ISSN: 0974-5823 Vol. 7 No. 1 January, 2022

International Journal of Mechanical Engineering

BIG DATA ISSUES AND CHALLENGES-A PROSPECTIVE SOLUTION

[1]*Ikhlas Ahmad Sheikh*

[1]*MCA,*

[2]*Firdoos Ahmad Wani*

[2]*MCA,*

[3]*Dr. Mujtaba Ashraf Qureshi*

Abstract: *With the emergence of cloud computing, internet of things (IOT), web 2.0 technologies there has been huge growth in the amount of data generated. In this research work the author tries to collect and capture the major challenges and issues faced to big data. Finally researcher proposed some of the suitable measures and solutions to the above mentioned problems of big data. Big data is an all-en compassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, and manage, sharing, storage, transfer, visualization, and privacy violations and process data within a tolerable elapsed time. The author also presents some prospective solutions for big data challenges.*

Key words: *Big data, Volume, Velocity, Hadoop.*

1. INTRODUCTION

Today's World is of digitalization where everyone uses the Internet and generates a huge amount of data. The usage of data in world has been increasing from a common man to multinational organizations. This data is generated from various sources like sensors used to gather climate information, digital pictures and videos, social media platforms (Facebook, Instagram, and YouTube etc) that led to the rise of big data. In nutshell big data refers to the collection of large, complex and unstructured amount of data that has the potential to be mined for information. Big Data decodes previously untouched data to derive new insight that gets integrated into business operations. Big data is an all-en compassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, and manage, sharing, storage, transfer, visualization, and privacy violations and process data within a tolerable elapsed time. Figure 1 *(source: Google images)* presents an overview of various steps followed in the mechanism of big data.

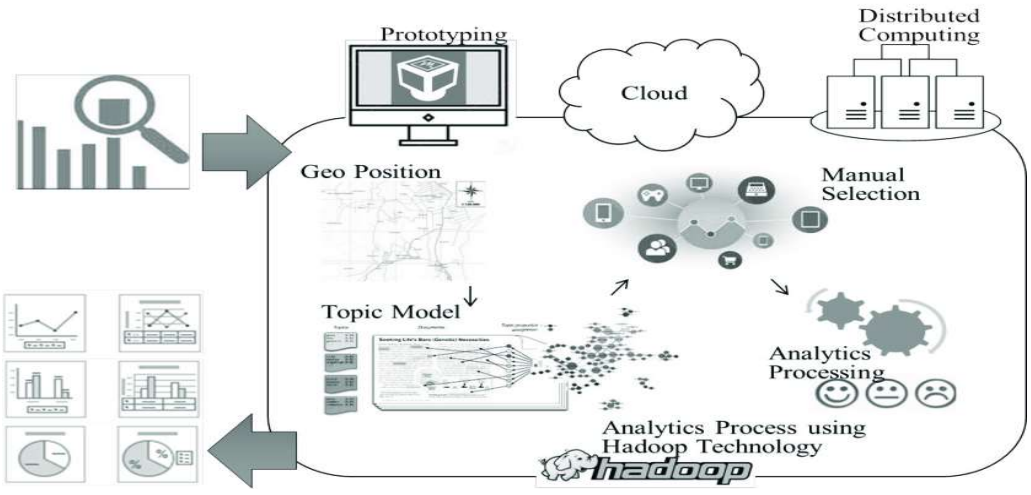

Figure 1: Big Data Steps.

Big data: Big data can be described as the voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.

Big Data can be simply defined by explaining the 3V's – volume, velocity and variety which are the driving dimensions of Big Data quantification. Gartner analyst, Doug Laney [1] introduced the famous 3 V's concept in his 2001 Metagroup publication, 3D data management: Controlling Data Volume, Variety and Velocity'. Figure 1 depicts the 3 V's of big data.

Figure 2: Three (03) V's of big data

Each of the characteristics of big data (shown in figure 1 is briefly described below.

Volume: It refers to the amount of data generated from business transactions, social media platforms, online applications etc. The amount of data generated is larger than the petabytes or even exabytes.

Variety: Data comes in all types of formats such as structures, semi-structured and unstructured forms (text, numbers, images, videos, sensor data etc.). Variety is all about the ability to classify the incoming data into various categories.

Velocity: Velocity refers the speed at which the data are being generated. It also means that how fast data is being produced and how fast data is being processed.

