

*Article**2025 2nd International Conference on Global Economics, Education and the Arts (GEEA 2025)*

Exploring the Application and Practice of Computer Science Technology in the Big Data Era

Zhangbo Shi ^{1,*}

¹ International Business Economic, University of Nottingham Ningbo China, Ningbo, Zhejiang, China

* Correspondence: Zhangbo Shi, International Business Economic, University of Nottingham Ningbo China, Ningbo, Zhejiang, China

Abstract: With the advent of the big data era, computer science technology faces unprecedented challenges and opportunities. This paper aims to explore the practical applications of computer science technology in the big data era, analyzing the fundamental concepts and characteristics of big data while delving into technological advancements and innovations in data storage, processing, analysis, and their applications, as demonstrated through real-world case studies. The paper begins by reviewing the foundational theories of big data and computer science, examining data storage and management technologies such as Hadoop and NoSQL databases, as well as data processing and analysis frameworks like MapReduce and Spark. It then provides an in-depth analysis of big data applications in industries such as healthcare, finance, smart manufacturing, the Internet of Things (IoT), and public safety, highlighting their critical role in improving efficiency, optimizing decision-making, and driving technological innovation. Finally, the paper addresses challenges related to privacy and security, data quality, and future trends in big data technology. This study offers valuable insights and references for understanding the practical applications and development of big data technology across various industries.

Keywords: big data; computer science technology; data storage; data analysis; practical applications; privacy and security

Received: 13 April 2025

Revised: 22 April 2025

Accepted: 06 May 2025

Published: 07 May 2025



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid advancement of information technology, we have entered a new era centered around big data. The arrival of the big data era has fundamentally transformed the applications and development of computer science technology. Traditional data processing methods struggle to handle the massive, diverse, and rapidly changing datasets of today. In response, the rapid development of big data technologies has introduced innovative approaches and solutions to address these challenges. Through advanced computer science techniques, the collection, storage, processing, and analysis of data have reached unprecedented levels, bringing transformative changes to industries worldwide. In the big data era, the applications of computer science technology extend beyond data storage and computation to encompass fields such as machine learning, artificial intelligence, and real-time data processing. Effectively managing, analyzing, and applying massive datasets has become a central focus for the technology sector and various industries. Sectors like healthcare, finance, smart manufacturing, and public safety have increasingly adopted big data technologies, driven innovation and industry growth while also introducing new challenges, such as privacy protection and data security. This paper explores

the practical applications of computer science technology in the big data era. It begins by introducing the core concepts and characteristics of big data, followed by an analysis of key technologies in data storage, processing, and analysis, along with their applications. The paper then examines the real-world applications of big data technology across industries, addressing the challenges faced and future development trends. Through a comprehensive analysis of big data technology, this study aims to provide readers with a deeper understanding of its applications and prospects, offering valuable references for research and practice in related fields [1].

2. Big Data and Computer Science: An Overview

2.1. Definition and Characteristics of Big Data

Big data refers to datasets that are too large, complex, or diverse to be efficiently stored, managed, or analyzed using traditional database management systems or data processing tools. The scale, complexity, and variety of these datasets exceed the capabilities of conventional computing methods, earning them the label "big data". Big data is typically defined by four key characteristics, often referred to as the "4V" features: massive volume, high variety, rapid velocity, and low value density. Massive Volume: Big data involves datasets that often reach petabyte (PB) or even exabyte (EB) scales. With the proliferation of the internet and sensor-equipped devices, the amount of data generated globally every minute surpasses the processing capacity of traditional systems. Every user action, device operation, or transaction generates vast amounts of data [2]. High Variety: Big data is characterized by its diversity, encompassing not only traditional structured data (e.g., tabular data in relational databases) but also unstructured data (e.g., text, audio, video, images, and sensor data) and semi-structured data (e.g., log files and social media posts). This diversity requires the ability to process, store, and analyze data in various formats using different technologies.

