Analysis of multidisciplinary limits, challenges, and considerations in big data storage: recommendations for fiscal surveillance and control

Análisis de límites, desafíos y consideraciones multidisciplinarias en el almacenamiento en big data: recomendaciones para la vigilancia y control fiscal

Orlando Zapata Cortes[1*] https://orcid.org/0000-0002-3720-466X
Martin Darío Arango Serna[2] https://orcid.org/0000-0001-8448-8231
Julian Andres Zapata Cortes[3] https://orcid.org/0000-0002-1270-3577
Jaime Alonso Restrepo Carmona[2] https://orcid.org/0000-0002-1104-5453

Recibido 22 de abril de 2024, aceptado 19 de enero de 2025
Received: April 22, 2024 *Accepted: January 19, 2025*

ABSTRACT

An increasingly digitized world demands efficient information management strategies. Implementing processes for data lifecycle management, from its collection to its end user consumption, is important, as it is a process that requires significant effort and challenges in all stages of Big Data (BD). As part of BD, Big Data Storage (BDS) allows for efficient retrieval and analysis of large amounts of data in terms of storing information efficiently. An efficient BDS can be supported by Scalability, high performance, flexibility, security, data governance, cost-effectiveness, and the capacity to handle multi-structured data. This article presents a multidisciplinary view of the limits, challenges, and considerations currently found in BDS conducted by a literature review. The systematic literature review identified current problems in BDS, such as volume, velocity, variety, veracity, and data security. Based on the limits, challenges, and considerations of BDS found in the review and synthesized in multiple areas, a list of BDS recommendations was identified and tailored to the field of Fiscal Surveillance and Control. Adopting modern technologies, including cloud computing, edge computing, and advanced technologies such as blockchain and artificial intelligence, can improve data security and efficiency. An organization intending to implement an internal or external Fiscal surveillance and control system may consider these recommendations in BDS to manage their information more efficiently.

Keywords: Big data, data management, data storage, fiscal control, fiscal surveillance.

RESUMEN

Un mundo cada vez más digitalizado demanda estrategias eficientes de gestión de la información. La importancia de implementar procesos para la gestión del ciclo de vida de los datos, desde su recopilación hasta su consumo por parte del usuario final, es un proceso que requiere esfuerzos significativos y enfrenta desafíos en todas las etapas del Big Data (BD). El Almacenamiento de Grandes Datos (BDS), como parte del

[1] Instituto Tecnológico Metropolitano. Facultad de Ingeniería. Medellín, Colombia. Email: orlandozapata@itm.edu.co
[2] Universidad Nacional de Colombia. Departamento de Ingeniería de la organización. Medellín, Colombia. Email: mdarango@unal.edu.co; jarestrepoca@unal.edu.co
[3] Fundación Universitaria CEIPA. Escuela de Negocios. Sabaneta, Colombia. Email: julian.zapata@ceipa.edu.co
* Autor de correspondencia: orlandozapata@itm.edu.co

BD, permite la recuperación y análisis eficientes de grandes cantidades de datos en términos de almacenar información de manera eficiente. La escalabilidad, alto rendimiento, flexibilidad, seguridad, gobernanza de datos, rentabilidad y la capacidad para manejar datos de múltiples estructuras pueden ser respaldados por un BDS eficiente. Este artículo presenta una visión multidisciplinaria de los límites, desafíos y consideraciones encontradas actualmente en el BDS realizado mediante una revisión de literatura. A través de la revisión sistemática de la literatura, se identificó problemas actuales en BDS, como el volumen, la velocidad, la variedad, la veracidad y la seguridad de los datos. Basándose en los límites, desafíos y consideraciones en el BDS encontrados en la revisión y sintetizados en múltiples áreas, se identificó una lista de recomendaciones de BDS y se adaptó al campo de la Vigilancia y Control Fiscal. La adopción de tecnologías modernas, incluyendo la computación en la nube, la computación en el borde y tecnologías avanzadas como blockchain y la inteligencia artificial, pueden mejorar la seguridad y la eficiencia de los datos. Una organización que tenga la intención de implementar un sistema interno o externo de vigilancia y control fiscal puede considerar estas recomendaciones en el BDS para gestionar su información de manera más eficiente.

Palabras clave: Grandes datos, gestión de datos, almacenamiento de datos, control fiscal, vigilancia fiscal.

INTRODUCTION

The volume of data is increasing, and the importance of extracting information from it has grown exponentially, making data analysis an imperative aspect for companies; this enables companies to identify patterns and trends to make strategic and efficient decisions [1]-[9]. According to [10] and [11], the total volume of data is projected to reach 175 ZB (Zettabytes) by 2025.

Given the importance of data management, BD has positioned itself as a multidisciplinary tool that extends across various fields, including healthcare, political decision-making, construction, renewable energy, agriculture, insurance, natural disasters, tax evasion, engineering, computer science, telecommunications, medical informatics, industries, and organizations, among others. More and more areas are embracing BD due to its numerous capabilities and benefits [9], [11].

The large volume of data and its complexity make data quality a complex issue in the context of BD. Moreover, other dimensions of BD pose specific challenges for improving data quality, such as the variety of sources, data formats, and the high speeds at which large volumes of data are collected. This landscape is known as the four characteristics of BD - 4 V's (Volume, Variety, Velocity, Veracity) [12]-[15]. With the increase in data volume due to the rise of mobile devices, process automation through the Internet of Things (IoT), and its application in industrial IoT (IIoT) [4], [16]-[24], the characteristics of BD have expanded [2], [9], [11], [25]-[29]. According to the authors [25]-[29], the most significant BD characteristics are those associated with the final visualization of data and results. These features are crucial in presenting a clear and precise analysis to the end users [30].

BD has several challenges, such as efficiently storing and converting this data into relevant information. Another challenge is the real-time visualization of this processed data, which is ready for decision-making without any delays or latency [5], [15], [29], [31]-[40]. Organizations must update how they store and manage their data to adapt to the dynamic environments of information capture and management; this implies that companies require investing in information technologies to enhance processes related to Big Data Storage (BDS), which means updating their technological infrastructure for storage and data management and training people to use these new technologies. The investment in BDS will enable companies to manage their data more efficiently, securely, and privately [5].

BDS is important in any organization because it allows for efficient retrieval and analysis of large amounts of data. Effective BDS strategies can support scalability, fault tolerance, high performance, flexibility, security, data governance, cost-effectiveness, and the ability to handle multi-structured data [32]. Nowadays, more architectures are needed to manage the data to reach low response

times compared to traditional relational databases [41]. BDS is necessary for efficient data acquisition in any decision-making domain with large volumes of data and to allow seamless integration of data from various sources [40].

Departments or offices dedicated to Fiscal Surveillance and Control in both the private and public sectors need continuous verification to ensure that internal processes are transparent; this is crucial to prevent the misappropriation of funds and anomalies in the regular channels of organizational processes [42]. Departments of this nature should be at the forefront of methods, techniques, and procedures related to BDS to enable effective monitoring and management of data.

This article presents a literature review to identify how BDS is employed in various contexts. Subsequently, based on this review, recommendations are proposed regarding methods, techniques, and procedures that may benefit Fiscal Surveillance and Control. This review and recommendations will provide academics and practitioners with input to guide their work, facilitating the quick identification of current frameworks, algorithms, and procedures for BDS implementation in fiscal surveillance and control processes.

