IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BIG DATA: Concepts, Applications and Challenges

Sobila Venkata Sai Chandravadan

B.Tech Student

BITS Pilani- Hyderabad Campus

Telangana

ABSTRACT

Big Data has gained much attention from the academia and the IT industry. In the digital and computing world, information is generated and collected at a rate that rapidly exceeds the boundary range. A huge repository of terabytes of data is generated every day from modern information systems and digital technologies such as Internet of Things and cloud computing. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The utilization of Big Data Analytics after integrating it with digital capabilities to secure business growth and its visualization to make it comprehensible to the technically apprenticed business analyzers has been discussed in depth. Aside this, the incorporation of Big Data in order to improve population health, for the betterment of finance, telecom industry, food industry and for fraud detection and sentiment analysis have been delineated. The challenges that are hindering the growth of Big Data Analytics are accounted for in depth in the paper.

Keywords: Big Data, Big Data Tools, Sources of Big Data, Big Data Classification, Benefits of Big data, Real Time Applications of Big Data

INTRODUCTION

The recent rapid surge in the abundance of big data, thanks to Internet-based technologies like social media platforms and mobile devices, has caught many industry leaders off guard when it comes to managing vast, unpredictable, and high-velocity data streams. Big data has become pervasive, with examples ranging from meticulously tagged and tracked library books to smartphones brimming with numerous data-collecting applications. Additionally, various devices, including healthcare machines, continuously record data like heartbeats, blood pressure, hemoglobin levels, and sleep patterns every minute. All these instances constitute big data, and companies are capitalizing on user preferences to drive commercial gains, potentially compromising user privacy.

Traditionally, technologies are developed and tested in controlled laboratory environments before being introduced to the public through media channels like press releases and advertisements.

Traditionally, data warehouses have been employed to manage these extensive datasets. In this context, the primary challenge lies in extracting precise knowledge from the available big data. Many existing data mining approaches often struggle to effectively handle such massive datasets. The central issue in analyzing big data stems from the lack of coordination between database systems and analytical tools like data mining and statistical analysis. These challenges typically arise when we aim to uncover and represent knowledge for practical applications.

DEFINITIONS

Gartner Glossary as defined "Big data is high-volume, high velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation".

Apache Hadoop defined big data as "Datasets which could not be captured, managed, and processed by general computers within an acceptable scope".

According to **Wikipedia** "Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them".

NIST (National Institute of Standards and Technology) has defined the term big data as "Big Data consists of extensive datasets –primarily in the characteristics of volume, variety, velocity, and variability– that require a scalable architecture for efficient storage, manipulation, and analysis."

CHARACTERISTICS OF BIG DATA

Big Data can be defined by five V's: i.e. Volume, Velocity, Variety, Veracity, and Value. The following figure depicts the overall five characteristics (Five V's) of big data. Each of the characteristics of big data shown in figure one is briefly described.

