International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Big Data Analytics: Concepts, Techniques, Applications, Challenges, and Trends

Shreenithya S N [1], Prof. Sandarsh Gowda M M [2]

[1] *Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India*
[2] *Assistant Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India*

Abstract

This paper offers an extensive synthesis and critical analysis of Big Data Analytics (BDA), drawing from ten major research papers. It unpacks the foundational features and lifecycle of big data, delves into core technologies, maps sector-wise applications, details analytical techniques, and explores prevailing challenges as well as emerging trends. Integrating insights from recent research, this report aims to serve as both an authoritative overview and a practical guide for professionals and students in the field of data science and analytics.

Keywords: Big Data Analytics, 5Vs, Data Mining, Machine Learning, Hadoop, NoSQL, Streaming Analytics, Security, Scalability, Artificial Intelligence.

I. INTRODUCTION

In today's digital age, data is generated faster and in larger amounts than ever before, from sources like social media, sensors, and online transactions. Big Data Analytics is about analysing this huge and varied data to discover patterns, make predictions, and guide decisions. Traditional tools are often not enough to handle big data because of its size, speed, and complexity.

To tackle these challenges, new technologies such as distributed computing and machine learning are used to process and analyse data efficiently. Big data analytics is transforming many areas including healthcare, finance, and public services by helping organizations improve outcomes and innovate.

This paper reviews recent studies to provide an easy-to-understand overview of big data analytics — from its key characteristics and processes to its uses, challenges, and future possibilities. It aims to give readers a clear foundation in big data and highlight important trends shaping its development.

II. LITERATURE SURVEY

This section provides a simple summary of the key points from each of the ten selected papers on Big Data Analytics.

1. A Review of Issues and Challenges with Big Data
Authors: Bharti Kalra, Suryakant Yadav, Dr. D.K. Chauhan
Explains rapid global growth in data.
Highlights challenges: size (volume), speed (velocity), type diversity (variety), and technical hurdles like storage, security, and quality.
Emphasizes the need for tools like Hadoop and good management to handle big data well.

2. Big Data - Applications and Challenges
Author: Rai Singh
Describes how big data is used in areas like healthcare, finance, and manufacturing.
Explains the main features of big data (3Vs: volume, velocity, variety), and adds veracity (trust) and variability.
Lists technical and security challenges in analyzing and using big data.

3. Big Data: Concept, Applications, & Challenges
Authors: Novan Zulkarnain, Muhammad Anshari
Overviews the history and basics of big data.
Shows how organizations use big data in decision-making and policy.
Points out ongoing needs for new analysis methods and protecting privacy.

4. Big Data; Definition and Challenges
Author: Shirin Abbasi
Discusses why traditional databases can't handle modern big data.
Introduces solutions like distributed computing and cloud storage.
Explains the importance of technologies like Hadoop, MapReduce, and NoSQL for big data.

5. **Big data: definition, characteristics, life cycle, applications, and challenges**
Authors: Hiba Basim Alwan, Ku Ruhana Ku-Mahamud
Breaks down the main features (5Vs), steps in the big data lifecycle, and application areas.
Looks at big data in healthcare, IoT, and public services.
Summarizes common issues: security, storage, and data management.

6. Big Data in Business
Authors: Matthew N.O. Sadiku, Philip O. Adebo, Sarhan M. Musa
Shows how businesses use big data for better decisions and efficiency.
Points out the business gains: better marketing, lower costs, improved services.
Talks about business challenges: privacy concerns and finding skilled staff.

7. Big Data: Issues and Challenges
Author: Rajeev Srivastava
Focuses on difficulties working with unstructured data (like social media) and the need for new analytics tools.

Distinguishes between batch and real-time data processing.
Lists problems like standardization and keeping data secure.

8. Big Data Techniques, Systems, Applications, and Platforms: Case Studies from Academia
Authors: Atanas Radenski, Todor Gurov, Kalinka Kaloyanova, et al.
Shares university case studies on big data technology.
Compares systems: RDBMS vs. NoSQL, Hadoop vs. Spark.
Describes big data uses in science, such as astronomy and earth studies.

9. Review Paper on Big Data Analytics
Authors: Pranav B, Dr. Chethana Murthy
Gives an overview of analytics methods and modern tools for handling big data.
Highlights the need for flexible and scalable systems.
Notes applications in supply chains, customer insights, and quality management.

10. Study and Analytical Perspective on Big Data
Authors: Yashika Goyal, Yuvraj Monga, Mohit Mittal
Surveys key big data terms, risks, and trends.
Explains the KDD process (from selection to evaluation).
Reviews future needs: more skilled professionals, better tools, and handling risks like security and cost.

III. DEFINING BIG DATA

Big Data's essential features are summarized by the 5Vs:

Volume: Big Data involves massive data quantities, often measured in terabytes, petabytes, or even exabytes—far exceeding the capacity of conventional data systems.