The other two V"s are Veracity and value.

Veracity*:* It refers to the uncertainty and inconsistencies shown by the data. Data in huge amount could create confusion.

Value*:* It is the ability to transform huge amount of data into business. The bulk of data having no value is of no good to the company until you turn it into something useful.

2. REVIEW OF LITERATURE

Big data is a growing technology due to the fact that huge amount of data is generated every day. About 2.5 quintillions bytes of data are generated in one day. According to a report about 90% of the world data was generated in last two years. [2] Some of the examples that how the amounts of data every single day.

In recent times social media generates data in exabytes per day. According to Domo's Data Never Sleeps 5.0 report [3] – these are numbers generated every minute of the day:

- Snap chat users share 527,760 photos
- More than 120 professionals join LinkedIn
- Users watch 4,146,600 YouTube videos
- 456,000 tweets are sent on Twitter
- Instagram users post 46,740 photos

The New York Stock Exchange captures 1 terabyte of information each day. By 2016, there were an estimated 18.9 billion network connections, with roughly 2.5 connect per person on Earth. [4]

According to GE (General Electric), each of its aircraft engines produces around one terabyte of data per flight. Added together, GE now has access to up to 50m variables from 10m sensors. [5]

3. EXISTING TECHNOLOGIES

Big data technology is defined as the technology and a software utility that is designed for analysis, processing, and extraction of the information from a large set of extremely complex structures and large data sets which are very difficult for traditional systems to deal with. Big data technology is used to handle both real-time and batch related data.

Hadoop: Hadoop helps in the processing of batch related jobs and process batch information which is based on map-reduce. It can be used to store and analyze the data present in various different machines with high storage, speed, and low costs. It was designed to store and process the data in a distributed data processing environment along with commodity hardware and a simple programming execution model.

Splunk*:* Splunk is used to capture, correlate and index real-time streaming data from a searchable repository from where it can generate reports, graphs, dashboards, alerts and data visualizations. It is also used for security, compliance and application management and also for web analytics, generating business insights and business analysis. It was developed by Splunk in Python, XML, and Ajax.

MongoD *:* Another very essential and core component of big data technology in terms of storage is the Mongo DB NoSQL database. It is a NoSQL database which means that the relational properties and other RDBMS-related properties do not apply to it. It is different from traditional RDBMS databases which make use of structured query language. Figure 3 *(source: Google images)* presents an overview of various technologies employed in big data.

Figure 3: Big Data Technologies.

4. SCOPE

"Big data absolutely has the potential to change the way governments, organizations, and academic institutions conduct business and make discoveries, and its likely to change how everyone lives their day-to-day lives," said Susan Hauser, a corporate vice president at Microsoft. [9]

a) Increasing enterprise adoption of Big Data- more and more organizations across the globe are adopting big data technologies to enhance their performance.

b) According to a study by Wanted Analytics (2015), the biggest significant demand for Big Data professionals is by Professional, Scientific and Technical Services (25%), Information Technology (17%), Manufacturing (15%), Finance and Insurance (9%), and Retail Trade (8%).[10]

c) By 2017 unified data platform architecture will become the foundation of BDA strategy. The unification will occur across information management, analysis, and search technology.

d) IDC Big Data and Analytics 2015 Predictions includes the following.[11]

d.1. Visual data discovery tools will be growing 2.5 times faster than rest of the business intelligence (BI) market.

d.2. Growth in applications incorporating advanced and predictive analytics, including machine learning, will accelerate in 2015. These apps will grow 65% faster than apps without predictive functionality.

d.3. By 2018, investing in this enabler of end-user self service will become a requirement for all enterprises. Over the next five years spending on cloud-based Big Data and analytics (BDA) solutions will grow three times faster than spending for on-premise solutions.

5. ISSUES AND CHALLENGES

Big data is the act of gathering and storing large amount of data and information for analysis. Big data challenges includes the easiest and best way of handling the large amount of data that involves the process of storing, analyzing huge set of information on various data sets. There are number of challenges that come into the way while dealing with big data which needs to be addressed.

