iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** II **Month of publication:** February 2018

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

Big Data: A Review

Sakshi Saini[1], Amita Dhankkar[2], Suryansh Dabas[3]
[1]*M. Tech (CSE) 4th Sem, UIET, M.D University, Haryana, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, UIET, M.D University, Haryana, India*
[3]*B.Tech, MSIT, Delhi*

Abstract: ***Big data plays a very crucial role in different fields of the modern world. Big data term is used for the data that is massive, varied and complex structure having the difficulties in collecting, storing, processing, analyzing and visualizing. Research which is to be processed in the direction of revealing the hidden patterns and the correlations between the different types of the data is named as Big Data Analytics or BDA. For the better decision making, for utilizing these useful information or for taking the better insights in the organizations or the company's big data analytics is used. For this reason the analysis and execution of the big data implementation is needed. This paper aims to provide overview about the contents of the big data, its characteristics, big data analytics phases and the tools and techniques used during the different phases of the analysis.***
Keywords: Big Data, Volume, Variety, Velocity, Variability, Veracity, Big Data Analytics phases.

I. INTRODUCTION

Big data and its analysis are at the centre of the modern science and business. In previous years the data was used to be stored on the hard disks, floppy disks, CDs, tape storages etc. But as the internet gains the popularity generation of the data becomes so huge that it becomes difficult to store, process and analyze. This huge amount of data can be generated from different types of sources like web, log files, sensors, multimedia files, social networking sites, online transactions etc. The volume of data increases by 5 exabytes (10^{18} bytes) in 2003 to 2.72 zettabytes in 2012. Today 5 exabytes amount of data is generated within 2 days. With the introduction of the social networking sites through which people usually communicates the amount of generation of the multimedia files grows enormously. The Human Face Of Big Data a global project carried out in 2012 whose objective is to visualize and analyze the large amount of data derives some statistics according to which Facebook has 955 million monthly active accounts using 70 languages, 140 billion photos uploaded, 125 billion friend connections, everyday 30 billion pieces of content and 2.7 billion likes and comments have been posted. Similarly the statistics were also derived from Youtube, Google, Twitter like sites [1]. Hence the growth of data rose from terabytes, petabytes to zettabytes. Big data is used in understanding and targeting customers, understanding and optimizing Business processes, personal quantification and performance optimization, improving health care and public health, improving sports performance, improving science and research, financial trading etc.[2]This article is worded as follows: Section II presents the brief concepts about the big data and the big data analytics. In Section III, the related works are discussed. Section IV concludes the work.

II. BIG DATA

Big data is defined as an extremely large data sets or the collection of the huge amount of the heterogeneous data that may be structured, unstructured or semi-structured. Big data term refers to the use of the user behavior analytics, predictive analytics or certain other analytics that extract the valuable data from large data sets [3]. IDC predicts that by 2025 there will be 163 zettabytes of generated data [3]. Big data can be represented by the **5v's**. These 5V's describes the characteristics of the big data. These 5V's are [4]:

1) Velocity: speed with which the data is generated and speed with which data moves around or the speed with which the data is processed. For example data generated on social networking site facebook on which 140 billion photos are being uploaded daily [1].
2) Volume: it represents the large amount of data generated from different sources. These sources can be multimedia files, log files, sensors etc.
3) Variety: it represents the heterogeneous data that is generated at a large scale. The data can be structured, unstructured or semi-structured.
4) Veracity: represents the quality of data that is collected can vary greatly that may affect the proper analysis of the data.

5) Value: it's not worthy if we can't convert the raw information into the some valuable or useful information. This characteristic describes about the quality of data that we retrieve from the raw information.

