International Journal of Computational Research and Development (IJCRD)
Impact Factor: 5.015, ISSN (Online): 2456 - 3137
(www.dvpublication.com) Volume 8, Issue 2, 2023

REVIEW: DATA ANALYTICS PROBLEMS, UNANSWERED RESEARCH CHALLENGES AND BIG DATA TECHNOLOGIES

K. Vinayakan* & V. Alamelu Mangayarkarasi**
* Department of Computer Science, Khadir Mohideen College (Affiliated to Bharathidasan University), Adirampattinam, Thanjavur, Tamil Nadu, India
** Department of Computer Applications, S.T.E.T Women's College (Affiliated to Bharathidasan University), Mannargudi, Tamil Nadu, India

Cite This Article: K. Vinayakan & V. Alamelu Mangayarkarasi, "Review: Data Analytics Problems, Unanswered Research Challenges and Big Data Technologies", International Journal of Computational Research and Development, Volume 8, Issue 2, July - December, Page Number 61-69, 2023.

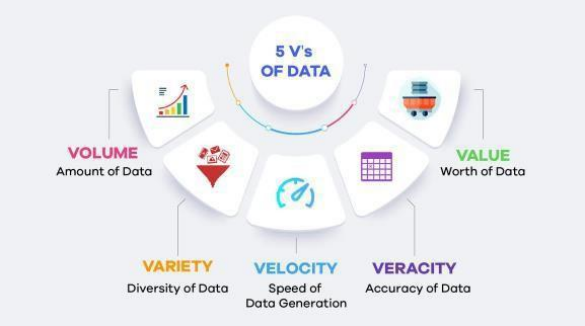
Abstract:

In today's world everything is based on massive amounts of data, which are collected from various resources. The data can be structured, unstructured and semi-structured data. The paper deals with the tools used for big data analytics. As a result, this paper explores a new platform to explore big data at several stages. Data is generated in the digital world from a variety of sources, and big data has grown as a result of the quick adoption of digital technologies. It offers evolutionary advancements with a vast dataset collection across numerous domains. It generally refers to the gathering of sizable and intricate datasets that are challenging to handle with conventional database administration software or data processing apps. These can be found in petabytes and larger formats in organized, semi-structured, and unstructured formats. It is defined formally as from 3Vs to 4Vs. Volume, Velocity, and Variety are referred to as 3Vs. While velocity refers to the rate of growth and the speed at which the data are gathered for analysis, volume alludes to the enormous amount of data that are generated every day. Variety tells us about the kinds of data.

Key Words: Big Data Analytics; Massive Data; Structured Data; Unstructured Data

Introduction:

Data is generated in the digital world from a variety of sources, and big data has grown as a result of the quick adoption of digital technologies. It offers evolutionary advancements with a vast dataset collection across numerous domains. It generally refers to the gathering of sizable and intricate datasets that are challenging to handle with conventional database administration software or data processing apps. These can be found in petabytes and larger formats in organized, semi- structured, and unstructured formats. It is defined formally as from 3Vs to 4Vs. Volume, Velocity, and Variety are referred to as 3Vs. While velocity refers to the rate of growth and the speed at which the data are gathered for analysis, volume alludes to the enormous amount of data that are generated every day. Variety tells us about the kinds of data.



The concept of big data is formally characterized by the 3Vs: volume, denoting the immense quantity of data generated daily; velocity, indicating the pace and speed of data collection for analysis; and variety, describing the diverse types of data. These factors create substantial obstacles to effectively leveraging big data for meaningful insights and decision-making. (Al-Mekhlal & Khwaja, 2019)

Big Data Analytics Challenges:

One of the primary challenges in big data analytics is the processing and management of unstructured data, which accounts for approximately 95% of the total data available (Tanwar et al., 2015). Unstructured data, such as text, images, videos, and audio, cannot be easily organized or analyzed using traditional database management tools and techniques. This heterogeneity of data formats poses significant hurdles in data integration, storage, and analysis. Additionally, the sheer volume of data being generated, often in the order of petabytes and beyond, necessitates the use of innovative technologies and architectures to handle the scale and complexity. Another key challenge in big data analytics is the speed at which data is being generated and the need for real-time or near-real-time analysis. Traditional data processing methods often struggle to keep up with the pace of data generation, leading to delays in extracting insights and making informed decisions (Rawat & Yadav, 2021). Moreover, the veracity of big data, which refers to the trustworthiness, accuracy, and reliability of the data, is another pressing concern. Dirty data, containing inaccuracies, missing values, or inconsistencies, can lead to flawed analysis and unreliable conclusions.

