IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Exploring The Boundless Power And Potential Of Big Data: A Comprehensive Review

Authors :
Saurabh R. Kalwar
Department of Computer Science
University of Mumbai,
Mumbai, India

Ravi S. Kalwar
Department of Computer Science
University of Mumbai,
Mumbai, India
Guides :
Dr. Harshali Patil
Associate Professor
MET Institute of Computer Science, Mumbai, India

Dr. Jyotshna Dongardive
Associate Professor
Department of Computer Science,
University of Mumbai, Mumbai, India

Abstract
Big data refers to huge, diverse, and complicated data collections that are challenging to store, analyze, and visualize for use in subsequent operations or outcomes.
Big data analytics is the practice of analyzing enormous volumes of data to uncover obscure patterns and hidden relationships.
These helpful details for businesses or organizations can assist them get richer, deeper insights and a competitive advantage. This calls for the most accurate analysis and execution of big data applications. Big data has gained recognition as a ground-breaking technical advancement in recent years.
However, our knowledge of how organizations convert potential into actual social and economic benefit is still limited or how big data exactly helps organizations.
This review examines the big data and its analytics and provides an overview of its characteristics, importance/benefits/opportunities, challenges, advancements, future trends/limitations.
We contend that in order for enterprises to effectively utilize big data, they must constantly realign their work processes, organizational structures, and stakeholder interests.

Keywords—Big data, Big data tools, Challenges, Volume, Variety, Velocity, Veracity, Value, Analytics, Literature review

1. INTRODUCTION

The basis of modern science, industry, and new-era technology is big data and its analysis. In the present day, these data are produced every second from online transactions, emails, videos, audios, photos, click streams, logs, postings, search queries, health records, social networking interactions, science data, sensors, mobile phones, and their applications.

They are kept in exponentially expanding databases that are challenging to gather, create, store, manage, distribute, analyze, and display using standard database software tools. [1]

Big data is a group of data sets that are so numerous and complex that they become challenging to analyse using typical data processing software or traditional database management solutions. Big data is data that is massive and it is beyond the processing capabilities of traditional database systems. The data does not match the constraints of traditional database designs because it is too large, moves too quickly, or both. [17]

Big Data refers to a variety of sources that provide a significant volume of data. Big Data is extremely challenging to handle and store using traditional information systems because of its diverse nature.[3]

Today the size of the big data is in exabytes which is 1024 petabytes of information, which includes trillions of individual records from millions of people. The global Big Data and analytics market is worth $274 billion. Every day, data of about 2.4 quintillion bytes is produced. Currently, the total amount of data in the digital cosmos exceeds 44 zettabytes. [5]

Organizations must handle all types of data in the modern day, including classic structured data, semi-structured data, unstructured data, and poly-structured data, which includes contents like emails, web page content, video, audio, and other types of media. Different data models were being used by conventional database management systems to store and handle data. Only structured data may be used with the majority of these data models. Additionally, it should be recognised beforehand when defining a schema for structured data. Consequently, when Big Data emerged, scaling and managing it turned into a laborious task. As a result, the technology known as "NoSQL" emerged, offering the characteristics and capabilities that the Big Data technology demands. [3]

Structured data is data which is to the point, factual and highly organized into formatted repository.It relates to any information that may be put into a table with rows and columns in a SQL database. They can be quickly mapped into pre-designed fields and have relational keys.[6]

Unstructured data is data which is not well organized in a predefined manner or does not have a predefined data model, which makes it a poor fit for a relational database.[6]

Unstructured data is information that has not been arranged according to a specified data model or schema and hence cannot be stored in a traditional relational database system, often known as an RDBMS.[7]

Semi-structured data refers to data that is not captured or formatted in conventional ways. Because semi-structured data lacks a set schema, it cannot be mapped to the format of a tabular data model or relational databases. However, the data is not fully raw or unstructured; it has certain structural features, such as tags and organisational information, that make it simpler to analyse.[8]

1.1 Characteristics of Big Data

Big data is characterized into five v's that are volume, velocity, variety, veracity and value (as shown in the Fig.1). Out of which the three main v's are. Volume: Huge volumes of digital data are continually generated by numerous systems and apps (ICTs, smartphones, products' codes, social networks, sensors, logs, etc). According to Andrew McAfee et al. (2012), In 2012, it is estimated that around 2.5 exabytes have been generated per day. This amount is doubling every 40 months. In 2013, the total digital data created, replicated, and consumed was estimated by the International Data Corporation as 4.4 Zettabytes (ZB). It doubles every two years. By 2015, digital data has increased to 8 ZB (Rajaraman, 2016). According to IDC research, the volume of data will exceed 40 ZB by 2020, with an increase of 400 times by the present (Kune et al., 2016).