METHODOLOGY

The following methodology was used to perform a systematic literature review on Big Data storage and generate recommendations for monitoring and control processes.

i. Initially, the question is posed: "How BD and BDS are currently contributing to the multidisciplinary context for data lifecycle management, and that are the limits, challenges, and considerations identified by the authors in recent years?" In this stage, a keyword-based search was conducted using "Big Data," "Data Acquisition," and "Data Storage" as keywords. The term "Fiscal Surveillance and Control" was not included in the initial search since this term limits the number of articles related to BDS.
ii. Articles were searched for in the following databases using the established keywords: Scopus, IEEE, Springer, and Taylor & Francis. The search period covered 2018 to 2023. A total of 94 BDS-related articles were found.
iii. A bibliometric analysis was conducted using the VOSviewer [43] software as a post-processing step for the article metadata in the databases. Figure 1 presents the bibliometric analysis

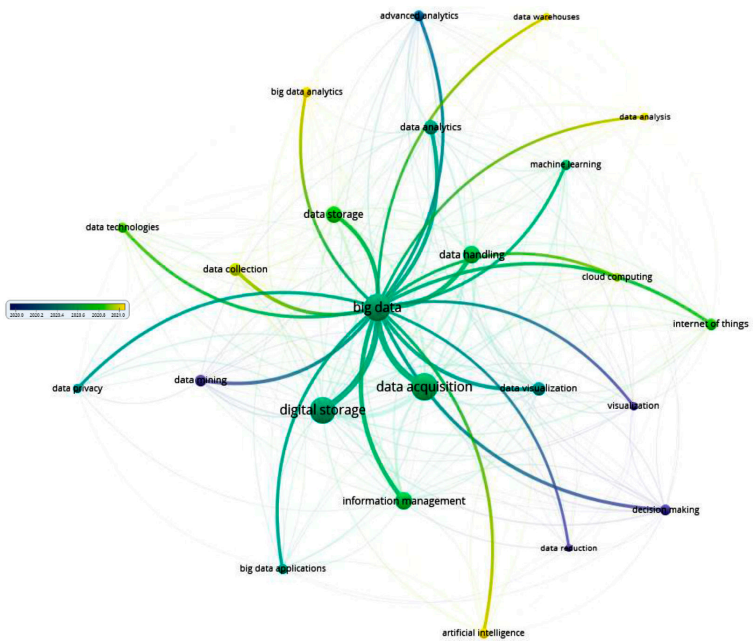

Figure 1. Bibliometric analysis.

that was conducted once the metadata search mentioned earlier was completed.

According to the bibliometric analysis based on the co-occurrence of words in the articles, other trending keywords were detected, such as 'Digital Storage,' 'Information Management,' and 'Data Handling.' A new search was performed by adding the new keywords, and based on the results obtained, only the term 'Digital Storage' was adopted, reducing the information to 82 articles. Only articles in the range of 2018-2019 with more than 10 citations were selected as a criterion for obtaining the most relevant articles in that period [4], resulting in a final count of 60 articles with relevant and precise information about BDS for the systematic literature review.

i. An analysis of the articles was carried out using an information systematization matrix to extract the necessary data that contributed to building this overview of the various applications that currently leverage the advantages of BD and BDS for effective data management.
ii. An analysis of the articles was carried out using an information systematization matrix to extract the necessary data that contributed to building this overview of the various applications that currently leverage the advantages of BD and BDS for effective data management.
iii. Based on the information obtained about the data storage processes reported in the reviewed articles, Limits, Challenges, and Considerations in BDS were identified. From there, a set of recommendations for managing BDS was proposed, which may apply to multiple areas but should be used explicitly in the Fiscal Surveillance and Control processes.

Big data states overview

The BD is consolidated as a comprehensive data lifecycle management process with different stages that range from Data Source, Data Acquisition, Data Collection, Data Preprocessing, Data Transformation, Data Storage, Data Management, Data Analysis, Data Visualization, and Data Presentation [5], [33], [34], [44]-[46]. Table 1 summarizes the stages of BD, their purpose in the data lifecycle management chain, and some applications found in the literature review.

Table 1 presents the identified DB states according to the literature review; each stage of the data lifecycle has a specific purpose and distinct applications important for efficient information management. The states depicted in Figure 2 summarily cover the DB process in Table 1, ranging from data acquisition to final presentation, highlighting the complexity and the need for a structured approach to DB management to address the challenges associated with data quality, volume, and diversity.

BD issues and states: open problems

The literature consulted reveals that some Big Data (BD) issues persist. The problems identified and mentioned by the authors in their articles are presented below:

i. Data generation volume [17]: Traditional and current methods still face challenges in storing the produced data. There is still a need for real-time aggregation or analysis methods for its management.
ii. Data generation speed [11]: There is still a need for higher information transfer speeds, and more proposed solutions are required in data sampling and system scalability.
iii. Variety in data [11]: Different sources (structured, unstructured, and semi-structured data) where there is a need for platforms that can handle the diversity of data sources and high complexity.
iv. Veracity in data [11]: Ensuring the authenticity and reliability of data. Protection against unauthorized access and modification. This issue increases when data is collected from multiple sources.
v. Value of data [11]: Managing data with no useful value. Advanced data preprocessing methods are needed to discard or transform such data into a suitable format for further analysis.
vi. Data storage [10]: Data growth in the order of Exabytes and Zettabytes requires more robust systems than traditional ones based on relational data management to handle structured data. There is a need to enhance distributed data management systems designed to handle large-scale unstructured data.
vii. Data visualization [11]: Data from different sources, structured and unstructured in different formats, is difficult to interpret, which is why more optimized and user-friendly interfaces are needed for end users.
viii. Security and Privacy [41], [49]-[55]: All important data is collected in the same system,

Table 1. States of Big Data - Big Data 10 D' states.

State	Objective	Applications
Data Source (Ds)	Identify the source	a) Coal mining systems and excavation systems [21]. b) Primary sources in social media data [10]. c) Cyber-Physical System (CPS) [28].
Data Acquisition (DA)	Identify data from various sources and apply data capture methods.	a) Real-time data acquisition from different data sources in agriculture [29]. b) Industrial environments [24]. c) Data acquisition from subsystems in building safety detection [13].
Data Collection (DC)	Collect data from different sources, which can be structured, semi-structured, or unstructured, and characterized by their multidimensionality.	a) Data collection for the healthcare applications. [1], [47]. b) Data collection in agriculture applications [48].
Data Preprocessing (DPp)	Clean data from noise, outliers, anomalies, filtering, extraction, cleansing, reduction, and validation.	a) Reduction of data volume IoT systems [19]. b) Preprocessing to send data to through smart sensors [15]. c) Preprocessing for industrial big data [20].
Data Transformation (DT)	Transform data into various formats or structures using operations such as normalization, merging, denormalization, and aggregation-linkage, among others.	a) Data transformation changing the format, structure, or values in Cyber-Physical systems [28]. b) Data integration and compression to data transfer to different platforms [33], [47].
Data Storage (DS)	Store, organize, and manage large amounts of data in a way that enables efficient retrieval and analysis.	a) Data storage in Frameworks like Hadoop, Apache Cassandra, MongoDB, Amazon S3, and Microsoft Azure [32] and Data Lakes cloud-based [20].
Data Management (DM)	Process data in parallel, resource sharing, communication mechanisms, query optimization.	a) Data ingestion, storage, and processing to data governance, security, and data visualization in multiples areas like banking, government and others [32].
Data Analysis (DA)	Analyze data collected through statistical procedures and transform it into information about historical patterns, current performance, or future trends.	a) Data mining frameworks IoT learning systems [24]. b) Descriptive and diagnostic data in climate evolution charts [31]. c) Social data in customer retention and customer loss [10].
Data Visualization (DV)	Visualize data in a visual and intuitive manner, often using images, graphics, animations, and other graphical representations.	a) Representation of climate patterns [5]. b) RFID-Cuboid model for visualizing large real-time data from the cloud [15]. c) eBay e-commerce uses data visualization tools to understand all the data generated by its customers [25].
Data Presentation (DP)	Display and convey information and insights derived from large and complex datasets using tools that are more understandable and accessible to users.	a) Data presentation Platforms like Microsoft Azure and Power BI used in business organizations [11].