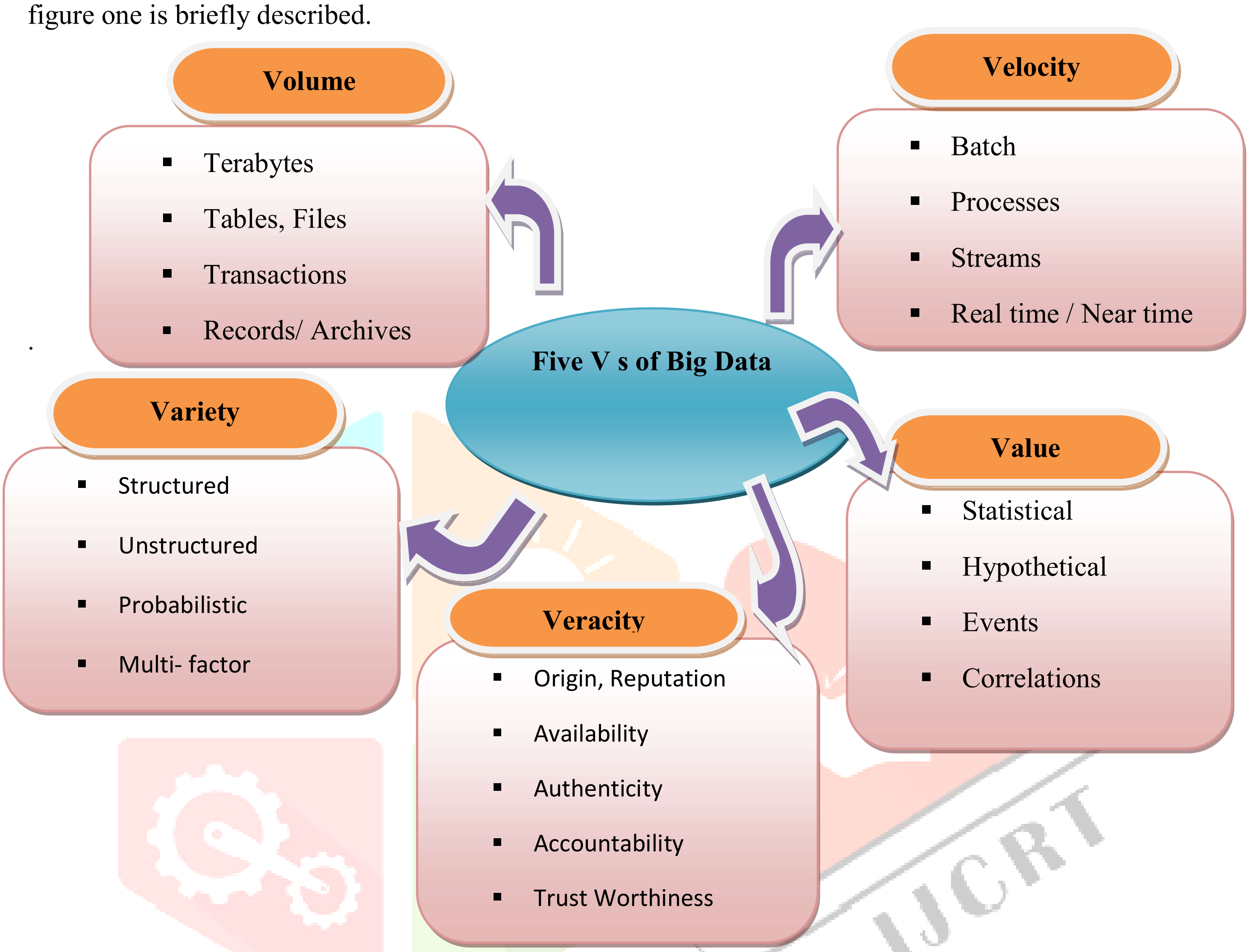

Fig. 1. Five V's of Big Data

Volume: "Volume" here means how much data we have, and it's growing rapidly because of different technologies. For example, the Internet of Things (IoT) is making a lot of data every second. In the last two years, the amount of data has doubled. Just think about all the emails, Twitter messages, photos, videos, and sensor data. Facebook and Instagram alone get more than 34,722 likes every second, and people upload more than 100 terabytes of data every day. There are more than 1.7 billion users on these platforms, and that's a lot of data!

Velocity: Velocity refers to the speed at which new data is generated and the speed at which data moves around. Report publish by clearly show that the speed of Internet have much higher than the developing country.

Value: As it name implies, The Value factor of Big Data define as related to a size which is enormous. Size of data plays very crucial role in determining value out of data.

Veracity: Data being stored in different database may have anatomies, noise and unfiltered. This is most important task of big data analyst then compare to other task

Variety: Variety is a measure of the richness of the data representation text, images, video, audio, and sensor data. In fact, 75 percent of world data are unstructured. It can be unstructured and it can include so many different types of data from XML to video to SMS. The main task to organized the useless data into meaningful data is big task for the analysis.

SOURCES OF BIG DATA

Big data comes from a wide variety of sources across different domains and industries. Here some of the examples of sources for large data and the different types of data are given below.

