International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 11, pp: 171-176 www.ijaem.net ISSN: 2395-5252

Study of Big Data: Applications, Challenges and Future Scope.

[1,]Harshit Kumar Tyagi, [2,]Mayank Kumar Tyagi,
ABES Engineering College , Ghaziabad, Uttar Pradesh
ABES Engineering College, Ghaziabad, Uttar Pradesh.

Date of Submission: 05-12-2020 Date of Acceptance: 20-12-2020

ABSTRACT: Big Data is a collection of data that is huge in volume and is still growing exponentially with time. This data can be in structured, unstructured or semi-structured form. According to an estimate, around 1.145 trillion megabytes of data is generated per day from various sources such as social networking, online shopping, searching etc. Big Data can be used by organizations for better understanding the information that is hidden and will help to identify the data that is most important to the business and future business decisions. Conventional techniques cannot be used to analyse such huge amount of data, and thus advanced analytical techniques are required to review such data efficiently. The main objective of this research paper is to talk about the application of Big Data in today's world, the challenges that are faced by Big Data industry as well as Big Data professionals. In the end, this paper also talks about the future scope of Big Data. The hurdles of securing the data and democratizing it have been discussed in this paper such as inability of sound professionals and software ability to efficiently process daily growing data. Through the paper, the author intends to decipher the notion in an intelligible manner embodying in text several use-case and illustrations.

KEYWORDS: Big Data, 3V's, 5V's , Analytics, Business Intelligence, Big Data Challenges, Big Data Applications.

I. INTRODUCTION

"Every day, we create 2.5 quintillion bytes of data- so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few"[1]. This large volume of information both structured and unstructured, that overwhelm business on a daily basis is what is termed as Big Data.

At the beginning of last decade, IDC estimated that 1.2 zettabytes (1.2 trillion gigabytes) of new data were created in 2010, up from 0.8 zettabytes. The amount of newly created data in 2020 was predicted to grow 44X to reach 35 zettabytes (35 trillion gigabytes). Two years ago, we were already at 33 zettabytes, leading IDC to predict that in 2025, 175 zettabytes (175 trillion gigabytes) of new data will be created around the world[2].In order to harness this colossal amount of data and to gain business insights from it, more and more organisations are moving to Big Data. According to a Forbes report, Big Data adoption reached 53% in 2017 for all the companies interviewed, up from 17% in 2015[3]. The global Big Data market size is expected to grow from USD 138.9 billion in 2020 to USD 229.4 billion by 2025, at a compound Annual Growth Rate (CAGR) of 10.6% during forecast period [4].

In the present era, almost all the organizations, be it a start-up or a multi-national corporation, everyone is fueled by data. Thus, to draw useful insights from this data and to unveil the hidden patterns and underlying connections within this gigantic amount of data becomes very crucial. Conventional data analyzing techniques cannot be used to analyze this massive amount of data and therefore, advanced analytics techniques and software like Hadoop Framework are used to study and analyze Big Data.

There are plethora of Big Data uses in the world. There are a plenty of applications that are done using Big Data when it is analyzed. Several business decisions and predictions are done using Big Data analytics. These predictive models are very useful in various sectors like Supply Chain Management, Healthcare, Logistics, Weather prediction and so on. Big Data is also used to better study and comprehend the customers, their behavior, and preferences.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 01-05www.ijaem.net ISSN: 2395-5252

II. ARCHITECTURE OF BIG DATA

A Big Data is designed to handle the ingestion, processing, and analysis of data that is too large or complex for traditional database system [5]. The basis for Big Data Analytics is the architecture of Big Data. It is a comprehensive mechanism used to store and process vast volumes of data so that it can be processed and monitored for business purposes. For Big Data infrastructures and solutions, the architecture framework provides a reference model describing how Big Data solutions can operate, information flow, and security features.