Velocity: Data are generated in a fast way and should be processed rapidly to extract useful information and relevant insights. For example, every hour, Wallmart generates over 2.5 PB of data from its transactions. YouTube is also a good example which shows the velocity of Big Data. [4]

Variety: Big Data is generated from a variety of sources and it is available in different formats (for example, videos, documents, comments, and logs). Massive data sets are made up of structured and unstructured data, public or private, local or distant, shared or confifidential, complete or incomplete, etc. [4]

Big data is defined as data whose size, dispersion, diversity, and/or timeliness require the adoption of new technological architectures, analytics, and tools to allow insights that unlock new sources of business value. Big data is defined by three primary characteristics: volume, variety, and velocity, or the three V's. The volume of data determines its magnitude and enormity. The pace at which data changes or is generated is referred to

as its velocity. Finally, variety includes the different formats and types of data, as well as the different kinds of uses and ways of analyzing the data. [24]
Data volume is the primary attribute of big data. Big data can be quantified by size in PBs or ZBs. Additionally, Big data is truly big, because of its multiple sources. There are different types of data, which is hard to categorize since it comes from audio, video, and other devices. Furthermore, multi-dimensional data can be drawn from a data warehouse to add historic context to big data. Thus, with big data, variety is just as big as volume. [2]
Furthermore, big data can be described by its velocity or speed. This is essentially the frequency of data creation or data transmission. Steaming data, which is collected in real-time from websites, is at the leading edge of big data. [25]
Many studies have discussed addition of a fourth V, which is veracity. Veracity refers to inconsistency or uncertainty in the data. veracity is the genuineness of the data whether the data is true or not. [26]
The fifth V, or value. Value focuses on the business value or usefulness of the data. [19]

1.2 Importance / Benefits / Opportunities with Big data

The importance of Big Data is not determined by the amount of data an organisation possesses. Its significance stems from how the organisation uses the obtained data.
Every organisation uses the data it collects in its own unique way. The more successfully a firm uses its data, the faster it expands.
Companies in the current market need to gather and analyze it since it aids in:

1.2.1. Cost Savings

Big Data tools such as Apache Hadoop, Spark, and others help organizations save money while storing big volumes of data. These tools assist organizations in discovering more successful business practices.

1.2.2. Time-Saving

Real-time in-memory analytics enables businesses to acquire data from a variety of sources. Tools like Hadoop enable businesses to analyse data quickly, allowing them to make swift decisions based on their findings.

1.2.3. Understand the market conditions

Big Data analysis assists organisations in better understanding market conditions.
For example, analysing client purchase behaviour enables businesses to discover the most popular items and develop those things appropriately. This allows businesses to stay ahead of their competition.

1.2.4. Social Media Listening

Companies may use Big Data technologies to perform sentiment analysis. These allow them to get feedback on their company, as namely who is saying what about it.

1.2.5. Boost Customer Acquisition and Retention

Customers are an essential commodity on which every company relies. No firm can be successful unless it has a strong consumer base. Even with a strong client base, businesses cannot avoid market rivalry.
Companies will fail if they do not understand what their customers desire. It will result in clientele loss, which will have a negative impact on business growth.
Businesses may use big data analytics to detect customer-related trends and patterns. Analysis of customer behavior leads to a prosperous business.

1.2.6. Solve Advertisers Problem and Offer Marketing Insights

All corporate activities are influenced by big data analytics. As it allows businesses to meet their customers' expectations. Big data analytics helps in the evolution of a company's product range. It guarantees effective marketing endeavours.

1.2.7. The driver of Innovations and Product Development

Big data enables businesses to innovate and remodel their products. [15]

Today, we are constantly bombarded with data. Data is being collected at incredible scales in a wide range of application areas. Decisions that were previously based on educated guesses or carefully crafted models of reality can now be based on the data itself. This type of Big Data analysis today drives practically every element of modern civilization, including mobile services, retail manufacturing, financial services, life sciences, and physical sciences. [12]

The use of Big Data will become an essential part of competition and growth for individual firms. From the viewpoint of competition and potential value extraction, all businesses must take Big Data seriously. In most industries, existing rivals as well as fresh competitors will use data-driven strategies to innovate, compete, and extract value from deep and up-to-the-minute data. [13]

The usage of Big Data will be important across industries, with certain sectors expected to benefit the most. Big Data has the potential to significantly benefit the computer and electrical products and information industries, as well as banking and insurance, and government. [13]

Some notable achievement involving Big Data

- The Los Angeles Police Department and the University of California are using Big Data to forecast crime before it occurs. Google Flu Trends forecasts the spread of the flu virus based on search phrases.
- Nate Silver, a statistician, predicted the outcome of the US election down to each individual state in 2012.
- MIT is studying how people's mobile phone data may be utilized for urban planning by establishing people's locations and traffic patterns. [14]