Open Research Issues in Big Data Analytics:

Despite the advancements in big data technologies, there are several open research issues that require further exploration. One critical area of research is the development of scalable and efficient data processing and storage solutions. Traditional relational database management systems often fall short in handling the volume, velocity, and variety of big data, leading to the rise of alternative data storage and processing frameworks, such as Hadoop, Spark, and NoSQL databases. Researchers are constantly exploring ways to enhance the scalability, performance, and reliability of these big data platforms to address the growing demands. Another important research topic is the integration and management of disparate data sources. Big data often comprises a heterogeneous mix of structured, semi-structured, and unstructured data, which requires advanced techniques for data extraction, transformation, and integration. Developing robust data integration and management strategies is crucial for enabling seamless data analysis and decision-making.

Furthermore, the issue of data quality and cleanliness is a significant research area in the field of big data analytics. Dirty data, containing errors, inconsistencies, or missing values, can severely impair the accuracy and reliability of the analytical insights. Researchers are exploring machine learning-based approaches for data cleansing and imputation to address these challenges and enhance the overall data quality. (Ridzuan & Zainon, 2019) Privacy and security concerns are also prominent research topics in the context of big data. As massive datasets are collected and stored, there is a growing need to ensure the protection of sensitive personal information and compliance with various data privacy regulations.

Big Data Analytics Tools:

To tackle the challenges and address the open research issues in big data analytics, a wide range of tools and technologies have been developed. One of the most prominent big data platforms is Apache Hadoop, which provides a scalable and fault-tolerant framework for distributed storage and processing of large datasets. Hadoop's core components, such as HDFS and Map Reduce, have become foundational in the big data ecosystem. Another widely adopted tool in the big data landscape is Apache Spark, which offers a fast and efficient in-memory data processing engine. Spark's ability to handle structured, semi-structured, and unstructured data, along with its support for a variety of programming languages, has made it a popular choice for advanced analytics and real-time data processing (Landset et al., 2015).

- The rate of data creation and movement is called velocity.
- The value that data offers is known as value.
- The different forms of data are what make up variety.
- The data's veracity refers to its accuracy and quality, while the data's volume gauges its status as big data.

Velocity:

Data creation and transfer rates are quantified by their velocities. This is an essential part for companies who need their data to be accessible quickly so they can make good judgments. There is a strong relationship between the value that big data may provide and the ways in which businesses can put the data they collect to use. Gaining good insights from big data increases its worth, making value extraction a crucial skill. Enterprises can employ big data tools for data collection and analysis, but they should develop unique strategies to derive value from the data. With the help of technologies like Apache Hadoop, companies can store, clean, and process these massive amounts of data rapidly.

Value:

Varieties of data are characterized by their diversity. Information quality can vary depending on the sources that a corporation receives from. Information could come to a company from a variety of sources, both inside and outside the company. An assortment of problems arises from the homogeneity and distribution of all the data being collected. There are three possible data formats: structured, unstructured, and semi-structured.

Veracity:

Data veracity refers to the data's quality, accuracy, integrity, and credibility. Data collected may not be comprehensive, may contain errors, or may not be able to provide useful insights. In most cases, the level of confidence in the collected information is what determines their truth. Data can get messy and difficult to handle at times. Misunderstandings rather than clearer understandings could result from a large amount of inadequate data. When it comes to medicine, for example, missing information about a patient's prescriptions can put their life in danger. The credibility and validity of data are defined by their value and truthfulness. In order to determine if the data is suitable, businesses often have and should have executive-level criteria for data authenticity.

Variability:

When it comes to defining the proper use of big data, the five qualities described above cover a lot of ground. Contrarily, variability is still another V that requires serious consideration. Rather of providing a specific definition, it stresses the importance of efficient administration of big data. Variability refers to discrepancies in the flow or use of big data. In the first case, a company might use a variety of terms to describe the same set of facts. As an example, in an insurance company, there may be two divisions, each using its own

set of risk levels. The second category includes data that is not centralized and is moving into company data repositories unfiltered and unchecked.