S.No	Big Data Source	Description
1	Research and Scientific Data	Data from Research Experiments, Academic studies, Scientific Observations, Digital twin simulations and Genomic sequencing.
2	Government and Public Data	Data provided by Government Agencies, Census data, Public Transportation data and Weather data.
3	Customer Data	Data collected through CRM systems, including Customer Profiles, Sales Records, and Customer Interactions.
4	Financial Transactions	Data obtained from Banking Systems, Credit Card Transactions, Stock Markets, and other Financial platforms.
5	Health and Medical Records	Data from Electronic Health Records (EHRs), Medical Imaging, Wearable Health Devices, Clinical Trials, and Patient Monitoring Systems.
6	E-commerce Transactions	Data generated from Online Retail Platforms including Customer Orders, Product Details, Payment Information and Customer Reviews.
7	Social Media Platforms	Data generated from Social Media Platforms like Facebook, Twitter, Instagram, and LinkedIn, including posts, comments, likes, shares, and User Profiles.
8	Internet of Things (IoT) Devices	Data collected from various IoT devices such as Intelligent Sensors, Smart Appliances, Wearable Devices, and Connected Vehicles.
9	Sensor Networks	Data gathered from Environmental Sensors, Industrial Machinery, Traffic Monitoring systems, and Other Wireless sensor networks.
10	Web and Mobile Applications	Data produced by users while interacting with Websites, Mobile Apps, and Online Services, including clicks, Page views and User behavior.

Table 1: Big Data's Information Sources

TYPES OF BIG DATA

Big data can be categorized into three main types. These are Structured big data, Unstructured big data and Semi-structured big data

Structured big data: It is highly organized and follows a pre-defined schema or format. It is typically stored in spreadsheets or relational databases. Each data element has a specific data type and is associated with predefined fields and tables. Structured data is characterized by its consistency and uniformity, which makes it easier to query, analyze and process using traditional database management systems.

Unstructured big data: It does not have a predefined structure and may or may not establish clear relationships between different data entities. Identifying patterns, sentiments, relationships, and relevant information within unstructured data typically requires advanced AI tools such as natural language processing (NLP), natural language understanding (NLU), and computer vision.

Semi-structured big data: contains elements of both structured and unstructured data. It possesses a partial organizational structure, such as XML or JSON files, and may include log files, sensor data with timestamps, and metadata.

In most cases, an organization's data is a mixture of all three data types. For example, a large data set for an e-commerce vendor might include structured data from customer demographics and transaction records, unstructured data from customer feedback on social media, and semi-structured data from internal email communication.

TOOLS OF BIG DATA

Dealing with large data sets that contain a mixture of data types requires specialized tools and techniques tailored for handling and processing diverse data formats and distributed data structures. Some of the popular tools are:

Azure Data Lake: A Microsoft cloud service known for simplifying the complexities of ingesting and storing massive amounts of data.

Beam: An open-source unified programming model and set of APIs for batch and stream processing across different big data frameworks.

Cassandra: An open source (free) database management system designed to handle huge amounts of data on a distributed system. This system was originally developed at Face book and is now managed as a project of the Apache Software foundation.

Databricks: A unified analytics platform that combines data engineering and data science capabilities for processing and analyzing massive data sets.

Elasticsearch: A search and analytics engine that enables fast and scalable searching, indexing, and analysis for extremely large data sets.

Business intelligence (BI): A type of application software designed to report, analyze, and present data. BI tools are often used to read data that have been previously stored in a data warehouse or data mart. BI tools can also be used to create standard reports that are generated on a periodic basis, or to display information on real-time management dashboards, i.e., integrated displays of metrics that measure the performance of a system.

Cloud Computing: A computing paradigm in which highly scalable computing resources, often configured as a distributed system, are provided as a service through a network.

Hadoop: A widely used open-source framework for processing and storing extremely large datasets in a distributed environment.

Hive: An open-source data warehousing and SQL-like querying tool that runs on top of Hadoop to facilitate querying and analyzing large data sets.

Kafka: An open-source distributed streaming platform that allows for real-time data processing and messaging.

KNIME Big Data Extensions: Integrates the power of Apache Hadoop and Apache Spark with KNIME Analytics Platform and KNIME Server.

MongoDB: A document-oriented NoSQL database that provides high performance and scalability for big data applications.

Pig: An open-source high-level data flow scripting language and execution framework for processing and analyzing large datasets.

Redshift: Amazon's fully-managed, petabyte-scale data warehouse service.

Spark: An open-source data processing engine that provides fast and flexible analytics and data processing capabilities for extremely large data sets.

Splunk: A platform for searching, analyzing, and visualizing machine-generated data, such as logs and events.

Tableau: A powerful data visualization tool that helps users explore and present insights from large data sets.