Rapid Velocity: The speed at which data is generated is another defining feature of big data. With the rise of IoT devices and social media, vast amounts of data are produced at an extraordinary pace. For example, on e-commerce platforms, every transaction, click, or search generates data. The widespread use of smart devices and sensors means data is created not only in traditional business processes but also in real-time by automated systems. These real-time datasets require efficient processing systems for rapid analysis and response. Low Veracity: Big data often has low value density, meaning much of it is not directly useful and may contain significant noise. Extracting valuable insights from massive, complex datasets is a critical challenge for big data technology [3]. Unlike traditional methods that focus on structured data, big data demands the ability to derive meaningful information from complex unstructured data using techniques such as sentiment analysis and image recognition. The rapid growth of big data stems from the swift development of various technologies. The widespread adoption of the internet, IoT, social media, sensors, and smart devices has led to an unprecedented surge in data generation. At the same time, the increasing proportion of unstructured data has rendered traditional database technologies and analytical methods inadequate. To address these challenges, modern computer science has evolved, giving rise to innovations like distributed storage, data mining, and artificial intelligence. These advancements provide robust support for extracting value from massive, complex datasets. Thus, big data extends beyond processing traditional structured data to include the analysis of unstructured data such as text, images, and audio. Efficiently handling and analyzing these diverse, large-scale datasets, and extracting valuable insights from them, are key priorities in data science and computer technology research today [4].

2.2. Foundations of Computer Science Technology

Computer science, as the backbone of modern information technology, encompasses a wide range of theoretical and technical domains, including algorithm design and analysis, data structures, computer networks, artificial intelligence, machine learning, database management, and information security. With the rise of big data, many traditional computer science technologies have undergone significant expansion and enhancement. The rapid growth of data science, in particular, has made big data storage, processing, and analysis a focal point of research. In terms of data storage, distributed storage technologies such as Hadoop HDFS and NoSQL databases provide a strong foundation for managing massive datasets. For data processing, technologies like MapReduce and Apache Spark have significantly improved the parallelism and efficiency of data handling. In data analysis, the application of machine learning and deep learning techniques enables the extraction of valuable patterns and trends from complex, unstructured data. Therefore, big data applications rely not only on traditional computer science technologies but also on the integration and innovation of new methods and approaches [5].

3. Core Applications of Big Data Technology

3.1. Data Storage and Management

Data storage and management technologies are critical components of big data architectures, designed to efficiently and reliably store vast amounts of diverse data. Traditional relational database management systems (RDBMS) face performance bottlenecks when handling large-scale data, prompting the development of distributed storage architectures. For instance, the Hadoop Distributed File System (HDFS) distributes data across multiple nodes in a cluster, ensuring high availability and fault tolerance. Additionally, with the rise of unstructured data, NoSQL databases such as MongoDB, Cassandra, and HBase have gained widespread adoption [6]. These databases can handle various data types (e.g., key-value pairs, documents, graphs) and offer flexible scalability, making them ideal for storing massive amounts of unstructured data. Key challenges in data storage and management include efficient data distribution, achieving load balancing, and ensuring high availability and fault tolerance. For big data systems, effective storage and management not only enhance system performance but also provide a solid foundation for subsequent data processing and analysis.

3.2. Data Processing and Analysis

Data processing and analysis are essential steps in transforming stored data into valuable insights. Big data processing technologies are typically categorized into batch processing and stream processing. MapReduce, a classic batch processing framework, enhances efficiency by dividing data into smaller chunks and processing them in parallel across multiple nodes. Apache Spark, an advanced alternative to MapReduce, supports both batch and real-time data processing, and leverages in-memory computing for greater efficiency. This feature makes it particularly advantageous for large-scale data processing. Unlike batch processing, real-time stream processing enables immediate data analysis, ideal for time-sensitive applications such as real-time monitoring or online recommendations. Stream processing frameworks like Apache Kafka, Apache Flink, and Apache Storm can handle rapidly generated real-time data, supporting event-driven analysis. These technologies allow businesses to analyze and respond to data as it is generated, improving decision-making efficiency and enabling swift responses to emerging situations [7].

3.3. Machine Learning and Artificial Intelligence

Machine learning and artificial intelligence (AI) play pivotal roles in big data analysis. As data volumes grow, traditional analytical methods struggle to address data complexity and diversity, but machine learning offers robust capabilities to automate the analysis

process. Machine learning is typically categorized into supervised learning (e.g., classification, regression) and unsupervised learning (e.g., clustering, dimensionality reduction) [8]. Supervised learning trains on labeled datasets to predict future trends or recognize patterns, while unsupervised learning uncovers hidden structures in unlabeled data. In big data contexts, deep learning, a subset of machine learning, uses neural networks to process complex data such as images, speech, and text, achieving remarkable results. Deep learning models have driven breakthroughs in natural language processing and computer vision, extracting deep insights from large datasets. Furthermore, AI integration extending the application of big data beyond traditional analysis and enabling more precise decision-making through intelligent decision-support systems.