Volume:

Data volume is the total amount of data. Data is being created at an unprecedented rate, and the amount is mind-boggling. Complex data makes processing more difficult when volume is high. Organizations face difficulties in utilizing big data due to all of these considerations. Most conventional analytics tools will be overwhelmed just by the sheer amount. Finding the right data and putting it all together becomes a real challenge due to the variability. Acceleration is a term that refers to the rate at which data is created and moved. "Value" refers to the benefits that can be derived from data Variability is comprised of the various types of data that are available. The volume of the data establishes whether or not it is considered to be big data, whereas the validity of the data refers to its accuracy and quality.

Speed or Velocity:

The velocities of data creation and transfer are used to quantify the rates of both processes. This is a vital component for businesses that require their data to be easily accessible in a relatively short amount of time in order to make sound decisions. There is a close connection between the potential value that big data can give and the various ways in which organizations might put the data that they collect to use. The value of large data may be increased by gaining valuable insights from it, which is why value extraction is such an important talent. Big data tools can be utilized by businesses for the purpose of data collecting and analysis; but, in order to gain value from the data, businesses should establish their own unique tactics. These enormous amounts of data may be stored, cleaned, and processed in a short amount of time by businesses with the use of technologies such as Apache Hadoop

A value their diversity is what distinguishes different types of data from one another. The sources from which a company obtains information can have an effect on the quality of the information that it receives. It is possible for a firm to obtain information from a wide number of sources, both within the company and from outside the company. A wide range of issues are brought about as a result of the uniformity and distribution of all of the data that is being gathered. Structured data, unstructured data, and semi-structured data are the three possible formats for digital information.

Reliability:

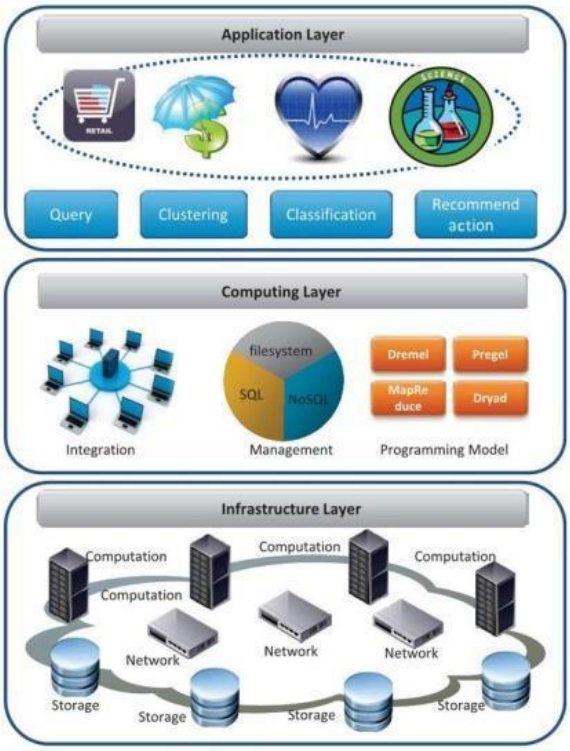


Figure 1: Layered architecture of big data system

When we talk about the validity of data, we are referring to its quality, accuracy, integrity, and credibility. There is a possibility that the data acquired is not exhaustive, that it may contain errors, or that it may not be able to provide insightful information. In the majority of instances, the reliability of the information that has been gathered is what determines whether or not it is accurate. There are instances when data might become disorganized and difficult to manage. When there is a huge number of insufficient data, it is possible that misunderstandings will occur rather than increased clarity of comprehension. When it comes to issues

concerning medication, for instance, the absence of information regarding a patient's prescriptions can put the patient's life in jeopardy. It is the worth and veracity of the data that determines the reliability and validity of the observations. For the purpose of determining whether or not the data is appropriate, organizations frequently have executive-level criteria for data authenticity, and they should have these criteria. The presence of variation a great deal of territory is covered by the five characteristics that were outlined above when it comes to establishing the appropriate application of big data. On the other hand, variability is still another kind of V that needs to be taken into serious mind. Instead of providing a particular description, it emphasizes the significance of effectively managing large amounts of data. Disparities in the flow or utilization of big data are what are meant by the term "variability." The first scenario involves a corporation that may employ a number of different phrases in order to describe the same collection of information. A good illustration of this would be the existence of two divisions inside an insurance firm, each of which employs its own unique risk ratings. The second category consists of data that is not centralized and is being sent into the internal data repositories of the organization without being filtered or reviewed.