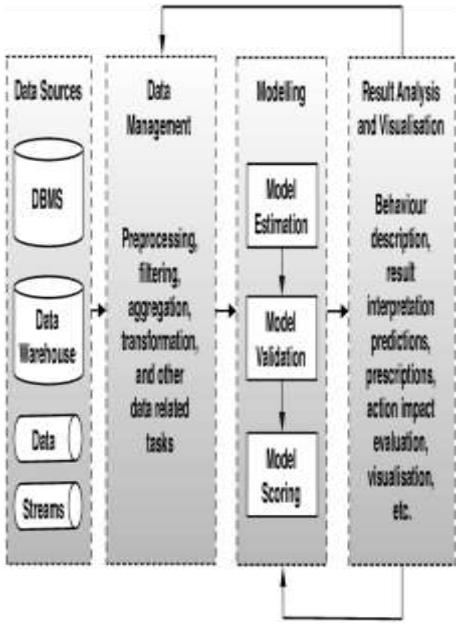

Figure 1. Overview of Big Data Analytics Workflow.

There are generally four logical layers in Big Data solution architecture. These layers are as following:-

Big Data Sources: There are several data sources for Big Data solution. Data can come through from company servers and sensors, or from third party data provider. A few data source example include enterprise applications like ERP or CRM, MS Office docs, data warehouses, mobile devices, social media, and email[6].

Data Messaging and Storage Layer: This layer receives data from the sources. If necessary, it converts unstructured data to a format that analytics tool can understand and store the data according to its format. The Big Data architecture might store structured data in a RDBMS, and unstructured data in a specialized file system like Hadoop Distributed File System(HDFS), or a NoSQL database [6]

Analysis Layer: The analytics layer interacts with stored data to extract business intelligence. Multiple analytics tools operate in the Big Data environment. Structured data supports mature technologies like sampling, while unstructured data needs more advanced (and newer) specialized analytics toolsets [6].

Consumption Layer: This layer receives analysis results and presents them to the appropriate output layer. Many types of outputs cover human viewers, applications, and business processes[6].

III. CHARACTERISTICS OF BIG DATA

Figure 2. Characteristics of Big Data

Figure 2. depicts the Big Data characteristics which is generally termed as a multi V model. Variety means different types of data records, Velocity represents the speed at which information is generated, and Volume defines the amount of information. Veracity is the trustability and reliability of information from the data sources.

- **Data Volume:** Volume represents the colossal amount of data that is generated continuously from various sources like sensors, social media sites, images, videos, transactions etc.

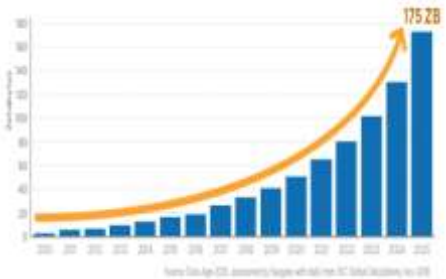

Figure 3. Expected Data Growth Over Years.

- **Data Velocity:** In terms of Big Data, velocity represents the speed at which information is generated. This speed of generating information is increasing every year due to technology advancements and more powerful

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 01-05www.ijaem.net ISSN: 2395-5252

hardware.

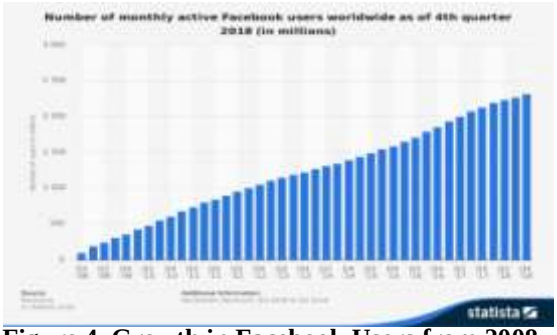

Figure 4. Growth in Facebook Users from 2008-18.

- **Data Variety:** Variety refers to the type of information (data) that is generated and stored. This data can be structured, unstructured or even semi-structured.

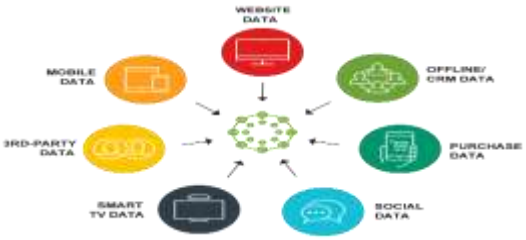

Figure 5. Variety of data collected from sources.

- **Data Veracity:** Veracity represents the extent to which the data source can be trusted for information. It refers to the quality of the information. With huge amount of data available in digital era, the accuracy of the data is always debatable.
- **Data Value:** It represents the end results of what can be done when all the data is analysed. It refers to useful business insights that are extracted from the data.