3.4. Real-Time Stream Processing

With the rapid proliferation of the Internet of Things (IoT), social media, and connected devices, real-time stream processing has emerged as a cornerstone of big data applications. This technology is designed to process and analyze dynamically changing data instantaneously, addressing the demand for low-latency, high-throughput solutions in modern business environments. By enabling immediate insights from continuously generated data, real-time stream processing supports critical applications across industries, enhancing responsiveness and operational agility. Apache Kafka, a leading platform for large-scale message passing, is widely adopted for its ability to handle massive data streams with high throughput and distributed architecture. It efficiently processes and transmits real-time data, ensuring seamless scalability. Complementary frameworks like Apache Flink and Apache Storm further enhance real-time stream processing by leveraging event-driven models and real-time computation. These tools excel in large-scale, low-latency environments, making them ideal for time-sensitive tasks [9]. Real-time stream processing supports a variety of applications, including continuous monitoring, personalized recommendations, and anomaly detection. In the financial sector, for instance, it enables institutions to monitor market fluctuations, detect potential risks, and execute trades within milliseconds, safeguarding investments. Similarly, e-commerce platforms harness real-time analysis of user behavior — such as clicks, searches, and browsing patterns — to deliver personalized product recommendations, boosting customer engagement and sales. In smart cities, real-time processing of sensor data optimizes traffic flow and enhances public safety by identifying incidents promptly. By providing instantaneous insights, real-time stream processing significantly improves operational efficiency and flexibility, meeting the growing need for timely data-driven decisions. As IoT devices and digital interactions continue to multiply, the importance of real-time stream processing will only increase, driving innovation and enabling businesses to stay competitive in fast-paced, data-intensive environments [10].

4. Applications of Big Data Technology Across Industries

4.1. Healthcare

Big data technology is revolutionizing healthcare, driving transformative advancements in disease prediction, precision medicine, and resource optimization. The widespread adoption of electronic health records (EHRs) has digitized vast amounts of patient data, enabling comprehensive long-term tracking and analysis. By integrating diverse data sources — such as medical histories, genomic data, lifestyle information, and real-time monitoring — big data provides deep insights that enhance the accuracy of diagnoses and the effectiveness of treatment plans. For instance, machine learning algorithms can analyze genomic data to identify risks of hereditary diseases, enabling personalized treatment strategies specifically tailored to individual patient needs, thereby improving outcomes. Wearable medical devices, including sensors and smartwatches, play a pivotal role in continuously monitoring vital signs, such as heart rate, blood pressure, or glucose levels, and transmitting this data to hospital systems in real time. This enables early detection of

health anomalies, allowing clinicians to intervene promptly and prevent complications. For example, real-time alerts from wearable devices can notify doctors of irregular heart rhythms, facilitating timely interventions that save lives. Big data also significantly contributes to public health surveillance, particularly in managing infectious disease outbreaks. From sources like hospital admissions, anonymized social media data, and mobile health apps, authorities can detect potential epidemic trends early and implement containment measures swiftly. During the COVID-19 pandemic, for instance, big data analytics helped track infection rates and predict hotspots, guiding resource allocation and policy decisions. Moreover, big data optimizes healthcare resource management by analyzing hospital workflows, patient admission patterns, and equipment usage, reducing inefficiencies and cutting costs. Collectively, these advancements improve service quality, enhance patient care, and increase the resilience of healthcare systems. Ultimately, big data's integration into healthcare fosters personalized and precision medicine, lowers operational costs, and drives systemic improvements, paving the way for a healthier future.

4.2. Financial Sector

Big data technology is extensively applied in the financial sector, impacting risk management, customer service, and fraud detection. Risk management is a key area where big data shines. Through the analysis of vast amounts of historical transaction data, financial institutions can more accurately predict market risks, conduct risk assessments, and inform investment decisions. For example, analyzing transaction behaviors, credit histories, and social network data enables precise credit risk evaluations, reducing default rates. Fraud detection is another critical application, with real-time monitoring of transactions and machine learning algorithms identifying anomalous activities to prevent fraud. Big data also supports precision marketing by analyzing customer behaviors and needs, allowing financial firms to offer tailored products and services. Thereby enhancing customer satisfaction and promoting revenue growth. Additionally, combining blockchain with big data enhances transaction transparency and traceability while ensuring data security. In summary, big data continuously enhances operational efficiency, mitigates risks, and drives innovation across the financial sector.