Quantity:

A whole amount of data is referred to as the data volume. The generation of data is occurring at a rate that has never been seen before, and the quantity is staggering. The processing of complex data becomes more challenging when the volume of the data is considerable. By taking into account all of these factors, organizations encounter challenges while attempting to make use of big data. By alone, the sheer quantity will be sufficient to overwhelm the majority of standard analytics systems. As a result of the variety, it becomes a particularly difficult task to locate the appropriate data and to put it all together.

Big Data Challenges:

Data Accessibility:

Big data has long held the promise of enabling organizations to make better decisions by revealing insights that were previously undiscovered in the massive amount of data. Making massive data usable and accessible is a difficult task, though. Big data is inaccessible due to three main factors: volume, variety, and velocity.

Data Quality Maintenance:

The sheer amount and diversity of data can be debilitating, and the quality of the data might deteriorate due to the lack of appropriate maintenance. Inaccurate analysis and decision-making may result from this, which could be expensive for companies. Make a plan for handling data. This entails outlining who will be in charge of preserving the accuracy of the data, establishing guidelines for its collection and processing, and developing procedures for fixing mistakes. Accurate and current knowledge of the source data is another essential component in maintaining data quality. This involves keeping track of the data's format, source, and any dependencies on other datasets.

Data Security:

Organizations are a more alluring target for cybercriminals as they gather ever-larger data repositories. Serious repercussions from data breaches could include losing clients, harming one's reputation, and incurring expenses. Develop a data security strategy with several defenses in place. Make sure your staff members are trained on sensitive information protection and are aware of the dangers posed by data theft. When transferring and storing data, use secure techniques. These covers employing secure networks, encrypting private data, and creating strong passwords. Continually evaluate your security posture and adjust as necessary to stay abreast of emerging threats.

Using the Right Tools and Platforms:

Businesses can benefit greatly from big data analysis, but we won't be able to take full advantage of your data sources and the insights they have to provide if you're not using the correct tools and platforms. Since data processing and analysis technologies are always evolving, a business must devote resources to identifying the best solutions that will integrate with your ecosystem. This frequently entails locating a solution that can expand and scale with you in response to changes in your infrastructure.

Big Data Analytics: Issues and Challenges

Big Data is facing a number of challenges in running the organization. The framework that works with big data must be able to understand both the client's and innovation's expectations. Big data presents challenges that are tough to overcome. Not only is the amount of information growing daily, but it is also aging faster than it has in recent memory, and the kind of information available is also expanding. The complexity of information introduced is beyond the capabilities of the equipment, advancements, engineering, executives, and investigation procedures currently in use.

Protection, Security, and Trust:

Big data-using organizations made a commitment to protect and secure their customers. They should also ensure that the association agrees on all security and protection-related demonstrations to improve the establishment of clear stopping points for the use of personal data, and that trust in the association grows as the volume of information increases. The amount of confidence that customers have in these companies and their

ability to securely store personal information can surely be affected by information or data leaks into the public domain.

Information Management and Sharing:

The agencies are aware that information needs to be discoverable, accessible, and useful in order to be valuable. While fulfilling these requirements, organizations should adhere to security regulations. Recent developments in the field of open information have placed a strong emphasis on making datasets widely accessible. Offices should prioritize establishing information that is open, normalized, and accessible both within and between enterprises. This will enable offices to collaborate and use information to the extent that privacy rules allow.

Innovation and Analytical Abilities:

Big data analytics places a great deal of pressure on ICT (information and communication technology) providers to develop new tools and innovate in order to handle complex data. Huge amounts of diverse information cannot be measured, stored, or broken down by the tools and inventions available today. Open source software developers and vendors of big data frameworks and arrangements are getting better at overcoming the challenges associated with big data analytics. Some specific challenges related to Big Data and Analytics are:

- Information Storage and Retrieval
- Information Growth and Expansion
- Speed and scale
- Organized and unstructured information
- Information proprietorship
- Data Skew
- Limited resources
- Data Variety
- Edge Data Processing
- Consumption of energy

As of right now, available innovations can address information sections and information collecting. Nevertheless, rather than searching through a massive amount of data, the devices meant for the exchange procedure might look for a smaller portion of it. It is still unclear whether approach is the most efficient for handling semi-structured or unstructured data.