IV. APPLICATIONS OF BIG DATA

- **The Third Eye- Data Visualization:** Organizations worldwide are slowly and perpetually recognizing the importance of big data analytics. From predicting customer purchasing behavior patterns to influencing them to make purchases to detecting fraud and misuse which until very recently used to be an incomprehensible task for most companies big data analytics is a one-stop solution. Business experts should have the opportunity to question and interpret data according to their business requirements irrespective of the complexity and volume of the data. In order to achieve this requirement, data scientists need to efficiently visualize and present this data in a comprehensible manner. Giants like Google, Facebook, Twitter, EBay, Wal-Mart etc., adopted data visualization to ease complexity of handling data. Data visualization has shown immense positive outcomes in such business organizations. Implementing data analytics and data visualization, enterprises can finally begin to tap into the immense potential that Big data possesses and ensure greater return on investments and business stability[7].

Integration- An exigency of the 21st century: Integrating digital capabilities in decision-making of an organization is transforming enterprises. By transforming the processes, such companies are developing agility, flexibility and precision that enables new growth. Gartner described the confluence of mobile devices, social networks, cloud services and big data analytics as the as nexus of forces. Using social and mobile technologies to alter the way people connect and interact with the organizations and incorporating big data analytics in this process is proving to be a boon for organizations implementing it. Using this concept, enterprises are finding ways to leverage the data better either to increase revenues or to cut costs even if most of it is still focused on customercentric outcomes.Such customer-centric objectives may still be the primary concern of most companies, a gradual shift to integrating big data technologies into the background operations and internal processes[8].

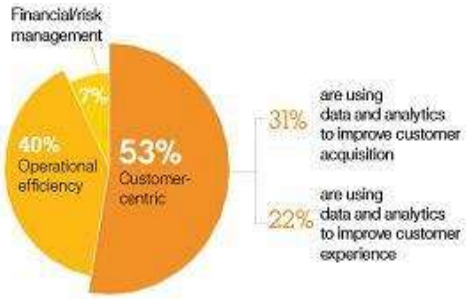

Figure 6.Analysis as generated by IBM institute of Business Value 2014 Anlytics study.

- **Big Data in Heathcare**: There are multiple advantages in the implementation of big data analytics in the healthcare database. The healthcare system sources its information from multiple sources which is comprised of various types of data assessment of the diverse and complex data is difficult during the process of database management and tools and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 01-05www.ijaem.net ISSN: 2395-5252

approaches. Since there is a need for fast development in computing technology, big data presents the optimum solution to efficiently use the significant value of the accumulated data. The big data is the most appropriate answer to the healthcare industry and facilitates increased quality of human life. The major objective of big data is not only to gain more profits but also minimize the time and wastages accruing from time and other wastages and predicting of the disease outbreaks and their cure. This, in turn, aids the society to lead a quality life [9].

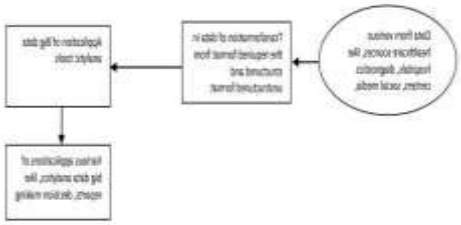

Figure 7. Conceptual model to utilize big data analytics in healthcare system

- **Big Data in Finance:** The finance industry generates lots of data. Structured data is information managed within an organization in order to provide key decision-making insights. Unstructured data exists in multiple sources in increasing volumes and offers significant analytical opportunities [10]. Big data is completely revolutionizing how the stock markets across the world are functioning and how investors are making their investment decisions. Machine learning – the practice of using computer algorithms to find patterns in massive amounts of data – is enabling computers to make accurate predictions and human-like decisions when fed data, executing trades at rapid speeds and frequencies [11]. The business archetype monitors stock trends in real-time and incorporates the best possible prices, allowing analysts to make smart decisions and reducing manual errors that arise due to behavioral influences and biases. In conjunction with big data, algorithmic trading is thus resulting in highly optimized insights for traders to maximize their portfolio returns [11]

  In general, big data is set to revolutionize the landscape of Finance and Economy. Several financial institutions are adopting big data policies in order to gain a competitive edge. Complex algorithms are being developed to execute trades through all the structured and unstructured data gained from the sources. The methods adopted so far has not been completely adept, however, extensive research ensures growing dependence of the stock markets, financial organizations and economies on big data analytics [8].