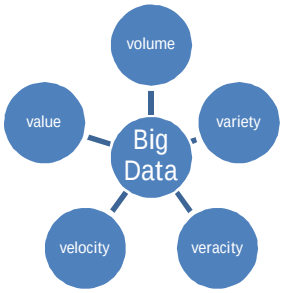


Figure 1: 5V's of Big Data

These 5 characteristics explain about the big data. Big Data Analytics or BDA helps in examining the large amount of data so that it can uncover the hidden patterns, correlations between the different data and other insights [5]. This examining of the data took place in different phases which requires various tools and techniques to do so. BDA helps in effective cost reduction, faster and better decision making in the business and new products and services can be designed and provided on the basis of the analysis of the data [5]. These different phases help in collecting, cleansing and processing of the data.

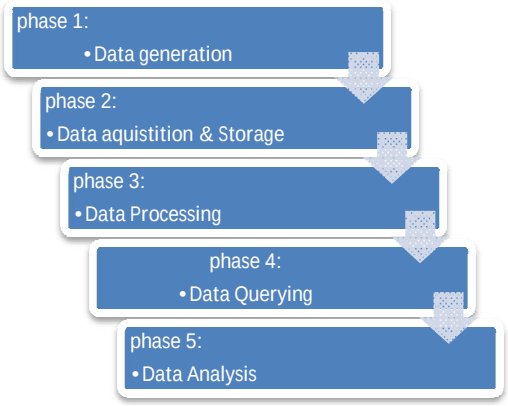


Figure 2: Big Data Analytics phases

6) *Phase 1: Data generation*: at this phase the data is being collected from the different sources like sensors, IoT devices, log files, web servers, a group of people or from the community. For this purpose Parallel Data Generation Framework tool is used to generate and distribute the data. This generation process of data is continuous.

7) *Phase 2*: *Data Acquisition & Storage:* Data acquisition is the process of gathering, filtering and cleaning of the data before putting it into a data warehouse or any other storage solution. Software tools used during data acquisition phase is Storm that consists of three nodes: Nimbus, Zookeeper, Supervisor Nodes. Other tool used are Kafka, Flume, Hadoop Common, Hadoop Distributed File System(HDFS), Hadoop YARN, Hadoop Map Reduce. Data Storage is a storage infrastructure which is specially designed for storing, managing and extracting the massive amount of data [6]. For the storage purpose of Big Data Hadoop, NoSQL and Cassandra analytics engines are used. Apache Hadoop Distributed File System is most used analytics engine which is combined with the flavor of the NoSQL database [7].

8) *Phase 3: Data Processing:* for the processing of the data Map Reduce component of the Apache Hadoop is used. It is the processing pillar of the Hadoop having the two functions Map and Reduce which splits the data into independent chunks to process, sort and retrieve. Big data techniques used to process the data are reporting, batch analytics, online analytical processing, data mining, text mining, complex event processing(CEP), predictive analysis etc. Tools used are Google Chubby, Apache Hadoop, HDFS (Hadoop Distributed File System), Hadoop YARN, MPI(Message Passing Interface), Spark, Kafka, Apache Flume, Apache Chukwa, Facebook Scribe etc.

9) *Phase 4: Data Querying:*the data which is stored and processed in the previous steps is retrieved. The data is gathered from various sources and aggregated with the help of the HDFS and Map Reduce. Data tools used are HIVE for data summation, querying and analysis, IMPALA allows user to perform the low latency queries effectively, HAWQ big tasks are divided into the smaller ones and these smaller tasks are distributed to the MPP SQL processing unit for execution, Drill it can handle up to 10,000 servers for the efficient querying it supports HBase, MongoDB, MapR-DB, HDFS, MapR-FS, Amazon S_3, Azure Blob Storage etc., Tajo designed for scalable ad hoc queries, online aggregation of data and ETL on large datasets that are stored on the HDFS and other data sources, Apache Pig designed to analyze large data sets that consists of the high level languages that expresses the analysis program.
10) *Phase 5: Data Analysis:* in this phase large and variety of data sets are examined so that we can uncover the hidden patterns, find the unknown correlations between the varying data, what are the market trends and some other useful information that helps to make the organizations better decisions for their firms. Tools that are used in this phase are Hadoop YARN, Kafka, Pig, HIVE, HBase, Spark, Hadoop Map Reduce etc. Hence these are the different phases that works during the analysis of the big data.