The Big data trend and growth

Companies all across the world are already putting strategies in place to store and analyze big data. The majority of IT decision-makers surveyed work for organisations in the manufacturing, healthcare, financial services, and other industries. More than half of its 64 percent are in manufacturing, 29 percent in production, 19 percent in healthcare, and 16 percent in financial services industries, but only 13 percent have a fully implemented big data solution, which will increasingly be a benchmark on which data marketplaces compete. [17]

1.3 Big data Challenges

However, upon deeper examination, just two or three major difficulties appear capable of making or breaking the Big Data promise, and they are connected to: solution strategy, personal privacy, and intellectual priority (IP). The first covers technology, deployment, and the organizational environment, whereas the latter two major concerns the type and relevant use of information or Big Data. Other possible barriers to full Big Data utilization include heterogeneity and incompleteness, scalability, and timeliness; another closely linked topic is data security. [10]

1.3.1 Heterogeneity

Humans can tolerate a lot of variability when they receive information. In reality, natural language's complexity and variety may add important depth. Machine analysis algorithms, however, do not comprehend sophistication and instead want homogeneous data. As a result, data must be carefully formatted before (or prior to) data analysis. Computer systems function best when they can store a number of identically sized and constructed things. Further research is needed to effectively represent, access, and analyze semi-structured data. [9]

1.3.2 Scale

Scale or volume is another significant difficulty for Big Data, as mentioned above. When it comes to big data, size is the first consideration for anybody. After all, the name itself contains the term "Big". For many years, managing vast and quickly growing amounts of data has been a difficult problem. In the past, this problem was lessened by faster processors, which followed Moore's law and gave us the resources we needed to handle the growing volume of data. However, a fundamental change is currently taking place: CPU speeds are stagnant, while data volume is increasing faster than computer resources. We need to reconsider how we design, construct, and use data processing components in light of these remarkable developments. [9]

1.3.3 Timeliness

The flip side of size is speed. The larger the size of the data to be processed, the longer it will take to analyze it. A system that can successfully handle size will probably also be designed to process a given size of data set more quickly. However, when someone uses the term "velocity" in relation to Big Data, it refers to more than simply this pace. Instead, there are issues with acquisition rate and timeliness. There are several circumstances when the analysis's findings must be available right away. For instance, it is preferable to alert a suspected fraudulent credit card transaction before the transaction is finalized possibly stopping the transaction from taking place at all. Obviously, it is difficult to be able to conduct a thorough study of a user's buying history in real-time. Instead, we must build predictions in advance in order to quickly make a decision using a limited amount of incremental computing with fresh information. [9]

1.3.4 Data integration

Data integration is essential for understanding the 5th V (value) of big data through cross-domain collaborations and integrative data analysis (Chen et al. 2013; Christen 2014).
Dong and Divesh (2015) summarised the difficulties in data fusion, record linking, and schema mapping.
Metadata is required for monitoring these mappings in order for the integrated data sources to be 'robotically' resolvable and for large-scale studies to be possible (Agrawal et al. 2011).
However, efficiently and automatically producing metadata from Big Data remains a difficult challenge (Gantz and Reinsel 2011).
Geo-data integration has spawned new prospects in the geospatial domain, fueled by an increasingly collaborative research environment. One example is the EarthCube programme, which was launched by the National Science Foundation's Geosciences Directorate in order to deliver unparalleled integration and analysis of geospatial data from a number of geoscience areas (EarthCube 2014).

1.3.5 Data quality

Data quality includes four aspects: accuracy, completeness, redundancy and consistency (Chen et al. 2014b).
The intrinsic nature of complexity and heterogeneity of Big Data makes data accuracy and completeness difficult to identify and manage, raising the danger of 'false discoveries' (Lohr 2012).
Social media data, for example, is significantly skewed in place, time, and demography, and location accuracy ranges from metres to hundreds of kilometres. Furthermore, data redundancy control and filtering should be performed in real-time at the moment of data collection (e.g. with sensor networks, Cuzzocrea, Fortino, and Rana 2013).
Finally, preserving data consistency and integrity in Big Data is difficult, particularly when the data changes often and is shared with several contributors. [11]

1.3.6 Data visualization

Big Data visualisation reveals hidden patterns and undiscovered relationships in order to enhance decision-making (Nasser and Tariq 2015).
Since Big Data is frequently diverse in terms of type, structure, and semantics, visualization is crucial for making sense of Big Data (Chen et al. 2014b; Padgavankar and Gupta 2014).
However, providing real-time visualization and human interaction for visually exploring and analyzing Big Data is challenging (Sun et al. 2012; Jagadish et al. 2014; Nasser and Tariq 2015).
SAS (2012) summarised five important Big Data visualisation features as follows: (i) highly interactive graphics that incorporate the latest techniques for data visualization; (ii) integrated, intuitive, and approachable visual analytics; (iii) web-based interactive interfaces for previewing, filtering, or sampling data prior to visualizations; (iv) in-memory processing; and (v) easily distributed answers and insights via mobile devices and web portals. Designing and creating these functions is difficult due to the various characteristics of Big Data, such as the amalgamation of many data sources and the high dimensionality and spatial resolution of geographic data (Fox and Hendler 2011; Reda et al. 2013).