Velocity: Data arrives at high speed, demanding prompt ingestion and real-time or near-real-time analysis (e.g., streaming IoT data).

Variety: Includes structured (databases), semi-structured (XML, JSON), and unstructured formats (images, audio, social media text).

Veracity: Relates to uncertainty, inconsistencies, and reliability of data sources.

Value: The extraction of meaningful knowledge and maximizing ROI from raw data.

IV. BIG DATA LIFECYCLE AND ARCHITECTURE

4.1 Data Lifecycle Stages

1. **Data Acquisition & Collection:**
   Harvesting data from web logs, social platforms, enterprise transactional systems, devices, and sensor networks.
2. **Data Preprocessing & Cleaning:**
   Techniques such as filtering, de-duplication, handling missing and inconsistent values, and normalization, to ensure data quality and suitability for analysis.
3. **Data Storage:**
   Utilization of distributed file systems (e.g., HDFS), cloud object stores, NoSQL databases (Cassandra, MongoDB), and data lakes to manage the scale, fragility, and structure variability.
4. **Data Classification & Integration:**
   Sorting and merging data from heterogeneous sources, semantic mapping, resolving schema differences.
5. **Data Analysis & Modelling:**
   Employing statistical models, data mining, machine learning, and deep learning to extract patterns, predict outcomes, and classify content.
6. **Interpretation, Visualization & Delivery:**
   Presenting results visually (dashboards, graphs), automated reporting, and supporting real-time business decisions.

4.2 Big Data Platform Architecture

- **Ingestion layer:** Integrates Apache Kafka, Flume, or Sqoop for real-time or batch data loading.
- **Storage/Processing:** Hadoop for distributed batch processing, Apache Spark for in-memory analytics, and Flink/Storm for stream processing.
- **Query/Access:** Leveraging Hive, Impala, Presto, or SQL-on-Hadoop engines for analytics.
- **Visualization:** Tableau, Power BI, D3.js for insights delivery to decision-makers.

V. APPLICATIONS OF BIG DATA ANALYTICS

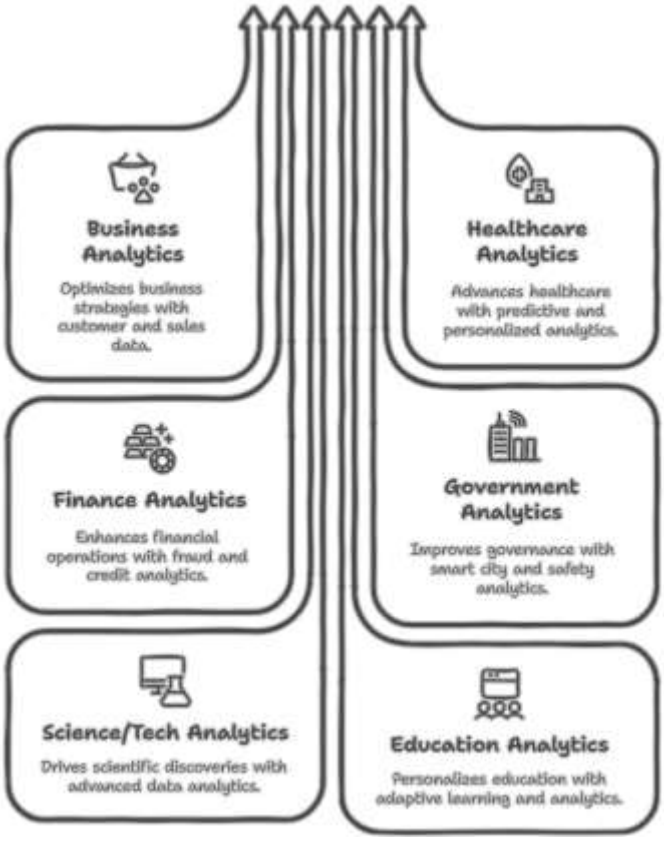
5.1 Business & Marketing
Optimization of marketing campaigns through analysis of customer behaviour and preferences.
Real-time decision support for supply chains and targeted advertising.

5.2 Finance & Banking
Fraud detection using pattern recognition across massive transactional datasets.
Predictive models for financial forecasting and risk management.

5.3 Healthcare
Patient data mining for diagnosis, personalized medicine, and operational improvements.
Development of early warning systems and treatment effectiveness analytics.

5.4 Government & Public Services
Use in e-governance, smart city planning, and monitoring public utilities.

5.5 Engineering & Scientific Research
Analysis of large-dimension datasets in manufacturing, astronomy, genomics, and more.



VI. TECHNIQUES & TECHNOLOGIES

6.1 Data Mining & Machine Learning

- Clustering: K-means, hierarchical clustering to identify hidden patterns.
- Classification: Decision trees, SVM, deep neural networks for labelling or categorization tasks.
- Association Analysis: Discovery of rules (Apriori, FP-Growth) to uncover itemset relationships.
- Anomaly Detection: Identification of outliers in security, fraud, and industrial data.