III. RELATED WORKS

Tiwarkhede et.al., 2013, stated about the concepts of the big data, its 3V's velocity, volume and variety. Paper provides a brief description about how the generated data can be divided into various big data applications such as Structured analysis, Text Analytics, Web Analytics, Multimedia Analytics and Mobile Analytics. These analytics applications describes about how the data is being generated from different fields. There are many techniques also through which we can analyze the datasets and some techniques are machine learning. Techniques that are discussed in this paper are A/B Testing in it control group is compared to the various test groups, Classification in it new data sets are categorized and assigned to the predefined classes, Crowd Sourcing data collected is submitted by the gropu of people or the community, Data Mining in it patterns of data are extracted [8].

Ahlawat et.al., 2016, stated about the various definitions of the big data given by the researchers, 5V's of the big data, importance of the big data, various data forms that are available in big data. Manyika et.al , 2011 describes the big data as the amount of data that is beyond the ability of the technology to store, manage and process the data efficiently. Tech America Foundation, 2014 describes the big data as the huge amount of data that is having high velocity, is complex, having variety, and having a huge volume that can be captured, stored, distributed and managed efficiently. Data Forms of the big data that are available are Structured Form where whole data is organized in the entity form, Semi-structured data may be available in many formats, Un Structured Format in this data has no format and sequence. various tools and techniques used in big data are also described and they are Association rule learning(discovering interesting relationships), Data Mining(searching or digging into a data file), Cluster Analysis(divides the group of people or community), Crowd Sourcing(info is gathered from large group of people), Machine Learning(algorithms are crafted), Text Analysis(unstructured text data is converted into meaningful data), EDWs(enterprise data warehouse), Visualization products(represents the result visually),Map Reduce (processing of the data), Hadoop(store and process big data in distributed environment), NoSQL(helps in analyzing and accessing massive amount of data)[9].

Thomas et.al., 2015, stated about the concepts of the big data, Parallel data flow model and Map Reduce and various analytics use cases. Parallel Data flow Model used for the parallel programming which makes the programming easy it works on the shared nothing cluster of the computers in the data centre and machines which are involved can communicate through the simple data messages stream without the need of expensive shared memory. Map Reduce is the heart of the Hadoop and provides the great scalability to work over the thousands of the servers. It allows the user to write the traditional code into C, Java, Python, Perl and requires a file system to read. The best big analytics use cases discussed in this paper are Semantic Analysis, 360^0 view of customer, Ad hoc Data Analysis, Real Time Analytics, Multi Channel Marketing, Customer Micro Segmentation, Ad Fraud Detection, Click Stream Analysis, Data Warehouse Modernization, Big Data and Predictive Modelling [10].

Beakta, 2015, have studied about the 4V's of big data, challenges of the big data , Hadoop and Map Reduce. This paper mainly concerns about the Hadoop and Map Reduce that are used for the storage and processing of the big data. In this the storage is associated with the HDFS(Hadoop Distributed File System) and processing is associated with the Map Reduce(Map and Reduce) these two functions divides the data into the independent chunks and reduce functions collects the answer from the different chunks and aggregate them to produce the useful information. Some applications of the big data are classification analysis, cluster analysis, evolution analysis and outer analysis [11].

Bhosale et.al., 2014, stated about the architecture of Hadoop and Map Reduce, and several other components of the Hadoop. Hadoop is a programming framework that is developed by the Google's map reduce that is a software where application is break down into various parts. Current system of hadoop is Apache Hadoop Ecosystem. Hadoop architecture is basically divided into two

layers i.e. HDFS layer and Map reduce layer. HDFS layer can store huge amount of information, it can also survive the failure of significant parts of the storage infrastructure without losing data. Hadoop creates the clusters of machines and coordinate the work among them. If one fails then it continues it work by shifting the work to remaining machines. Map Reduce is the processing pillar of hadoop system. This framework allows the specification of operation to be applied on a huge data set, it divides the data and problem and run it in parallel [12].