1.3.7 Personal Privacy

Consider all of the personal information held and sent by Internet Service Providers, mobile network providers, supermarkets, local governments, medical and financial service organizations (e.g., hospitals, banks, insurance companies, and credit card companies). Not to mention information posted and maintained on social media by religious organizations, educational institutions, and/or employers. Every organization has the challenge of organizing, safeguarding, and using its business, operational, and consumer data. [10]

The unprecedented networking of smart gadgets and computer platforms benefits Big Data but raises privacy issues because an individual's location, behaviour, and transactions are digitally recorded (Cukier 2010; Tene 2011; Michael and Miller 2013; Cheatham 2015).

For example, Personal and health information is contained in social media and individual medical records, presenting privacy concerns (Terry 2012; Kaisler et al. 2013).

Companies are now employing Big Data to analyze worker performance by tracking employees' mobility and productivity (Michael and Miller 2013).

These privacy problems highlight a gap between conventional laws/regulations and Big Data, necessitating the development of new policies to handle privacy concerns fully (Khan et al. 2014; Eisenstein 2015).

1.4 Advancements

Big Data in the Twenty-First Century

Big data as we know it arrives and the explosion of creativity that it provides cannot be overstated. Everybody and everything is affected.

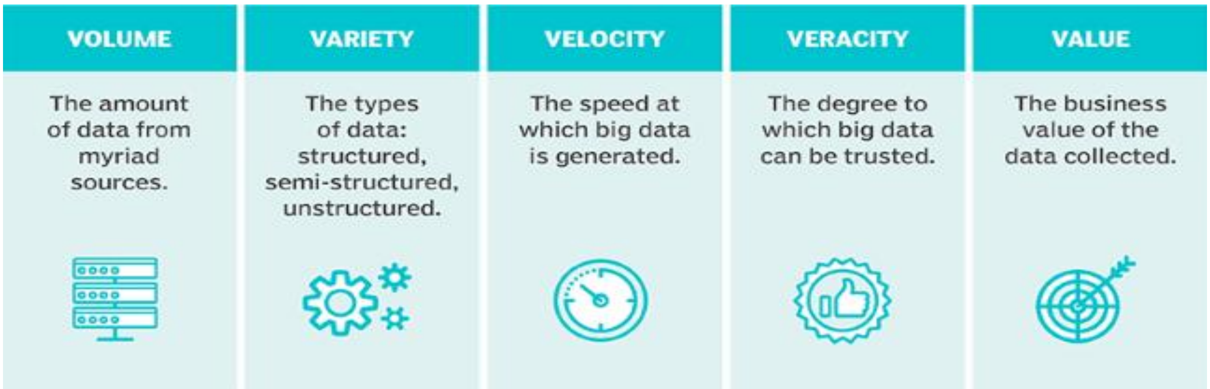
2001

Gartner analyst Doug Laney coined the 3Vs (volume, variety, and velocity) to define the characteristics and features of big data. The Vs capture the fundamental essence of big data and symbolize a new era in which big data is considered as a major element of the twenty-first century. Other Vs have since been added to the list, including variability, veracity, and value.

2005

With a team of engineers broken out from Yahoo, computer scientists Doug Cutting and Mike Cafarella build Apache Hadoop, the open source framework used to store and handle enormous data volumes.

The five v's of big data(Fig.1)



2006

Amazon Web Services (AWS) began providing web-based computer infrastructure services, which are today referred to as cloud computing. AWS now leads the cloud services business, accounting for around one-third of the worldwide market.

2008

Over 9.57 zettabytes (or 9.57 trillion gigabytes) of data are processed by the world's CPUs, roughly equivalent to 12 gigabytes per person. The global generation of fresh information is predicted to be 14.7 exabytes.

2009

According to Gartner, the top priority for CIOs is business intelligence. As businesses suffer economic instability and uncertainty as a result of the Great Recession, extracting value from data becomes critical.

2011

According to McKinsey, the United States will experience a scarcity of analytical skills by 2018. There are between 140,000 and 190,000 people who have deep analytical abilities, and another 1.5 million analysts and managers who can make reliable data-driven judgments. Facebook also launched the Open Computer Project, which will publish specs for energy-efficient data centers. The initiative's objective is to boost energy efficiency by 38% while lowering costs by 24%.