Talend: An open-source data integration and ETL (Extract, Transform, Load) tool that facilitates the integration and processing of extremely large data sets.

SQL: Originally an acronym for structured query language, SQL is a computer language designed for managing data in relational databases. This technique includes the ability to insert, query, update, and delete data, as well as manage data schema (database structures) and control access to data in the database.

Relational Database: A database made up of a collection of tables (relations), i.e., data is stored in rows and columns. Relational database management systems (RDBMS) store a type of structured data. SQL is the most widely used language for managing relational databases.

REAL TIME APPLICATIONS OF BIG DATA

Real-time applications ensure that responses, actions, or decisions are delivered promptly, adhering to specific time constraints. They earn the label "real-time" due to their ability to process responses or necessary decisions in a timely manner. To achieve swift and timely responses, real-time applications necessitate the availability of all essential resources during processing. Various domains require real-time big data analytics applications are needed and here are some diverse examples of such applications are discussed.

Transportation: In the realm of transportation, real-time data analytics plays a pivotal role, primarily due to the pressing need for rapid data processing to serve various purposes or offer services. Consider, for instance, the significance of real-time data analytics in assessing current traffic conditions. It swiftly delivers valuable information to end-users, enabling them to make efficient decisions within a brief timeframe such as

- Route selection for the destination
- Estimate time to reach to the destination
- Changing route because of any kind of sudden incidents like accident, or roadblocks.
- Quick delivery of orders for any kind of goods, like pizza delivery, or emergency postal delivery.
- Dynamic time calculation for emergency vehicles like ambulance, fire service car, police van for the quick arrival to the destination

Recent time sensor technology is developed a lot which could be used for monitoring traffic condition associated with the communication technology. Lists some sensor technologies categorized into road sensor and vehicle sensor and list of communication mechanisms for reliable and time efficient communication. Detailed categorized technologies are given Table 2

Road Sensor Technologies	Inductive-Loop Detectors
	Monitoring Cameras
	Capacitance Mats
	Road Tube Axie
	Inductive Loop
	Piezoelectric Axie
Vehicle Sensor Technologies	On-board Cameras
	Proximity Sensors
	GPS Systems
	Speedometers
Communication Technologies	Satellite
	GSM
	Wi-Fi
	Bluetooth

Table 2. Sensor and Communication Technology

Stock Market: The stock market is like a big marketplace where people trade company shares. Every workday, a ton of data gets generated in the stock market. This data is not only vast but also constantly changing. When we analyze this data in real-time, it can help both buyers and sellers. It can also catch fraud and illegal activities, making the stock market work better. Here are some things real-time data analysis in the stock market can do:

- Prediction of share prices before actual changes occur in share prices. So that timely selling or buying of shares can be done for higher profit margin.

- Earlier decision making ability for buying or selling shares. —Financial threads detection in quick time.
- Detection of illegal activities in market which helps to improve market performance.
- Automated trading of shares and threads detection system, which could increase number of buyer and seller in the market.

All the benefits associated with the stock market can be realized through real-time data analytics. Without this capability, manual processes would be much slower. Consequently, this delay would not be advantageous for either buyers or sellers, nor would it enable the market to identify potential threats and enhance its overall performance.

Clinical Care: Clinical research in the realm of real-time big data aims to make accurate predictions promptly. This enables physicians to deliver better and quicker decisions, enhancing the quality of patient care by analyzing medical data in a timely and dependable manner. The volume of data generated in the medical field and clinics has grown substantially, and the ability to analyze this data and provide timely responses can greatly improve patient care.

Recognizing the significance of real-time big data analytics in clinical care, numerous research efforts are underway. For instance, Zhang et al. have developed a clinical support system for real-time prognosis and diagnosis, prioritizing speed. Thommandram et al. have designed the Artemis system to detect cardiorespiratory events in real-time, utilizing IBM's InfoSphere Streams for stream processing.

However, research in this field is still in its early stages, and further development and investigation are required to consistently generate real-time results from medical data. Challenges arise when dealing with vast quantities of data, such as how to reliably analyze it and produce real-time results to administer timely treatment to patients. Achieving this goal can minimize the risks to human life. The primary objective of clinical care-based applications is to provide real-time healthcare to end-users, the patients, by collecting real-world medical data from various sources and analyzing it reliably and promptly.