4.3. Smart Manufacturing and the Internet of Things

In smart manufacturing and the Internet of Things (IoT), big data technology supports decision-making by collecting, transmitting, and analyzing device data in real time, improving production efficiency and product quality. IoT devices and sensors interconnect production lines, equipment, and factory operations to form comprehensive intelligent systems, monitoring operational status and production data in real time. This data is transmitted to cloud platforms for analysis. For example, real-time analysis of production line data can detect equipment issues early, enabling predictive maintenance to prevent disruptions or damage, reducing costs, and enhancing efficiency. Big data also supports predictive maintenance by analyzing historical equipment data to forecast potential failures and schedule repairs proactively. Additionally, supply chain optimization benefits from big data, as real-time analysis enables companies to dynamically adjust production plans, streamline inventory management, minimize overstock, and ultimately reduce costs, improving supply chain responsiveness. The integration of IoT and big data enhances automation and intelligence in manufacturing, promoting more efficient and sustainable practices.

4.4. Public Safety and Smart Cities

Big data technology is playing an increasingly indispensable role in public safety and smart city development, especially in areas such as urban management, traffic optimization, and security monitoring. A wide array of sensors, surveillance systems, and smart devices deployed across urban environments continuously collect and transmit real-time

data for analysis. In traffic management, big data analyzes factors like traffic flow, vehicle speed, and weather to optimize traffic signal timing, reduce congestion, and improve road efficiency. In public safety, video surveillance systems paired with intelligent analytics monitor street activities in real time, identifying potential criminal behavior and triggering timely alerts to enhance security. In environmental management, big data monitors air quality, temperature, and humidity, using historical data to predict environmental trends and inform government policies. In times of public emergencies — such as natural disasters or health crises — big data delivers real-time insights that support rapid decision-making, efficient resource deployment, and effective crisis management. Through advanced data analytics, smart cities can optimize resource allocation, enhance residents' quality of life, and promote safety and sustainable urban development.

5. Challenges and Future Trends of Big Data Technology

5.1. Data Privacy and Security

While big data technology offers immense opportunities, it also presents significant challenges related to data privacy and security. The collection and processing of vast volumes of personal, corporate, and governmental data has made data privacy a central concern in the application of big data technologies. First, robust encryption measures must be implemented during data collection, storage, transmission, and processing to prevent unauthorized access. Second, with the introduction of privacy regulations such as the General Data Protection Regulation (GDPR), organizations must operate within legal and compliant frameworks, ensuring data collection and usage respect user privacy rights. Additionally, data anonymization techniques are widely used to protect user privacy, allowing sensitive data to be processed for analysis without compromising confidentiality. Big data platforms are also susceptible to data breaches and misuse, especially within cloud-based multi-tenant architectures, which may lead to unauthorized access or data leaks. To address these challenges, organizations must invest heavily in data security, enhance technical capabilities, and establish stringent data access controls and protection mechanisms to ensure the secure application of big data technology.

5.2. Data Quality Issues

Data quality is a pervasive challenge in big data applications. One defining characteristic of big data is its heterogeneity and complexity, often originating from diverse sources such as sensors, social media, or enterprise systems, resulting in inconsistent data quality. Common data quality issues include incomplete data, noise, duplication, and bias, all of which can undermine the accuracy of analytical outcomes. Ensuring high data quality remains a critical challenge in the processes of data analysis and decision-making. For incomplete or missing data, techniques like data imputation and completion are used to fill gaps. For noise and redundancy, data cleaning methods help remove irrelevant information, improving data accuracy. Furthermore, effective data quality control relies on advanced data governance frameworks to ensure data standardization and reliability. As businesses increasingly depend on big data, efficient data cleaning and quality assurance will directly impact the technology's effectiveness across industries. Thus, data quality management will remain a key focus in the evolution of big data technology.

5.3. Future Development Directions

The future trajectory of big data technology presents both significant opportunities and complex challenges. Advancements in artificial intelligence (AI) and machine learning will drive more intelligent big data analytics, enabling the automatic discovery of hidden patterns and insights from massive datasets. In particular, deep learning applications will enhance the extraction of valuable information from unstructured data, such as images, videos, and audio, significantly boosting analytical capabilities. The rollout of 5G

technology will further accelerate data collection and transmission, increasing the demand for real-time data processing. The integration of big data with cloud computing will also democratize access, allowing businesses to leverage on-demand computing and storage resources, reducing costs and improving scalability. Edge computing, an emerging technology, will profoundly influence big data by decentralizing data processing to locations closer to data sources, reducing latency and benefiting IoT applications requiring real-time processing. Furthermore, as data volumes continue to expand, achieving a balance between computational efficiency, storage optimization, and energy sustainability will become a key priority in the evolution of big data technologies. In summary, big data technology will become more intelligent, efficient, and adaptable, driving deeper digital transformation across industries.