Bendre et.al., 2016, this paper describes about the big data, big data analytics, cloud computing and Apache Hadoop. Different phases of the BDA have been discussed that are data generation, data acquisition & storage, data processing, data querying and data analytics. It includes the brief description about the tools or techniques used during different phases. Data generated from different sources like sensors, log files, multimedia files etc. Data acquisition and storage phase includes Kafka, flume, HDFS, Hadoop Common, Hadoop YARN, Map Reduce etc. Data Processing phase includes Apache Kafka, Apache Flume, Apache Chukwa etc. Data querying phase includes IMPALA, HIVE, PIG, DRILL, HBase, Google Cloud Storage, Tajo, Azure Blob Storage etc. Big Data Analytics classes are also described in this paper that are Structured Data Analytics, Text Analytics, Multimedia Analytics, Network Analytics, Web Data Analytics and Mobile Analytics [13].

Zhuming Bi et.al, 2014, in this paper the concepts of the big data, big data analytics , how the data is being collected from the different sources and how the IOT is making possible for the cloud computing so that it can acquire the data from the different sources. As the data is increasing day by day therefore to manage the data cloud computing is offering the reliable services or the technologies like NoSQL, Map Reduce like technologies are needed to tackle with the big data and to retrieve the big data. Data is being collected, data is managed and then data is utilized and all this happens with the help of the different tools that are developed to analyze and retrieve the big data. BDA has been explained as the process of inspecting, cleaning, transforming and modeling the big data. The BDA tools have been designed so that it can take into account the increase in volume of the requests, size of the data, computational load, the type of the user and the locality. Talia (2013) the BDA tools can be discussed with respect to the following (1) Programming abstracts (2) interoperability and openness of the data and tools (3) System Integration (4)Annotation Mechanisms. Software and the platforms are the driving forces of the BD that is big data. Four primary technologies for the processing of the big data are Grid Computing, in-database processing, in-memory analytics And the Hadoop. Two architectures have been discussed that deals with the BDA they are the

RDBMS and Map Reduce/ Hadoop. Hadoop is being described that it being used for the distribution, storage, query processing and management of the data. Hadoop have two components (1) HDFS i.e. hadoop distributed file system used for the storage purpose of the data (2) Map Reduce which is also known as the processing pillar of the Hadoop and it also consists of the two functions known as the Map function and Reduce function. The BDA tools helps in the efficient capture of the system information, processing and utilization of the information. The ten core technologies stated in this paper for the processing are Google Refine, data serialization such as Avro, data storage such as Amazon S3, the cloud such as Azure, NoSQL such as Hypertable, Map Reduce such as Pig, data processing such as Mechanical Turk, natural language processing such as National Language Toolkit, Machine learning such as Mahout and visualization such as Graph Viz [14].

Kaur et.al, 2017 stated about the algorithms used in data mining for the big data, the types of data mining system , about the issues and challenges and problems of big data in data mining. Algorithms that are used are classification tree, logistic regression, neural networks, clustering techniques. The types of data mining system are categorized according to the types of data sources mined, according to data model used, according to sort of knowledge discovered, according to excavation techniques used. The problem about the big data is how the huge amount of data get explored so that we can explore useful information from that large data. Issues are poor data quality, security, higher cost, less flexibility etc. Solutions to the big data are Hadoop which allow massive amount of storage for any kind of data, Cloud era it allows companies to access the data from large databases, Monod it manages the data that is unstructured or changes frequently [15].