Defense: In the defense and intelligence sectors, real-time data analytics plays a crucial role in making timely decisions that can save human lives. It's not just about having military strength; making the right decisions at the right time is equally vital. To achieve this, a vast amount of information needs to be analyzed, such as details about different vehicles used in warfare, the strength and movements of the opposition, current and historical war-related data, the number of soldiers, and various other relevant resources.

In a war scenario, data generation is extensive and constantly changing. This dynamic data, combined with static information, must be collected and analyzed promptly to plan the next steps or make decisions on the spot. These actions and decisions are essential in defense, military, and national security centers to safeguard human lives. Therefore, having a real-time data analytics application or system capable of handling such large and dynamic datasets is critically important for addressing security sector challenges.

Events/Festivals: In today's world, large-scale events occur regularly worldwide, ranging from open-air concerts and various sports games to New Year celebrations and religious or traditional festivals. These events draw substantial crowds, making crowd control essential for security and the smooth execution of the events.

The Control Center responsible for overseeing these events must make real-time decisions based on crowd movements. This may involve actions such as expanding parking facilities, managing traffic, providing medical support, or adjusting security force presence in specific event areas. To monitor and manage these situations, various technologies such as GPS, satellite tracking, and traffic or vehicle sensing systems (refer to Table 1 for Sensor and Communication Technologies) are employed.

Once the necessary information is collected, the Control Center can use it to offer real-time services in the overcrowded areas. Since these large events attract substantial crowds, the volume of data collected is extensive. Consequently, managing this vast amount of data in real-time becomes crucial for making immediate decisions or responding to emergencies, which can potentially save lives in case of accidents or disasters. However, it's important to note that handling such a significant volume of data in real-time is a challenging task due to its sheer size, diverse characteristics, and high velocity as it constantly changes and moves.

Natural Disasters: Throughout history, the world has experienced many natural disasters like earthquakes, floods, tsunamis, cyclones, and volcanoes. These disasters have caused the loss of countless lives, health problems, economic damage, and resource depletion.

To mitigate these devastating effects, early prediction and warning systems are crucial. These systems provide advance notice and critical information such as shelter locations, necessary actions, and emergency support. They rely on collecting real-time data, often using sensors, GPS, or satellite technology. However, the real challenge lies in processing this vast amount of data quickly to issue timely warnings.

Notable examples of such systems include the Indian Tsunami Early Warning System and the global early warning system for wildfires. These applications must provide accurate responses as they directly impact the lives of thousands of people. While it's a difficult task, the development of an early warning system capable of handling extensive data and processing it rapidly can significantly improve the accuracy of natural disaster warnings. This, in turn, can save countless lives, protect economies, and preserve valuable resources by enabling people to take timely action.

Daily Resources: In this paper, the term "daily resources" refers to essential utilities like electricity, gas, and water, which are necessary for everyday life. The production of these resources is typically based on daily consumption patterns and anticipated future needs. However, usage of these resources can vary on specific days or during particular time periods, and sudden spikes in demand can occur.

Real-time monitoring of resource usage can effectively address such situations. Consequently, the production of these resources can be managed efficiently and in a timely manner, ensuring that consumers always have an adequate supply. Additionally, this real-time data analytics application can be used to predict future usage rates and plan accordingly. It also facilitates the efficient allocation of resources as needed.

Moreover, the relevant authorities responsible for managing these resources can use this type of data analytics application to efficiently oversee and manage resources in real-time.

BENEFITS OF BIG DATA

While the vast scale of big data may seem daunting, it actually offers a treasure trove of information that professionals can harness to their benefit. These extensive data collections can be explored to uncover patterns related to their sources, yielding valuable insights for enhancing business operations or making predictions about future business results.

Cost Reduction: Hadoop is a framework for storing huge amount of data on distributed clusters. In Hadoop cluster, one year storage cost for one terabyte is $2,000. That is 800 times less than the traditional relational databases.

Time Reduction: Macy's merchandise pricing optimization application calculates data sets in seconds or in minutes which actually can take hours for calculation.

Support in Internal Business: Decisions The main idea of big data is to assist in the interior company decisions like, what kind of new products should be offered to people? , How much stock should be detained? And what must be the cost of our item?

