Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

An Extensive Exploration of Big Data Analytics and Related Technologies

[1]Mohammad Asif A Raibag, [2]Veeresh, [3]Irshad Ahmad Gorikhan, [4]Rashel Sarkar, [5]Savitha Kulematagi

[1]Professor, Department of CSE (IOT, CS & BT), YIT, Moodbidri
[2]Assistant Professor, Department of CSE, Govt. Engineering College, Talakal
[3]Assistant Professor, Department of CSE, AGMRCET, Varur
[4]Associate Professor, Department of CSE, The Assam Royal Global University, Assam
[5]Assistant Professor, Department of CSE, YIT, Moodabidri

Abstract: **-** Big data is a term for massive data sets having large, more varied and complex structure with the difficulties of storing, analysing and visualizing for further processes or results.Big data has emerged as a transformative force reshaping industries, research, and everyday life. This abstract explores the evolution of big data from its conceptualization to its current status as a critical asset in decision-making processes. It delves into the diverse applications of big data across sectors such as healthcare, finance, retail, and transportation, highlighting its role in revolutionizing business strategies, enhancing customer experiences, and informing policy-making. Moreover, it discusses the challenges associated with big data, including privacy concerns, data quality issues, and the need for robust infrastructure and analytics tools. By elucidating these aspects, this abstract aims to provide a comprehensive understanding of big data's significance, potential, and future directions. As a result, this article provides a platform to explore big data at numerous stages. Additionally, it opens a new horizon for researchers to develop the solution, based on the challenges and open research issues.
Keywords*: Big Data, Hadoop, Quantum Computing, Veracity.*

1. Introduction

Massive data collections, including intricate, ever-growing datasets gathered from diverse data sources, are rapidly evolving as a result of advancements in digital resources. Such large amounts of data are referred to as big data coined by John R. Mashey. The core of contemporary research in industry and academics is the analysis of this ever growing mammoth datasets. These data are produced by science data, sensors, through electronic communications, click streams, logs, posts, snaps, videos, audios, health records, social networking activities, mobile phones and associated apps. The tremendous growth of databases makes it challenging to acquire, structure, archive, handle, distribute, analyze, and visualize them using standard database software technologies. Annual digital data generation is predicted to reach 147 zettabytes in 2024 and it will reach 181 zettabytes at the end of 2025. The below figure 1.1 depicts the exponential growth of data over the years.

Fig1.1 Data Growth Worldwide

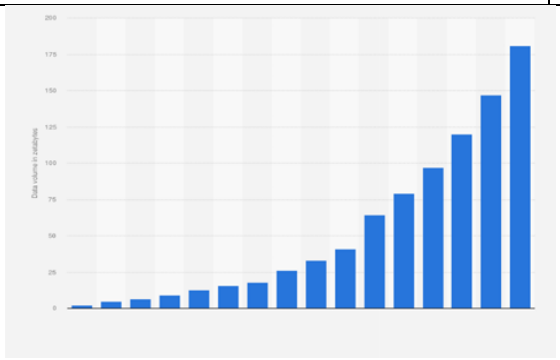
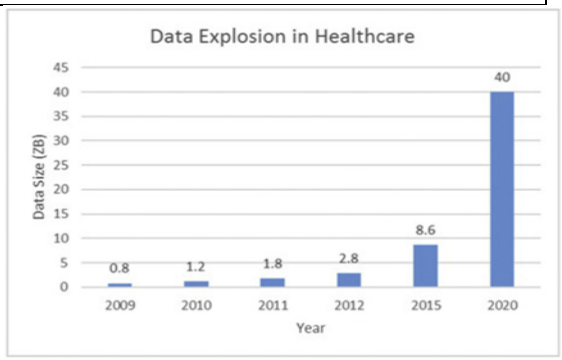


Fig:1.2 Rise in Healthcare Data



The authors in [15] exhibited how handling large amounts of data, or "big data," with RDBMS technologies was easier in the past but more challenging now. They expressed the five dimensions—volume, velocity, variety, value, and complexity where the big data differentiates from other types of data. The Hadoop architecture, which handles massive data and the scalable algorithms for log management was demonstrated here. The focus was also on the difficulties businesses must overcome while managing large data, such as search analysis and data privacy. Tasks related to data mining have expanded dramatically as the datasets continue to grow and is another significant obstacle dealing with such data. The following activities like data reduction, data selection, and feature selection are applied when working with huge datasets. This poses an unprecedented challenge when working with very high dimensional data as the current algorithms might not always react quickly enough. One of the main challenges in recent years has been automating this procedure and creating new ML algorithms to guarantee consistency. Moreover to all of these the primary focus is on clustering enormous datasets that facilitate big data analysis [5]. A significant amount of data, both structured and unstructured, can potentially be gathered swiftly through the use of Hadoop and MapReduce. The main architectural problem is figuring out the best way to use these data for more insightful analysis. Converting semi-structured or unstructured data into structured data and then using data mining techniques to retrieve insights is a common procedure to achieve this goal [16].

Here a novel MapReduce scheduling method to improve the data localization of map tasks is introduced which is integrated into Hadoop default FIFO scheduler and Hadoop fair scheduler. According to experimental results, this method frequently yields the lowest response time and the highest data locality rate for map jobs. Furthermore, it does not necessitate a complex process of parameter adjustment, in contrast to the delay algorithm [17]. The framework of the big data application architecture—which involves data collection and storage—was addressed here [18]. The emphasis is on real-world application situations involving Flume, Kafka, GFS, HDFS, and various data handling techniques. A large-scale information production for mobile and the widely used data analysis methods was presented.

2. Characteristics of Big Data

"The datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" is the way McKinsey describes big data [19]. In order to gain a better understanding from the ever-growing mammoth datasets, we must have to appropriately categorise this data. Therefore, 8V‘s can be used to describe the characteristics of big data: volume, variety, velocity, value, veracity, validity, volatility and visualization. These qualities help us not only understand big data but also how to handle massive, fragmented data at a manageable speed in a reasonable amount of time so that we can extract value from it, do real-time analysis, and react quickly.

Table1. Characteristics of Big Data

Volume	One of the major features of big data is volume. As we already know, big data refers to the enormous "volumes" of data that are produced every day from a variety of sources, including online social networking platforms, corporate procedures, gadgets, sensors, the networks, and human interactions. Data warehouses are used to store such vast amounts of data. The utilisation of sophisticated processing technology—much more potent than a standard electronic system is required to handle these enormous volumes of data. With all of this constantly growing data, there is a tonne of opportunity for analysis, pattern recognition, and much more.
Variety	Structured, unstructured, and semi-structured data collected from various domains are referred to as variety of big data. Data may now be gathered from a wide variety of sources. Conventional data formats are well-structured and compatible with relational databases. New unstructured class of data has emerged with the emergence of big data. Further pre-processing is required for these unstructured and semi-structured data formats in order to enable metadata and derive meaning.