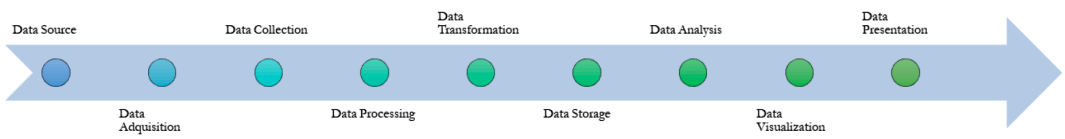

Figure 2. States of Big Data - Data lifecycle management stages in BD.

which must be highly secure. Additionally, data can be vulnerable and exposed during its transmission; this raises challenges in using decentralized data management models and efficient encryption [13], [38], [56], [57], or sensitive or highly confidential data processing, ensuring its privacy.

Table 2 presents some of the open problems in the different states in BD.

Current issues for each open problem in every BD state presented in Table 2, are:

In Data Source and Data Acquisition, some current issues for Data Quality are Quality errors, inconsistencies, or inaccuracies in data that can occur for various reasons, such as sensor failures, human errors, or issues in the data collection process. Additionally, high data generation rates come with requests. In Data Volume, an important issue is the rapid growth of data and its multiformat characteristics, as well as how to manage it [5], [21].

In Data Collection, Data Fragmentation data is often stored in separate data warehouses from different departments, leading to data fragmentation. For Data Format Diversity, there is a need to implement new technologies and procedures to improve the data formats from different sources to avoid data irregularities and lack of coherence, which affects the quality, accuracy, and availability of collected data [25].

In Data Preprocessing, due to heterogeneous Data produced by the diversity of original data formats and different sources, it is required an improvement of data preprocessing technology to avoid irregularities and the lack of coherence of the data, which affects the quality, accuracy, and availability of preprocessed data [23], [25].

Data Transformation and Data Complexity require transforming complex continuous attributes or converting raw data from different formats or structures into the same format or structure for transmission [29].

In Data Storage, for Data Volume in traditional and current methods, there is a lack of the capacity to handle the rapid growth of data and the diverse characteristics of data types. Data Fragmentation of data from the same organization stored in different locations requires the association and integration of different data types and databases while sharing technologies and tools to eliminate the information island phenomenon and Data. Better procedures are required in data complexity to handle unstructured and semi-structured databases [12], [30], [33].

Table 2. Big Data 10 D' states open problems.

State	Open problems	The article's ref#
Data Source (Ds) and Data Acquisition (DA)	Data Quality Data Volume	[5], [21]
Data Collection (DC)	Data Fragmentation Data Format Diversity	[24]
Data Preprocessing (DPp)	Heterogeneous Data	[22], [24]
Data Transformation (DT)	Data Complexity	[28]
Data Storage (DS)	Data Volume Data Fragmentation Data Complexity	[11], [29, [32]
Data Management (DM)	Security and Data Protection Policies	[23]
Data Analysis (DA)	Model Prediction Accuracy Data Speed growth Model Scalability	[22], [24]
Data Visualization (DV)	High Dimensionality and Large Data Size	[15], [25]
Data Presentation (DP)	Selection of Appropriate Tools for Data presentation to the final user	[5], [34]

In Security and Data Protection Policies in the Data Management State, insufficient defense technology and various forms of network attacks are recurrent. Additionally, there is a lack of laws, regulations, and policies for data privacy protection [23].

In Data Analysis, Model Prediction Accuracy is affected by poor analyses during the preprocessing stage, leading to incorrect or biased predictions that affect the model's reliability and validity. Data Speed growth requires more time for analytics models to calculate and analyze the data, which leads to high computational complexity and longer decision-making times or real-time recommendations. In Scalability, models are often focused on specific aspects of the application in which they are implemented. In this sense, scaling the same model to another application to obtain good results remains a persistent challenge [22], [24].

In Data Visualization, the creation of visualization procedures is a difficult task due to the High Dimensionality and Large Size of the Data in the databases and DB applications [16], [26].

Selecting the appropriate tools for data presentation for the final user in the Data Presentation Stage is important to find the best way to visualize the collected data and the results of the BD procedures. This decision should be based on the target audience [6], [35].

According to the previous information, data storage is one of the most important problems that can be solved in current BD applications. This issue is analyzed in the following sections.

APPLICATIONS RELATED TO BIG DATA STORAGE: LIMITS, CHALLENGES AND CONSIDERATIONS

The rapid adoption of BD has permeated multiple areas of knowledge, and each of these has autonomously embraced the use of the different stages of BD, leading to the emergence of particular characteristics and problems in their implementation; this has allowed new boundaries and challenges induced by the unrestricted use of the processes implemented in BD.

Different areas have fully or partially implemented the data lifecycle management process in the BD and BDS literature review. According to the literature review, seven areas were identified. The limits, challenges, and important considerations found in the literature review are presented for each area.

Healthcare

The administrative management of information associated with BDS in the healthcare sector is critical because these data consolidate a relevant and crucial knowledge base for clinical care; this enables continuous improvement in service quality and more accurate patient care recommendations. The applications presented in Table 3 are related to general medical care and administrative management associated with clinical data.

The administrative management of BDS-related information is essential to consolidate a relevant knowledge base to improve clinical care. This management allows for improving the quality of service and providing more accurate recommendations on patient care. Table 3 highlights the need to improve the integration and quality of clinical data to optimize healthcare delivery.

Security

In data security, BDS has made communication networks more reliable in transmitting data between different levels of organizations that handle large volumes of data locally or externally. One of the major challenges in security applications is data protection and the prevention of potential information attacks. The applications presented in Table 4 focus on security use for intrusion detection and how artificial intelligence algorithms contribute to information network security.

In data security, the rapid growth of data volume and complexity presents significant challenges for effective data collection, storage, and management; this highlights the necessity for robust security measures and efficient data processing techniques to mitigate risks associated with data handling and to ensure the integrity and confidentiality of sensitive information.