Developing New Big Data-Based Offerings: Big data must be used to create new products and offerings. LinkedIn is the top example, which has used big data to develop products and offerings, including jobs you may be interested in, who have viewed my profile, people you may know, and numerous others. These ideas have pulled people to LinkedIn.

CHALLENGES OF BIG DATA

Security and Privacy: Cloud security alliance big data working group identify top security and privacy problems that need to capture for making the big data computing and infrastructure more secure. Most of these issues are related to the big data storage and computation. Some of the challenges are secure data storage. Various security challenges related to data security and privacy are discussed in which include data breaches, data integrity, data availability and data backup.

Dynamic Provisioning: A service of the cloud computing is infrastructure as service in which it provides computation resources on demand, many cloud related companies are implementing this concept and to making it easy for customers to access these services. Current frameworks do not have the property of the dynamic provisioning. Here is an issue that Compute resources can be insufficient for the submitted job, some process may requires more resources. Another issue is scheduling and protection algorithm, current algorithms does not consider these aspects

Algorithms Organizations were granting the papers by capturing key words from the abstract and titles. Analyzing the science with hand was difficult task. After that work was done by program analyst. They use algorithms to do this work. These algorithms can be varying from each other. This difference can reduce the effectiveness and reliability of the final result. Improvement in the data management will result in better technology but it will face many issues.

Misuse of Big Data: Challenges including potential misuse of big data are here, because information is power. Types of the data which people will produce in the future are unknown. To overcome these challenges we have to strengthen and increase our intent and capacity

Data Management: Data management is also a critical issue for corporation and industries. Data warehouse has efficient data management techniques. In two data warehouse management strategies are discussed; which are Immediate Incremental Management (IIM), Deferred Incremental Management (DIM) but the favor is given to IIM because of its algorithmic implementation

CONCLUSION

Big data has emerged as a transformative force affecting economies, industries, and organizations across the board. By delving into the analysis of vast datasets that are increasingly accessible, there exists the potential for rapid progress in various technical domains. Big data analytics has become indispensable for automatically uncovering valuable insights, detecting patterns, and revealing concealed rules. Prominent companies that have been built around the concept of big data include Google, eBay, Instagram, Amazon, LinkedIn, and Facebook. Even large-scale organizations are now embracing the data-driven approach by integrating big data analytics alongside traditional methods. This shift carries significant implications for an organization's skill sets, leadership, organizational structure, and technological infrastructure. This paper delves into the challenges and complexities associated with big data while also exploring the tools used to extract valuable knowledge from extensive datasets.

REFERENCES

1. Big data challenges and Opportunities http://blogs.wsj.com/experts/2014/03/26/six-challenges-of-big-data/

2. Complete Introspection on Big Data and Apache Spark http://www.ijsdr.org/papers/IJSDR1604006.pdf

3. Nobody tells you- 5 things big data 'CAN' and 'CANNOT' DO https://www.analyticsvidhya.com/blog/2015/11/5-big-data-can-cannot/

4. The Limitations of big data. https://www.elevatedthird.com/article/big-data-limitations

5. http://www.excelacom.com/resources/blog/the-5-vs-of-big-data-predictions┐ for-2016

6. https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5- vs-everyone-must-know.

7. Big data challenges and Opportunities http://blogs.wsj.com/experts/2014/03/26/six-challenges-of-big-data/

8. Open source Tools for Big data http://www.happiestminds.com/blogs/top-10-open-source-big-data-tools/

9. Big data Analytics Tools https://www.qubole.com/big-data-analytics/

10. Tools that ditch the big data http://www.infoworld.com/article/3128344/analytics/7-big-data-tools-to -ditch-in-2017.html

11. Five Things you need to know about Hadoop vs. Apache Spark http://www.infoworld.com/article/3014440/big-data/five-things-you-need-to -know-about-hadoop-v-apache-spark.html

12. The documentation of big data by Oracle https://docs.oracle.com/cd/E63064_01/

13. The Limitations of big data. https://www.elevatedthird.com/article/big-data-limitations

14. J. P. Dijcks, "Oracle: Big data for the enterprise," Oracle White Paper, 2012.

15. "Big Data Survey Research Brief," Sas White Paper,2013.

16. M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano, "Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report," IBM Institute for Business Value and Said Business School at the University of Oxford, 2012.

17. "Turn Big Data into Big Value, A Practical Strategy,"Intel White Paper,2013.