Figure 8. Applications of Data in Finance Industry.

- **Big Data in Fraud Detection:** Big data fraud detection is a cutting-edge way to use consumer trends to detect and prevent suspicious activity. Even subtle differences in a consumer's purchases or credit activity can be automatically analyzed and flagged as potential fraud. Using data analytics todetect fraud requires expert knowledge and computer resources, but is easier than ever, due to improvements in programming languages and server technology[12].

A survey finds that 42% of companies with revenues between US$100 million to US$1 billion are reviewing less than 10,000 records. And 71% companies with more than US$1 billion in sales report examining just one million records or fewer….Companies know there are high risk numbers in book entries, such as round thousands or duplicates, but they're only just starting to analyze descriptions for those book entries. Looking at both the numbers and words can mean the difference between uncovering fraud, and falling victim to it. The combination of appropriate data and big data analytics can help combat fraudulent activities. Though several companies are mining big data for this purpose there are still limitations in their approach. They are either keeping the data siloed, limiting the analysis to be performed or only taking into consideration the structured data thus only giving a subset of information. A more holistic approach to the implementation of big data analytics is required. Companies such as Pactera is developing solutions which will process massive amounts of structured and unstructured data and develop varied models and algorithms to find patterns of fraud and anomalies and predict customer behaviour [8].

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 01-05www.ijaem.net ISSN: 2395-5252

- **Big Data in Education:** Big data is used quite significantly in higher education. For example, The University of Tasmania. An Australian university with over 26000 students has deployed a Learning and Management System that tracks, among other things, when a student logs onto the system, how much time is spent on different pages in the system, as well as the overall progress of a student over time[13]. With the help of big data, customized programs for each individual student can be created. Even if colleges and universities have lakhs of students, customized programs can be created for each of these students. This is possible with the help of what is called as 'blended learning' – a combination of online and offline learning. This gives students the opportunity to follow classes that they are interested in and also work at their own pace, while still having the possibility for offline guidance by professors. We already see this happening in the case of MOOCs that are developed and delivered around the world now. For example, when the Machine Learning class at Stanford University was taught by Andrew Ng, only 400 students participated. However, when the same course was delivered as a MOOC, it attracted 100,000 students[14].

Figure 9.Application of Big Data in Education[14]**.**

V. CHALLENGES IN BIG DATA IMPLEMENTATION

The four major challenges that are faced by organisations in implementation of Big Data are:-

- **Maintaining Data Integrity, Security and Privacy:** An enterprise can use big data tools to collect structured and unstructured data from varied sources. The tools even enable businesses to store, process and analyze efficiently. But the Facebook-Cambridge Analytica data scandal makes it essential for businesses to focus on keeping the entire big data lifecycle secured to retain customers and avoid legal hassle. The businesses must implement a robust security strategy to collect, store, analyze, manage, and utilize large volumes of real-time data by eliminating risks of data breaches and exposure. They also need to keep the big data security strategy flexible and dynamic enough to address the new security and privacy issues being generated due to the constant increase in data volume[15].
- **Handling a Large Amount of Data**:There is a huge explosion in the data available. Look back a few years, and compare it with today, and you will see that there has been an exponential increase in the data that enterprises can access. They have data for everything, right from what a consumer likes, to how they react, to a particular scent, to the amazing restaurant that opened up in Italy last weekend[16].
  This data exceeds the amount of data that can be stored and computed, as well as retrieved. The challenge is not so much the availability, but the management of this data. With statistics claiming that data would increase 6.6 times the distance between earth and moon by 2020, this is definitely a challenge[16].
- **Uncertainty of Data Management Landscape:** With the rise of Big Data, new technologies and companies are being developed every day. However, a big challenge faced by the companies in the Big Data analytics is to find out which technology will be best suited to them without the introduction of new problems and potential risks[17].
- **Shortage of Skilled Worker**:There is a definite shortage of skilled Big Data professionals available at this time. This has been mentioned by many enterprises seeking to better utilize Big Data and build more effective Data Analysis systems. There is a lack experienced people and certified Data Scientists or Data Analysts available at present, which makes the "number crunching" difficult, and insight building slow[16]. Again, training people at entry level can be expensive for a company dealing with new technologies. Many are instead working on automation solutions involving Machine Learning and Artificial Intelligence to build insights, but this also takes well-trained staff or the outsourcing of skilled developers[16].