Agriculture

The application of BD in agriculture goes hand in hand with automation. In agriculture, automation refers to IoT devices for continuously monitoring and controlling various crop conditions. The continuous

Table 3. Healthcare limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[1], [9]	Lack of availability of high-quality and abundant data for healthcare and lack of validation with independent data.	The challenge is to compare Healthcare architectures with existing ones that may have similar features.	Machine learning enhances healthcare predictions and decision-making, but robust data management architecture is needed to handle large health data volumes and ensure integrity and security.
[9], [46]	Data privacy, Data security, accuracy, and validity of collected data.	Collect and analyze a large amount of health data from various sources to provide personalized health management.	Utilizes HDFS data storage technologies. Uses relational databases, allowing for fast and secure access to information.
[47]	Limitations handling the complexity and dynamics of security and privacy risks.	Establishing and improving security workflows in urban computing and enhancing data security risk prevention and privacy protection measures.	Hospitals need risk management strategies to ensure data security and privacy, especially in the lifecycle of big medical data, due to issues in storage, usage, and transmission stages.

Table 4. Security limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[14], [23], [56]	The rapid growth of data volume and complexity creates problems and the need for efficient data collection and storage.	Data processing Data transfer Data management Data analysis Data visualization.	Encryption and privacy are critical aspects of data protection in big data. Use scalable algorithms and machine learning techniques to obtain the best performance and reduction.
[23], [34]	Managing data privacy and security requieres significant computational resources.	Data mining technology for network intrusion detection requires high data privacy and security.	K-means clustering and Apriori algorithms enhance network intrusion detection efficiency, while adequate data storage is crucial for analyzing, classifying, and storing massive amounts of information.
[9], [49]	A large amount of data is required to train the artificial intelligence model.	Limited scalability and fault tolerance in the centralized monitoring framework.	AI-based solutions can expand the network security perspective, provide intelligent security analysis, and address new security threats.

monitoring of all these variables must be in real-time, which has required data lifecycle management systems like those offered by BDS to manage the information and provide recommendations to farmers effectively; this makes their harvests more efficient, aids in pest control processes and offers recommendations for identifying and exploring new business opportunities. The applications presented in Table 5 focus on the IoT areas of communication networks for data management and data analytics in agriculture.

The agricultural sector faces significant challenges related to the need for low latency in real-time data

Table 5. Agriculture limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[22], [29]	There is a need for low latency for real-time analysis and consistent protocols for data collection.	Strategies for storing large volumes of high-resolution data are required. Ensuring information security in the event of malicious attacks is necessary.	IOT-HSQM can meet the requirements for build and query performance and statistical analysis.
[9], [14], [48]	Data volume, frequency, and Data variety make traditional methods less effective.	Handling diverse data sources and the high complexity of data structures in agriculture.	Advanced technologies like Hadoop, Hive, HBase, and Spark are used for data cleaning, consolidation, and persistent storage.
[31]	Scalability problems and lack of hyperparameter calibration in the implemented machine learning model.	System scalability or handling long-running batch processing or ETL.	Weather data storage is essential for implementing innovative and sustainable agriculture.

analysis and the establishment of consistent protocols for data collection. Additionally, strategies are urgently required to effectively store large volumes of high-resolution data while ensuring information security against malicious threats. These factors are important for enhancing the efficiency and effectiveness of agricultural practices.

Blockchain

Blockchain and DB technology enable secure and reliable data collection, processing, and analysis to address current data management challenges. Additionally, this enables establishing a more effective and realistic system for assessing and storing security information. Table 6 presents some applications of BDS and Blockchain to data management in healthcare, construction, and administrative information management.

Effective cooperation and participation from all departments and employees are essential for successful data collection and recording. The lack

Table 6. Blockchain limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[54]	Cooperation and participation of all departments and employees in the company in data collection and recording.	Manual information from workers is not standardized, hindering effective use for occupational safety decision-making, real-time identification, and recording of safety information.	Information management, such as machinery and material tracking, can provide new insights for construction management and reduce safety accidents.
[16], [18]	Real-time processing and data security with IoT devices. Standard security solutions like firewalls and other peripheral security measures are often ineffective.	Attackers often change policies to acquire data illegally, which can challenge most data owners to verify if a policy is legitimate.	The framework that provides a highway protocol enhances security, scalability, utilization, and efficiency in big data storage.
[18], [50]	Handling high-frequency transactions and storing large blocks of data.	Management and security of electronic health records.	Efficient systems that outperform existing methods in terms of efficiency, delays, attack level, security level, encryption and decryption time, and memory usage.

of standardized manual information from workers poses a significant challenge, hindering the effective use of blockchain technology for occupational safety and data integrity; this highlights the need for improved standardization and collaboration to maximize the benefits of blockchain in various applications.

Internet of things - IoT

IoT is one of the most representative areas in which the implementation of processes in BDS has been explored due to the high volume of data generated when automating a process with data generation sensors. Generating large volumes of data has posed significant challenges in all stages of BDS, especially in data acquisition, collection, preprocessing, and storage. All these challenges have led to developing techniques for optimizing, transforming, and efficiently transmitting data. The applications presented in Table 7 are framed in IoT and IIoT for industrial purposes, cloud computing, and edge computing.

Due to the accelerated digitization of the data IoT sector, there are significant hurdles related to the processing and storage capacity of devices. The challenges include managing large volumes of data, addressing latency issues, and dealing with data heterogeneity and inconsistent data patterns. There is a need to overcome these for more efficient

Table 7. Internet of things limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[15], [17], [58]	Processing and storage capacity of the devices used	Storage and processing of a large volume of data, latency, data heterogeneity, and inconsistent data patterns.	Frameworks are more efficient in parallel data ingestion, parallel processing at each edge, efficient cluster resource management, and model parallelism.
[36]	Need for adequate information technology infrastructure and data privacy and security issues.	Traditional methods of Internet data processing are not fully applicable to IoT data, requiring efficient processing and storage methods.	Three-tier architectures for data collection and analysis. Uses of Smart big data and a distributed data collection architecture.
[14], [21]	Data formats are unclear, and data quality is not adequately described, which can lead to errors in data collection, transmission, and storage.	Data normalization in smart mining, acquired from various automation systems, can lead to data collection, transmission, and storage errors.	The industry's rapid evolution and explosive data growth size pose high requirements for data management.
[20], [33]	Need for adapting to changes in technology and industry practices in IIoT and complex computational capability for handling and storing volumes of data generated by IIoT sensors.	Real-time data processing in production environments ensures integrity for diagnostics using IoT devices in innovative manufacturing environments, encompassing data acquisition, compression, storage, quality, and processing.	Apache Kafka and Apache Spark in an Apache Hadoop cluster are real-time data acquisition, compression, storage, and processing solutions. Additionally, Apache NiFi allows for the creation of data pipelines for efficient and secure data flow in production processes.
[4]	Unstructured IoT data is a problem in the IoT ecosystems (multiple protocols, platforms, security, and technologies). Another limitation is the lack of standards and best practices for data management in IoT.	Requirements for specialized techniques and tools and high-performance hardware and software with massive storage capacity to process, analyze, and visualize large volumes of IoT data quickly, efficiently, and accurately.	The lack of standards and best practices for IoT data management and machine learning for processing and analyzing IoT data is an emerging trend with great potential to enhance the efficiency and accuracy of data management systems.

frameworks that facilitate parallel data ingestion, processing at the edge, and effective cluster resource management.