Velocity	In essence, velocity is the rate at which new data emerges in real time. More broadly, it includes the rate of change, the linking of incoming data sets at different rates, and bursts of activity. Velocity is the general term for the speed at which data is received and processed. The classification of data as large data or regular data is also influenced by the rate at which it is gathered. Systems must be able to handle the speed and volume of data creation since a large portion of this data needs to be assessed in real-time. The velocity of data processing must be at least twice as fast as the processing speed of data, which implies that there will be an increasing amount of data accessible than the prior data.
Value	Another important characteristic is value; the majority of data is useless to the business unless it can be transformed into something valuable. Therefore, in order to extract information, it must be transformed into something valuable. It goes beyond the volume of data that we can handle or save. The quantity of important, dependable, and trustworthy data that must be processed, saved, and examined in order to uncover insights is the real issue.
Veracity	In essence, veracity refers to the level of dependability that the information possesses. Big Data must discover a different method to filter or translate the majority of unstructured and irrelevant data because it is essential to extract any meaningful information from the datasets. Therefore, we must apply big data technology to eliminate the ambiguous, partial, and uncertain data in order to obtain huge veracity.
Validity	Surprisingly a significant quantity of big data—also referred to as "black data" remains useless. Validity refers to how closely the data matches the desired idea or information. It entails determining if the gathered data is useful and trustworthy for analysis, as well as whether it corresponds with the stated objectives. By ensuring that the data utilized for evaluation is adequate and fit for the intended purpose, validity lowers the possibility that decisions made based on incomplete or insignificant information will be faulty.
Volatility	Volatility is a measurement of the lifetime or rate of change of the data. It establishes a time limit for storing pertinent information. Frequent variations, a large volume, variety, and velocity of data are among its characteristics. To obtain significant insights, real-time processing and analysis are frequently necessitated. Volatility is influenced by elements such as data sources, processing methods, and storage setups.
Visualization	The difficulty of visualizing big data is another vital characteristic. Because of the constraints in technology and its low scalability, functionality, and response time, current big data visualization faces significant technological challenges. Traditional graphs are not feasible for plotting billions of data points; therefore adopting alternative methods of data representation is essential. Hence creating a meaningful visualization is difficult when you take into account the plethora of factors that come from the variety, velocity, and due to intricate interactions of big data.
Vulnerability	Big data raises unexpected privacy issues. A massive data leak is, after all, a big data theft. Big data system flaws can result from a number of things, such as weak authentication rules, ineffective data encryption, and unsecured repository of information. Unauthorized accessibility, privacy infringements, and data breaches can all result from these vulnerabilities. To reduce these threats, organizations must put robust safety mechanisms in place.
Variability	In the context of big data, variability can mean several distinct things. The quantity of data discrepancies is one. Any useful analytics will require using anomaly and outlier detection techniques to locate these. Big data is unpredictable due to multiplicity of data dimensions arising from several dissimilar data kinds and sources. Another factor is the erratic rate at which large amounts of data are inserted into the database. In order to ensure data quality,

data cleaning and validation procedures are usually used to address this unpredictability.

3. Applications of Big Data

3.1 Healthcare Data

Any type of data about health issues, quality of life, and health outcomes is considered health care information. Everything from patient records and medical imaging technology to gadgets like smart watches can provide health care data. Therefore, among other things, health data can be used to monitor risk variables that might be indicators that a patient may acquire a specific medical disease, measure the quality of care delivered by healthcare systems, and offer recommendations for clinical decision making.

Electronic Health Records/ Electronic Medical Records

The way healthcare professionals handle information regarding patients has been completely revolutionized by the introduction of two particularly important forms of healthcare data: electronic health records (EHRs) and electronic medical records (EMRs). EMR data is the electronic equivalent of a patient's physical document in a doctor's office, which includes all of the patient's medical and treatment histories from a single practice. However, EHR data expands on this idea by providing a more thorough and integrated picture of the patient's wellbeing. In order to guarantee that a patient's wellness details is available and carefully divulged across multiple healthcare organisations, electronic health records (EHRs) are made to share information with other healthcare providers and organisations, including laboratories, specialists, medical imaging facilities, pharmacies, emergency facilities, and clinics. This easy access to the patient's medical history for all authorized clinicians participating in their care not only increases the accuracy of diagnosis and treatment plans but also improves the overall quality of healthcare.

The following are the most widely acknowledged advantages of big data in healthcare. More patient data is a chance to enhance care and gain a deeper understanding of the patient experience. Enhanced investigation of such large data provides medical researchers with previously unheard-of access to vast amounts of data and data collection techniques. These data may therefore lead to significant medical advancements that ultimately save lives. More intelligent treatment strategies can be developed and also even better treatment plans for patients in the future by analyzing the treatment plans that worked/didn't work for previous patients can be rigorously analyzed. Medical expenses can be high. By determining the best course of medication, distributing resources wisely, and spotting any medical conditions prior to than they arise, big data presents an opportunity to lower the expense of receiving and offering treatment.

Administrative Information

Every engagement with the healthcare sector, including doctor visits, diagnostic tests, hospital stays, encounters with long-term care institutions, home care, and prescription drug fillings, generates healthcare administrative data. In addition to being collected for administrative or billing purposes, this data can be used to examine the benefits, drawbacks, and efficiency of healthcare service. Details on hospital discharge information that will subsequently be reported to the government or other bodies are included in some portions of this information. Several kinds of health-related data mostly consist of monetary and activity-tracking information, which are essential elements of many observation plans.

Insurance Data

The financial exchanges that take place between insured patients and the healthcare delivery system mostly that deal with the reimbursement or coverage are included in claims data. Claims include details on patient diagnoses, treatments and tests performed, dates of services, expenses, and locations of services. Usually, the information comprises the following the standardized codes that represent patient conditions and help with resource allocation and trend analysis, the procedure codes that help with effective invoicing, resource allocation, and treatment evaluation for medical services, unique IDs that ensure healthcare providers are paid correctly and are made clear to deal with and finally the accurate records of the times services were provided in order to monitor patient care, optimize processes, and enhance health outcomes. Exploiting claims data in

conjunction with other real-world data creates new opportunities for medical research. This opens up a world of creative uses, like monitoring disease outbreaks, assessing therapy efficacy, and enhancing pharmaceutical safety. These revelations offer a more profound comprehension of patients' personal encounters and medical procedures.