- Volume of data- due to the volume of data the first challenge is storage. As the data increases the amount of storage needed to store the data also increases.
- For retrieving the stored data at fast speed to get better results. Networking, bandwidth, cost of storage are other issues that needs to b looked after. [6]
- As the data volume of the data increases its value tends to be decrease in proportion of quality, type, richness etc.[7]
- Being new to big data and its management is the biggest challenge users of big data face. As organizations are new to big data it typically has inadequate data analysts and IT professionals having the skills to help interpret digital marketing data[8]
- Data processing is another big challenge that requires analytic algorithms and parallel processing in order to generate rapid information. It also includes finding out data points that are really important.
- Lack of proper understanding and knowledge professionals also needs to be addressed.
- Dealing with data integration and preparation issues.

6. PROSPECTIVE SOLUTIONS

The speed at which data is generated is surprising. It is quite problematic to handle unstructured data using traditional data bases e.g. relational database, spreadsheet etc. The solution is using unstructured data analytic tools. *Top 6 unstructured data analytics tools:[12]*

I. MonkeyLearn | All-in-one data analytics and visualization tool

II. Excel and Google Sheets | Organize data and perform basic analyses

III. RapidMinder | All-around platform for predictive data models

IV. KNIME | Open-source platform for advanced, personalized design

V. Power BI | Business intelligence leader from Microsoft

VI. Tableau | Visualization tool with a number of business integrations

- Big data workshops and seminar needed to be conducted by an organization for everyone.
- Basic training programs must be arranged for all the employees who are handling data regularly and are a part of the Big Data projects.
- In order to handle these large data sets, companies are opting for modern techniques, such as compression, tiering, and deduplication. Compression is used for reducing the number of bits in the data, thus reducing its overall size. Deduplication is the process of removing duplicate and unwanted data from a data set. Companies can also use big data tools such as Hadoop, NoSQL etc.
- Big data security can be achieved by hiring more cyber security professionals in order to secure their data. They can also use some security mechanism such as-

I. Data encryption

II. Identity and access control.

III. Big data security tools[13]

- IBM security Guardium is used to monitor Big data and NoSQL environments.
- CDAP (Cloudwick Data Analytics Platform) is a managed security hub that integrates security features from multiple analytics toolsets, traditional IDS and IPS, and machine learning projects.
- Gemalto SafeNet protects big data platforms. Usually, it protects the big data platforms in the cloud, data center, and virtual environments. The toolset of security includes digital signing solutions, data encryption, strong authentication, and cryptographic key security management.

7. CONCLUSION AND FUTURE WORK

From the last few years the data has been generated in huge amount. Processing and analyzing the data is challenging for an individual person. In this paper we addressed some challenges, scope and prospective solutions. Big Data is an evolving field, where much of the research is yet to be done. Big data at present is handled by the software named Hadoop. However the rapid growth of data makes Hadoop insufficient. To harness the potential of Big Data completely in the future, extensive research needs to be carried out and revolutionary technologies need to be developed.

8. REFERENCES

1. http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/\
2. https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=46d7bf9860ba
3. Data never sleep-https://www.domo.com/learn/infographic/data-never-sleeps-5
4. Available at-https://insight.harlandclarke.com/2017/12/5-challenges-for-financial-institutions-to-overcome-when-it-comes-to-big-data/GE data analytics https://www.aerosociety.com/news/digital-takeover
5. www.coursera.org, Introduction to Big Data, University of California, San Diego. https://www.coursera.org/learn/big-data-introduction
6. http://www.slideshare.net/HarshMishra3/harsh-big-data-seminar-report. Published: 4th January 2014 in Technology, Education Harsh Kishore Mishra. Center for Computer Science and Technology. School of Engineering and Technology, Central University of Punjab, Bhatinda
7. http://www.dataversity.net/common-big-data-management-issues-solutions/ The Most Common Big Data Management Issues (And Their Solutions). By: A.R. Guess. July 15 2014.
8. The Big Bang: How the Big Data Explosion Is Changing the World https://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/
9. Available at- https://www.forbes.com/sites/louiscolumbus/2015/06/25/where-big-data-jobs-are-in-2015-midyear-update/?sh=3cc017b42050
10. https://www.cfo-india.in/cmsarticle/spending-on-cloud-based-big-data-and-analytics-solutions-will-grow-three-times-faster/
11. Unstructured data analysis tools- https://monkeylearn.com/blog/unstructured-data-analytics/
12. Big data security tools- https://techvidvan.com/tutorials/big-data-security/