Energy

The field of energy efficiency is also an important area for BD, which is closely related to process automation through IoT. The implementation of BD and BDS in the energy sector is focused on energy efficiency using IoT devices. BD and BDS aim to make data management and visualization more efficient and effective for process analysis and recommendations to make the most of the data and optimize processes, reduce operating costs, and lower energy consumption in companies. The applications presented in Table 8 are framed in the area of energy efficiency, focusing on wireless networks, irrigation assessment, security, and intelligent energy efficiency.

The energy field is facing challenges in using artificial intelligence algorithms to handle large volumes of data and identify complex patterns. Key challenges include data acquisition, storage, quality, security, and privacy, and the presence of anomalies. These obstacles are compounded by the current energy transition, which calls for more efficient and sustainable resource management. As demand for renewable energy and low-carbon solutions grows, data storage and processing capacity becomes increasingly important. Solving these problems involves improving the efficiency of energy data identification methods, facilitating a smooth energy transition, and ensuring effective information management in a context that demands speed, accuracy, and sustainability.

Data base, process and storage

The BDS application in the database and information processing field is focused on creating new procedures for integrating and adapting massive data. One of the objectives of creating these new information management methods in BDS is to create new mechanisms and procedures for the collection and creation of databases with characteristics that combine elements of structured and semi-structured data or parallelism in processing, generating novel methods that include

Table 8. Energy limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[3]	Artificial intelligence algorithms have limitations in handling large amounts of data or identifying complex patterns.	The efficiency of intelligent energy data identification methods requires addressing challenges such as data acquisition, storage, quality, security, privacy, algorithm limitations, anomalies, and real-time issues.	Efficient data transmission is challenging due to high bandwidth consumption and energy efficiency concerns, requiring reliable storage while balancing cost.
[52]	The accuracy of the ANN analytics model depends largely on data quality.	The energy sector faces increasing security and privacy risks in big data, particularly in a Cloud Computing environment, due to the growing industry's distribution and diverse data types.	The proposed method, uses a risk assessment index system and neural network model, offers high prediction accuracy, objective evaluation results, and practical value in cloud data storage.
[55]	Data storage solutions require a financial investment and specialized infrastructure, which can be challenging to implement for some organizations.	Wireless technologies can increase power consumption and decrease energy efficiency due to high data volumes, making optimizing energy efficiency a complex and ever-changing challenge.	Virtualization, cloud computing, low-power wireless networks, and advanced mobile caching techniques can enhance energy efficiency by reducing overhead, optimizing resource demand, and improving data transmission.

artificial intelligence and statistics to retrieve and repair data or generate missing data.

It is also recommended to implement these new methods for intelligent filtering processes to clean and purify data that contains noise, makes sense, or is redundant; this is relevant because BDS permeates and makes it possible to analytically analyze information without bias, facilitating better training of analytical models and better predictions or recommendations. The applications presented in Table 9 are framed databases area and the generation of new data management and processing methods.

Increasing complexity and volume of data necessitate advanced systems capable of efficiently handling high-dimensional and rapidly expanding datasets. High processing rates and managing heterogeneous data types and sources are critical for effective data processing and storage solutions. In Data Base, Processes, and Storage, the need to improve security, performance optimization, and data management and reduce the costs associated with infrastructure and human resources continues. Security and Performance challenges, optimization processes, data management inefficiencies, and access barriers for small organizations are current issues continually being faced in this area. More efficient solutions to

Table 9. Data base, process and storage limits, challenges and considerations.

The article's ref#	Limits	Challenges	Considerations
[10], [12], [49]	Despite its potential for efficiently handling high-dimensional data, the final handling rapidly expands the network data and increases calculations.	The complexity of processing and storing large volumes of transformer state data due to heterogeneity, multiple types, and sources necessitating high processing rates.	The deep hash algorithm for performing massive data retrieval. Data standardization can enhance quality.
[32], [53]	Apache Hadoop platform faces security issues and lower performance with small data, while Apache Spark lacks optimization, highlighting security concerns in cloud data storage.	These issues pose challenges in various applications, including data management structure, integration, indexing, and security.	Future directions of research in data variety, data volume, data velocity, data security, and data governance.
[23], [57]	Traditional data storage technology struggles to eliminate data redundancy among files, posing speed, capacity, storage efficiency, and security challenges.	There is high redundancy in the data stored by application systems, which can reach up to 60% and increase over time, along with deficiencies in current models of private data management.	Methods that achieve high duplication rate and low storage space requirements using the Huffman algorithm, enhancing data query speed. Efficient data optimization schemes require a combination of multidisciplinary expertise.
[6]	Data storage infrastructure and human resources can be expensive and complex, which may limit access and usage of these technologies to larger organizations.	Without proper infrastructure, the capacity for processing and storing data is limited, which can restrict the quantity and type of information that can be handled and analyzed. No infrastructure allows for significant errors and hampers organizations' ability to harness the full potential of these technologies.	The right people and technical expertise may be necessary to ensure the effectiveness and sustainability of big data projects. Financial and technical limitations should be carefully considered before starting a big data project and must be balanced with the project's objectives and potential benefits.

processing and storage complexity, data redundancy, and infrastructure limitations are needed to enable organizations to leverage the full potential of data, optimizing their ability to analyze and use information effectively for decision-making.

BD is venturing into emerging areas where the analyzed literature intuitively envisions significant efforts to adopt all these processes and leverage these tools to provide a higher level of security, efficient management, process optimization, and analytics and visualization tools for more effective decision-making. BD works in harmony with areas in Education [8], [45], Smart Production [19], [27], [35], emerging fields of industry 4.0 such as Virtual Reality [55], implementation of artificial intelligence to network services in real-time to detect failure based on Cloud Computing and Edge-Computing [59], [60]. In these areas, there are limits, challenges, and important considerations in BDS due to the volume of data the authors have reported in their research work.

RECOMMENDATIONS OF BIG DATA STORAGE TO FISCAL SURVEILLANCE AND CONTROL

The development of modern Information and Communication Technologies (ICT) has significantly impacted organizations' audit and control. Modern audits require organizations to update their traditional data management systems with intelligent management systems [42]. Surveillance and fiscal control in organizations need constant updating in their data collection, data cleaning, data management, data conversion, data analysis, and data visualization processes to achieve faster, more efficient, and secure control. The need for process updating arises because, in this area, there is more confidence in documenting and tracing processes on paper than digitally, which implies delays in handling, updating, and visualizing real-time data [42]. It will be important for both government and private organizations to adopt processes such as BD to keep up with the current demands of modern audits.

The following recommendations emerge for Fiscal surveillance and control organizations within the literature review framework. Although these recommendations are oriented to Fiscal Surveillance and Control, they also can be applied to BDS processes in other application domains.

Novel data storage methods, such as Cloud Computing and Edge Computing, applied to BDS enable the implementation of more efficient and secure data management systems, allowing for the more reliable collection, recording, and analysis of large volumes of data. Technologies such as blockchain and data encryption methods provide high levels of security and confidentiality to information. The inclusion of new advances and techniques in artificial intelligence allows scaling models and frameworks through transfer learning to achieve more precise predictions, recommendations, and effective data storage processes.