6. Conclusion

The arrival of the big data era has catalyzed profound changes across industries through the application of computer science technologies. In-depth research into big data storage, processing, and analysis reveals that big data not only enables the handling of massive information but also transforms traditional industry operations, enhancing operational efficiency, reducing costs, and fostering innovation. In healthcare, big data advances precision medicine and intelligent diagnostics; in finance, it strengthens risk management and customer service; in smart manufacturing and IoT, it elevates production line intelligence and supply chain efficiency; and in public safety and smart city development, it optimizes traffic management and resource allocation. However, big data applications face significant challenges. Data privacy and security remain pressing issues requiring urgent solutions, while data quality constraints affect the reliability and accuracy of analyses. Looking forward, big data technology is expected to evolve toward increased intelligence, real-time processing capabilities, and operational efficiency, with emerging technologies like AI, 5G, and edge computing driving further innovation and adoption. As technological maturity advances and application scenarios diversify, big data is poised to unlock transformative potential across numerous sectors, propelling digital transformation. In conclusion, the continued advancement and widespread adoption of big data technology will deliver substantial economic and social benefits, empowering businesses with greater competitiveness. By proactively addressing technical challenges and aligning with emerging innovations, big data technologies will continue to exert a profound and far-reaching impact on global digital transformation.

References

1. E. Mohammadi and A. Karami, "Identifying controversial topics in social media: A case study of Twitter," *J. Inf. Sci.*, vol. 48, no. 1, pp. 44–56, 2022, doi: 10.1177/0165551520932855.
2. R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," *Technol. Forecast. Soc. Change*, vol. 153, p. 119253, 2020, doi: 10.1016/j.techfore.2018.03.024.
3. L. Liu, "Exploration on the application of electronic information technology in signal processing based on big data," *Open Comput. Sci.*, vol. 13, no. 1, p. 20220272, 2023, doi: 10.1515/comp-2022-0272.
4. N. Thuethongchai, T. Taiphapoon, A. Chandrachai, and S. Triukose, "Adopt big-data analytics to explore and exploit the new value for service innovation," *Soc. Sci.*, vol. 9, no. 3, p. 29, 2020, doi: 10.3390/socsci9030029.
5. A. Setiawan, A. Muhammad, A. P. Sandi, and N. R. Pebrianti, "Utilization of Big Data in Educational Technology Research," *Int. J. Educ. Comput. Stud.*, vol. 4, no. 1, pp. 10–20, 2024, doi: 10.35870/ijecs.v4i1.2643.
6. F. Gaiseanu, "Information: From cognitive-sentient exploration of reality to predictive big data assisted informational era," *Rom. J. Inf. Sci. Technol.*, vol. 26, no. 1, pp. 78–99, 2023. ISBN: 14538245.
7. X. Jia, "Research on the role of big data technology in the reform of English teaching in universities," *Wirel. Commun. Mob. Comput.*, vol. 2021, no. 1, p. 9510216, 2021, doi: 10.1155/2021/9510216.
8. D. Li and W. Gao, "An exploration of CDIO-based big data analytics and applications course for chemistry students," *IJAEDU-Int. E-J. Adv. Educ.*, vol. 6, no. 18, pp. 300–305, 2021, doi: 10.18768/ijaedu.819010.
9. J. C. Huang and K. M. Ko, "Cultivation of innovative talents in higher education based on big data," in *Data Process. Tech. Appl. Cyber-Phys. Syst. (DPTA 2019)*, Springer Singapore, 2020, pp. 177–182. ISBN: 9789811514678.

10. R. Dwivedi, S. Nerur, and V. Balijepally, "Exploring artificial intelligence and big data scholarship in information systems: A citation, bibliographic coupling, and co-word analysis," *Int. J. Inf. Manag. Data Insights*, vol. 3, no. 2, p. 100185, 2023, doi: 10.1016/j.jjimei.2023.100185.

Disclaimer/Publisher's Note: The views, opinions, and data expressed in all publications are solely those of the individual author(s) and contributor(s) and do not necessarily reflect the views of CPCIG-CONFERENCES and/or the editor(s). CPCIG-CONFERENCES and/or the editor(s) disclaim any responsibility for any injury to individuals or damage to property arising from the ideas, methods, instructions, or products mentioned in the content.