Disease Registries

Patient/disease registries are databases of supplementary information linked to individuals who have received a diagnosis of a certain ailment, illness, or treatment. For the long-term monitoring of medications, these data are crucial. Vital data for long-term illnesses like Alzheimer's, cancer, diabetes, heart disease, and asthma is tracked via diagnostic information systems. These sources of information provide essential perspectives for managing and tracking patient problems. In general, registers range in intricacy from simple spreadsheets accessible only by a small number of clinicians to complicated web databases that are available to several organisations. These registries can also serve as a helpful reminder for medical professionals (and even patients) to order particular tests in response to particular patient needs.

Health Surveys

National health surveys determine the prevalence of diseases and evaluate the overall wellness of the populace. These widely accessible data sources are carefully chosen to support specific investigation goals. The National Health & Nutrition Examination Survey, the Medicare Current Beneficiary Survey, and the National Medical Expenditure Survey are a few examples of this type of data. Data from health surveys are used by public health researchers to examine psychological wellness and health-related behaviors. Health surveys are especially useful for pinpointing specific health issues or risk factors, such as alcohol and tobacco use, poor eating habits, and physical inactivity, in the neighborhoods surrounding healthcare facilities. When it comes to making decisions about health plans, health surveys are crucial because they provide precise information about individuals' experiences with healthcare services, health patterns, lifestyle choices, and epidemiological status.

Clinical Trial Data

Data from clinical trials are derived from investigations that assess patient therapies, whether they are behavioral, surgical, or medical. These data points are the main means via which researchers can investigate ways to identify diseases early on, frequently before symptoms appear, learn how to prevent some health complications, even in those who appear to be in good condition, improve the standard of living for individuals with severe illnesses or long-term medical issues and finally different kinds of data that is gathered during the clinical study can be transformed into datasets that may be analyzed to address specific research questions.

3.2 Big Data in Finance

Large, varied, complex structured and unstructured data collections that can be leveraged to address persistent business problems are referred to in the financial industry as "big data." Since the financial services industry processes, analyses, and uses data in a variety of valuable ways, it is widely regarded as one of the most data-intensive industries. In the past, humans supervised the number-crunching and made decisions based on conclusions from patterns and estimated risks. However, computers have recently supplanted such capacity. Consequently, the finance sector presents a highly promising industry with immense potential for big data technology.

Personalized Service for Clients

Similar to enterprises operating in many sectors, banks leverage big data to get insights into their customer base and, consequently, devise novel approaches to enhance customer service, establish deeper connections, and provide greater value. It is possible to optimize customer experience by using data to gain insightful knowledge about user behavior.

User segmentation and classification

ML algorithms enable efficient client segmentation through massive data analysis. Financial sector is able to use big data analytics to classify customers according to a variety of factors, including credit card spending and their assets. This makes it possible to create smart and efficient marketing strategies that are more precisely customized to the demands of each individual customer.

Business Process Optimization and Automation

Banking firms may maximize their business processes and gain a lot of advantages by utilizing AI, ML and Big Data analytics. The majority of issues with business process optimization can be resolved with the careful and precise application of these technologies. One such serious issue in this particular sector is the ability to identify vulnerabilities and take appropriate action to address them by utilizing observations developed following the processing of reliable datasets. The processing and analysis of data is another important area that will enable them to determine which areas resource allocation will yield the highest level of profitability. Additionally, by utilizing data sciences for employee recruitment and management, they can assist in hiring talent in line with the needs of the company and further utilize predictive data models to evaluate their staff's performance.

Fraud Detection and Prevention

Fraud identification and prevention are among the most urgent problems facing the financial industry. Big data analytics can stop illegal transactions by keeping an eye on consumer spending trends and spotting odd behavior. It also helps to improve the general security of the financial sector by recognizing anomalous behavior and identify fraud's unlawful conduct.

Smart Lending Decisions

Analytical techniques using big data allow banking to see a customer’s databases economic condition in greater detail, which enables them to make more sophisticated loan decisions. Businesses can even go so far as to employ non-traditional models that evaluate prospective borrowers' credibility by fusing big data from sources like social media platforms.

Enhanced risk management and cybersecurity

Technologies like AI and big data are essential for spotting fraud and averting internal hazards. Big data analytics provides cybersecurity teams with more information and more intelligent methods to identify risks early and take swift action. Challenges can be identified early and prevented from causing significant harm by gathering data from every machine and applying clever techniques like identifying unusual behavior and mapping links. Large-scale attack strategies can be seen by connecting relevant threats with the use of big data techniques. Analysts can create a model with predictive abilities that can send out an alarm instantly as it detects a potential entry point for a cybersecurity attack by using advanced big data analytics. An important part of creating such a mechanism is ML and AI. Analytics-based solutions assist in anticipating potential events and preparing for them.

Comprehensive Evaluation of Stock Prices

The stock market is a vital component of the country's economy and has played a significant part in carrying out the evolution of financial developments up to this point. The stock market is a very complex, challenging, fluid, and unexpected field of expertise. Predicting stock price is a difficult task that can provide the greatest return in any business, regardless of size. The financial industry is currently seeing a rise in algorithmic trading, and machine learning enables computers to analyze data quickly. Big data analytics' real-time image has the ability to enhance trading companies and individuals investing chances. To put it briefly, everyone from giant financial institutions to novice traders can use big data to improve their decision-making regarding investments. Information is provided in an accessible way so that traders can make profitable decisions.

3.3 Big Data in Education

The field of Big Data in Education pertains to the systematic collection, manipulation, and analysis of vast and diverse data sources in the educational domain. It entails the use of enormous datasets produced by academic establishments, learners, instructors, and associated systems. These databases frequently contain a vast range of data, including the statistics of students, details of their academic performance, information on their attendance, online activities, assessment outcomes, and much more. This abundance of data is analyzed using advanced analytics and data mining techniques to find relevant patterns, correlations, and trends. These discoveries set the stage for data-driven decision-making in the educational sector, allowing establishments to customize their approaches and improve the quality of their student's education as a whole.

3.4 Big Data in Communications, Media and Entertainment

To analyze viewer behavior and enhance their offerings in a way that would lead to success and establish them as the audience favorite, the Media and Entertainment Industry also aggregates and gathers similar types of data from other sources. It's a well-known truth in marketing and business that the more you understand your target audience, the easier it is to cater to their preferences and adjust the user experience and content appropriately. Big data facilitates the retrieval of up-to-date insights into consumer preferences and enables the organization to adjust content in response to platform demand in mass. Big data not only helps consumers and businesses create appropriate commercials, but it also helps businesses create successful marketing techniques based on factors like the climate, scheduling, and dual screen usage.