Data storage

- Storage systems like Hadoop Distributed File System (HDFS) [1] are used to store large amounts of data efficiently. Proper data storage facilitates access and management of the information necessary for tax surveillance and control.
- Adaptive data integration [49] can help consolidate and unify data from different sources and formats, which is crucial for fiscal surveillance and control. Through more effective data integration and analysis, patterns, trends, and anomalies relevant to tax fraud detection and decision-making can be identified.
- Frameworks like Hadoop and NoSQL provide efficient data acquisition tools [56]. These technologies can be useful for fiscal surveillance and control since they allow large datasets' rapid and efficient processing.
- Real-time data acquisition: Frameworks like Apache Kafka and Apache Spark in an Apache Hadoop cluster enable real-time data acquisition [33]; this can be valuable for fiscal surveillance and control as it allows real-time monitoring of financial transactions and operations.
- Hybrid storage [41] platforms that efficiently handle large volumes of data can improve response times, which is crucial for fiscal surveillance and control, as these activities often involve analyzing large amounts of financial and transactional data.
- Advanced platforms like Hadoop, Hive, HBase, and Spark for data cleaning, data consolidation, and persistent storage [23], [29], [36], [48].

These technologies can handle large volumes of fiscal data and provide fast and secure access to information.

- Frameworks like Apache Hadoop, Apache Spark, Apache Flink, and Apache Kafka can provide the necessary tools and algorithms for processing and analyzing large volumes of fiscal data [32].
- Platforms like MongoDB can offer solutions for efficient data storage and fast data access [27], [31].

Data storage using cloud computing and edge computing

- Cloud-based technologies such as AWS, Microsoft Azure, GPC, IBM Cloud, Oracle Cloud, Alibaba Cloud, Salesforce, Red Hat OpenShift, and VMware Cloud are platforms developed to handle BD, distributed file systems, and scalable data storage [14], [15], [17], [55]. They can overcome the limitations of traditional local data storage and management technologies, which can benefit fiscal surveillance and control. These systems must ensure data integrity, availability, and confidentiality in the cloud.

Data security in big data storage

- Use encryption and privacy techniques to protect fiscal data stored in massive data storage systems. These techniques can help prevent unauthorized access to the data and ensure its confidentiality [50].
- Blockchain can enhance the security and management of fiscal records in BD analysis [16], [18], [54].

Scalability and machine learning techniques in big data storage

- Scalable algorithms and machine learning techniques can be used to analyze fiscal data and detect possible irregularities or fraud [15]. These techniques can help identify patterns and anomalies in fiscal data, facilitating the detection of potential fiscal evasion or fraudulent activities.
- Implement software that allows integration with artificial intelligence algorithms like Random Forest [1], [42], K-means [22], [34], Apriori [34], ANN, SVM [1], [51], which can lead to greater efficiency in detecting tax anomalies or irregularities.

An organization intending to implement an internal or external Fiscal surveillance and control system should consider all the mentioned features to ensure that data lifecycle management is effective in terms of security, storage, data analysis, and data visualization.

CONCLUSIONS

In the areas identified by the literature review, there are ongoing efforts to improve data processing from BDS. The unique characteristics of each area lead to variations in data processing, highlighting specific limitations and challenges. Issues such as quality, integration, and security of clinical data in healthcare; robust security measures to ensure data integrity and confidentiality in data security; real-time data analysis in agriculture; effective cooperation and standardization of information in blockchain; data processing, storage capacity, data heterogeneity, and latency issues in IoT; integration of artificial intelligence-based methods in the energy sector and the increasing complexity and volume of data in data Base, process, and storage require advanced systems. All the problems mentioned above are still being studied. Improvement directions to address these issues in BDS include data standardization by establishing data collection protocols and improving data management systems that facilitate interoperability and data integration across systems. Enhanced security measures by implementing robust security frameworks to protect sensitive data and ensure compliance with privacy regulations. Additionally, invest in a scalable IT infrastructure that can handle the growing volume and complexity of data.

The applications of BDS are currently numerous, and as the field evolves, more areas are joining in due to the advantages it offers for data management. An increasing number of sectors continue to recognize the importance of improving their processes with BD. However, many challenges need to be addressed, such as standardization, data interoperability, improvement in related BD frameworks, real-time data processing and latency-free result visualization, data management, optimized searches, accurate and unbiased exploration, and data analytics with precise predictions, scalability to different processes, data security, and privacy mechanisms in BD. The organizations that succeed in this process must

develop dynamic capabilities to continuously adapt and harness the possibilities offered by data-driven innovation, promoting a culture open to change and experimentation.

Technologies related to BDS, such as Cloud Computing and Edge Computing, are increasingly positioned as efficient processes to handle the accelerated growth of data generation due to the digitalization of information through IoT and IIoT, social networks, emails, digital format generation, and reports, among others. It is important to note that these systems are networked and susceptible to cyberattacks; therefore, it is important to implement these processes hand in hand with computer security. This article suggests that by using data analytics and artificial intelligence, organizations can improve decision-making when a holistic approach to data and its sources is required; this facilitates the effective integration of the various states of BD.

The successful implementation of BDS depends on selecting the right tools, identifying the proper use, building suitable and trained teams, fostering a data culture within the organization, and integrating the results into existing processes. Additionally, it is important to understand that implementing a BD process can be costly due to acquiring cutting-edge technologies for each application.

As future work, the search scope of this review can be expanded to cover the BDS strategies implemented in companies and organizations oriented to implementing BD processes, users of big data, and government entities, which can generate another point of view different from the literature review.

ACKNOWLEDGEMENTS

This paper was funded by the project titled "Diseño e implementación de soluciones integrales para modernizar la infraestructura tecnológica que permita fortalecer los sistemas de información y analítica de datos de la Contraloria General de la República", with code H: 59943 of the Universidad Nacional de Colombia.

REFERENCES

[1] A. Sisodia and R. Jindal, "An effective model for healthcare to process chronic kidney disease using big data processing," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 10, pp. 1-17, 2023, doi: 10.1007/s12652-022-03817-w.

[2] T. Ho *et al.*, "A new model for collecting, storing, and analyzing big data on customer feedback in the tourism industry," *Interdisciplinary Journal of Information, Knowledge, and Management*, no. 18, pp. 225-249, 2023, doi: 10.28945/5107.

[3] L. Luo *et al.*, "Intelligent identification over power Big Data: Opportunities, solutions, and challenges," *Computer Modeling in Engineering and Sciences*, vol. 134, no. 3, pp. 1565-1595, 2023, doi: 10.32604/cmes.2022.021198.

[4] A. Naghib, N. Jafari Navimipour, M. Hosseinzadeh, and A. Sharifi, "A comprehensive and systematic literature review on the big data management techniques in the internet of things," Wireless Networks, vol. 29, no. 3, pp. 1085-1144, abr. 2023, doi: 10.1007/s11276-022-03177-5.

[5] M. Petrova, P. Popova, V. Popov, K. Shishmanov, and K. Marinova, "Potential of Big Data analytics for managing value creation," in *2022 International Conference on Communications, Information, Electronic and Energy Systems (CIEES)*, Veliko Tarnovo, Bulgaria, 2022, pp. 1-6, doi: 10.1109/CIEES55704.2022.9990882.