2012

The Obama administration announced a $200 million commitment to the Big Data Research and Development Initiative, citing the need to improve the ability to extract valuable insights from data, accelerate the pace of STEM (science, technology, engineering, and mathematics) growth, enhance national security, and transform learning. The term has now been changed to STEAM, with an A added to represent the arts. The most appealing job of the twenty-first century, according to Harvard Business Review, is data scientist. Demand for data scientists increased as more businesses saw the need to filter and obtain insights from unstructured data.

2013

The global big data market has grown to $10 billion.

2014

In the United States, mobile devices now outnumber desktop PCs for the first time. Two years later, in 2016, the rest of the globe follows suit.

2016

Ninety percent of the world's data was produced in the previous two years alone, and IBM says that 2.5 quintillion bytes (that's 18 zeroes) of data is created every day. [16]

2017

In 2017, the world operates in a more complex analytics environment. The 42 V's of the Big Data.

1. Vagueness: Regardless of how much data is provided, the meaning of the data is frequently quite ambiguous.
2. Validity: For valid forecasts, rigor in analysis (e.g., Target Shuffling) is required.
3. Valor: In dealing with big data, we have to tackle huge challenges head-on.
4. Value: As more data becomes accessible and new approaches are discovered, data science tends to bring ever-increasing value to users.
5. Vane: Data science may help decision-makers by directing them on the appropriate path.
6. Vanilla: Even the most basic models, when built with attention to detail, may bring value.
7. Vantage: Big data gives us a unique perspective on complicated systems.
8. Variability: Variable data sources are frequently modeled in data science. Models in production may meet extremely unpredictable data.
9. Variety: There is a wide range of data types (flat files, relational databases, graph networks) along with data completeness levels.
10. Varifocal: When big data and data science work together, we are able to see both the forest and the trees.

11. Varmint: As big data grows in size, so do software problems.
12. Varnish: How end users engage with our work is important, and polish is important.
13. Vastness: The "bigness" of big data is increasing with the rise of the Internet of Things (IoT).
14. Vaticination: Predictive analytics allows the systems to forecast.
15. Vault: Because many data science applications rely on huge, confidential data sets, data security is becoming increasingly crucial.
16. Veer: With the emergence of rapid data science, we must be able to cope with the demands of the client and shift courses immediately as needed.
17. Veil: The capacity to see behind the mask and evaluate the impact of unknown variables in data is provided by data science.
18. Velocity: The speed at which the data is generated is increasing rapidly.
19. Venue: Data science activity happens in a variety of environments, including locally, on customer desktops, and in the cloud.
20. Veracity: Reproducibility is required for proper analysis.
21. Verdict: As the number of individuals affected by model decisions grows, veracity and validity become more essential.
22. Versed: Data scientists frequently need to know a little bit about a wide range of topics, including mathematics, statistics, programming, databases, and so on.
23. Version Control: You are utilizing it, right?
24. Vet: Data science enables us to vet our convictions by supplementing intuition with data.
25. Vexed: Part of the interest in data science comes from its promise to shed light on huge and complex problems.
26. Viability: It is difficult to construct strong models, and it is even more difficult to develop systems that will be sustainable in production.
27. Vibrant: A vibrant data science community is essential for providing findings, perspectives, and support in all of our pursuits.
28. Victual: Big data is the gasoline that powers data science.
29. Viral: How does data propagate broadly across other users and applications?
30. Virtuosity: If data scientists must know a little bit about many things, we must also learn a lot about one thing.
31. Viscosity: Connected to Velocity; how challenging is it to work with the data?
32. Visibility: Data science gives insight into complicated large data problems.
33. Visualization: how customers engage with models.
34. Vivify: Data science has the ability to visually represent a wide range of decisions and processes, from marketing to fraud detection.
35. Vocabulary: Distinct modeling approaches address distinct problem areas, and different validation procedures harden these approaches in various applications.
36. Vogue: The evolution of big data is happening every second.
37. Voice: Data science gives the ability to communicate with knowledge on a wide variety of topics.
38. Volatility: Data volatility must be anticipated, particularly in production systems. Data that should "never" go missing or change its data type.
39. Volume: The size of the big data, which is growing at a rapid pace.
40. Voodoo: Data science and big data aren't voodoo, but how can we encourage potential clients of the data science in that big data is capable to delivering desired results.
41. Voyage: There is no end to learning with data science as the big data grows.
42. Vulpine: It describes how the big data analytics is used in elections. [51]

According to IDC, the big data analytics business will be worth $203 billion by 2020. [16]

2020

According to Allied Market Research, the big data and business analytics market was worth $193.14 billion in 2019 and is expected to expand to $420.98 billion by 2027 at a compound yearly growth rate of 10.9%. [16]