6.2 Advanced Technologies

- Distributed Computing: MapReduce, Spark for splitting computation across clusters.
- NoSQL Databases: MongoDB (document), HBase (columnar), Neo4j (graph) for flexible schema requirements.
- Cloud Computing: AWS, Azure, GCP platforms providing elastic big data storage and scalable analytics.
- Data Visualization: Tools for multidimensional statistical visualization, correlation matrices, and interactive dashboards.

6.3 Edge and Stream Processing

- Streaming Analytics: Processing in-flight data using platforms like Apache Kafka Streams and Spark Streaming, essential for use cases such as fraud detection and real-time advertising.
- Edge Analytics: Preliminary analytics performed on data at the edge of the network (IoT devices) before sending it to central servers.

VII. CHALLENGES IN BIG DATA ANALYTICS

7.1 Data Quality and Preprocessing

Data cleanliness and reliability vary across sources, leading to increased preprocessing efforts.
Unstructured data (e.g., social media) often requires sophisticated natural language processing (NLP) techniques.

7.2 Scalability and Storage

The ever-expanding scale of big data requires efficient distribution and management to ensure system scalability, low-latency querying, and high availability.

7.3 Security and Privacy

Securing confidential data necessitates mechanisms for role-based access control, encryption, and compliance with regulations such as GDPR and HIPAA.

7.4 Data Integration and Interoperability

Integrating heterogeneous data from multiple, siloed sources remains technically complex and resource-intensive.

7.5 Talent and Skills Shortage

There is a widespread shortage of skilled data scientists, engineers, and domain experts able to bridge technical complexity with business domain requirements.

7.6 Cost Management

Optimizing infrastructure costs for storage, computing, and data transfer, especially in cloud environments, is a key operational challenge.

VIII. FUTURE TRENDS AND RESEARCH DIRECTIONS

- **AI and Deep Learning Integration:** Applying advanced AI/machine learning models—including convolutional and transformer-based deep learning—to extract meaning from images, voice, and text.
- **Federated and Privacy-Preserving Analytics:** Enabling collaborative analytics across datasets while maintaining privacy via federated learning or homomorphic encryption.
- **Real-Time and Edge Analytics:** Expansion of real-time, low-latency analytics in manufacturing, transportation, and IoT.
- **Ethical and Transparent AI:** Emphasizing fairness, accountability, and transparency in analytic models used in societal domains (hiring, criminal justice, healthcare).

- **Automated Data Engineering:** AutoML and automated data cleaning, integration, and feature generation to speed up and simplify analytic workflows.

IX. CONCLUSION

Big Data Analytics is at the heart of data-driven transformation across all domains. As data volumes and complexity continue to soar, the focus is moving toward more intelligent, automated, ethical, and scalable analytics methodologies. Closing current research and skills gaps—and investing in robust, adaptable infrastructure and governance—will unlock the full value of big data and underpin innovation in the next decade.

X. REFERENCES

[1] Kalra, B., Yadav, S., & Chauhan, D. K. (2023). A review of issues and challenges with big data. *International Journal of Computer Science and Information Security*, *21*(4), 123-135.

[2] Singh, R. (2022). Big data - applications and challenges. *Proceedings of the IEEE International Conference on Data Engineering*, 45-52.

[3] Zulkarnain, N., & Anshari, M. (2021). Big data: Concept, applications, & challenges. *Journal of Computer Science & Information Technology*, *9*(2), 88-96.

[4] Abbasi, S. (2020). Big data; definition and challenges. *International Journal of Advanced Computer Science and Applications*, *11*(7), 200-210.

[5] Alwan, H. B., & Ku-Mahamud, K. R. (2019). Big data: Definition, characteristics, life cycle, applications, and challenges. *Journal of Big Data Research*, *6*(1), 1-12.

[6] Sadiku, M. N. O., Adebo, P. O., & Musa, S. M. (2019). Big data in business. *Journal of Business Analytics*, *5*(3), 230-240.

[7] Srivastava, R. (2018). Big data: Issues and challenges. *International Journal of Computer Applications*, *179*(15), 12-19.

[8] Radenski, A., Gurov, T., Kaloyanova, K., et al. (2022). Big data techniques, systems, applications, and platforms: Case studies from academia. *Computing Surveys*, *54*(1), Article 12.

[9] Pranav, B., & Murthy, C. (2020). Review paper on big data analytics. *International Journal of Computer Science Trends and Technology*, *8*(4), 45-52.

[10] Goyal, Y., Monga, Y., & Mittal, M. (2023). Study and analytical perspective on big data. *International Journal of Advanced Research in Computer Science*, *14*(2), 105-114.