Singh et.al, 2017 stated about the tools that are used in big data . These tools are Apache Hadoop is an open source framework based on java developed and maintained by apache foundation it's used for the massive analysis of data and storage of data in a cluster. Microsoft HD Insight this tool is provided by the Microsoft for big data solution and this is also powered by Apache Hadoop, NoSQL it is used to handle the unstructured data that does not follow any particular schema and provides improved performance in storing huge amount of data, HIVE it's an associated library of hadoop and it also supports query language known as HiveSQL that provide query solutions on big data, Sqoop it's a tool that connects Hadoop with various relational databases to transfer data, Polybase it allows data analysts to use the very commonly known T-SQL, in a very commonly used development environment - SQL Server Management Studio to query data stored in a Hadoop cluster[16].

Kaur et.al,2017 stated about the tools and techniques for the big data.

Hadoop: important technique and it's a programming framework developed by Google's Map reduce. Hadoop is used to handle the data with the help of divide and conquer method. Hadoop includes two steps: map (divides the data into number of sub parts) and reduce(collects all the answers from sub parts and combine them to get an appropriate output).
HDFC: its hadoop distributed file system it have client-server architecture and process the large amount of data.
HPCC: its high performance computing cluster used to manage complex problems. It's a single platform system, having single architecture and single programming language to process the data. Some components of HPCC are HPCC data refinery, HPCC data delivery, enterprise control language.
Grid computing: it's the technique in which computers are interconnected and share resources to each other. This technique is used with the help of hadoop.
Data mining: technique that is used to extract useful information from large datasets.
R tool: R is free software programming language which is used for statistical computing and graphics.
KEEL: It is Knowledge Extraction based on Evolutionary Learning and its application software of machine learning tools. It helps to solve the data mining problems with the use of evolutionary algorithms.
WEKA: Waikato Environment for Knowledge Analysis. For solving the data mining problems WEKA works on the machine learning algorithms [17].
Chen et.al., 2014, stated about the general background of the big data, related technologies such as cloud computing, Internet Of Things, data centers and Hadoop. This paper also reviews about the four phases of the value chain of big data i.e. data generation, data acquisition, data storage and data analysis. Relationship between Cloud Computing and Big Data is that development of the cloud computing provides solution for the storage and processing of the big data. The distributed storage technology based on the cloud computing can effectively manage the big data. Relationship between IoT and Big Data is that big data is generated by the IoT devices. Report was given by the Intel that pointed out that IoT has three different features that conform the big data diagram(1) various terminal generate massive amount of data (2) data that is generated by IoT is generally structured or semi-structured (3) data generated by IoT is useful only when it is analyzed. But the data processing capacity of IoT has fallen so it becomes necessary to accelerate the big data technologies to promote the development of IoT. Data centers in the big data provides a back stage support, the growth of big data applications accelerates innovation and revolution of data centers, data centers also strengthens the soft capacities like capacities of acquisition, processing, organization, analysis and application of big data. Hadoop is used in big data for the storage and processing purpose and for this different components of hadoop are used [31].
TRIFU et.al., 2014, stated about the big data characteristics given by 4V's volume, velocity, veracity, variety. Different tools have been briefed used for the big data efficient processing, storage and analysis. These are NoSQL Databases means "Not Only SQL" which uses wide column store, document, key value structures or other type of structure. MongoDB can manage the large number of data sets with low maintenance. Cassandra is a key and column oriented and used for the storage purpose of the big data. Big Table is a distributed store system used for managing structured data designed for a very large scale. HBase is known as hadoop database also used for the storage of the massive amount of data and is a open source clone to the big table. Map Reduce Model helps in processing large data sets in parallel. Hadoop is a Map Reduce system developed by YAHOO after Google's Map Reduce infrastructure. Various uses of big data are in Healthcare, Marketing, Education, Transportation etc.[32].
Gencer et.al., 2015, stated about the scope of the big data i.e. what was the past of the big data, what is present situation of big data and what will be the future of the big data. The paper shows that the increase in the big data violently increases in the year 2011. At present the search for the big data is on its peak or we can say that the big data becomes the most important term of the IT industry. The increase in interest of the big data is increasing day by day but the decrease in interest of data mining day by day. In today of the Big data they have shown the work of the various authors and their researches like Quian et.al. first introduces the "granular computing". After that they have defined encoded decision table and discussed some criteria. Google presented the another big data study that "Google Flu Trends" which was helpful in analyzing worldwide flu trends by using Google Search Terms. The interest in big data is increasing day by day. There will be increase in the studies of the big data in the field of industries from automotive and communication to finance and health will increase in future. Hence it will become most important in the future to manage the big data very efficiently [33].
Chong et.al., 2015, presents an overview of big data analytics, programming model, storage and application of big data. The paper states about the infrastructure of the big data which includes different phases like data acquisition and storage, data processing, data analysis, data querying. The programming model of the big data are Map Reduce , Graph processing Model, Stream Processing Model etc. The big data analytics means that we have to extract the meaningful information from the bunch of the data collected from different sources. Data analytics can be done in different classes like Descriptive Analysis, Predictive analysis, Prescriptive