It is predicted that within a few years while discussing sensor data, we will have to speak in Bronto bytes (1027). The Internet of Things, also known as machine-to-machine (M2M) connectivity, links billions of devices and creates an incomprehensible quantity of data. M2M data will account for 40% of all data in the globe by 2020. To have any significance and to drive your business propositions, this data must be processed, stored, analyzed, and visualized. Sensor data, also known as M2M data, is information derived from readings taken by machine sensors that detect pre-set conditions at regular intervals. Data gathered includes log data, geolocation data, CPU utilization, temperature, regulations, and so on. These data may be connected to Key Performance Indicators, which can send an alert if a specified threshold is exceeded and an action is required. Nearly half of respondents rated managing data growth as extremely difficult when managing their company's business intelligence (BI), 41% rated integrating Business Intelligence tools as extremely difficult, and 40% rated the need for tools to analyze data and glean insights as extremely difficult. In practice/Practical, we investigated the nature of big data and assessed the big data world from a high level. As is customary, there are factors to consider beyond tool selection when it comes to deployment. [17]

1.5 Future Trends / Future Research Scope / Limitations

Further research can include more parameters that affect the performance of MapReduce jobs with varying input data sizes in terms of GB, TB, and PB, and empirical evaluation can be done to improve this model to take into account more parameter combinations.
In terms of new technologies, machine learning may be used to improve the accuracy of predicting the execution time of MapReduce tasks and various combinations of characteristics that impact performance and the execution time.
Live data from social media, IoT, financial market, healthcare, weather, EGovernance, and E-Commerce may also be benchmarked.
By taking into account performance-affecting aspects, a jobs time and category perdition model for Apache Spark may be constructed. [18]

Data Analytics Trends for the Future
Banking, retail, manufacturing, finance, healthcare, and government are among the businesses making considerable investments in big data analytics. Data and business analysts in these sectors use cutting-edge data analytics tools to create better-educated business suggestions, forecast trends, and increase their revenues. Careers in big data appear to have a bright future. According to the U.S. Bureau of Labour Statistics (BLS), computer and information technology professions will rise by around 13% between 2016 and 2026. Database administrators, information security analysts, network and computer system administrators, and other occupations that deal directly with big data will have a great boost. [49]

Future of data analytics in Financial Fraud Detection
Over the last decade, technological innovations in the banking industry have affected how customers bank. From mobile banking to fast peer-to-peer money transfers via smartphone applications, these developments have had a significant influence on how people manage their finances. As a result, banks have increased their investments in big data and analytics solutions. According to IDC, banks have invested around 13% of all money invested in big data solutions worldwide.
Because of technological improvements, the financial industry can now include big data analytics into its business strategy. While the future of big data includes enhancing the banking customer experience, Big data helps in the financial sector to prevent fraud. According to the United States Department of Justice, financial institutions not only lose billions of dollars each year due to fraud, but they may also lose their brand name as a result of a serious security breach. According to the BLS, there will be a 28% increase in information security analysis positions across industries in the United States. Typically, professionals in this field are engaged to monitor networks for security breaches such as fraud. Credit card numbers, loan information, and other sensitive information may be hacked if these systems are not maintained securely.
Big data when combined with AI is capable of machine learning which can detect patterns of fraud in advance.

Furthermore, big data enables banks to monitor and analyze customer behavior, discovering suspicious or foreign activity more quickly. [49]

The majority of big data specialists think that the quantity of data collected in the future will expand significantly. IDC predicts that the global datasphere will reach 175 zettabytes by 2025.
As organizations acquire access to real-time big data analytics, they will be able to produce and manage 60% of big data in the near future.
Individual consumers, on the other hand, play a key role in data growth. IDC projects that by 2025, 6 billion people, or 75% of the world's population, would engage with internet data on a daily basis. In other words, every 18 seconds, each connected user will have at least one data transaction.
Large datasets are difficult to manage in terms of storage and processing. but the use of open-source ecosystems such as Hadoop and NoSQL was used to overcome huge data processing difficulties. However, open-source solutions need manual configuration and debugging, which may be difficult for most businesses. Businesses began to shift huge data to the cloud in pursuit of more flexibility. [50]

Big Data will at least dominate the market through 2030 and has shown to be a game changer in technology. Since new technologies are being launched on a daily basis, data will be generated in any way. As a result, data will continue to expand. It will never come to an end. As a result, the volume of data grows, causing a problem for organizations. [52]

In the future, major publications and conferences will cover topics such as big data comparison in diverse learning contexts, ethical and cultural values, support from the government, and training in big data adoption in higher education. [20]

Conclusion

Current Big Data platforms include a variety of processing, analytical, and dynamic visualization capabilities. Such platforms allow for the extraction of knowledge and value from a complex and dynamic environment. They also aid decision-making by providing suggestions and automatically detecting abnormalities, odd behavior, or new trends.
In this work, we examined Big Data features and analyzed in depth the issues posed by Big Data computing systems. Furthermore, we have discussed the importance of Big Data in a variety of disciplines. We have also highlighted the big data challenges, what are challenges that we face while handling big data.
Our review work has also given an insight on the advancement of the big data with its timeline. As a result, this article gives a complete understanding of the architecture, techniques, and practices now used in Big Data computing.
The future trends and limitations have also been highlighted in the terms of how we can improve the big data and the use of big data.
As a result, further effort is required in various areas, including data organization, domain-specific tools, and platform tools, in order to construct next-generation Big Data infrastructures. As a result, technical challenges in various Big Data fields may be researched further and serve as a significant study topic.

References

1. Data, B. (2013). A Review. Seref SAGIROGLU and Duygu SINANC, Gazi University. Department of Computer Engineering, Faculty of Engineering, Ankara, Turkey.

2. Elgendy, N., & Elragal, A. (2014). Big data analytics: a literature review paper. In *Advances in Data Mining. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16-20, 2014. Proceedings 14* (pp. 214-227). Springer International Publishing.

3. Chaudhari, M. M. M., & Chaudhari, M. D. LITERATURE REVIEW ON BIG DATA, BIG DATA MODELS AND BIG DATA ANALYTICS.

4. Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, *30*(4), 431-448.

5. IT Chronicles. (2023, February 7). *What is Big Data - Comprehensive Guide Covering All Aspects*. ITChronicles.https://itchronicles.com/what-is-big-data/#:~:text=%E2%80%9CBig%20data%E2%80%9D%20is%20a%20term,records%20from%20millions%20of%20people.

6.GeeksforGeeks. (2023). Difference between Structured Semi structured and Unstructured data. *GeeksforGeeks*.https://www.geeksforgeeks.org/difference-between-structured-semi-structured-and-unstructured-data/

7. MongoDB. (n.d.-b). *What is unstructured data?* https://www.mongodb.com/unstructured-data

8. *What is Semi-Structured Data? | Teradata*. (n.d.-c). https://www.teradata.com/Glossary/What-is-Semi-Structured-Data

9. Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., ... & Widom, J. (2011). Challenges and opportunities with Big Data 2011-1.

10. Brian Runciman(2013), "IT NOW Big Data Focus, AUTUMN 2013", www.bcs.org, retrieved 03/02/14.

11.Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13-53. DOI:https://doi.org/10.1080/17538947.2016.1239771.

12. Chris Deptula(2013), "With all of the Big Data Tools, what is the right one for me", www.openbi.com/blogs/chris%20Deptula, retrieved 08/02/14.

13. James M, Michael C, Brad B, Jacques B, Richard D, Charles R, Angela H.B(2011), "Big Data: The next frontier for Innovation, Competition, and Productivity", www.McKinsey.com, retrieved 23/02/14.

14. Oguntimilehin, A., & Ademola, E. O. (2014). A review of big data management, benefits and challenges. *A Review of Big Data Management, Benefits and Challenges*, *5*(6), 1-7.

15. Team, T., Team, T., Team, T., & Team, T. (2021b). Why Big Data – Benefits and Importance of Big data. *TechVidvan*. https://techvidvan.com/tutorials/why-big-data/

16. Phillips, A. (2021). A history and timeline of big data. *WhatIs.com*. https://www.techtarget.com/whatis/feature/A-history-and-timeline-of-big-data

17. Syed, A., Gillela, K., & Venugopal, C. (2013). The future revolution on big data. *Future*, *2*(6), 2446-2451.

18. IDENTIFYING THE PERFORMANCE CHALLENGES WITH BIG DATA PROCESSING AND PARTIAL SOLUTIONS MR. BRIJESH YATINDRA JOSHI http://hdl.handle.net/10603/408862

19. Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191-209.

20. Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: a state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education*, *17*(1), 1-23.

21. Rajaraman, V. (2016). Big data analytics. *Resonance*, *21*, 695-716.

22. Andrew McAfee, A., Brynjolfsson, E., et al., 2012. Big Data: the management revolution. Harvard Bus. Rev. 90, 60–68.

23. Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R., Buyya, R., 2016. The anatomy of big data computing. Software: Pract. Experience 46, 79–105.

24. EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)

25. Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)

26. TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: TechAmerica Reports, pp. 1–40 (2012)

27. Chen, J., Y. Chen, X. Du, C. Li, J. Lu, S. Zhao, and X. Zhou. 2013. "Big Data Challenge: A Data Management Perspective. " Frontiers of Computer Science 7 (2): 157–164.