3.5 Big Data in Manufacturing and Natural Resources

Big Data analytics is causing a revolution in the manufacturing sector. Manufacturers may save costs, enhance product quality, optimize their supply chain, and obtain insights into their manufacturing processes by employing data from several sources. Additionally, big data analytics can help factories save downtime, forecast maintenance requirements, and provide safer working conditions. Furthermore, producers may now make data-driven decisions that can spur growth and profitability thanks to the usage of big data in the manufacturing sector. Manufacturers may decrease waste and increase operational efficiency by adopting the proper Big Data strategy. Big Data analytics may help manufacturers stay ahead of the curve in a world that is becoming more and more data-driven while also driving growth and profitability when the proper systems and procedures are in place.

3.6 Big Data in Government

Governments may benefit greatly from big data analytics, which has a plethora of uses and applications. Aside from enhancing operations, preventing fraud, and comprehending the demands of citizens, smart use of big data analytics can help central, state, and local governments save billions of rupees every year. The advantages of big data for government agencies is the higher productivity, better decision-making, and improved citizen services with real-time data insights. Analyze development programme trends and results to forecast future spikes in issues related to health, food aid requests, education, and other areas. In order to take the appropriate corrective action, it is imperative to promptly identify any abnormalities or disparities that may indicate fraud attempts, possible resource waste, or harmful actors abusing government resources. It will also help to control the crime rates and security risks as a result of proactive measures taken to stop suspicious or aberrant behavior. Finally, there will be enhanced operational results in medical or natural catastrophe situations because data-driven insights can help with effective emergency response preparation.

3.7 Big Data in Retail and Wholesale trade

Big data analytics gives traders access to so much useful and practical information that it is becoming indispensable for businesses in practically every decision. The idea behind retail analytics is to leverage big data to analyze customer behavior and optimize the price and supply chain. In order to identify patterns, trends, human behavior, and their interactions, a massive amount of data is used. The retail sector needs to gather a lot of information, including past purchases made by customers, in order to increase product sales. The amount of data being gathered keeps increasing since the company may operate online and may broaden its customer base.

3.8 Big Data in Transportation

Businesses in a wide range of travel and transportation industries, including airlines, airports, freight logistics, hotels, railroads, and others, are benefiting from big data's ability to handle vast amounts of data. Nowadays linked and configured world collects an incredible quantity of data from every industry, including the transportation sector. Consequently, the advantages of big data and analytics enable transportation companies to accurately improve a variety of model parameters, including capacity, demand, revenue, pricing, customer sentiment, and cost. Through the analysis of historical data, current and previous environmental trends, and real-time traffic conditions, big data analytics algorithms will assist in determining the most efficient routes. This suggests that businesses can reduce their use of fuel, speed up deliveries, and increase customer satisfaction. Next, devices installed in their vehicles and containers may provide a regular stream of data to the system on wear and tear, possible flaws, and other performance metrics. This suggests that companies may prolong the life of their most valuable assets and decrease unanticipated breakdowns, which could otherwise result in costly delays, by performing more proactive management.

3.9 Big Data in Energy and Utilities

Identifying, utilizing, and implementation of energy are the topics that are discussed the most these days on a global scale. Reusable and renewable energy sources are crucial for consumers as well as companies. Energy is used extensively in modern times. Currently, the energy industry supports and powers all other operations in some way. More energy than ever is required by every entity, and they all want it cheaply. In the past, this was an impossible endeavor, but with the advancement of big data and analytics, it is now feasible. Businesses can gather, store, and analyze enormous amounts of data—terabytes and petabytes—thanks to big data. The power and energy sector has been processing large amounts of data on a regular basis and has worked with big data for years.

Big data is being utilized to make strategic investment decisions and to address complicated corporate issues. Energy businesses filter the data they gather from thousands of sensors using sophisticated statistical techniques, and then utilize this information to make wise investment choices. These statistics offer all the required information and aid in assessing consumer demand. In order to adapt their operational model to the difficulties ahead, power generation businesses utilize advanced analytics and modeling to estimate future prices. Energy makers may make more strategic decisions by using big data to evaluate the risk profiles of their portfolios and potential opportunities. It has been demonstrated that big data analytics is a key enabler for attaining the best possible business outcomes in the energy and power industry. Energy companies are becoming more competitive as a result of data analytics. They are able to produce greater economic value and consumer happiness because to innovative distribution patterns, improved business interaction models, and massive amounts of data.

3.10 Big Data & Auto Driving Car

Big data is the new engine propelling innovation in the automotive sector, influencing market trends and the life cycle of vehicles. Large volumes of data are produced by connected automobiles using IoT technology, allowing for more safety, maintenance predictions, and customized driving experiences. This analysis greatly improves predictive maintenance and vehicle reliability, enabling automakers to maximize technology and boost output. The industry is revolutionized by on-board sensors, which makes driving more dependable and efficient. By spotting possible risks and enhancing safety features with sophisticated analytics, big data makes enhanced security achievable. Data analysis revelations provide preventive security measures that may avert collisions and injuries; analytics-driven insights also improve fuel efficiency, navigation systems, and driver assistance features that provide personalized driving experiences and real-time updates. The automotive enterprises could benefit greatly from big data in a number of ways, including supply chain optimization, loyalty building, and comprehending consumer preferences.

Big data and analytics have unquestionably had a significant impact on the automotive sector, resulting in advancements in efficiency, safety, and personalization. The company's ability to remain competitive allows it to improve safety and offer predictive maintenance, which lowers the amount of failures and accidents and helps

to improve safety in next car models. Improvements in vehicle dependability, safety features, fuel efficiency, navigation systems, and driver assistance systems have resulted from this data-driven approach to the driving experience. Furthermore, big data has been crucial in determining consumer preferences, cultivating client loyalty, and streamlining the automotive supply chain, all of which have improved operations' efficiency and profitability.

3.11 Big Data in IoT

Our interactions with the physical world are changing fundamentally as a result of the convergence of big data and the internet of things. Distributed networks are used to connect physical items, which are referred to as the Internet of Things (IoT). Numerous sensors collect data and distribute it to systems that handle, organize, filter, and evaluate the information. Anything from wearables to healthcare devices to manufacturing machinery can be referred to as an IoT device. Organisations now have unmatched real-time visibility into what's happening across all of their connected devices because of to the Internet of Things. From linked IoT devices, a massive amount of real-time information are gathered and sent across the internet for assessment and retention. IoT acts as a constant source of data for Big Data analytics, while Big Data offers the facilities and resources to manage the massive volumes of data produced by IoT devices. This collaboration is propelling progress across multiple sectors and producing a more interconnected and smarter future. Big Data and IoT integration will further spread as technology advances, creating even more revolutionary applications and significant implications.

3.12 Big Data in Marketing

Successfully addressing various consumers is the core of marketing, and Big Data reveals whether it is productive or not. Today's marketing teams and businesses have access to more data than they know what to do with, which presents enormous potential provided the data is used and understood correctly. A vast amount of data about their customers from many sources, such as social media, online transactions, and browsing patterns can be accumulated by using big data analytics. The clients can be more precisely and usefully grouped using the data based on their demographics, behaviors, and interests. By using this valuable information marketers can then tailor their offers and content to each segment, the end result will be higher engagement and better conversion rates. Another advantage of using Big Data in this field is to monitor market trends and competitors activities. Companies can use client feedback, pricing data, and social media comments to uncover the advantages and disadvantages of their competitors. Further, the corporations may anticipate threats and obstacles by using data analysis. For example, social media sentiment analysis can be utilized to help detect and manage unfavorable marketing issues before they get worse.

4. Challenges of Big Data

In our increasingly computerized environment, the organizations are generating enormous amounts of data every minute. The volume of data generated every minute makes data management, storage, utilisation, and analysis difficult. Even major corporations are having difficulty figuring out how to use this enormous volume of data. As previously indicated, the volume of data generated by major corporations is increasing at a rate of 40 to 60% annually today. Organisations are looking at the availability of big data analysis tools that will greatly assist them in handling big data because simply storing this enormous amount of data will not be very helpful.

Data Transfer and Availability

The separation of data sets from external resources is among the most common problem encountered in large-scale data initiatives. The exchange of information might provide serious obstacles. Information obtained from open businesses presents a variety of challenges. Data must be available in a fast, accurate, and comprehensive manner. Information must be available in this way if it is to be used in the organization's data structure to help decision-makers make precise decisions at the right moment.

Addressing the Massive Growth in Data Volumes

Appropriately storing the enormous data volumes i.e. generated at a very rapid space is one of the biggest problems with big data. The companies databases and data centers are storing an ever-increasing volume of

data. These data sets become very challenging to manage as they grow enormously over time. The majority of data is unstructured and originates from text files, audio files, movies, and other sources. This indicates that databases do not have them. This can provide significant Big Data analytics difficulties that need to be addressed right away to avoid impeding the company's expansion.

Poor Data Produces Disappointing Results

One of the main issues with big data that organisations are facing when making decisions is the poor quality of data. Inaccuracies, inefficiencies, and deceptive insights are caused by poor data effectiveness, and they ultimately result in costs to the organization. It might be as little as when a business makes a mistake in entry and cannot match one customer with the correct order. Inconsistent, obsolete, missing, erroneous, illegible, and duplicate data might lower the whole set's quality. Serious big data issues can arise from even little errors and inconsistencies. For this reason, monitoring its quality is crucial. If not, there can be more harm than good.

Big Data Protection

Data is gathered from a variety of sources, not all of which can be trusted to be secure and compliant with company policies. Therefore, the most cumbersome task with big data is securing these enormous data volumes. Integrating data sources that were not meant to be merged can compromise safety as well as security. Companies frequently put off data security until later because they are too busy comprehending, storing, and analyzing their data sets. However, this is a bad idea because vulnerable data repositories can serve as havens for malevolent hackers.

Scalability in Big Data

Considering the massive quantity, velocity, and variety of data involved in Big Data, the idea of data scalability is essential. Scalability is the capacity of a system or infrastructure to support expanding tasks and volumes of data while preserving effectiveness and efficacy. Large volumes of data must be able to be stored in big data systems. Such enormous data sets might be too much for conventional storage systems to handle. Also working with enormous volumes of data, big data processing entails calculations and analysis. Conventional processing methods could find it difficult to manage the workload and complete calculations in a timely manner. Therefore, traditional systems with limited resources fail to cope with these large data sets.

5. Big Data Analytics and Decision Making

Developing intelligent choices is critical in the hectic corporate world where clients expectations are always changing. It aids businesses in becoming more customer-oriented, promptly adjusting to shifts in the market, and keeping hold of important clients. Here, AI-powered data analytics is the ideal answer since it can improve the way we make our decisions. The first one is increasing efficiency and speed where we can process large amounts of data quickly by utilizing the capabilities of AI algorithms. Consequently, we are able to accelerate our decision-making process, react promptly to shifting market conditions, and make well-informed decisions.

These systems can retrieve hidden patterns in large, complicated data sets that are typically invisible to human observers. These patterns can highlight links, correlations, and market trends that have a big influence on our business choices. Finally, one of the most important quality of the system is its capacity to approach decision-making objectively and logically, where the judgments are fact-based, error-free, and accurate by doing away with human biases and emotions.

6. Big Data Analytics Tools and Methods

A variety of applications and architectures that are designed to manage, process, and evaluate massive amounts of data are referred to as big data tools. Data processing, data storage, data analysis, and data visualization are made easier by these instruments. Some of the important Big Data tools are discussed in the following section. Since each tool has advantages of its own, they are frequently used to solve certain big data problems.

Table 2. Big Data Analytics Tools

APACHE HADOOP	Massive dataset analysis is made possible by Apache Hadoop, an open-source Java platform that uses distributed storage and parallel programming. A Hadoop cluster is made up of thousands of readily accessible, cost-effective units that are connected using Hadoop. Each node in a Hadoop cluster is an individual computer with its own dedicated storage and disc space; that is, it shares no resources with other machines other than a shared network. Nodes function as a single, centralized system focused on only one task. Actually, nodes are divided into three separate groupings, each with a specific role. Distribution of data and resources, as well as parallel processing coordination, is handled by a master node. A client or edge node acts as a gateway between a Hadoop cluster and external systems and applications, while a slave or worker node carries out tasks assigned by the master node. It loads data and retrieves the processed results by avoiding the master-slave hierarchy. Large volumes of data from many sources can be processed and stored using Apache Hadoop. With the growing amount of unstructured data from social media, smart devices, photos, and other sources, therefore Hadoop is the most effective approach to extract valuable insights from it.
APACHE STORM	Modern big data architectures are increasingly requiring real-time data analysis. It meets the need for real-time analytics and insights and enhances conventional sequential processing solutions. An open-source, free distributed real-time computing system is called Apache Storm. Processing infinite streams of data securely is made simple by Apache Storm, which performs real-time processing in the same way as Hadoop performed batch processing. Using Apache Storm is effortless and pleasant, and it works with any kind of programming platform. Considering its unique functionality and applicability, it offers many benefits. It provides an effective strategy for capability building by analysing information in real-time that is often utilised in real-time applications. Real-time processing is possible because Apache Storm is an exceptionally real-time analytic framework. Because it is user-friendly and open-source, it can be integrated with both high and small industries. It produces real and legitimate results and is very fast and valid. It possesses a robust processing capability and the operational potential of intelligence. It is highly compatible with large datasets due to its enormous volume and pace of data absorption. It supports all programming languages and is easily achievable and adaptable.
CASSANDRA	A Java-written NoSQL database is called Apache Cassandra. In addition to processing large volumes of data and unstructured data types, it provides excellent dependability and scalability. Compared with different databases, Apache Cassandra can handle tasks like replication considerably more easily because it does not require an established set schema. Cassandra's main architecture consists of a collection of nodes. Fundamental to Cassandra's architecture is the idea that each node is equal and has the same amount of significance. Every node is the precise location where a certain piece of data is kept. A data center is made up of a collection of interconnected nodes. A cluster is the entire collection of data centers that are able to store data for processing. Major corporations and freelance programmers alike embrace Cassandra as their go-to database management solution due to its robust functionalities and distinctive distributed design.
KAFKA	Given that data must first be linked and made available to enterprise users before analytics can be performed on it, analytics is frequently referred to as one of the

	most complex processes related to big data. Data may be gathered, processed, stored, and integrated at scale using Apache Kafka, an event streaming platform. Distributed streaming, stream processing, data integration, and messaging are just a few of its many application cases. Streams are important to Kafka's architectural design. An unchanging, ordered succession of records that are processed as they come in is called a stream. A timestamp, a value, and a key make up each record in a stream. Partitions, or ordered subsets of records, can be created from streams. Partitions enable several clients to process data in simultaneously. All of these indicate that Apache Kafka preserves both the sequence of events and a very high degree of accuracy in its data records. It supports horizontal scalability and is made to be extremely scalable. Additionally fault-tolerant, Kafka can dynamically recuperate from node malfunctions. Kafka has these wonderful qualities. It has a decentralised architecture and is capable of storing large amounts of data on inexpensive devices. The framework is made to support multiple users. It is possible to use the same published data set more than once. It can send messages to batch and real-time customers simultaneously without sacrificing efficiency, and it stores data on discs. Since it has diversity built in, it can be utilised to give vital information the dependability it needs.
LUMIFY	Lumify offers an extensive range of analytical tools to help customers find connections and explore correlations within their data. It is a big data aggregation, evaluation, and conceptual framework. It is based on proven, robust big data technology and includes specific data intake processing for text, video, and image interface elements. With a range of automated layouts, it offers graph visualizations in two and three dimensions. Not only this it offers a wide range of choices for examining the connections between the graph's entities and also it is based on scalable, well-proven big data technology.
MONGO DB	A prominent and extensively used NoSQL (Not Only SQL) database management system (DBMS) is MongoDB. It is categorised as a document-oriented database, which stores data in a flexible way, in contrast to conventional relational databases. This tool improves security, streamline database administration, expedite requests, and provide insightful data on database efficiency. Similar to other NoSQL databases, MongoDB doesn't need schemas to be predefined. It keeps all kinds of data. Because of this, users can define as many fields as they like in a document, which makes scaling MongoDB databases simpler than scaling relational databases. Another important feature is its horizontal scalability, which helps businesses using big data applications to leverage the database. Furthermore, data can be distributed throughout a cluster of machines by the database by creating zones of data using a shared key in MongoDB.
QUBOLE	QUBOLE Data Service is an all-inclusive self-governing big data platform that uses a blend of machine learning and heuristics to self-optimize, self-manage, and learn from usage. Access to insights is made possible by creating clusters and whenever required expanding and discontinuing them dynamically in response to user queries sent through requests from outside applications. QUBOLE enhances a variety of applications, including weblog analytics, linked cars, fraud detection, gaming analytics, and operational monitoring. Without having to write any code, advanced users can utilise the programme to construct a data pipeline interface.

SPARK	An integrated, open-source engine for massive data analytics is the Apache Spark. Because of its adaptability, it can run on both large clusters and single-node computers, acting as a multilingual platform for data science, ML and data engineering operations. Spark's built-in support for fault tolerance and in-memory distributed processing allows users to construct intricate, multi-stage data pipelines with a surprisingly high degree of ease and performance. Spark's speed has made it possible for the big data industry to choose some technologies over others, which is one of its most important features. Because Spark has RDD (Resilient Distributed Dataset), it may run nearly several times quicker than Hadoop by reducing the amount of time required for data entry and retrieval processes. A plethora of ML, sophisticated analytics, SQL queries, and other features are combined in Spark. It might be used to implement analytics more effectively by utilising each of these features.
TALEND	Big data, data integration, data quality, and data preparation are just a few of the solutions that may be achieved with the open source software integration platform i.e. TALEND. It is one of the most potent cloud computing, big data integration, and data integration ETL tools available in the market. Because it has all the plugins needed to integrate with big data effectively, it is specialized in big data. The metadata is stored and reused in a unified repository that is powered by Talend. The tools data integration feature is able to merge data from multiple sources into a single, fully functional and provide an advanced view of the data. Another major characteristic of this tool is the improved data accessibility, quality, and speed of transfer of the required outcome to the target systems, it is acknowledged as the leading provider of big data integration software for the cloud. When automating a task, the tool provides faster development and deployment. Talend offers a single platform that satisfies all of the companies requirements on one base.
TABLEAU	Tableau is an end-to-end data analytics platform that facilitates the preparation, analysis, sharing, and collaboration of the companies big data insights. An organization can gain a great deal by being able to analyze more data more quickly, as this will enable it to use data more effectively to address key questions. There are many advantages to using Tableau that improve data analysis and decision-making. The principal benefit is in its capacity to manage substantial data sets with ease, allowing users to expeditiously execute intricate investigations. Users without substantial technical expertise may create sophisticated visualizations with Tableau thanks to its drag-and-drop interface. By enabling real-time data analysis, the platform guarantees that consumers always have access to the most recent information. Furthermore, Tableau's scalability and flexibility may accommodate expanding data requirements due to its ability to interact with several data sources.
XPLENTY	Another option for data integration, processing, and preparation for cloud analytics can be performed by using the Xplenty platform. Users may gather, store, and get ready data for cloud analytics with this software. ETL, ELT, and replication solution implementation is facilitated by its intuitive visual design. Xplenty possesses a complete toolset for building data pipelines with low-code and no-code techniques. Few of the notable characteristics of this platform are that it is scalable and elastic. It also provides rapid access to several data sources.

7. Open Research Issues in Big Data Analytics

Big data analytics open research concerns include interpretability of models, scalability of algorithms, real-time analytics, privacy-preserving approaches, data integration from diverse sources, and ethical considerations around data utilisation and bias reduction. Prominent study fields also include tackling issues with quality of data and reliability, guaranteeing security in distributed contexts, and investigating effective methods to handle streaming data.

IoT for Big Data Analytics

The internet has transformed cultural revolutions, international relations, business practices, and an astounding array of individual attributes. The development of facts, network, and communication devices down the road will be greatly impacted by the Internet of Things, both socially and economically. IoT is an extensive system made up of physical objects that have sensors, electronics, and software embedded in them so they can communicate with one another. Due to innovations in mobile devices, cloud computing, data analytics, embedded and ubiquitous communication technologies, the Internet of Things is starting to make more sense in the real world. IoT also poses difficulties in terms of volume, velocity, veracity, value, vulnerability and variety. In a more general sense, the Internet of Things, like the internet, makes it possible for devices to exist in a variety of locations and supports applications ranging from minor to vital.

Effective management of large-scale data that is collected from the large chunk of interconnected electronic devices is crucial. The data derived from these devices is loosely organized and offers a constrained insight. For the purpose of creating the context needed for wise decision-making, this data has to be properly timestamped, indexed, and connected with other data sources. It is challenging to handle data from such wide variety of devices efficiently due to the combination of data volume and complexity. Many technologies made to handle intricate datasets are unable to handle the amount of data generated by IoT devices. However, large-scale data handling systems might not provide the necessary level of in-depth analysis and might not be able to keep up with IoT devices latency needs. Massive data collection is meaningless without procedures in place for cleaning, organizing, and processing it. IoT data management is crucial because it helps businesses to get the conclusions they require from the data that their connected gadgets gather. In an IoT context, using real-time analytics is essential. Only when real-time analytics are applied to stored data can the benefits of IoT be observed. Therefore one of the important approaches to manage large amounts of data from IoT perspective is to use ML algorithms and computational intelligence approaches.

Cloud Computing for Big Data Analytics

A potent tool for handling sophisticated and large-scale computation is cloud computing. It does away with the requirement for costly computer gear, dedicated storage requirements, and software. There has been a noticeable explosion in the amount of large data or data scale that is produced by cloud computing. Handling of such huge and complex datasets is a difficult and time-consuming operation that necessitates a substantial computational infrastructure for effective data processing and analysis. Despite the widespread adoption of cloud computing by numerous organisations, big data in the cloud research is still in the beginning stages. A number of current problems have not received enough attention. Additionally, new problems are still arising from organizational applications. Some of the major research challenges are discussed in the following sections.

Scalability is the storage system's capacity to manage growing volumes of data in a suitable way. The facilities for cloud computing have been crucially dependent on scalable distributed data storage systems. Because big data analysis on the cloud involves large volumes of data, scalability is a major difficulty. It is imperative to guarantee that systems can effectively manage growing data loads without compromising their dependability and efficiency. In order to effectively use resources across numerous nodes, distributed processing frameworks like Hadoop or Spark must be optimized. Furthermore, size complicates the management of data storage and retrieval, necessitating the use of scalable storage solutions like cloud-based object storage or HDFS. In addition, cloud-based big data environments continue to have issues with scaling both horizontally and vertically, as well as dynamically assigning resources to suit changing demands.

A significant obstacle in cloud computing for big data analysis is resource availability, or making sure that customers can regularly access and use data and services. In order to guarantee continuous service availability, methods that can tolerate failures at several levels—such as hardware, software, and network problems—must be developed. In a normal scenario, resources should still be made available even in the event of a node failure or network-related problems. The ability for clients to maintain and retrieve their data in cloud data centers is made possible by the increasing number of cloud-based applications. These applications must guarantee data integrity. Nevertheless, one of the primary obstacles that must be dealt with is ensuring the validity of user data in the cloud. Given that users may not be physically able to retrieve the data, the cloud must offer an interface for the client to validate how well the data is maintained. Another major concern here is the variety that the big data possesses, which arises from a growing number of nearly infinite data sources. Big data's varied nature results from this growth. Multiple sources of data typically have disparate types, representations, and levels of interconnectedness; they also have incompatible formats and inconsistent representations.

Agriculture for Big Data Analytics

Big data is frequently understood to be a statistical and technological hybrid that can gather, aggregate, and analyse new data and use it to support decision-making processes. It can be challenging to combine data from numerous sources, such as machines, satellites, and sensors, while maintaining consistency and accuracy. Studies conducted on the next generation of agricultural design models demonstrate that data is the most crucial factor when it comes to supporting decisions made on farms, investing in research, and formulating policy. For the purpose of making decisions about farm management, gathering trustworthy agricultural data is crucial. With the use of high-precision algorithms, advances in the notion of smart farming increase the effectiveness and efficiency of agriculture. ML in tandem with high-performance computing and big data technologies has opened up new avenues for simplifying, measuring, and comprehending data-intensive operations in agricultural operating environments.

Bio-inspired Computing for Big Data Analytics

The term "bio-inspired computing" refers to a broad range of recent computer science, mathematics, and biology disciplines. Another developing method called "bio-inspired computational optimising algorithms" draws inspiration and ideas from the biological evolution of ecology to create innovative and resilient diverse strategies. There are a number of issues that need to be resolved for bio-inspired algorithm-based big data analytics, including handling resources, communication, adaptability, durability, diversity in correlated clouds, and security of information, confidentiality, and finally accessibility.

For big data analytics, the cloud leverages virtual resources to process user data efficiently and affordably. Utilising bio-inspired algorithms, virtualization technology effectively manages cloud resources to increase user happiness and resource efficiency. Optimising cloud resource provisioning is necessary for big data analytics using the current bio inspired algorithms. In order to overcome this difficulty, an algorithm-based resource management method that is bio-inspired and quality of service is needed for the effective management of large data in order to maximise quality of service parameters. Data processing that occurs geographically presents a problem for data synchronisation in bio-inspired algorithms, leading to under provisioning of resources. The size of cloud data centres (CDCs) is growing, and this is increasing the complexity of computing systems, which raises the risk of resource failures during big data analytics. Data corruption, early execution termination, and service level agreement (SLA) violations are examples of resource failures and also resources that are overburdened must be found quickly enough to manage the data coming in from various Internet of Things devices. To increase the system's dependability, more information regarding the malfunctions must be discovered. These bio-inspired algorithms may eventually be combined with other strategies and techniques like quantum computing and chaotic theory to improve the efficiency of bio-inspired optimisation algorithms when attempting to solve truly difficult applications.

Healthcare for Big Data Analytics

Big data in healthcare refers to the vast quantities of health-related information gathered from a variety of sources, including wearables, genomic sequencing, electronic health records, clinical trials, and diagnostic

imaging, to name a few. This data is too complicated for ordinary data processing tools and has a volume and diversity of formats that make it impossible to store in regular databases. The kind of insights and "smartness" of statistical information are more crucial for the healthcare industry. A variety of sources and forms, including structured data, images, videos, paper, digital, multimedia, and more, are used to compile health care data. For organisations, gathering data that is clear, accurate, complete, and organized correctly for use in multiple systems is a significant difficulty.