VI. FUTURE SCOPE AND DEVELOPMENT

Today, Big Data is influencing IT industry like few technologies have done before. The massive data generated from sensor-enabled machines, mobile devices, cloud computing, social media, satellites help different organizations improve their decision making and take their business to another level. "Big data absolutely has the potential to changethe way governments, organizations, and academic institutions conduct business and make discoveries, and its likely tochange how everyone lives their day-to-day lives," - Susan Hauser, corporate vice president of Microsoft.

As the technology becomes more complex and integrates with other technology, the industry will continue to evolve and require specialized experts for niche job roles and responsibilities. Coders will be required to upskill to Big Data analytics and managers will be asked to upskill to business intelligence and orient themselves with anything to do with data for business and career growth. This sector of IT has just exploded, and the major expansion after this big bang is yet to happen. If you are looking at a career in Big Data – a rewarding one – it is the perfect time to pull up your socks and get started with it because, in the coming years, it will need experts like you[18]. According to IDC analysts, "**Total revenues from big data and business analytics will rise from $122 billion in 2015 to $187 billion in 2019**." Business spending on big data will surpass $57 billion dollars this year. Although, the business investments in big data might vary from industry to industry, the increase in big data spending will remain consistent overall. Manufacturing industry will spend the most on big data technology while health care, banking, and resource industries will be the fastest to adopt[19].

IDC predicts that half of the business analytics software will incorporate prescriptive analytics build on cognitive computing functionality. This will help businesses to make intelligent decisions at the right time. With intelligence built into the software, you can sift through large amounts of data quickly and get a competitive advantage over your competitors[19].

VII. CONCLUSION

This literature survey discusses Big Data from its very basic definition, to the challenges faced by Big Data industry as well as its future scope and development. The field of Big Data is only going to evolve. The ability of Big Data to store and analyze information is only going to better in the future. With new technologies and tools coming in, the field of Big Data will surely be revolutionized in future.

REFERENCES

[1]. https://medium.com/@JJThaeler/the-evolution-of-data-solutions-1885a9fccb80
[2]. https://www.forbes.com/sites/gilpress/2020/01/06/6-predictions-about-data-in-2020-and-the-coming-decade/?sh=164a0a094fc3
[3]. https://onlinemasters.ohio.edu/college-of-business/masters-business-analytics/revenue-growth-through-business-analytics-the-ultimate-guide
[4]. https://www.marketsandmarkets.com/PressReleases/big-data.asp
[5]. https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data
[6]. https://www.datamation.com/big-data/big-data-architecture.html
[7]. Samiddha Mukherjee, Ravi Shaw. Big Data-Concepts, Applications, Challenges. IJARCCE, Volume 5, Issue 2, 2015.
[8]. T. Karunakar, M. Nagarjuna, K. Hema. Concepts, Applications and Future Scope of Big Data. International Journal of Scientific & Engineering Research. Volume 8, Issue 5, 2017.
[9]. https://www.ukessays.com/essays/information-technology/big-data-in-the-healthcare-sector.php
[10]. https://www.talend.com/resources/big-data-finance
[11]. https://corporatefinanceinstitute.com/resources/knowledge/other/big-data-in-finance/
[12]. https://www.syntelli.com/using-big-data-financial-fraud-prevention
[13]. https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications
[14]. https://www.allerin.com/blog/4-ways-big-data-is-transforming-the-education-sector
[15]. https://www.bluepiit.com/blog/challenges-in-big-data
[16]. https://www.dataversity.net/four-common-big-data-challenges
[17]. https://elearningindustry.com/big-data-analytics-challenges-faced-business-enterprises-7-top
[18]. https://www.digitalvidya.com/blog/big-data-future/
[19]. https://www.smartdatacollective.com/future-big-data-predictions/