28. Christen, P. 2014. Privacy Aspects in Big Data Integration: Challenges and Opportunities. In Proceedings of the First International Workshop on Privacy and Security of Big Data, 1–1. ACM.

29. Dong, X. H., and S. Divesh. 2015. "Big Data Integration." Synthesis Lectures on Data Management 7 (1): 1–198.

30. Agrawal, D., P. Bernstein, E. Bertino, S. Davidson, U. Dayal, M. Franklin, J. Gehrke, et al. 2011. Challenges and Opportunities with Big Data 2011–1. Cyber Center Technical Reports. http://docs.lib.purdue.edu/cctech/1.

31. Gantz, J., and D. Reinsel. 2011. "Extracting Value from Chaos." IDC iView 1142 (2011): 1–12.

32. EarthCube. 2014. EarthCube Enterprise Governance Draft Charter. Accessed November10 2015. http://workspace.earthcube.org/.

33. Chen, Z. K., S. Q. Yang, S. Tan, H. Zhao, L. He, G. Zhang, and H. Y. Yang. 2014b. "The Data Allocation Strategy Based on Load in NoSQL Database." Applied Mechanics and Materials 513–517: 1464–1469.

34. Lohr, S. 2012. The age of big data. New York Times, 11

35. Cuzzocrea, A., G. Fortino, and O. Rana. 2013. Managing Data and Processes in Cloud-Enabled Large-Scale Sensor Networks: State-Of-The-Art and Future Research Directions. In 2013 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 583–588. IEEE.

36. Nasser, T., and R. S. Tariq. 2015. "Big Data Challenges." Journal of Computer Engineering & Information Technology 4 (3): 1–10.

37. Padgavankar, M. H., and S. R. Gupta. 2014. "Big Data Storage and Challenges." (IJCSIT) International Journal of Computer Science and Information Technologies 5 (2): 2218–2223.

38. Sun, M., J. Li, C. Yang, G. A. Schmidt, M. Bambacus, R. Cahalan, Q. Huang, C. Xu, E. U. Noble, and Z. Li. 2012. "A Web-Based Geovisual Analytical System for Climate Studies." Future Internet 4: 1069–1085.

39. Jagadish, H. V., J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi. 2014. "Big Data and its Technical Challenges." Communications of the ACM 57 (7): 86–94.

40. Fox, P., and J. Hendler. 2011. "Changing the Equation on Scientific Data Visualization." Science, Washington 331 (6018): 705–708.

41. Reda, K., A. Febretti, A. Knoll, J. Aurisano, J. Leigh, A. Johnson, M. E. Papka, and M. Hereld. 2013. "Visualizing Large, Heterogeneous Data in Hybrid-Reality Environments." IEEE Computer Graphics and Applications 33(4): 38–48.

42. Cukier, K. 2010. Data, Data Everywhere, Economist. Accessed November 25, 2015. http://www.economist.com/node/15557443.

43. Michael, K., and K. Miller. 2013. "Big Data: New Opportunities and New Challenges [Guest Editors" Introduction]." Computer 46 (6): 22–24.

44. Cheatham, M. 2015. Privacy in the Age of Big Data. In The 2015 International Conference on Collaboration Technologies and Systems (CTS), 334–335.

45. Kaisler, S., F. Armour, J. A. Espinosa, and W. Money. 2013. Big Data: Issues and Challenges Moving Forward. In 2013 46th Hawaii International Conference on System Sciences (HICSS), 995–1004.

46. Terry, N. 2012. Protecting Patient Privacy in the Age of Big Data. Accessed November 25, 2015. http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 2153269.

47. Khan, N., I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. Mahmoud Ali, M. Alam, M. Shiraz, and A. Gani. 2014. "Big Data: Survey, Technologies, Opportunities, and Challenges." The Scientific World Journal 2014: 1–18.

48. Eisenstein, M. 2015. "Big Data: The Power of Petabytes." Nature 527 (7576): S2–S4.

49. *The Future of Big Data in Business: How Data Analytics Provide insight*. (2023c, March 1). Maryville Online.https://online.maryville.edu/blog/future-big-data/#:~:text=The%20future%20of%20big%20data%20points%20to%20more%20digital%20growth,turning%20it%20into%20actionable%20insights.

50. Khvoynitskaya, S. (2020, January 30). The future of big data: 5 predictions from experts for 2020-2025. *Big data future*. https://www.itransition.com/blog/the-future-of-big-data

51. Shafer, T. (2020, November 11). *The 42 V's of Big Data and Data Science | Elder Research*. Elder Research. https://www.elderresearch.com/blog/the-42-vs-of-big-data-and-data-science/

52. Patgiri, R., & Ahmed, A. (2016, December). Big data: The v's of the game changer paradigm. In *2016 IEEE 18th international conference on high performance computing and communications; IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS)* (pp. 17-24). IEEE.