18. T. H. Davenport and J. Dyche, "Big Data in Big Companies," May 2013, 2013.

19. W.-H. Weng and W.-T. Lin, "A Scenario Analysis Of Big Data Technology Portfolio Planning," in International Journal of Engineering Research and Technology, 2013.

20. J. Lin and D. Ryaboy, "Scaling big data mining infrastructure: the twitter experience," ACM SIGKDD Explorations Newsletter, vol. 14, pp. 6-19, 2013.

21. A. Bifet, "Mining Big Data in Real Time," Informatica (03505596), vol. 37, 2013.

22. G. De Francisci Morales, "SAMOA: A platform for mining big data streams," in Proceedings of the 22nd international conference on World Wide Web companion, 2013, pp. 777-778.

23. X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," Knowledge and Data Engineering, IEEE Transactions on, vol. 26, pp. 97-107, 2014.

24. “Mining big data in the enterprise for better business inteligence.”Intel White Paper,2012.

25. “The Compelling Economics and Technology of Big Data Computing, For Big Data Analytics There’s No Such Thing as Too Big.” 4syth White Paper,2012.

26. A. Hems, A. Soofi, and E. Perez, "How innovative oil and gas companies are using big data to outmaneuver the competition," 2013.

27. A. Bensrhir, "Big data for geo-political analysis: Application on Barack Obama's remarks and speeches," in Computer Systems and Applications (AICCSA), 2013 ACS International Conference on, 2013, pp. 1-4.

28. F. Zulkernine, P. Martin, Y. Zou, M. Bauer, F. Gwadry-Sridhar, and A. Aboulnaga, "Towards Cloud-Based Analytics-as-a┐ Service (CLAaaS) for Big Data Analytics in the Cloud," in Big Data (BigData Congress), 2013 IEEE International Congress on, 2013, pp. 62-69.

29. P. Costa, A. Donnelly, A. Rowstron, and G. O’Shea, "Camdoop: Exploiting in-network aggregation for big data applications," in USENIX NSDI, 2012.

30. F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, et al., "Bigtable: A distributed storage system for structured data," ACM Transactions on Computer Systems (TOCS), vol. 26, p. 4, 2008

31. K. Crawford, G. Faleiros, A. Luers, P. Meier, C. Perlich, and J. Thorp, “Big Data, Communities and Ethical Resilience, A Framework for Action,” Big Data, Communities and Ethical Resilience White paper,2013.

32. N. Laptev, K. Zeng, and C. Zaniolo, "Very fast estimation for result and accuracy of big data analytics: The EARL system," in Data Engineering (ICDE), 2013 IEEE 29th International Conference on, 2013, pp. 1296-1299.

33. “Top ten big data security and privacy challenges,” Cloud Security Alliance White paper,2012.

34. A. A. Soofi, M. I. Khan, R. Talib, and U. Sarwar, "Security Issues in SaaS Delivery Model of Cloud Computing," 2014.

35. M. N. Vijayaraj, M. D. Rajalakshmi, and M. C. Sanoj, "Issues and challenges of scheduling and protection algorithms for proficient parallel data processing in cloud."

36. G. Halevi and H. Moed, "The evolution of big data as a research and scientific topic: overview of the literature," Research Trends, Special Issue on Big Data, vol. 30, pp. 3-6, 2012.

37. U. G. Pulse, "Big Data for development: challenges & opportunities," Naciones Unidas, Nueva York, mayo, 2012.

38. U. Rasheed, M. U. Sarwar, and R. Talib, "A Review on Data Warehouse Management.

39. L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," pp. 1–10, 2020.

40. M. Adam and I. Fakharaldien, "Big Data Analysis and Storage," no. March, 2017.

41. A. Salina, "A Study on Tools of Big Data Analytics," no. October 2016, 2018, doi: 10.15680/IJIRCCE.2016.

42. A. Ramasamy and S. Chowdhury, "Big Data Quality Dimensions: a Systematic Literature Review," J. Inf. Syst. Technol. Manag., vol. 17, no. 0, 2020, doi: 10.4301/s1807-1775202017003.

43. B. T. Hazen, C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications," Int. J. Prod. Econ., vol. 154, pp. 72–80, Aug. 2014, doi: 10.1016/j.ijpe.2014.04.018