analysis etc. The benchmarking of the big data drawn the attention of the researchers and practitioners. Benchmarking can be grouped in two types: component benchmarks and system benchmarks. Various applications of the big data are also discussed in the field of Business, Social Application, Scientific application [34].

O. Chan, 2013, stated about the concepts of the big data and its characteristics. This paper also gives overview about the Big Data Analytics, NoSQL, Hadoop, Distributed File System and Map Reduce. This paper describes about the characteristics of the big data that are volume, velocity, variety, veracity, value. An overview about the architecture of the big data is given that describes that architecture is based on the client server architecture. HBase/ Hadoop Cluster Architecture for big data also described which states that it consists of the master and slave nodes. This architecture is used for the storage and processing purpose of the big data. Big Data Analytics architecture is also described which consists of the different components like Map Reduce Analytics, Hadoop Cluster HDFS, Real-Time NoSQL, ETL, BI Analytics etc. all having their own functionality. It also explains how different type of data captured through different systems, how data is being captured or collected through different sources, cleansed, processed and analyzed. Hence this paper reviews about the concepts of the Big Data Analytics and its architecture [34].

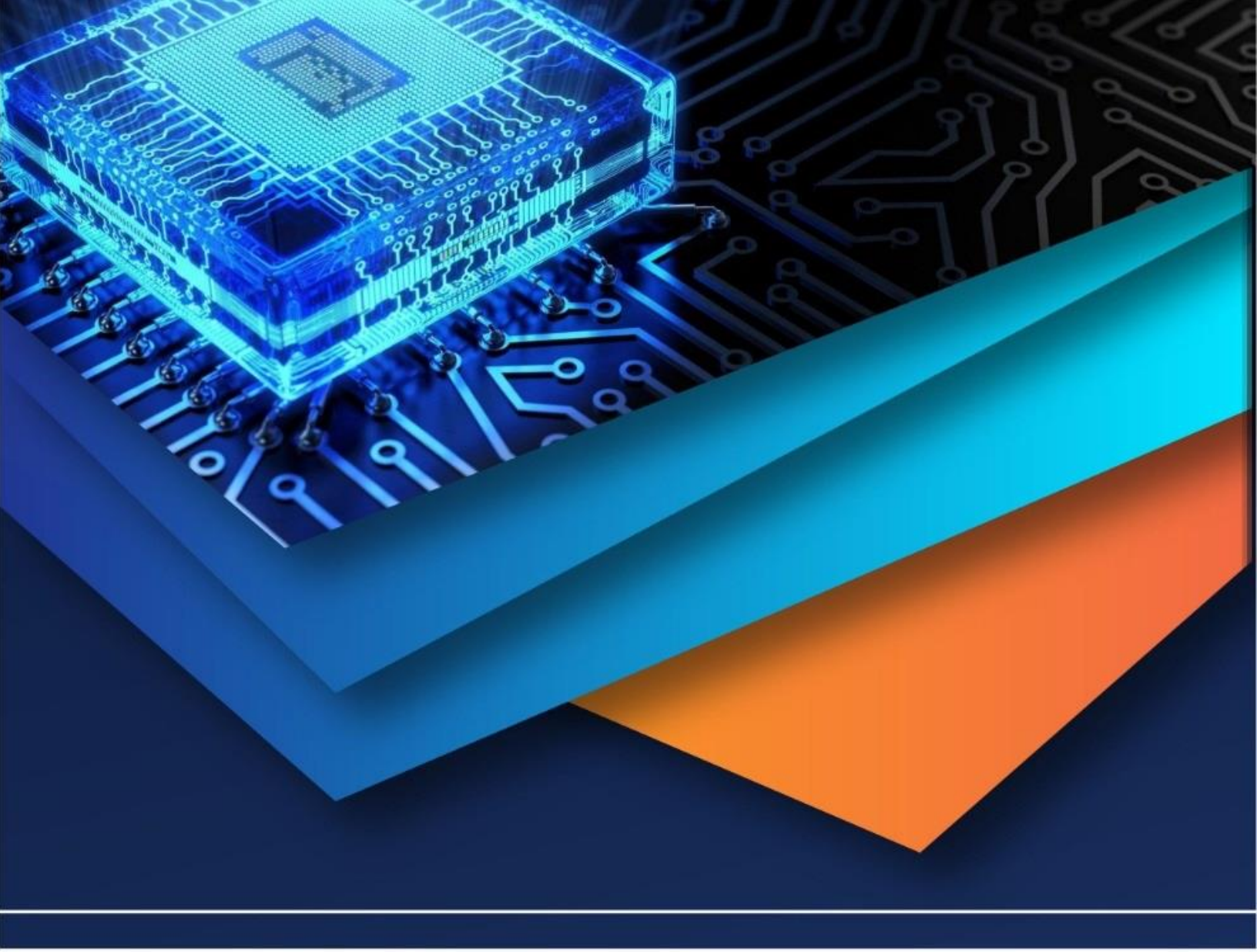
IV. CONCLUSION

Today, all the IT professionals, engineers and researchers are working on big data. Big data is term of concerning about large volume of complex data sets. In order to solve problems of big data challenges, many researchers proposed a different system models, techniques for big data. The high performance computing paradigm is required to manage the huge amount of data being generated in different fields. In the coming years the existing tools and techniques will not cooperate with the increasing size of the data hence in future the alternates will be needed for the existing tools and techniques. In future growth rate of the data is going to be very huge therefore new discoveries for the tools and techniques will be needed to manage the enormous growth of the data.