Information Growth and Expansion:

It is anticipated that the organizations' information would grow as their administrations do. The association takes into account the advancement of information because it is richly detailed and employs innovative tactics.

Speed and Scale:

Learning to become an information-literate person during a time when data volume is increasing is challenging. Learning about information is more important than managing the entire information arrangement. Computing near-continuous data will always need planning ahead in order to provide results that are acceptable.

Organized and Unstructured Information:

The flow of information from structured data stacked in neat tables to unstructured data (images, audio files, and text) required for analysis will impact the information preparation process from start to finish. The development of new non-relative innovations will lead to more adaptability in information depiction and computing.

Information Proprietorship:

Employees at web-based media specialized companies own an enormous amount of knowledge. They store information about their clients, even though they do not have this information. The page or record was created by the page's true owner. In the world of modern media, information proprietorship is becoming more and more difficult.

Data Skew:

Data skew, which only applies to Parallel Processing architectures where Data Distributed Processing takes place, is the term used to describe the unequal distribution of data in a dataset. The benefit of data distributed processing is that a job can be finished more quickly overall if it is split up into several parallel smaller jobs that are handled by various processors rather of being completed by a single processor. By doing this, execution time is shortened and performance is enhanced.

Data Variety:

Variation in data is one of the main elements of big data. Data diversity is produced through combining data from multiple sources and distributing the data in various ways.

Limited Resources:

A distributed computing paradigm known as "edge computing" moves data storage and processing closer to the data sources. It is anticipated to reduce bandwidth usage and speed up response times.

Edge Data Processing:

Applications that analyze incoming data, logs, and queries to generate aggregated results or dashboards might benefit from the usage of approximation techniques in big data. Since the output of these applications is much lower than the input, they make the greatest use of resources-money, time, and energy.

Consumption of Energy:

The efficiency of data centers varies; a large amount of power supplied to a data center might be utilized for cooling and other auxiliary purposes instead of the IT equipment. When work may be scheduled in data centers, utilization rates have to be high. It is more difficult to maintain high utilization in other areas unless demand is fairly predictable.

Big Data Tools:

The extensive analysis of big data is necessary for the commercial activities to continue advancing. Big data analysis plays a significant role in the decision-making process to improve the association's growth and success. In any case, merely computing information using standard information computing devices as a guide does not yield useful results, and the gadget is unachievable. As a result, a number of big data tools have been created recently to help associations and data scientists make informed decisions that are both economical and productive. Experts work with information storage, information management, information purging, information mining, data forecasting, and data endorsement usinga variety of big data analytics tools. This section briefly explains the instruments.

NoSQL:

Structured query language, or SQL, is typically widely used to constrain and analyze structured data. On the other hand, the tremendous development of infinite data has given rise to unstructured information-analytical tools. Eventually, SQL (NoSQL) evolved as a useful tool for handling disorganized data models. NoSQL databases don’t adhere strictly to architecture when storing infinite amounts of data. The table shift's segment values are then determined by each information record (line). The system's architecture-less design allows it to balance accessibility, internal failure tolerance, and consistency over momentum. Despite the fact that NoSQL has been extremely popular in recent years, low-level query dialects provide problems, and the lack of standard interfaces has not yet been taken into account.

Cassandra:

Cassandra is the type of NoSQL, non-proprietary, and disseminated dataset that manages the exceptionally enormous datasets across various providers. The magnificence of Cassandra is regarding the low failure, guaranteeing high accessibility under any conditions. Tus, the group of experts prefers Cassandra when versatility and accessibility highlights are pivotal without affecting the execution performance. In addition, Cassandra enables information replication across various clouds or server farms to guarantee lower dormancy and adaptation to internal failure. The kind of non-proprietary, NoSQL distributed dataset that handles the extraordinarily large datasets from several providers is called Cassandra. Cassandra's greatness lies in its low failure rate, which ensures high accessibility no matter what. However, the panel of experts favors Cassandra in situations when accessibility and versatility are essential without compromising execution performance. Furthermore, Cassandra facilitates the replication of data among multiple clouds or server farms in order to provide reduced dormancy and internal failure tolerance. Cassandra was designed to handle big data workloads across multiple nodes without a single point of failure. It has a peer-to-peer distributed system across its nodes, and data is distributed among all the nodes in a cluster.

Hadoop:

Hadoop is a system that comprises an assortment of programming libraries, which integrates the different programming models to enable the disseminated computation of enormous datasets. Versatility is the significant advantage associated with the Hadoop system, where the dispersed repository is independently called the Hadoop Distributed File System (HDFS). Tis Hadoop framework is a highly accessible analytical tool independent of the hardware equipment. Instead, it consolidates programming libraries to distinguish, recognize, and restrain the insufficiency at the implementation layer. Hadoop system comprises the normal libraries, stockpiling libraries, Hadoop YARN, and Hadoop Map Reduce to compute the vast datasets. The Hadoop system is made up of a variety of programming libraries that combine various programming models to allow for the distributed processing of large datasets. The major benefit of the Hadoop system, where the distributed repository is known separately as the Hadoop Distributed File System (HDFS), is its versatility. This Hadoop framework, which is not dependent on hardware, is an extremely accessible analytical tool. Rather, it combines programming libraries in order to identify, identify, and limit the inadequacy at the implementation layer. To compute the enormous datasets, the Hadoop system consists of Hadoop Map Reduce, Hadoop YARN, and standard libraries.

Spark:

An open-source distributed cluster processing tool known as Spark is becoming increasingly popular. Spark is a computational framework that ensures increased productivity in cluster and stream information processing. MLib, GraphX, SQL, and Spark streaming segments are all areas in which Spark helps to collaborate. Among the many platforms that Spark works on, Yarn, Hadoop, and Mesos are some of the systems that it uses to verify data from several sources. The spark is considered productive from the perspective of the storm assessment due to the fact that the same code regulation can be utilized for both cluster computation and

constant computation. Despite this, the storm has gotten more frequent in terms of inactivity, and there are fewer constraints on its behaviour.

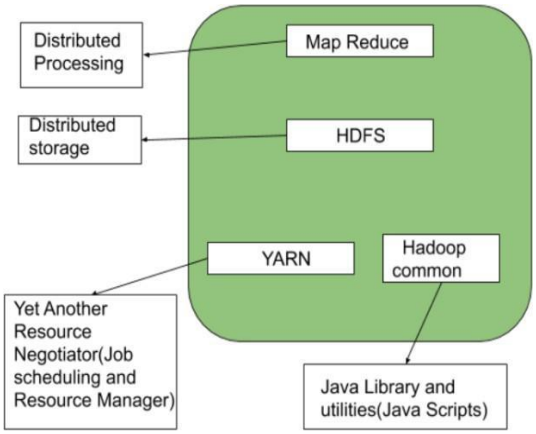


Figure 2: Hadoop architecture

Hive:

Hive is a cross-stage information distribution programming center that works with Hadoop. It provides information evaluation and querying across several storage spaces and document frameworks that are integrated with Hadoop. Also, it works with Hadoop. A query language that is similar to SQL is called HiveQL, and it is used to query delivered document frameworks. Hive provides SQL reflection for this purpose. A consequence of this is that solicitations are able to make use of Java APIs of lower quality without necessitating the processing of the query. In addition, Hive offers lists that can be used to boost the speed at which queries are computed.

Differences between noSQL, Cassandra, Hadoop, Spark, Hive:

Differences between noSQL, Cassandra, Hadoop, Spark, Hive

S.No	noSQL	Cassandra	Hadoop	Spark	Hive
What is it?	Type of database that stores data in a non-tabular format.	an open-source, distributed, NoSQL database management system that stores data for applications that need fast read and write performance	Opensource framework for distributed data storage and processing	Opensource framework for in-memory data storage and app development.	an open-source data warehouse system that allows users to query, analyze, and summarize data stored in large datasets.
Initial release	1998	2008	2006	2014	2010
Supported languages	Java, Python, Node. JS	Java,Python, C#, NodeJS,PHP	Java	Java, Scale, Python, R	Hive translations,Hive QL,Hive, Translation API
Processing methods	Hierarchy modeling Benchmarks Key-value stores Transaction models	Logging Memtable SSTables	Batch Processing	Batch Processing and Micro batch processing	HiveQL Hive metastore Hive compiler MapReduce
Built-in Capabilities	Flexible data models Scalability,High Availability	Incremental Scalability High Availability	File System(HDFS) Resource Management (Yarn) Processing engine(Mapreduce)	Processing engine(Spark Core) Near real-time processing(Spark Streaming)	Dataware house Query language
Real life use cases	E-commerce Big data analytics Social media	Logging and event streaming Fraud detection and risk management Real-time analytics	Enterprise archived data processing Sentiment analysis Predictive maintenance Log file Analysis	Fraud Detecting Telematics Analytics User behaviour analysis Risk Management	Data warehousing Business intelligence Machine learning

Conclusion:

A number of problems that are linked with big data applications are enumerated in this study, which also examines big data terminologies, challenges, and tools. In this study, we will be focusing on the numerous methods that have been described in the literature to control the problems that are associated with big data. This paper then goes on to cover the prospects and applications that are associated with big data, as well as providing an overview of the big data processing frameworks, such as NoSQL, Cassandra, Hadoop, Spark, and Hive.

References:

1. Al-Mekhlal, M., & Khwaja, A. A. (2019). A Synthesis of Big Data Definition and Characteristics. https://doi.org/10.1109/cse/euc.2019.00067
2. Hariharakrishnan, J., Mohanavalli, S., Srividya, & Kumar, K. B. S. (2017). Survey of pre-processing techniques for mining big data (p. 1). https://doi.org/10.1109/icccsp.2017.7944072
3. Hashem, M., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2014). The rise of "big data" on cloud computing: Review and open research issues. In Information Systems (Vol. 47, p. 98). Elsevier BV. https://doi.org/10.1016/j.is.2014.07.006
4. M Vasuki, PS Kumar, N Rajesh, On Anti-Q-Fuzzy Deductive Systems of Hilbert Algebras, International Journal of Analysis and Applications, Vol 21, No. 42, 2023, 1-15
5. M Vasuki, PS Kumar, S Broumi, N Rajesh, On Radical of Neutrosophic Primary Submodule, International Journal of Neutrosophic Science, Vol 22, No. 3, 2023, 36-52
6. AD Kumar, R Sivaraman, Asymptotic Behavior of Limiting Ratios of Generalized Recurrence Relations, Journal of Algebraic Statistics, Vol 13, No.2, 2022, 11-19
7. R Sivaraman, J Suganthi, AD Kumar, PN Vijayakumar, R Sengothai, On Solving an Amusing Puzzle, Specialusis Ugdymas, Special Education, Vol 1, No. 43, 2022, 643-647
8. AD Kumar, R Sivaraman, Analysis of Limiting Ratios of Special Sequences, Mathematics and Statistics, Vol 10, No. 4, 2022, 825-832
9. AD Kumar, R Sivaraman, On Some Properties of Fabulous Fraction Tree, Mathematics and Statistics, Vol 10, No. 3, 2022, 477-485
10. MS Kumar, AD Kumar, Effect of Mental Training on Self Confidence among Professional College Students, International Journal of Recent Research and Applied Studies, Vol 4, No. 12, 2017, 51-53
11. MS Kumar, AD Kumar, A Statistical Approach towards the Effect of Yoga on Total Cholesterol of Overweight Professional College Students, International Journal of Recent Research and Applied Studies, Vol 4, No. 2, 2017, 126-128
12. Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. In Journal Of Big Data (Vol. 2, Issue 1). Springer Science+Business Media. https://doi.org/10.1186/s40537-015-0032-1
13. Rawat, R., & Yadav, R. (2021). Big Data: Big Data Analysis, Issues and Challenges and Technologies. In IOP Conference Series Materials Science and Engineering (Vol. 1022, Issue 1, p. 12014). IOP Publishing. https://doi.org/10.1088/1757-899x/1022/1/012014
14. Ridzuan, F., & Zainon, W. M. N. W. (2019). A Review on Data Cleansing Methods for Big Data [Review of A Review on Data Cleansing Methods for Big Data]. Procedia Computer Science, 161, 731. Elsevier BV. https://doi.org/10.1016/j.procs.2019.11.177
15. Tanwar, M., Duggal, R., & Khatri, S. K. (2015). Unravelling unstructured data: A wealth of information in big data (p. 1). https://doi.org/10.1109/icrito.2015.7359270
16. Sharma, A., and Chen, Y. (2018) highlighted the synergy of deep learning and Zero Trust principles, emphasizing their potential to redefine security protocols for cloud data.
17. VA Mangayarkarasi, M.V.Srinath. " A Novel Prioritized Deciding Factor (PDF) Approach for Directed Acyclic Graph (DAG) Based Test Case Prioritization using Agile Testing Methodology", International Journal of Computing Algorithm, Dec 2016, Vol 05, No.02 72-78.
18. VA Mangayarkarasi, M.V.Srinath. "Big data management using NOSQL", International Journal of scientific transactions in environment and Technovation, July 2016, Vol. 10, No.1, 37-42
19. VA Mangayarkarasi, A.Karthiga," Web Refining Validation Thought Users Session Timing for Web Search Result Optimization", International Journal of Scientific Research in Computer Science Applications and Management Studies, July 2019, Vol 8, No.4
20. K Vinayakan, M V Srinath, A Secured On-Demand Routing Protocol for Mobile Ad-Hoc Network, A Literature Survey, Vol 6, No 6, 2015, 598-604
21. K Vinayakan, M V Srinath, Reinforcing Secure on-Demand Routing Protocol in Mobile AD-Hoc Network Using Dual Cipher based Cryptography, International Journal of Control Theory and Applications, Vol. 10, No 23, 2017, 103-109
22. K Vinayakan, M V Srinath, Security Mandated Analytics based Route Processing with Digital Signature [SMARPDS] - Pseudonymous Mobile Ad Hoc Routing Protocol, Indonesian Journal of Electrical Engineering and Computer Science, Vol 10, No 2, 2018, 763-769

23. K Vinayakan, M V Srinath, A Adhiselvam, Security for Multipath Routing Protocol using Trust based AOMDV in MANETs, Vol. 2 No. 43, 2022, 1640-1654
24. K Vinayakan, M V Srinath, A Adhiselvam, Reinforced Securing of Data Leakage in Mobile Ad hoc Network (MANET) by Hybrid Mechanism of Identity Based Encryption (IBE), International Journal of Health Sciences, Volume 6. No S8, 2022, 3622-3635
25. S Sujatha, K Vinayakan, The Role of Collaborative Learning in Mathematics Education: A Review of Research and Practice, Indo American Journal of Multidisciplinary Research and Review, Vol 6, No. 2, 2022, 200-206
26. S Sujatha, K Vinayakan, Mathematical Literacy for the Future: A Review of Emerging Curriculum and Instructional Trends, International Journal of Applied and Advanced Scientific Research, Vol 7, No. 2, 2022, 65-71
27. S Sujatha, K Vinayakan, Assessing the Impact of Math Competitions and Challenges on Student Learning: A Review, International Journal of Advanced Trends in Engineering and Technology, Vol 8, No 2, 2023, 62-67
28. S Sujatha, K Vinayakan, Integrating Math and Real-World Applications: A Review of Practical Approaches to Teaching, International Journal of Computational Research and Development, Vol 8, No. 2, 2023, 55-60
29. S Sujatha, K Vinayakan, Engaging Students with Mathematics: A Review of Motivation and Engagement Strategies, International Journal of Interdisciplinary Research in Arts and Humanities, Vol 8, No. 2, 2023, 55-60
30. VA Mangayarkarasi, A.Indhuja, " Effective Pattern Discovery for Text Mining Using Hidden Pattern Filter Sorting Techniques", International Journal of Scientific Research in Computer Science Applications and Management Studies, July 2019, Vol 8, No.4
31. VA Mangayarkarasi, "An Capable Re-Cluster Based Panel Collection Using Mst And Heuristic System", International Journal of Research and Analytical Reviews (IJRAR), October 2020, Vol 7, No.4, 94-100.
32. VA Mangayarkarasi, "A Real Time Big Data Analysis Using R" International Journal of Research and Analytical Reviews (IJRAR) February 2021, Vol 8, No.1, 384-389.
33. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00659-3
34. https://www.researchgate.net/publication/320771893_Big_Data_Analytics_Applications_Prospects_and_Challenges