Data analytics is now a crucial tool for enhancing health outcomes and promoting innovation within the industry. Nevertheless, putting good data analytics into practice for healthcare organisations is fraught with difficulties. Realizing the prospective advantages derived from data analytics, safeguarding patient privacy and data security in the healthcare sector, and possessing the requisite knowledge and experience to put into practice efficient data analysis techniques all depend on an understanding of the difficulties associated with using patient data in clinical data analytics.

Quantum Computing for Big Data Analytics

Quantum computing solves challenging tasks that conventional computers or supercomputers cannot solve, or cannot solve quickly enough, by utilizing specialized technologies, such as computer hardware and methods that take advantage of quantum mechanics. Large-scale data analysis is an extremely complex process. In fact, it can be quite difficult for people to sort through massive amounts of data and locate relevant information that can be put to use. A significant amount of the data generated by organisations is ignored. Quantum computing provides high-speed detection, analysis, integration, and diagnosis capabilities when working with huge dispersed data sets. By concurrently scanning every entry in a massive database, quantum computers are able to swiftly identify patterns in massive, unorganised data sets. Extremely complex calculations could take a non-quantum computer hundreds of years to complete, while quantum computers can complete them in a matter of seconds. Extremely complex calculations could take a non-quantum computer hundreds of years to complete, while quantum computers can complete them in a matter of seconds.

These days, large data is handled by artificial intelligence systems that also assist in dataset analysis to find patterns. Even with the speed at which technology is developing, traditional computers can only process a finite amount of data. This constraint, however, does not impede quantum computers. From datasets, artificial intelligence can be utilized to extract relevant historical and contemporary data. When coupled with quantum computing, more data may be analyzed, producing pertinent data that can subsequently be utilized for forecasting. But occasionally a predictive model which has to take into consideration a plethora of choices, variables, and factors cannot handle the massive volume of data that is readily available. Therefore without slowing down the process, quantum computing helps create predictive models that are more scalable.

Conclusion

Over the last 20 years, technological advances have brought about a new era of progress. Unprecedented opportunities and unanticipated challenges have arisen with this advancement. The current investigation used a variety of approaches to conceptualize the big data importance analysis of both structured and unstructured data. It has evolved into a crucial element of most company activities and is recognized as the cornerstone of all decision-making processes. The significance of big data analytics is in its ability to leverage vast quantities of data in many forms from various sources to detect possibilities and threats, enabling enterprises to act swiftly and enhance their profitability. The qualities of big data bring opportunities as well as obstacles. With strong big data management and analysis capabilities can spur innovation and obtain insightful knowledge and maintain their competitive edge in today's data-driven world by comprehending these essential traits. It is not feasible to qualify and validate every item in big data; therefore, new strategies need to be created. Big Data's main security concerns are integrity, reliability, accessibility, and privacy with regard to outsourced data. Using the analysis of performance metrics involving many different types of enterprise processes, big data technology can offer insights into efficacy and efficiency by enabling firms to make well-informed decisions and to modify their plans by analyzing future patterns and outcomes.

Refrences

[1]. T. R. Rao, P. Mitra, R. Bhatt, and A. Goswami, “The big data system, components, tools, and technologies: a survey,” Knowledge and Information Systems, vol. 60, no. 3, pp. 1165–1245, 2019.

[2]. S. Ketu, P. K. Mishra, and S. Agarwal,“Performance analysis of distributed computing frameworks for big data analytics: Hadoop vs Spark,” Computacion y Sistemas, vol. 24, no. 2, pp. 669–686, 2020.

[3]. T. Kolajo, O. Daramola, and A. Adebiyi, “Big data stream analysis: a systematic literature review,” Journal of Big Data, vol. 6, no. 1, pp. 1–30, 2019.

[4]. N. Khan, A. Naim, M. R. Hussain, Q. N. Naveed, N. Ahmad, and S. Qamar, “The 51 v’s of big data: survey, technologies, characteristics, opportunities, issues and challenges,” in Proceedings of the international conference on omni-layer intelligent systems, pp. 19–24, Heraklion, Crete, Greece, 2019.

[5]. L. Belcastro, F. Marozzo, and D. Talia, “Programming models and systems for big data analysis,” International Journal of Parallel, Emergent and Distributed Systems, vol. 34, no. 6, pp. 632–652, 2019.

[6]. M. Bansal, I. Chana, and S. Clarke, “A survey on IoT big data,” ACM Computing Surveys, vol. 53, no. 6, pp. 1–59, 2021.

[7]. J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, “Big data service architecture: a survey,” Journal of Internet Technology, vol. 21, no. 2, pp. 393–405, 2020.

[8]. M. B. Ozcan, B. Konuk and Y. M. Yesilcimen, "Big Data Analytics in Industry 4. 0", Industry 4. 0, pp. 171-199, 2022.

[9]. G. Manikandan, S. Abirami, K. Gokul and G. Deepalakshmi, "Big data analytics in healthcare", Big Data Analytics for Healthcare, pp. 3-11, 2022.

[10]. A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman and R. Maskat, "The state of the art and taxonomy of big data analytics: view from new big data framework", Artificial Intelligence Review, vol. 53, no. 2, pp. 989-1037, 2019.

[11]. Lundberg, L., & Grahn, H. “Research Trends, Enabling Technologies and Application Areas for Big Data”. Algorithms, 15(8), 2022.

[12]. Anwar, M. J., Gill, A. Q., Hussain, F. K., & Imran, M. “Secure big data ecosystem architecture: challenges and solutions”. In Eurasip Journal on Wireless Communications and Networking (Vol. 20, Issue 1). Springer Deutschland, 2021.

[13]. Sen R, Jindal A, Patel H, Qiao S. AutoToken: predicting peak parallelism for big data analytics at Microsoft. Proceedings of the VLDB Endowment, 2020; 13 (12), 3326–3339.

[14]. Said F, Zainal D, Azlina AJ. Big data analytics capabilities (BDAC) and Sustainability reporting on Facebook: Does tone at the top matter. Cogent Business & Management, 2023.

[15]. S. V. Phaneendra et. al. “Big Data- solutions for RDBMS problems- A survey” In 12th IEEE/IFIP Network Operations & Management Symposium, Japan, 2013.

[16] T. K. Das and P. M. Kumar, Big data analytics: A framework for unstructured data analysis, International Journal of Engineering and Technology, 5(1) (2013), pp.153-156.

[17]. Chen He Ying Lu David Swanson “Matchmaking: A New MapReduce Scheduling” in 10th IEEE International Conference on Computer and Information Technology (CIT’10), pp. 2736–2743, 2010.

[18]. Xie, Z. Song, Y. Li et al., “A survey on machine learningbased Mobile big data analysis: challenges and applications,” Wireless Communications and Mobile Computing, vol. 2018, Article ID 8738613, 19 pages, 2018.

[19]. McKinsey, "Big data: The next frontier for innovation, competition, and productivity," May 2011.