[6] H. Li, "Research on Big Data analysis data acquisition and data analysis," in *2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)*, Xi'an, China, 2021, pp. 162-165, doi: 10.1109/CAIBDA53561.2021.00041.

[7] S. Sultana, S. Akter, E. Kyriazis, and S. F. Wamba, "Architecting and developing big data-driven innovation (ddi) in the digital economy," *Journal of Global Information Management*, vol. 29, no. 3, pp. 165-187, 2021, doi: 10.4018/JGIM.2021050107.

[8] W. Xiao, "Quality-Centered mobile terminal teaching assessment system under the background of Big Data," in *2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, Zhangjiajie, China, 2020, pp. 446-449, doi: 10.1109/ICSGEA51094.2020.00102.

[9] M.K. Saggi and S. Jain, "A survey towards an integration of big data analytics to big

insights for value-creation," *Information Processing and Management*, vol. 54, no. 5, pp. 758-790, 2018, doi: 10.1016/j.ipm.2018.01.010.

[10] S. Benjelloun *et al.*, "Big Data Processing: Batch-based processing and stream-based processing," in *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Fez, Morocco, 2020, pp. 1-6, doi: 10.1109/ICDS50568.2020.9268684.

[11] M. Mahmoudian, S.M. Zanjani, H. Shahinzadeh, Y. Kabalci, E. Kabalci, and F. Ebrahimi, "An overview of Big Data concepts, methods, and analytics: challenges, issues, and opportunities," in *2023 5th Global Power, Energy and Communication Conference (GPECOM)*, Nevsehir, Turkiye, 2023, pp. 554-559, doi: 10.1109/GPECOM58364.2023.10175760.

[12] P. Zhang, F. Xiong, J. Gao, and J. Wang, "Data quality in big data processing: Issues, solutions and open problems," in *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, San Francisco, CA, USA, 2017, pp. 1-7, doi: 10.1109/UIC-ATC.2017.8397554.

[13] J. Xu, C. Yan, Y. Su, and Y. Liu, "Analysis of high-rise building safety detection methods based on big data and artificial intelligence," *International Journal of Distributed Sensor Networks*, vol. 16, no. 6, 2020, doi: 10.1177/1550147720935307.

[14] H.N. Dai, R.C.W. Wong, H. Wang, Z. Zheng, and A.V. Vasilakos, "Big data analytics for large-scale wireless networks: Challenges and opportunities," *ACM Computing Surveys*, vol. 52, no. 5, 2019, doi: 10.1145/3337065.

[15] S. Khare and M. Totaro, "Big Data in IoT," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944495.

[16] A. Sasikumar, L. Ravi, K. Kotecha, A. Abraham, M. Devarajan and S. Vairavasundaram, "A secure Big Data storage framework based on blockchain consensus mechanism with flexible finality," in *IEEE Access*, vol. 11, pp. 56712-56725, 2023, doi: 10.1109/ACCESS.2023.3282322.

[17] M. Babar, M.A. Jan, X. He, M.U. Tariq, S. Mastorakis and R. Alturki, "An optimized IoT-enabled Big Data analytics architecture for Edge-Cloud computing," *IEEE Internet of Things Journal,* vol. 10, no. 5, pp. 3995-4005, 2023, doi: 10.1109/JIOT.2022.3157552

[18] N. Deepa et al., "A survey on blockchain for big data: Approaches, opportunities, and future directions," *Future Generation Computer Systems*, vol. 131, pp. 209-226, 2022, doi: 10.1016/j.future.2022.01.017.

[19] H. Shi, G. Cao, G. Ma, J. Duan, J. Bai, and X. Meng, "New progress in artificial intelligence algorithm research based on Big Data Processing of IOT systems on intelligent production lines," *Computational Intelligence and Neuroscience*, vol. 2022, Art. no.3283165, 2022, doi: 10.1155/2022/3283165.

[20] W. Yu, Y. Liu, T. Dillon, W. Rahayu, and F. Mostafa, "An integrated framework for health state monitoring in a smart factory employing IoT and Big Data techniques," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 2443-2454, 2022, doi: 10.1109/JIOT.2021.3096637.

[21] H. Wang *et al.*, "A management specification for Big Data sharing in smart mine," in *2022 International Conference on Computing, Communication, Perception and Quantum Technology, (CCPQT*), 2022, pp. 313-318. doi: 10.1109/CCPQT56151.2022.00060.

[22] L. Hu and X. Xia, "5G-Oriented IoT Big Data analysis method system," *Mobile Information Systems*, vol. 2021, no. 1, Art. no. 3186696, 2021, doi: 10.1155/2021/3186696.

[23] X. Feng, "The optimization of privacy data management model in Big Data era," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, 2021, pp. 2656-2660, doi: 10.1109/IAEAC50856.2021.9390675.

[24] J. Cao, M. Lin, and X. Ma, "A survey of big data for IoT in cloud computing," *IAENG International Journal of Computer Science*, vol. 47, no. 3, pp. 585-592, 2020.

[25] H.B. Abdalla, "A brief survey on big data: technologies, terminologies and data-intensive

applications," *Journal of Big Data*, vol. 9, no. 1, Art. no. 107, 2022, doi: 10.1186/s40537-022-00659-3.

[26] M. Tanisha and M.R. Gupta, "An Intensive review on implementation of Big Data in different applications and its associated issues and Challenges," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India, 2022, pp. 670-673, doi: 10.1109/IC3I56241.2022.10072480.

[27] D. Pevec, H. Vdovic, I. Gace, M. Sabolic, J. Babic, and V. Podobnik, "Distributed data platform for automotive industry: A robust solution for tackling big challenges of Big Data in Transportation Science," in *2019 15th International Conference on Telecommunications (ConTEL)*, Graz, Austria, 2019, pp. 1-8, doi: 10.1109/ConTEL.2019.8848542

[28] S. Luo, H. Liu, and E. Qi, "Big data analytics - enabled cyber-physical system: model and applications," *Industrial Management and Data Systems*, vol. 119, no. 5, pp. 1072-1088, 2019, doi: 10.1108/IMDS-10-2018-0445.

[29] M.P.S. Gopal and B.R. Chintala, "Big data challenges and opportunities in agriculture," *International Journal of Agricultural and Environmental Information Systems*, vol. 11, no. 1, pp. 48-66, 2020, doi: 10.4018/IJAEIS.2020010103.

[30] 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 *ACM International Conference Proceeding Series*, New York, USA, 2019, pp. 19-24. doi: 10.1145/3312614.3312623.

[31] C.E. Hachimi, S. Belaqziz, S. Khabba, B. Sebbar, D. Dhiba, and A. Chehbouni, "Smart weather data management based on artificial intelligence and big data analytics for precision agriculture," *Agriculture*, vol. 13, no. 1, Art. no. 95, 2023, doi: 10.3390/agriculture13010095.

[32] P.H. Ahmad y M. Rai, "Analysis of optimization strategies for Big Data storage management: a study," in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2023, pp. 1747-1753, doi: 10.1109/ICESC57686.2023.10193738.

[33] Ö. Akın, H.F. Deniz, D. Nefis, A. Kızıltan, and A. Çakır, "Enabling Big data analytics at manufacturing fields of Farplas automotive," in *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions. Advances in Intelligent Systems and Computing, vol. 1197, C.* Kahraman, S. Cevik Onar, B. Oztaysi, I. Sari, S. Cebi, and A. Tolga, Eds. Cham, Switzerland : Springer, 2021, pp. 817-824, doi: 10.1007/978-3-030-51156-2_94.

[34] J. Bi and X. Chen, "Technology and system of network intrusion detection based on Big Data," in *2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang*, China, 2023, pp. 251-255, doi: 10.1109/ACCTCS58815.2023.00087.

[35] Q. Chen, J. She, W. Min, Z. Zhang, and Z. Liao, "Design and development of big data system for blast furnace ironmaking," in *2023 IEEE 6th International Conference on Industrial Cyber-Physical Systems (ICPS)*, Wuhan, China, 2023, pp. 1-5, doi: 10.1109/ICPS58381.2023.10127866.

[36] H. Yan, "News and public opinion multioutput IoT intelligent modeling and popularity big data analysis and prediction," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, Art. no. 3567697, 2022, doi: 10.1155/2022/3567697.

[37] W. Liu, S. Li, J. Pan, L. Xu, Z. Gu, and L. Hai, "Design of management platform architecture and key algorithm for massive monitoring big data," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, Art. no. 3111844, 2021, doi: 10.1155/2021/3111844.

[38] D. Miltiadou *et al.*, "A Big Data intelligence marketplace and secure analytics experimentation platform for the aviation industry," *Big Data Technologies and Applications. BDTA WiCON 2020. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, vol. 371, Z. Deze, H. Huang, R. Hou, S. Rho, and N. Chilamkurti, Eds. Cham, Switzerland: Springer, 2021, pp. 48-62, doi: 10.1007/978-3-030-72802-1_4.

[39] J. Li and X. Xu, "A study of big data-based employees' public opinion system construction," *Journal of Industrial Integration and Management*, vol. 5, no. 2, pp. 225-233, 2020, doi: 10.1142/S2424862220500062.

[40] D. Geng, C. Zhang, C. Xia, X. Xia, Q. Liu, and X. Fu, "Big data-based improved data acquisition and storage system for designing industrial data platform," *IEEE Access*, vol. 7, pp. 44574-44582, 2019, doi: 10.1109/ACCESS.2019.2909060.

[41] Z. Lv, X. Li, H. Lv, and W. Xiu, "BIM big data storage in WebVRGIS," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2566-2573, 2020, doi: 10.1109/TII.2019.2916689.

[42] Y. Hao and F. Qiu, "Research on the application of DM technology with RF in enterprise financial audit," *Mobile Information Systems*, vol. 2022, no. 1, Art. no. 4051469, 2022, doi: 10.1155/2022/4051469.

[43] "VOSviewer - Visualizing scientific landscapes," VOSviewer.com. Access: Jan. 6, 2024. [Online]. Available: https://www.vosviewer.com//

[44] S. Pan, "Research on E-Learning resource push system based on computer Big Data technology," in *2023 International Conference on Networking, Informatics and Computing (ICNETIC)*, Palermo, Italy, 2023, pp. 167-171, doi: 10.1109/ICNETIC59568.2023.00041.

[45] H. Li, L. Gao, y X. Yang, "Research and application of application system architecture based on big data," *Journal of Physics: Conference Series*, vol. 482, Art. no. 042086, 2021, doi: 10.1088/1742-6596/1852/4/042086.

[46] L. Ai, "artificial intelligence system for college students' physical fitness and health management based on physical measurement big data," *Wireless Communications and Mobile Computing*, vol. 2021, Art. no. 4727340,2021, doi: 10.1155/2021/4727340.

[47] R. Jiang, M. Shi, and W. Zhou, "A privacy security risk analysis method for medical big data in urban computing," *IEEE Access*, vol. 7, pp. 143841-143854, 2019, doi: 10.1109/ACCESS.2019.2943547.

[48] X. Wang, K. Yang, and T. Liu, "The implementation of a practical agricultural big data system," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC),* Chengdu, China, 2019, pp. 1955-1959, doi: 10.1109/ICCC47050.2019.9064475

[49] Y. Lu and F. Lu, "Adaptive integration method of transformer state big data based on deep hash algorithm," *Multimedia Tools and Applications*, vol. 82, no. 30, 2023, doi: 10.1007/s11042-023-15578-5.

[50] V. S. Marichamy and V. Natarajan, "Blockchain based securing medical records in big data analytics," *Data and Knowledge Engineering*, vol. 144, Art. no. 102122, 2023, doi: 10.1016/j.datak.2022.102122.

[51] X. Lu, "Research on the construction system of computer big data security platform based on artificial intelligence," in *2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, Hamburg, Germany, 2022, pp. 865-869, doi: 10.1109/AIAM57466.2022.00175.

[52] Z. Li, W. Xu, H. Shi, Y. Zhang, and Y. Yan, "Security and privacy risk assessment of energy Big Data in cloud environment," *Computational Intelligence and Neuroscience*, vol. 2021, no.1, Art. no. 9896475, 2021, doi: 10.1155/2021/2398460.

[53] Z. Qichen, "Cotton textile equipment data management system based on big data technology," in *2020 International Conference on Robots & Intelligent System (ICRIS), Sanya, China, 2020,* pp. 257-260, doi: 10.1109/ICRIS52159.2020.00071.

[54] Y. Wei and E. Cui, "Application of blockchain and big data technology in the safety information management of architecture employees," in *IOP Conference Series: Earth and Environmental Science*, vol. 567, Art. no. 012042, 2020, doi: 10.1088/1755-1315/567/1/012042.

[55] X. Cao, L. Liu, Y. Cheng, and X.S. Shen, "Towards Energy-Efficient wireless networking in the Big Data era: a survey," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 1, pp. 303-332, 2018, doi: 10.1109/COMST.2017.2771534.

[56] D. Demirol, R. Das, and D. Hanbay, "A key review on security and privacy of big data: issues, challenges, and future research directions," *Signal, Image and Video Processing*, vol. 17, no. 4, pp. 1335-1343, 2023, doi: 10.1007/s11760-022-02341-w.

[57] T. Yu, "Unbalanced Big Data-Compatible cloud storage method based on redundancy elimination technology," *Scientific Programming*, vol. 2022, Art. no. 1371778, 2022, doi: 10.1155/2022/1371778.

[58] D. Geng, C. Zhang, C. Xia, X. Xia, Q. Liu, and X. Fu, "Big data-based improved data acquisition and storage system for designing industrial data platform," *IEEE Access*, vol. 7, pp. 44574-44582, 2019, doi: 10.1109/ACCESS.2019.2909060.

[59] H. Dong, X. Liang, Y. Liu, and D. Wang, "5G virtual reality in the design and dissemination of contemporary urban image system under the background of Big Data," *Wireless Communications and Mobile Computing*, vol. 2022, no.1, Art. no. 8430186, 2022, doi: 10.1155/2022/8430186.

[60] Y. Wang *et al.*, "Virtual network fault management platform and mechanism based on Big Data," in *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Beijing, China, 2023, pp. 1-4, doi: 10.1109/BMSB58369.2023.10211122.