security systems, wireless inventory trackers, biometric cyber security scanners etc., generate variety of humongous amount of data. The big data analytics is closely associated with statistics, data mining, machine learning, artificial intelligence, data science and systems. It helps the several organizations to make precise decisions to meet several benefits. The big data analytics can help the police departments to anticipate and identify the criminal activity before it may occur. This helps the healthcare sector to find new cures and predict disease patterns. The enormous amount of digital information in the big data may help the businesses to make right decision in order to improve business related outcomes, effective marketing, new revenue opportunities and it allows optimizing the traffic flows based on real time traffic information.

REFERENCES

[1] C. Liang, F. R. Yu, and X. Zhang, "Information-centric network function virtualization over 5G mobile wireless networks," IEEE Network, Vol. 29, pp. 68–74, May 2015.

[2] J. Liu, F. Liu, and N. Ansari, "Monitoring and analyzing big traffic data of a large-scale cellular

 DOI : 10.35291/2454-9150.2022.0079

network with Hadoop," *IEEE Network*, Vol. 28, no. 4, pp. 32–39, 2014.

[3] Amine Belhadi, Karim Zkik, Anass Cherrafi, Yusof M. Sha'ri, "Understanding the capabilities of Big Data Analytics for manufacturing process: insights from literature review and multiple case study".

[4] Mohammad Masoud, Yousef Jaradat, Ahmad Manasrah,and Ismael Jannoud, "Sensors of Smart Devices in the Internet of Everything (IoE) Era: Big Opportunities and Massive Doubts", Hindawi Journal of Sensors Volume 2019.

[5] Arkady Zaslavsky, Charith Perera, Dimitrios Georgakopoulos "Sensing as a Service and Big Data" Proceedings of the International Conferences on Advances in Cloud Computing (ACC), Bangalore, India, July 2012, Pages 21-29

[6] https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-Data-Analytics.html

[7] https://www.ericsson.com/en/about-us/company-facts/ericsson-worldwide/india/authored-articles/ushering-in-a-better-connected-future

[8] https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/

[9] https://www.cloverdx.com/blog/how-much-data-will-the-world-produce-in-2021

[10] https://financesonline.com/how-much-data-is-created-every-day

[11] https://seedscientific.com/how-much-data-is-created-every-day

[12] Muhammad Anshari, Mohammad Nabil Almunawar, Syamimi Ariff Lim, Abdullah Al-mudimigh, "Customer Relationship Management and Big Data Enabled: Personalization & Customization of Services", Applied Computing and Informatics Volume 15, Issue 2, July 2019, Pages 94-101.

[13] https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data

[14] Petr Skoda, Fathalrahman Adam, "Knowledge Discovery in Big Data from Astronomy and Earth Observation"

[15] https://acuvate.com/blog/understanding-the-5vs-of-big-data/

[16] https://www.ibm.com/blogs/internet-of-things/big-data-crime-prevention/

[17] https://www.sygmatechnology.com/news-events/how-big-data-helps-in-decision-making

[18] https://bernardmarr.com/big-data-in-practice/

[19] https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics

[20] Pejman Lotfi-Kamran, Hamid Sarbazi Azad, "Data Prefetching Techniques in Computer Systems", Volume 125

[21] Iqbal H. Sarker, "Data Science and Analytics: An Overview from Data Driven Smart Computing, Decision-Making and Applications Perspective", SN Computer Science (2021) 2:377

 DOI : 10.35291/2454-9150.2022.0079