REFERENCES

[1] http://asperasoft.com/the-human-face-of-big-data/
[2] https://www.bernardmarr.com/default.asp?contentID=1076
[3] https://en.wikipedia.org/wiki/Big_data
[4] https://www.slideshare.net/BernardMarr/140228-big-data-volume-velocity-variety-varacity-value
[5] https://www.sas.com/en_us/insights/analytics/big-data-analytics.html#dmhistory
[6] https://link.springer.com/chapter/10.1007/978-3-319-21569-3_4
[7] http://www.computerweekly.com/podcast/Big-data-storage-Defining-big-data-and-the-type-of-storage-it-needs
[8] Ankita S. Tiwarkhede, Prof. Vinit Kakde," A review paper on big data analytics", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438.
[9] Tanvi Ahlawat and Dr. Radha Krishna Rambola School Of Computing Science and Engineering Galgotias University, Greater Noida," Literature review on Big Data", International Journal of Advancement in Engineering Technology , Management and Applied Science ,Volume 3, Issue 5 May 2016, ISSN No :2349-3224.
[10] Mrs. Mereena Thomas Assistant Professor, department of computer applications , SNGIST college, N.Paravur, Kerala, India,"A Review Paper on Big Data", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 09 | Dec-2015 www.irjet.net p-ISSN: 2395-0072
[11] Rahul Beakta," A review on big data and hadoop", Research Gate
[12] Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar," A review paper on big data and hadoop", International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014 1 ISSN 2250-3153
[13] Analytics, challenges and applications in big data environment: a survey Mininath R. Bendrea* and Vijaya R. Thool Department of Information Technology, SGGSIE&T, Vishnupuri, Nanded, India; bDepartment of Instrumentation Engineering, SGGSIE&T, Vishnupuri, Nanded, India, 2016.
[14] Big Data Analytics with Applications Zhuming Bi* and David Cochran† Department of Engineering, Indiana University Purdue University Fort Wayne, Fort Wayne, IN 46805, USA (Received 1 May 2014; Revised 15 June 2014; accepted 25 October 2014)
[15] Nirmal Kaur, Gurpinder Singh," A review paper on data mining and big data,"Volume 8, No. 4, May 2017 (Special Issue) International Journal of Advanced Research in Computer Science REVIEW ARTICLE Available Online at www.ijarcs.info
[16] Gurvinder Singh, Anurag Rana , Joyti," A review paper on big data tools and techniques," International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 9 September 2017, Page No. 22463-22468 Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i9.1
[17] Simranjot Kaur, Er. Sikander Singh Cheema,"A review paper on big data tools and techniques," IJARCCE ISSN (Online) 2278-1021 ISSN (Print) 2319 5940 International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 6, June 2017
[18] https://datafloq.com/read/understanding-sources-big-data-infographic/338
[19] https://mixpanel.com/topics/what-is-mobile-analytics/
[20] https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper
[21] https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data
[22] https://intellipaat.com/blog/7-big-data-examples-application-of-big-data-in-real-life/
[23] Big Data And Hadoop: A Review Paper Rahul Beakta CSE Deptt., Baddi University of Emerging Sciences & Technology, Baddi, India rahulbeakta93@gmail.com,2015.

[24] Nada Elgendy and Ahmed Elragal," A review paper on Big data analytics", Department of Business Informatics & Operations, German University in Cairo (GUC), Cairo, Egypt {nada.el-gendy,ahmed.elragal}@guc.edu.eg
[25] Nawsher Khan, Ibrar Yaqoob,Ibrahim Abaker Targio Hashem,Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani," A review paper on big data technologies, oppurtunities and challenges," Hindawi Publishing Corporation Scientific World Journal Volume 2014, Article ID 712826, 18 pages
[26] http://dx.doi.org/10.1155/2014/712826
[27] https://www.simplilearn.com/big-data-applications-in-industries-article
[28] https://data-flair.training/blogs/big-data-applications-various-domains/
[29] https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.pdf
[30] https://www.google.co.in/search?q=5+vs+of+big+data&dcr=0&source=lnms&tbm=isch&sa=X&ved=0ahUKEwitu_3dm fYAhXFso8KHWMdBcQQ_AUICigB&biw=1366&bih=662#imgrc=LPWYSfsQ9Ze7OM:
[31] Big Data: A Survey Min Chen · Shiwen Mao · Yunhao Liu © Springer Science+Business Media New York 2014
[32] Big Data: present and future Mircea Răducu TRIFU, Mihaela Laura IVAN University of Economic Studies, Bucharest, Romania trifumircearadu@yahoo.com, ivanmihaela88@gmail.com
[33] World Conference on Technology, Innovation and EntrepreneurshipYesterday, Today and Tomorrow of Big DataHakan Özköse, Cevriye Gencerb, TurkeybGazi University, Maltepe, Ankara, Turkey
[34] Big data analytics: a literature review Dazhi Chonga* and Hui Shiba Department of Information Technology & Decision Sciences, Old Dominion University, Norfolk, USA; bDepartment of Management and Information Sciences, University of Southern Indiana, Evansville, USA (Received 15 January 2015; revised 21 March 2015; accepted 10 July 2015)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY