REVIEW ARTICLE

Blockchain for Big Data: Approaches, Opportunities and Future Directions

Amrita Jyoti[1], Vikash Yadav[2,*], Ayushi Prakash[1], Sonu Kumar Jha[3] and Mayur Rahul[4]

[1]ABES Engineering College Ghaziabad, Uttar Pradesh, India; [2]Government Polytechnic Bighapur Unnao, Board of Technical Education, Uttar Pradesh, India; [3]Krishna Engineering College, Ghaziabad, Uttar Pradesh, India; [4]Department of Computer Applications, CSJM University, Kanpur, Uttar Pradesh, India

ARTICLE HISTORY

Received: November 13, 2022
Revised: March 05, 2023
Accepted: March 07, 2023

DOI:
10.2174/2352096516666230607104537

Abstract: The last several years have seen a significant increase in interest in big data across a range of scientific and engineering fields. Despite having several benefits and applications, big data still has some difficulties that must be overcome for a higher level of service, such as big data analytics, big data management, and big data privacy and security. Big data services and apps stand to greatly benefit from blockchain decentralisation and security features. In this article, we present an overview of blockchain for big data with an emphasis on current methods, possibilities, and upcoming trends. We begin by providing a succinct explanation of big data, blockchain, and the purpose of their integration. After that, we look at different types of blockchain assistance for big data, such as blockchain for security in big data collection, data privacy protection, storage, and collection. Next, we examine the latest work on the utilization of blockchain applications for big data across different industries, including smart grid apps and applications, smart city applications, and smart healthcare applications. A few illustrative blockchain-big data initiatives are given and discussed for a good understanding. Finally, difficulties and potential directions are examined to advance research in an exciting field.

Keywords: Big data, blockchain, blockchain services, distributed ledger, amazon, grid apps.

1. INTRODUCTION

Over the past ten years, data traffic has spread at an unmatched rate internationally, which is why "big data" has received so much attention. The big data business is expected to reach around 250 billion dollars before 2025, Big Data is a new generation technology being studied to evaluate vast amounts of data and identify its key features like analytics, knowledge discovery, and high velocity [1, 2]. Big data is regarded from a comparative perspective as datasets with very high dimensions and sizes that are unable to be managed, stored, analysed, or collected by present database methods. From an architectural point of view, big data is defined as a database with exceedingly high volumes, speeds, and representations that demand significant horizontal scaling techniques for effective processing [3].

Difficulties, privacy, and security are crucial concerns since big data usually incorporates many forms of sensitive information, like names, addresses, ages, and banking information. Many different techniques and solutions have been researched for protecting data privacy and confidentiality, and reinforcement learning's potential applicability was examined in [4]. To improve the data standards and handle the computationally in-depth activities required by IoT tools while providing security and privacy guarantees [5]. Blockchain can drastically change the way that present big data systems are run. In this survey, we look at blockchain for big data from all angles, including how it works, what it can do, and what the future holds for it.

1.1. The State of the Arts Today and our Contributions

There have been several reviews released in related fields over the past few decades due to the significance of big data and blockchain. One of the very first studies on blockchain was conducted to examine the security and privacy concerns with blockchain applications [6-9]. The survey presented game theory applications for blockchain systems, such as games for managing mining operations, games for addressing security and privacy concerns, and games for blockchain applications [10]. The use of blockchain in various technologies has been studied through several surveys. For instance, the potential use of blockchain for IoT systems in addition to related issues (including security, scalability, and data management) discussed [11-14]. The incorporation of blockchain with 5G and edge computing platforms was investigated [15-20]. Reviews of blockchain's uses and potential for smart

*Address correspondence to this author at the Government Polytechnic Bighapur Unnao, Board of Technical Education, Uttar Pradesh, India;
E-mail: vikas.yadav.cs@gmail.com

2352-0965/24 $65.00+.00

grid networks were conducted for the polls [17, 18]. Recently, surveyed the uses of blockchain in the present healthcare system [19]. A thorough analysis of the interoperability of blockchain technology is provided [20]. A thorough analysis of several experimental techniques, analytical models, and theoretical blockchain modelling was published in other intriguing studies [21, 22].

The concepts and applications of big data analytics have also been the subject of numerous surveys and given a survey of methods and tools for large data management [23]. Representative surveys may be found and big data analytics have also been used in intelligent transportation and smart grid systems [24, 25]. In conclusion, blockchain may enhance big data by improving data integrity, security, and privacy, enabling real-time data analytics, enhancing data sharing, and enhancing big data quality.

The blockchain application in developing private and secure drone big data solutions was highlighted in a recent article [26]. This review also put forth a secure plan for big data from drones that were built on a four-layer architecture comprising layers for users, data, clouds, and blockchain. However, this study only discusses blockchain drones for big data and is unable to go into great detail about the function of blockchain for big data. Similar to our paper, other surveys address the interaction between big data and blockchain but they only offer cursory overviews rather than an in-depth survey [12, 16, 27].

We specifically aim to present a thorough assessment of blockchain for big data, covering foundational information, contemporary methodologies, prospects, research problems, issues, and future perspectives. This review's main objective is to examine the most recent research and conduct an assessment of how well blockchain fits into applications of big data.

We demonstrate how blockchain holds enormous promise for advancing big data analytics, including the administration of data sharing, improved security and privacy, control over unclean data, and improved data quality.

2. AN OVERVIEW OF BLOCKCHAIN AND BIG DATA

This section provides background information on blockchain, big data, and the drivers behind their integration.

2.1. Blockchain

To put it simply, the blockchain is a "distributed ledger" of comparable data items, or "blocks," that are linked together. Each block of this ledger is connected through cryptography, and it is continuously growing. A blockchain's stored data is a shared database that is regularly updated. The fact that this database is not stored or gathered in one place is one of the powerful advantages of the blockchain that contributes to its high level of security. Because there are a few duplicate copies of the ledger and a lot of personal computers (PCs) on the network, it would require a lot of processing power to get into the network and corrupt the records. Although it is theoretically conceivable to quantify the amount of computing power required to carry out a hack, this is practically impossible [28].

The blockchain's "blocks" are made up of PC code that contains information and can be altered to represent anything from money to a birth certificate. Through secure encryption, every "block" is linked to other blocks, creating the "chain." This "chain" can be compared to any representation of a conventional database because it combines data. When viewed as a whole, the blockchain can be compared to an accounting ledger, which contains a record of transactions. The blockchain's programming relies on a "distributed ledger" rather than keeping track of transactions on a local ledger. In practical terms, the distributed ledger is also known as the blockchain. A synchronised database that is stored simultaneously on thousands of PCs all around the world is the distributed ledger. The ledger is widely distributed and has had numerous undetectable copies at various times [29]. To grasp the impact of this invention, one must understand distributed ledger rules.

Furthermore, as stated, "Blockchains are distributed digital ledgers that record, authenticate, and prevent duplication of transactions using algorithms or an exact set of instructions without the need for a central authority," a more thorough definition of the blockchain is provided. This definition provides a more in-depth look at the blockchain by highlighting its key characteristics and reiterating that it is both a distributed technology and a decentralised system.

2.1.1. Constructional Elements of Block Chain Frameworks

Frameworks can be built with the help of blockchain technology [30]. The primary building blocks of blockchain frameworks are depicted in Fig. (**1**).

2.1.2. Data source

A database is a type of information structure used to store data. By fusing information from several databases, it uses a social model to provide increasingly composite techniques for querying and information collection. Using a database management system (DBMS), the stored data can be sorted. The database is one of the core elements of the blockchain. Considering that this is unquestionably not your ordinary database with lines and parts, Instead, it is a record of all previous transactions for each participant client in the blockchain network. High-throughput, decentralised control, changeless information storage, and implicit security are characteristics of this type of database.

2.1.3. Miner

A miner is a central processing unit (CPU) that tries to solve a computationally challenging numerical problem in search of a new block [31]. To try to find the answer to the numerical puzzle, the miners can either work alone or in groups called pools. By informing all users of the blockchain network of new connections, the process of discovering another block is started. As the minor who discovers the block purchases the expenditures or charges of all the transactions in that block, in a few instances, the transactions with the highest expenses are first selected from minors.

2.1.4. Mechanism for Consensus

Trust frameworks involve using the strength of the scheme to monitor transactions. These are the foundations of

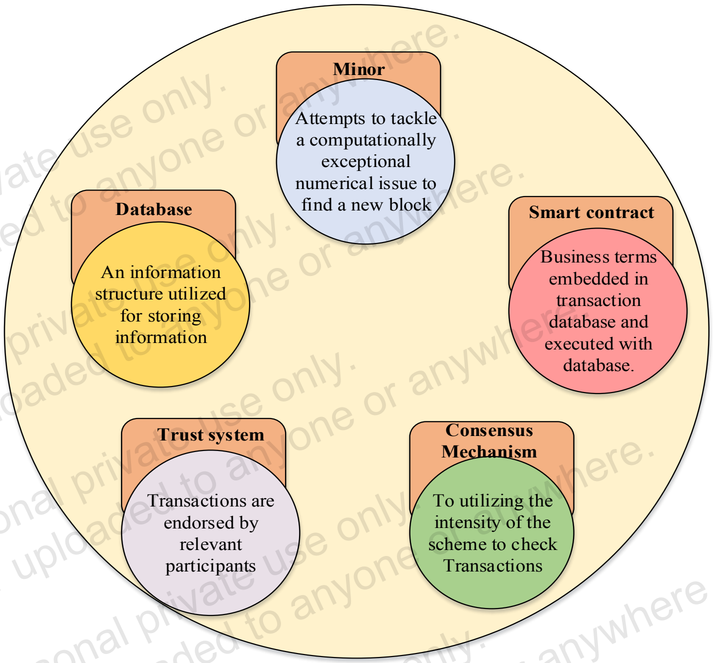
Minor
Attempts to tackle a computationally exceptional numerical issue to find a new block
Database
An information structure utilized for storing information
Smart contract
Business terms embedded in transaction database and executed with database.
Trust system
Transactions are endorsed by relevant participants
Consensus Mechanism
To utilizing the intensity of the scheme to check Transactions

Fig. (1). Blockchain building blocks [7].

the blockchain network scheme to monitor transactions. These are the foundations of blockchain networks. They provide the basis of blockchain applications, and since not all approval is obtained by consensus, we accept that the term "trust framework" is preferred over that of "consensus mechanism." This fundamental tenet of trust guides the shared goals and enthusiasm in blockchain architecture. The trust architecture changes with each new player in the blockchain market, creating variants that are specific to the stated blockchain use cases [32]. The three pillars of blockchain innovation are trust, exchange, and possession. The trust framework facilitates transactions for trade between cooperating organisations and exchanges between organisations. The ideal trust framework for explicit use cases, such as P2P and sharing economy models with several models, still need to be characterised.

2.1.5. The Smart Contract

Smart contracts are PC protocols that support, validate, or carry out the arrangement or execution of an agreement or that obviate the need for a formal declaration. Most smart contracts also have a user interface and frequently mimic the logic of authoritative provisions. Many different types of legally enforceable clauses can be made partially or entirely self-executing, self-authorising, or both, according to smart contract proponents. Smart contracts aim to provide security that is superior to that conventional smart contracts and to reduce other contracting-related exchange expenses.

2.2. BIG DATA

The background detail on big data's primary properties, uncertainty, and the analytics techniques that address big data's intrinsic uncertainty is covered in this part.

Big data was mentioned to be the upcoming frontier for innovation, competition and productivity in May 2011. Over 3.7 billion individuals used the Internet in 2018-a 7.5% increase from 2016 [33-37]. The quantity of data formed globally increased from 1 zettabyte (ZB) in 2010 to 7 ZB in 2014 [38]. Variety, Velocity, and Volume (the three Vs) were found to recognise the developing features of big data in 2001 [39]. The four Vs-Value, Velocity, Variety, and Volume-were also used by IDC to define big data in 2011. Veracity was added as the fifth attribute of big data in 2012 [40-42]. Although there are numerous other V's, we focus on the five traits of large data, which are shown in Fig. (**2**).

Volume characterises the scope and extent of a database and alludes to the huge quantity of information generated per second. Because the amount of data and its type might affect how it is defined, it is difficult to establish a common criterion for large data volumes [43]. Currently, exabyte (EB) or ZB-sized datasets are typically regarded as big data, although problems remain for small-sized datasets [35, 44]. For example, every hour, Walmart collects 2.5 PB from around a million shoppers [45]. When attempting to study and comprehend the data and information at scale, many of the currently used data analysis tools may fail since they aren't intended for massive datasets [35, 45].

Variety mention to the various data types that are likely to be found in a database, such as multimedia and text information, which is random and challenging to analyse, but structured data, such as that kept in a relational database, is typically ordered and can be quickly sorted out. The semi-structured data contains tags to divide data items (such as in NoSQL databases), but it is up to the database user to enforce this structure [43, 46]. The presentation of mixed types of data, switching between different data types (like from unordered to ordered data), and conversion to the database's key structure during run time may cause uncertainty. Traditional big data analytics algorithms confront difficulties when managing multi-modal, imperfect, and noisy data from

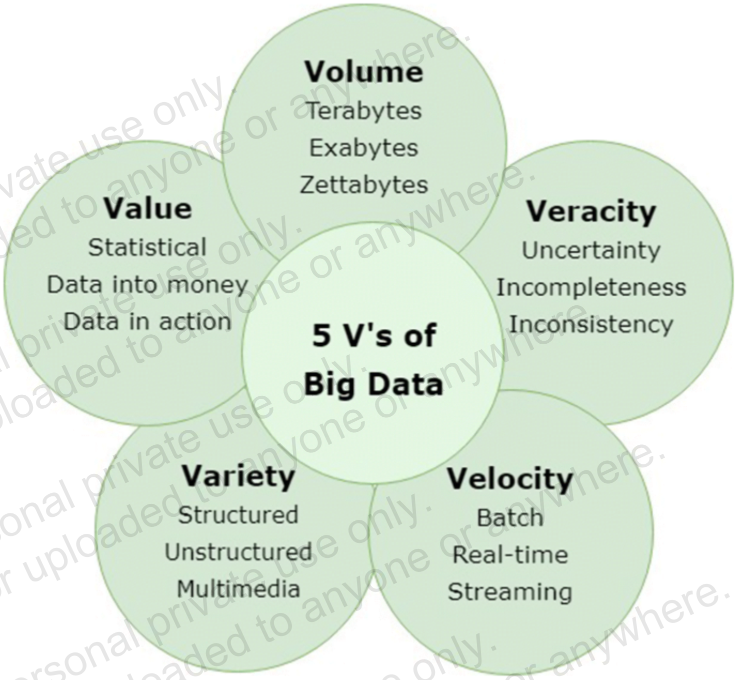

Fig. (2). Common big data characteristics [36].

the perspective of variety. Such techniques (such as data mining methods) can't take well-formatted input data [34]. Although the dataset itself can also be affected by uncertainty, this paper targets how uncertainty affects big data analysis.

It might be hard to analyse semi-structured and unstructured data effectively since they originate from sources with a diversity of data kinds and representations. Real-world datasets, for instance, are adversely affected by noisy, conflicting, and insufficient data. To eliminate noise from data, different pre-processing methods, such as data manipulation, data integration, and data cleaning, are utilised. Techniques for cleaning data address issues with data uncertainty and quality created by variance in large data sets (*e.g.*, inconsistent and noisy data). These methods for minimising distracting elements from the analysis method can considerably improve the effectiveness of the analysis of the data.

Velocity is defined as the rate of data processing (expressed in respect of batch, streaming, real-time, or near-real time), show up the demand that the rate of processing of the data match with the production rate of the data [8]. For instance, IoT gadgets constantly produce a lot of sensor information. A pacemaker which reports emergencies to a facility or doctor is an example of a gadget that monitors medical information and may cause patient damage or death if data processing and delivering the results to physicians is delayed [40]. Similar to how real-time OS enforces rigorous execution timing rules on devices in the cyber-physical world, big data applications may run into issues when their data is not supplied on time.

The level of data quality is represented by veracity (*e.g.*, imprecise or uncertain data). For instance, according to IBM's estimate, impoverished data standards cost the US financial system $3.1 trillion each year [41]. Veracity is categorized into three types: undefinable, terrible, and good, terrible, since data may be noisy, inconsistent, unclear, or incomplete. Trust and accuracy are difficult to start in big data analysis because of the amount and variety of data sources. For example, a worker may occasionally use an identical account to post personal ideas on Twitter while still utilizing it to publish official company information, making it hard for any techniques built to leverage the Twitter database to function properly.

Far from the preceding V's, which focused on indicating the problems in big data, the value represents the utility and context of data for a conclusion. For example, Amazon, Google, and Facebook have applied analysis to enhance the quality of big data in their products. To give product suggestions and boost customer satisfaction and sales, Amazon inspects a huge database of customers and their purchases. To improve its position in Google Maps, Google gathers location data from Android customers. Facebook keeps an eye on user interest to produce friend-tailored ads and recommendations. These organisations have grown remarkably as an outcome of data analysis, analysing huge amounts of information and gaining and reclaiming insightful knowledge that helps them make wiser business decisions [47].

2.3. Motivations for Integrating Big Data with Blockchain

The following is a discussion of the reasons for merging blockchain with big data.

Enhancing Big Data Privacy and Security: As the quantity of devices attached to the Internet rises daily, so does the volume of data being kept on external sites like the cloud. This produces new problems, like data attacks or breaches from interested outsiders [48]. Because of the reality that big data is not stored within a company's web area, existing privacy solutions such as firewalls aren't able to tackle this is-

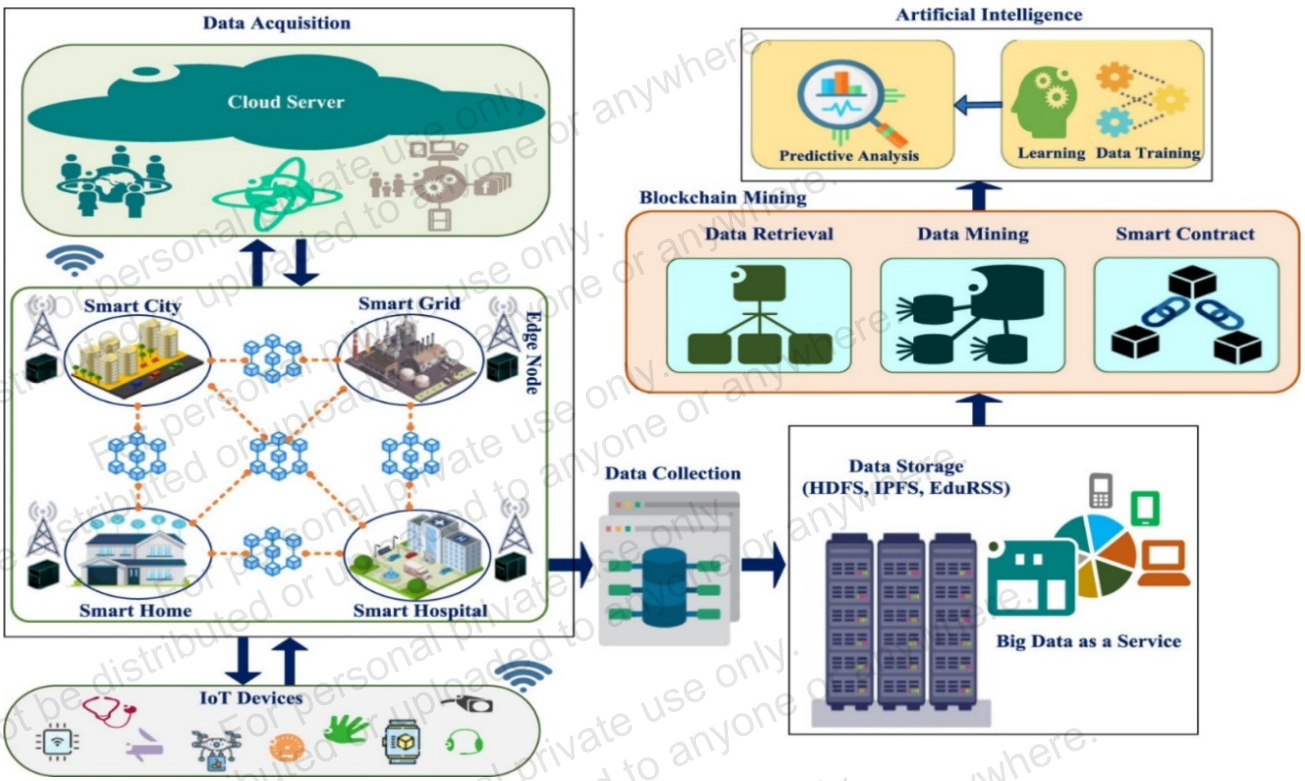

Fig. (3). Big data environment: blockchain services [84]. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

sue. This is by enterprises that don't have authority over the information. Blockchain, a warehouse of huge amounts of information, can resolve this issue. It is exceedingly challenging for any unwanted access to the data due to the data's encrypted and decentralised storage in the blockchain network.

Enhancing Data Integrity: There is a possibility that somebody will interfere with big data information to affect the analysis forecast. It is virtually not possible to interfere with the information acquired through the blockchain connection thanks to the immutability feature of the blockchain. It is nearly possible to render the information in blockchain since doing so will require replacing it in not less than half of the network's nodes. Moreover, the blockchain's stability characteristic guarantees the correctness of the information stored there.

Real-time data analytics is now possible thanks to the blockchain's ability to store each transaction. Financial and banking organisations may resolve cross-border business, plus sizable sums, by enclosing real-time credit to a blockchain combination with big data analysis. Financial organisations may also look at data interchange in actual time, which gives them the ability to respond immediately to block transactions, for example.

Improvement in Data Sharing: By combining blockchain technology with big data, business providers can share information with stakeholders while minimising the risk of information leakage. Moreover, since each observation is recorded on the given blockchain, the inspection of the massive amount of data stored from different origins does not require to be redone.

Enhancement of Big Data Quality: Data technologists consume a lot of time integrating data because different origins use various formats when collecting data. Since the data is structured and comprehensive when stored on a blockchain, its quality can be improved. As a result, information scientists can use high-quality information to produce more precise forecasts in real-time.

3. BLOCKCHAIN APPLICATIONS FOR BIG DATA

Blockchain applications for big data are covered in this area, as shown in Fig. (**3**).

3.1. Using Blockchain to Acquire Huge data Securely

Big data applications are becoming more popular nowadays, but they still confront significant security challenges. In the process of processing data, gathering data is a crucial step. Untrustworthy data sources and communication channels make data collection vulnerable to different dangers and hostile attacks. As a result, safe data collection techniques are essential for many data applications. To date, some research projects have been conducted to offer secure data collection. For mobile crowd sensing (MCS), for instance, a safe massive data collection method based on blockchain is presented [49]. As a result of a grouping of cloud servers and MTs, an MCS system is formed. The MCS server records various sensing-based tasks and picks MTs in the region to execute them. Finite energy resources in MTs, the variety of sensing equipment, and securing the sharing of data across MTs are the biggest problems with data acquisition.

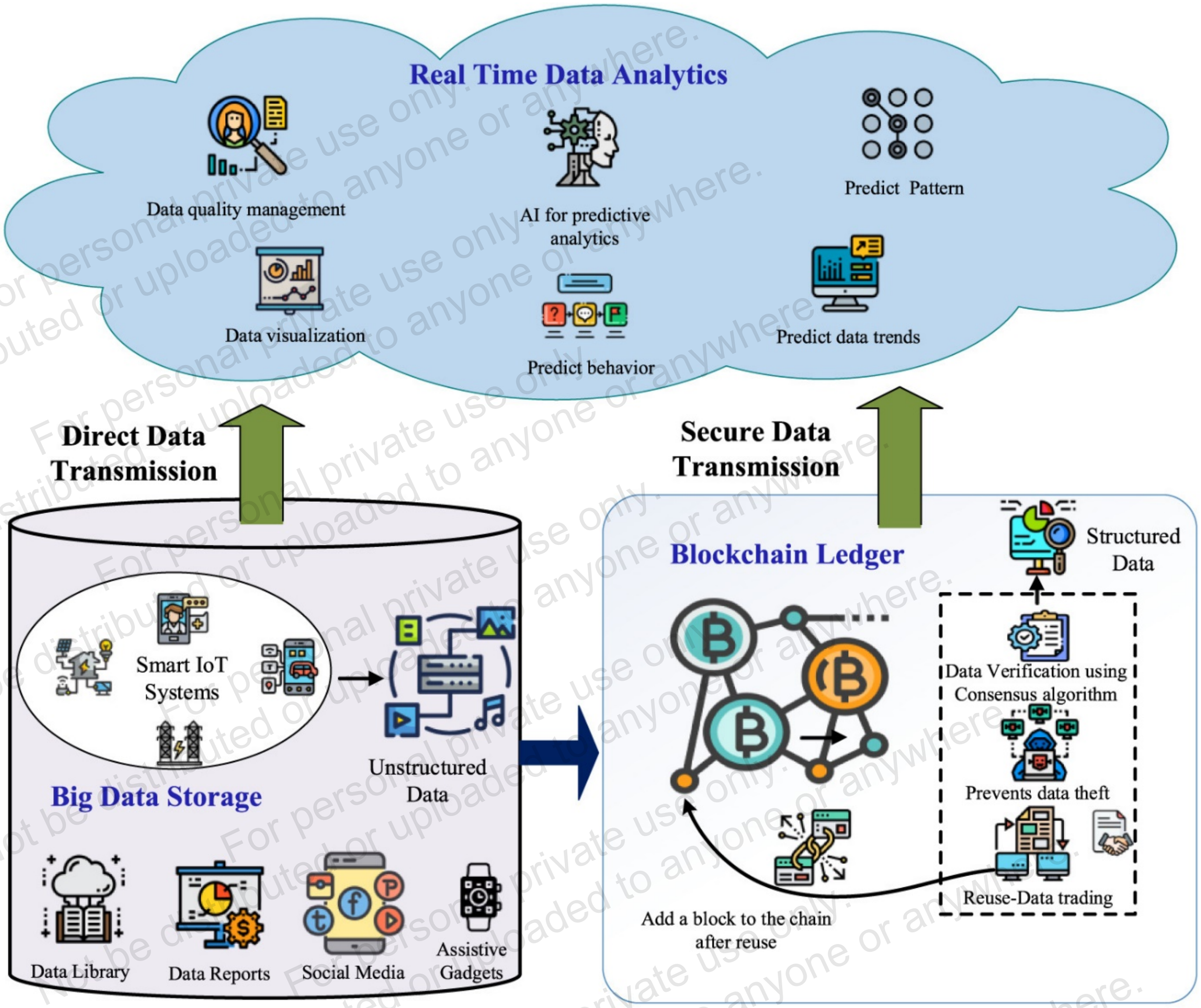

Fig. (4). Secured blockchain services: Big data environment [74]. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

3.2. Blockchain Technology for Secure large-scale Data Sharing and Transmission

The application of blockchain to handle massive data transmissions and sharing is the subject of research; an example model is shown in Fig. (**4**). To share trustworthy data at the edge node, for instance, presents in unique blockchain approach [50]. The authors focus on leveraging proof-of-collaboration to reduce the computational load at the edge nodes throughout this procedure. Finally, hollow blocks and explicit transactions were recommended by scientists as a method to enhance the network's efficacy in the suggested design. The authors propose a smart contract-based blockchain system for licence-free spectrum access that enables safe non-real-time data transport [51].

3.3. Blockchain as a Secure file System

Presently, the most common usage of electronic data and information systems is in healthcare. Every day, an enormous amount of data is generated, including medical photos, records, diagnostic reports, *etc.* When shared with other medical academies, electronic medical information can have an impact on how patients are treated and their experiences. The patient's privacy is threatened if the shared medical information is unlawfully used. To regulate access to medical data, some controls should be put in place.

These kinds of security issues can be solved by using the blockchain in conjunction with IPFS. A decentralised storage system called IPFS was formed to deal with the issue of file duplication. Before being stored in the cloud, the medical information is encrypted using an attribute-based method. Users' traits and policies are linked together by their private keys and ciphertexts, respectively. Any user who has a personal key that complies with the access policy specified in the ciphertext can decrypt the ciphertext. Blockchain is also used to document the process of storing and retrieving data. To prove the validity of user verification, the hash function of medically acquired data is kept on the blockchain. A contact point of disaster is avoided, and privacy for file storage is helped by the decentralised blockchain architecture [52].

3.4. Training with Secure Data using Blockchain

The volume of data has expanded in many scenarios as a result of the growth of cloud and edge computing. Several machine learning and deep learning techniques are used for efficient data analysis. The most commonly used ML algorithm is SVM, which is known for its accuracy and effectiveness. Data and information for auto social platforms are acquired from different sources, including vehicle management corporations, automakers, and social network providers. The characteristics of information from different data origins typically change. Due to different sources, entities face the problem of data with insufficient characteristics while training an SVM technique.

A blockchain-based SVM training approach that protects privacy has been presented. This approach uses a homomorphic cryptosystem and blockchain consortium to construct a secured teaching programme without the aid of a reliable third party. Teaching procedures are carried out locally over original information, and the blockchain consortium and homomorphic cryptosystem safeguard interactions between the parties.

3.5. Blockchain in AI Algorithms for Safe data Learning

The development of AI approaches was facilitated by the vast creation and production of data from sensing, IOT, social media, and the web. For data analysis, the data may be used with ML and DL techniques. These techniques are based on a centralised server for training, which compromises data security. Accordingly, the interpretation drawn from AI is incorrect and dangerous. Centralised AI, when combined with blockchain, was formed to focus on this problem. By merging AI and blockchain, different restrictions are reduced. Data and information are benefitted from AI applications to understand, collect data, and make predictions. When incoming data is collected from different trusted data repositories, esteemed, and safe, these methods do better.

Blockchain suggests different secure surroundings for transactions and data collection through different distributed ledger techniques. Here, information remaining in the blockchain with good integrity and resilience can't be altered [53]. The results are unquestionable and reliable when smart constructors are applied to the learning process in AI techniques to gain insights and conclusions. Accordingly, combining AI and blockchain can produce an environment that is secure, decentralised, and immutable for learning extremely complex methods. Significant improvement is made achievable by a combined system in various fields, including trading, finance, medicine, and banking.

3.6. Blockchain for Large-scale data Processing Privacy Protection

Nowadays, preserving privacy has become a top preference due to ascending rise in data production. In this era of big data, information is continuously acquired and reviewed to produce progress in innovation and business. Large organisations and businesses use the gathered data to enhance decision-making, better serve customers and forecast future trends. As a result, data has recently become a valuable resource. Big data is frequently used in smart cities to ensure the quality of the air and water, monitor city traffic, and perform comprehensive maintenance. To protect privacy in ITS for in-car navigation systems in smart cities, a blockchain-based solution is suggested [54]. Users have access to different security mechanisms, including the capacity to share data and information about location, speed, and other aspects [54].

3.7. Blockchain for Huge data Storage and Privacy protection

User security is being targeted by the era of big data in different digital issues. By obtaining, analysing, and managing the enormous volume of user personal data, third-party organisations benefit from managing user data. Without the consumers' knowledge, these third-party services are vulnerable to security lapses and data misuse. Blockchain offers several solutions to the problems user data faces. Blockchain-using businesses don't raise security issues, and users have few options for managing personal data. The specifics of each transaction's disclosure of personal information, including when, by whom, which, and what Blockchain privacy-preserving solutions are becoming available, based on crypto-privacy techniques, to provide users with control over their data while enabling anonymity during digital transactions [55]. To ensure the safe exchange of drone big data, a blockchain-based privacy protection mechanism was created [26]. The proposed system uses a cryptosystem based on number theory to encrypt the data, resulting in cheap processing costs for key generation, decryption, and encryption.

4. BLOCKCHAIN FOR BIG DATA INITIATIVES AND APPLICATIONS

With its numerous uses across a wide range of industries, blockchain technology has experienced tremendous growth. The use of blockchain in conjunction with big data can be seen in the data management and data analytics sectors. Blockchain technologies, which are secure and widely used, are used to store critical data in instances of data management. Additionally, it can detect data tampering and assess the validity of the data. Blockchain is used in data analytics applications to examine trading trends and forecast future clients, diseases, or business partners.

4.1. Blockchain Big Data in Smart Cities

Rapid urbanisation has resulted in the development of smart cities, which call for efficient and thoughtful solutions for the improvement of their administration, environment, energy, and transportation systems. Such infrastructure solutions are necessary for living in smart cities. Inadequate security, dependability, maintenance, flexibility, and prices, however, are the source of a wide range of issues. Blockchain technology satisfies requirements for IoT device maintenance, energy efficiency, space, and transparency.

The research describes the usage of Hash functions, encryption, a consensus protocol, and a Merkle root to guarantee a transaction is not tampered [13]. The use of a Merkle tree makes it even safer for carrying out smooth transactions.

Third-party auditors (TPAs) are big data-based auditing tools that have grown significantly in popularity in recent years. The centralised TPA frameworks are vulnerable to security problems in the cloud environment. Decentralised TPAs for smart cities with improved security and dependability have been developed using blockchain technologies. The complete audit history is tracked by this system, known as the data auditing blockchain (DAB). A study [56] presents a blockchain-based infrastructure for safe spatiotemporal smart contract services. In smart megacities, the framework offers a shared economy based on sustainable IoT. The massive production of big data has prompted the need to gather, analyse, and use it to automatically predict any risky or extraordinary events. To support the blockchain and other offline processes, the architecture includes fog nodes that are installed locally and device-to-device (D2D) communication mechanisms. The MEC tower controls the load well by hosting the cloudlet apps, relevant databases, data storage clients, and blockchain nodes. To do complex analyses like digital forensics, emotion extraction, and other tasks, the data from the blockchain, IoT, and social networks is finally put into the AI engine.

The processing of transactions on the blockchain is, in fact, far faster than it is on conventional financial systems. Money may be transferred to foreign workers and international tourists more quickly thanks to this technology. Faster settlement of contractual accounts is also made possible by smart contracts that run on the blockchain [57].

4.2. Blockchain Big Data in Smart Healthcare

A dramatic increase in the generation of medical data has resulted from recent developments in the healthcare industry. As an illustration, a common tool for treating patients who are elderly in particular is the remote patient monitoring system. However, despite their many advantages, these technologies have some problems that could seriously violate both data security and privacy. The study [58] presents a system that addresses these problems. Thus, the framework offers patients access to and control over their medical records with enhanced privacy and security.

A secure smart healthcare system employing blockchain is suggested [59]. Furthermore, this kind of data is kept in scattered formats rather than centralised cloud storage systems, allowing only those with patient consent to access it. Wireless sensor networks (WSN) keep all the entities in this framework connected to conduct seamless yet secure communication, including IoT and EHRs.

Proposes an Ethereum protocol-based private blockchain framework for connecting smart devices and sensors, that was proposed [60]. These "smart" gadgets, also known as "smart contracts," log every event on the blockchain. As a result, these smart contracts assist in the real-time monitoring of patients and also alert healthcare professionals when interventions are necessary. The link in the blockchain ensures authentication and removes any chance of EHR data tampering, which secures the recorded records.

4.3. Blockchain Big Data in Smart Transportation

Due to its potential to improve ITS, blockchain technology has grown quickly. With the best possible utilisation of pertinent infrastructure and resources, such innovations can be exploited to build safe, dependable, and autonomous ITS ecosystems. As an illustration, provided a seven-layer conceptual model for ITS that may be used to describe the architecture and key elements of a blockchain-based system [61, 62]. The physical layer is where the various ITS-related vehicles, equipment, and assets are kept. The blocks of data and related hashing, encryption, and Merkle tree methods are provided by the data layer. The smart contracts and algorithms that make up the contract layer turn on the blockchain's data storage function.

In vehicle communication systems, security is frequently a top priority. For achieving network security, the study developed a secure key management structure [63]. The study makes use of security managers, who record information on when vehicles leave the guarded area and package the blocks that contain the keys.

Blockchain technologies, notably those used in car navigation, not only provide security but also significantly contribute to the privacy protection of ITS. The client app, which is installed on the user's smartphone, and the main app, which is installed on the server, are the two main applications that the system uses. It is believed that the server and smartphone have been set up securely and that simple and complicated security measures, depending on the user type for data exchange, are utilised. To maximise the utilisation of computational resources and minimise network delays and overheads, all clients in the network are grouped into clusters according to their location. Users of the system can choose their privacy policies, which are then immediately applied to the client application that offers precise transit routes.

The different issues relating to auto insurance that blockchain may be able to resolve The insurers will be able to trace their claims easily with the use of this technology by searching the reliable ledger. By deploying mobile sensors to collect streaming IoT data, this data was gathered. This innovative approach established transparent insurance and encouraged responsible driving to get insurance rewards.

4.4. Blockchain Big Data in the Smart Grid

The efficiency of procedures and operations in the energy sector can be greatly increased with the use of blockchain technologies. By accelerating the development of IoT platforms and digital applications, big data and blockchain technology have the potential to advance P2P energy trading and decentralised services. The control and management of decentralised energy systems and microgrids present problems that the blockchain architecture is capable of resolving [64].

The blockchain architecture implements an encryption technique for multidimensional data capture and reception [65]. The control centre, the grid operator, and the grid supplier all receive predefined blocks as part of the regulation process, and after the decryption step, they receive plaintext. The appropriate receivers first analyzed the multidimensional data. Control pieces are produced as a result of this. The security and data integrity aspects are handled by these control components, which lowers transmission costs.

4.5. Blockchain Big Data Initiatives

It's also crucial to keep in mind that the usage of large, diverse, and rapidly changing data is a key component of the big data field of data science. But to draw the right conclusions, the data must be examined to spot intriguing trends, correlations, and patterns. But in this process, data security is a crucial requirement. A distributed ledger system like blockchain, which records transactions that cannot be altered or falsified, helps in this area.

4.5.1. Storj

Storj is an end-to-end decentralised storage project that makes use of unused hardware and bandwidth capacity to enable peer-to-peer storage contract authentication between service providers and customers [66]. At the client's end of the process, files are encrypted before being divided into "shards" for convenience. For the farmer's side backups, these shards are later saved three times. In contrast to more established centralised cloud services, the customer only has access to the data, increasing security. Renters can monitor the farmer's files and cover the cost of this storage system's upkeep using the Storj money. Without any additional charges for setup or user requirements, the renters just pay for the space that is used.

4.5.2. Omnilytics

Insights for the sales, marketing, and merchandising industries are provided by the blockchain-based big data analytics platform Omnilytics [67]. Data from diverse industries are combined using a variety of technologies, including blockchain, Big data, Machine Learning, Artificial Intelligence and others.

4.5.3. Provenance

A blockchain platform called Provenance is primarily employed in supply chain management, where it assists in gathering crucial product information and sharing it in a reliable, safe, and open way [68-84]. Customers of the protocol can access information on the items' place of origin, their travel through multiple supply chain nodes, their quality, and their environmental impact.

4.5.4. FileCoin

With the help of a decentralised storage network, traders would be able to purchase and sell storage on a free market. Using the cryptocurrency FileCoin, customers can rent storage on machines with extra storage space. Customers pay in cryptocurrency for the sharing or retrieval of data, and miners make Bitcoin by providing storage and data services. A proof-of-space-time (PoST) is a document that miners must provide to the network when they mine a specific block to verify whether a storage provider is carrying out the necessary duties to store outsourced data for the allotted amount of time.

4.6. Trading Platforms Using Cryptocurrency

As was previously said, Bitcoin was the first asset cryptocurrency and the first decentralised payment system that operated without a central bank or system administrator. Transactions take place directly between users on a peer-to-peer network, without the need for any middlemen. Blockchain is a distributed ledger where transactions are recorded and validated by network nodes using cryptography. Other cryptocurrencies, like Ethereum, are used in addition to Bitcoin. Ethereum makes it possible for programmers to use the Ethereum network to pay for services and transaction fees. Other cryptocurrencies that assist in converting currencies or units of value, such as US dollars, euros, British pounds, and sterling, include Bitcoin cash and Ripple. Cryptocurrencies like Dash and Bitcoin also allow for quick payments from regular people [64].

4.7. Cross-border Payments

Cross-border payments are the regional and international transfers of money between different nations and regions. These transactions take place as a result of international trade, the settlement of foreign debts and claims, and other settlement tools or payment systems. The high costs, high fund requirements, and security flaws of traditional banking systems frequently prevent them from accomplishing their goals. Such delays in cross-border payment systems are reduced by the use of blockchain. For instance, using the SWIFT and VISA blockchains to start cross-border transactions.

4.8. Blockchain-Enabled Industrial Internet of Things Technology

In IIoT-based applications, the IIoT system plays a primary role that may produce smart connected assets, enable IIoT, and connect operations with the following accomplishments: application development, big data analysis, and connectivity. To resolve the above issues, the IIoT system would consider the following conditions: (1) Cyber security; (2) Data islands (3) ageing workforce; (4) integration of technologies; and (5) visibility of assets For now, the IIoT may also notably affect users and customers behaviours. Most common industrial facilities, such as vehicular ad-hoc networks (VANETs), smart-grid IoT, micro-grids, etc., are not able to connect to IIoT because they have internal intelligence, and need interfaces to connect with IIoT. Instead, the operators present in IIoT are accommodated with new techniques like augmented reality (AR), which can produce better forecasting and interacting method behaviours, which consequently become easier to operate and simplify at enhanced efficiency [85].

4.9. Distributed Blockchain-Based Trusted Multidomain Collaboration for Mobile Edge Computing in 5G and Beyond

Mobile edge computing (MEC) delivers power to the edge of networks and combines internet service and mobile access networks in 5G and beyond. With the routine enhancement of services, security protection is very important in a different MEC system for multi-server cooperation. Moreover, most of the present techniques only consider the security of users or applications, other than the security of the network topology. For the sake of topology privacy and security protection, this article utilises blockchain to construct different MEC systems and takes on the role of a bloom filter as a bearer for multi-domain cooperative routing agreements without revealing topology security. Blockchain is used to implement multiplex mutual trust networking and collaborative routing verification through the membership service and consensus mechanism [86].

4.10. Blockchain-Based Hierarchical Trust Networking for Joint Cloud

The Internet of Things (IoT) is gradually maturing and has already entered our daily lives, interconnecting more machines and making communication more convenient and intelligent. Massive IoT devices produce innumerable pieces of data that need to be analysed in joint cloud computation (JointCloud) with diversified services. However, due to the weak security of IoT devices, the existing JointCloud architecture hardly provides a secure, trusted trading environment for users, which severely affects the applications in the IoT network. By introducing the credit bonus-penalty strategy (CBPS), HTJC can solve the trust problem and provide users with a secure, trusted trading environment [87].

5. RESEARCH ISSUES

Here, we examine several pressing issues and suggest many lines of inquiry for large data blockchain research.

5.1. Blockchain Security

IoT applications also contain data transfers from heterogeneous devices. Data security and privacy, which need greater processing power (about 50% more) and where hostile users attempt to falsify the block information, are some of the significant services provided by the blockchain through decentralisation [73]. 51% attack is the name given to this kind of assault. The ecosystem is strengthened by blockchain smart contracts to fend off double-spend assaults [57]. Blockchain's distributed architecture, which distributes each transaction across the entire network, makes fraudulent block transactions more challenging. Even though blockchain and big data are a fantastic match, there is still a security concern around data analysis and big data methods used to handle large data.

Social and economic effects of blockchain security: users who are part of macro-level organisations are subject to severe restrictions on the private blockchain, while users who are part of micro- and Meso-level organisations are subject to restrictions on the public blockchain. Because of these unproven assumptions and constraints, blockchain must adhere to stringent security requirements. Academicians have lost motivation as a result of the lack of knowledge systematisation. Incorporating this into Bitcoin modelling has created several unanswered questions because cryptocurrencies are specifically dealt with as an incentive. The main cause is a disconnect between distributed system security, game theory, and cryptography for security and incentive-based solutions, respectively. The blockchain POS protocol always favours the wealthier nodes, which are then chosen to participate in mining, leading to economic inequalities across the network's various participating nodes. Therefore, for the advantage of all network users, a security evaluation with an economics-based focus must be conducted. A methodical framework that depends on the idea of fairness in node choice, producing opportunities for every node to address this problem, is introduced [74].

5.2. Blockchain Sharding for Huge Data

Blockchain uses sharding as a technique to grow its chain without sacrificing its decentralisation or hash power. The vast data repository is divided into shards by sharding, and each shard is kept in a different place. Blockchain uses a variety of sharding techniques, including network, computational, transactional, and sharding. The sharding strategies can eliminate the storage, bandwidth, and computational power limitations [75]. The management of the time series becomes a difficult undertaking since different processes on different shards have different timelines. Protecting the data and information in shards from different attacks is another major concern. Additionally, using just one shard for the majority of cross-shard transactions causes that shard to overheat. It must be watched over to prevent this data overload at any given shard. Therefore, to prevent hindering efficient big data communications from diverse different platforms, many interoperability difficulties about cross-shard transactions must be taken into consideration.

5.3. Standardization

Blockchain was created to monitor the problem of digital currency (cryptocurrency named Bitcoin). This makes it easy to secure the transfer of digital assets between different banks. Within hours, blockchain technology has completely automated overseas payments made over the Internet, regardless of topographical limitations. The old financial system, in contrast, requires a remarkable amount of time to process any financial communication from anywhere on the globe. But due to interoperability issues, blockchain use has been limited in breadth. These difficulties include not just changes across different cryptocurrencies but changes in various communications as well. Hence, cooperating and integrating with previous systems with blockchain automation is hard. Thus, it harms blockchain's regulatory approval. Incorporating and implementing standardisation to establish ordinary technical law for every industry is one potential remedy for this kind of open organization and analyzed blockchain terminologies and several non-profit organisations' efforts to standardise blockchain [69].

5.4. Big Data Complexity

The rise of the Internet of Things and cloud computing applications has caused a significant data buildup. The massive expansion of data in the information era has been accompanied by an increase in data management studies, like unclean and dirty data and data privacy. The emergence of big data has made data quality management considerably more difficult. Furthermore, the companies must guarantee the integrity of the data origin, data cleanliness, and breach while managing larger and more complicated information. This is because the entire digital modification of all heritage data still poses a difficult problem. Blockchain can guarantee data management security, but huge data management complexity needs to be taken into account at the integration level.

The type of data, traditional analysis paradigms, and ineffective data rectifying tools are the three main problems with big data. Due to large data's inherent complexity, it is difficult to describe and understand it, which raises the computational complexity. Big data advances from heterogenic origins with heterogeneity of patterns and behaviours. Complex types of data, their design, more nuanced connections, and broad quality are a few of the main qualities. Big data mining tasks, including data retrieval, topic analysis, and mining of text (semantics and sentiment extraction), are more difficult than classical data mining [70]. Computational models are created that are inefficient due to ignorance of these traits and domain-based data-rectification methods. For the highest level of abstraction in computational model creation, a good grasp of the characteristics of large data, which is inherently complicated, is essential. In other words, while combining big data with blockchain in future applications, it is important to manage the complexity of the information.

6. FUTURE DIRECTION

The system performs poorly as a result of an unforeseen rise in computing complexity brought on by the combination of sophisticated large data having a blockchain. Therefore, adaptive blockchain architectures ought to be favoured to reduce the burden on computing resources for 5G networks and blockchain connectivity, which can then be used for quicker services. The management for integrating blockchain having big data in the future is presented in this section.

6.1. Big Data-specific Adaptive Blockchain Design

The methodology was created for big data applications that leverage blockchain for access control and are applied in large-scale, real-time cyber-physical social networks [71]. The framework processes the local data dynamically at the edge nodes using fog computing. Data communications that need encryption to manage security are encrypted with a simple symmetric technique. The blockchain is applied to acquire and manage data and information connected to authorising access, including authorization and authentication. Experiments proved that the system is effective and workable, but that managing anonymity takes more time since each authorization takes place on the blockchain. For a better result, the retrieval techniques must be strengthened.

A user may join the chain using the blockchain network, depending on the consensus. PoS and PoW are fundamental consensus approaches that have developed to let anyone join the network based on stake (digital currency) and hash rate, respectively. BigChainDB offers petabytes of storage, subsecond latency (less than a second), and one million writes per second. Additionally, it mandates granting access to platforms that may be used with both private and public blockchains. Similar to this, the Hadoop system hosts the HBasechainDB structure, an ascendable blockchain for big data repositories. The Hadoop Hbase database serves as the foundation for the blockchain's immutability and decentralisation. However, this paradigm is highly suited for the adoption of blockchain by businesses using the Hadoop environment. Big data requires investigating a common adaptive framework supporting many forms of blockchain.

6.2. Blockchain for Large Data based on the Fog, Cloud, and Edge

The cloud's topmost layer, which is found at the peak of a fog-cloud-edge hybrid system, is largely responsible for handling high-level, extremely complex operations that demand a lot of computational and storage power. The volume of data created by connected devices grows with the number of linked devices. Due to their sensitivity to latency, several applications demand speedy responses when processing and transferring data. Several technological problems develop if the responses are not received by the deadline. Fog and edge computing was developed to offer effective services, control vast volumes of information, and lessen computer complexity in response to these problems. To coordinate resource allocation between edge devices and cloud servers, the fog computing layer serves as a middleman. The fog computing layer serves as a bridge between edge devices and cloud servers. Fog computing makes use of powerful computers called fog nodes. The fog nodes are placed at the edge of the network and have a faster rate of data collection, analysis, and processing. As a result, there is less traffic, and the services are provided faster. The fog computing layer would alert the cloud server for assistance when it is necessary to store and analyse huge amounts of data [76].

Instead of sending data to an area gateway, the edge layer calculates data locally on the sensor or end device. The system can easily become overloaded due to its limited capacity for storing and processing. To deal with the problems posed by high computational and storage operations, fog and edge computing are coupled. Fog nodes ensure better services and lower latency for the user because they can manage massive amounts of computing data. The device layer employs blockchain technology to create a decentralised trust model, enhancing the cloud-fog-edge architecture's overall validity. A block is created at the fog computing layer in response to a transaction request from an end device and broadcast to all participants for validation and verification. A block is added to the integrated blockchain once it has been verified by every participant. When the blockchain is kept on a cloud server, the transaction is finally finished [77].

Blockchain for cloud-based big data: A system for improving the present building information modelling (BIM) by fusing tamper-resistant blockchain and mobile cloud with big data sharing was proposed in [78]. BIM collects a remarkable amount of information during the project and uses historical information to make some decisions. For data provenance and auditing in BIM as a cloud-based service, the framework leverages a personal blockchain that has been approved by a trusted centre (computation and outsourced storage). The block size, security, hashing time, and packaging period are applied to calculate the system [84]. The results illustrate that blockchain could help BIM progress forward by solving data quality and security challenges.

Blockchain for large data based on mobile edge computing: A good way to obtain low-latency responses while reducing computer power through local data processing is to combine mobile edge computing (MEC) and blockchain. Data transfers joining a single demander, a single supplier, many providers, and many demanders could take numerous forms in a smart toy domain. The ids are acquired and authenticated for every data interchange with smart contracts (handling accounting). The complex bookkeeping in the market for smart toys is managed by Chaincode, a blockchain-based smart contract. The local client calculations are processed *via* edge-based computing to make sure minimum response times. The framework ensures that members of smart toy organisations may communicate information in a way that is private, scalable, adaptable, and secure. In the same way, introduced an insubstantial RFID-enabled recognition system based on blockchain for supply chains with 5G MEC [72].

6.3. Blockchain-based Big Data and Software-defined Networking

By improving network authorization and producing a quicker response to changing requirements, software-defined networking (SDN) makes the network agile and flexible. Both techniques gain from the seamless combination of SDN and big data [79]. Big data might help with many issues, including resource efficiency, data scheduling, transmission, cloud processing, and data delivery. Big data could also be helpful for SDN-managed interactions, security flaws, and traffic data inside and across data-centric networks. Even though they complement each other very well, there are a few unresolved issues. The SDN controller occasionally suffers from overloaded queries that impair its performance, rendering it unable to handle increasingly extensive large data entries, providing a scaling difficulty and creating a single point of failure. Unintelligent network switches with limited resources have the potential to overwhelm the SDN controller with raw data packets, adding to the burden of processing. It's crucial to keep in mind that the creation of large data applications cannot be done in a high-level programming environment. To be more specific, SDN has greater security flaws. Big data and SDN could complement one another more effectively when combined with blockchain and faster 5G, which makes them irreversible and scalable. While 5G allows for quicker processing, blockchain could help SDN's scalability and security problems.

6.4. Integration of Big data and Blockchain in Industry 5.0

The 5th industrial revolution (Industry 5.0), which will be necessary for balanced innovation in the digital age, will only require human labour on manufacturing floors to interact with collaborative robots and AI (COBOTS). The cobots would be tasked with labour-based tasks, while creativity and human intelligence would be utilised for manufacturing. Mass personalization to meet the expectations of clients for customization and waste reduction through the bio-economy are the main priorities of Industry 5.0 [80]. In Industry 5.0 standards, blockchain will offer a safe transfer of data for large data aggregation from intelligent tools. The blockchain will be used by Industry 5.0 to safeguard massive amounts of data and harness extreme automation. Digital twins could be developed to manage a digital model of the actual world and ensure information transparency as the number of entities in the hyperconnected network keeps rising dramatically. The digital twin, a feature of Industry 5.0, is the digital presentation of physical devices used by an organisation for its intricate business requirements. This includes people, equipment, and business processes. Industry 5.0's digital twins include a digital representation of the product in its designed, manufactured, and maintained states, as well as real-time data on precise production methods, product configurations, and equipment. Because digital twins are so dependent on information, blockchain could be used to manage huge amounts of data securely for complicated process representation. As a result, Industry 5.0 is confident in its ability to use blockchain as an enabling technology to tackle high automation with Big data and a secured digital twin [81].

6.5. Big Data from Blockchain in Federated Learning

Federated learning (FL) is a collaborative machine learning technique that uses edge devices for decentralised training models. In FL, a central server aggregates locally trained forecast models using on-device intelligence to create a global shared model, which is subsequently, distributed to all edge devices [82]. FL gains knowledge from a large number of diverse local datasets. As a result, algorithms for big data analysis can retrieve relevant data with market trends and potential correlations needed for modelling. Without sharing the data, FL enables several entities to develop a viable shared model. Therefore, a privacy-focused approach is made possible by FL learning solely from the shared prediction model while the user's edge device keeps the private data. Although FL guarantees privacy-preserving data exchanges, the centralised aggregator may still be vulnerable to attacks from hostile middlemen (customers or manufacturers). Because both blockchain and FL are decentralised, they can work together to create a trusted analytical chain in which blockchain records interactions between central servers and edge devices. FL globally trains the scattered data in parallel. This collaboration ensures that interaction logs cannot be altered, making it possible to track down any malicious breaching behaviour by the intermediates [83].

CONCLUSION

A lot of people are interested in using blockchain, a revolutionary ledger technology, to enable large data systems with very effective network management. We have undertaken a cutting-edge analysis of the use of blockchain for big data in this paper. A lot of people are interested in using blockchain, a revolutionary ledger technology, to enable large data systems with very effective network management. A cutting-edge analysis of the use of blockchain-based big data has been completed. The limitations in applications of blockchain-based big data were also discussed, and they will serve to spur further study in this field. To be more precise, the crucial problem with applications of blockchain-based big data is maintaining a balance between security and privacy. Additionally, it was noted from the surveys examined that there has been a delay in the extensive testing, deployment, and commercialization of combined technology, which unquestionably needs more attention. To validate and get the full benefits of this technology, scalable commercial implementation in contemporary smart setups represents a promising future topic.

Big data and blockchain do indeed work well together. Who will be the first to create the most suitable and skilled AI/machine learning framework for application at the peak of blockchain-produced data layers that are immutable, transparent, and distributed in the true question at this point? The business that accomplishes this will make a lot of money.

We might anticipate seeing greater development in the interaction between Big Data analytics and blockchain as this industry's breakthroughs continue. As technology advances and more innovations occur, more real-world use cases for managing data analysis and Big Data will be created and investigated.

LIST OF ABBREVIATIONS

AR	=	Augmented Reality
BIM	=	Building Information Modelling
CBPS	=	Credit Bonus-Penalty Strategy
CPU	=	central Processing Unit
D2D	=	Device-to-Device
DAB	=	Data Auditing Blockchain
DBMS	=	Database Management System
FL	=	Federated Learning
MCS	=	Mobile Crowd Sensing
MEC	=	Mobile Edge Computing
PoST	=	Proof-of-Space-Time
SDN	=	Software-Defined Networking
TPAs	=	Third-party Auditors
VANETs	=	Vehicular ad-Hoc Networks
WSN	=	Wireless Sensor Networks
ZB	=	Zettabyte

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The author(s) declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Big data market worth $229.4 billion by 2025, Available from: https://www.marketsandmarkets.com/PressReleases/big-data.asp

[2] J. Gantz, and D. Reinsel, "Extracting value from chaos", *IDC Iview,* vol. 1142, pp. 1-12, 2011.

[3] S. Pouyanfar, Y. Yang, S.C. Chen, M.L. Shyu, and S.S. Iyengar, "Multimedia big data analytics: A survey", *ACM Comput. Surv.,* vol. 51, no. 1, pp. 1-34, 2019. http://dx.doi.org/10.1145/3150226

[4] J. Wu, K. Ota, M. Dong, J. Li, and H. Wang, "Big data analysis-based security situational awareness for smart grid", *IEEE Trans. Big Data,* vol. 4, no. 3, pp. 408-417, 2018. http://dx.doi.org/10.1109/TBDATA.2016.2616146

[5] X. Xu, X. Zhang, H. Gao, Y. Xue, L. Qi, and W. Dou, "Become: Blockchain-enabled computation offloading for IoT in mobile edge computing", *IEEE Trans. Industr. Inform.,* vol. 16, no. 6, pp. 4187-4195, 2020. http://dx.doi.org/10.1109/TII.2019.2936869

[6] Z. Zheng, S. Xie, H.N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey", *Int. J. Web Grid Serv.,* vol. 14, no. 4, pp. 352-375, 2018. http://dx.doi.org/10.1504/IJWGS.2018.095647

[7] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems", *Future Gener. Comput. Syst.,* vol. 107, pp. 841-853, 2020. http://dx.doi.org/10.1016/j.future.2017.08.020

[8] Q. Feng, D. He, S. Zeadally, M.K. Khan, and N. Kumar, "A survey on privacy protection in blockchain system", *J. Netw. Comput. Appl.,* vol. 126, pp. 45-58, 2019. http://dx.doi.org/10.1016/j.jnca.2018.10.020

[9] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. Nyang, and D. Mohaisen, "Exploring the attack surface of blockchain: A comprehensive survey", *IEEE Commun. Surv. Tutor.,* vol. 22, no. 3, pp. 1977-2008, 2020. http://dx.doi.org/10.1109/COMST.2020.2975999

[10] Z. Liu, N.C. Luong, W. Wang, D. Niyato, P. Wang, Y.C. Liang, and D.I. Kim, "A survey on blockchain: A game theoretical perspective", *IEEE Access,* vol. 7, pp. 47615-47643, 2019. http://dx.doi.org/10.1109/ACCESS.2019.2909924

[11] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, "On blockchain and its integration with IoT. Challenges and opportunities", *Future Gener. Comput. Syst.,* vol. 88, pp. 173-190, 2018. http://dx.doi.org/10.1016/j.future.2018.05.046

[12] H.N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for internet of things: A survey", *IEEE Internet Things J.,* vol. 6, no. 5, pp. 8076-8094, 2019. http://dx.doi.org/10.1109/JIOT.2019.2920987

[13] M.S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, and M.H. Rehmani, "Applications of blockchains in the internet of things: A comprehensive survey", *IEEE Commun. Surv. Tutor.,* vol. 21, no. 2, pp. 1676-1717, 2019. http://dx.doi.org/10.1109/COMST.2018.2886932

[14] L. Da Xu, Y. Lu, and L. Li, "Embedding blockchain technology into IoT for security: A survey", *IEEE Internet Things J.,* vol. 8, no. 13, pp. 10452-10473, 2021.

[15] R. Yang, F.R. Yu, P. Si, Z. Yang, and Y. Zhang, "Integrated blockchain and edge computing systems: A survey, some research issues and challenges", *IEEE Commun. Surv. Tutor.,* vol. 21, no. 2, pp. 1508-1532, 2019. http://dx.doi.org/10.1109/COMST.2019.2894727

[16] D.C. Nguyen, P.N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for 5G and beyond networks: A state of the art survey", *J. Netw. Comput. Appl.,* vol. 166, p. 102693, 2020. http://dx.doi.org/10.1016/j.jnca.2020.102693

[17] P. Zhuang, T. Zamir, and H. Liang, "Blockchain for cybersecurity in smart grid: A comprehensive survey", *IEEE Trans. Industr. Inform.,* vol. 17, no. 1, pp. 3-19, 2021. http://dx.doi.org/10.1109/TII.2020.2998479

[18] M.B. Mollah, J. Zhao, D. Niyato, K.Y. Lam, X. Zhang, A.M.Y.M. Ghias, L.H. Koh, and L. Yang, "Blockchain for future smart grid: A comprehensive survey", *IEEE Internet Things J.,* vol. 8, no. 1, pp. 18-43, 2021. http://dx.doi.org/10.1109/JIOT.2020.2993601

[19] E.J. De Aguiar, B.S. Faiçal, B. Krishnamachari, and J. Ueyama, "A survey of blockchain-based strategies for healthcare", *ACM Comput. Surv.,* vol. 53, no. 2, pp. 1-27, 2021. http://dx.doi.org/10.1145/3376915

[20] R. Belchior, A. Vasconcelos, S. Guerreiro, and M. Correia, "A survey on blockchain interoperability: Past, present, and future trends", *ACM Comput. Surv.,* vol. 54, no. 8, pp. 1-41, 2022. http://dx.doi.org/10.1145/3471140

[21] H. Huang, W. Kong, S. Zhou, Z. Zheng, and S. Guo, "A survey of state-of-the-art on blockchains: Theories, modelings, and tools", *ACM Comput. Surv.,* vol. 54, no. 2, pp. 1-42, 2022. http://dx.doi.org/10.1145/3441692

[22] D. Berdik, S. Otoum, N. Schmidt, D. Porter, and Y. Jararweh, "A survey on blockchain for information systems management and security", *Inf. Process. Manage.,* vol. 58, no. 1, p. 102397, 2021. http://dx.doi.org/10.1016/j.ipm.2020.102397

[23] A. Siddiqa, I.A.T. Hashem, I. Yaqoob, M. Marjani, S. Shamshirband, A. Gani, and F. Nasaruddin, "A survey of big data management: Taxonomy and state-of-the-art", *J. Netw. Comput. Appl.,* vol. 71, pp. 151-166, 2016. http://dx.doi.org/10.1016/j.jnca.2016.04.008

[24] M. Ghorbanian, S.H. Dolatabadi, and P. Siano, "Big data issues in smart grids: A survey", *IEEE Syst. J.,* vol. 13, no. 4, pp. 4158-4168, 2019.
http://dx.doi.org/10.1109/JSYST.2019.2931879

[25] L. Zhu, F.R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey", *IEEE Trans. Intell. Transp. Syst.,* vol. 20, no. 1, pp. 383-398, 2019.
http://dx.doi.org/10.1109/TITS.2018.2815678

[26] Z. Lv, L. Qiao, M.S. Hossain, and B.J. Choi, "Analysis of using blockchain to protect the privacy of drone big data", *IEEE Netw.,* vol. 35, no. 1, pp. 44-49, 2021.
http://dx.doi.org/10.1109/MNET.011.2000154

[27] D.C. Nguyen, P.N. Pathirana, M. Ding, and A. Seneviratne, "Integration of blockchain and cloud of things: Architecture, applications and challenges", *IEEE Commun. Surv. Tutor.,* vol. 22, no. 4, pp. 2521-2549, 2020.
http://dx.doi.org/10.1109/COMST.2020.3020092

[28] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, USA, 2017, pp. 557-564.

[29] D. Vujičić, D. Jagodić, and S. Ranđić, "Blockchain technology, bitcoin, and Ethereum: A brief overview", *2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH), March 21-23, 2018,* East Sarajevo, Bosnia and Herzegovina, 2018.
http://dx.doi.org/10.1109/INFOTEH.2018.8345547

[30] G. Zyskind, and O. Nathan, "Decentralizing privacy: Using blockchain to protect personal data", *2015 IEEE Security and Privacy Workshops,* 2015, pp. 180-184.

[31] K Kotobi, and SG Bilen, "Secure blockchains for dynamic spectrum access: A decentralized database in moving cognitive radio networks enhances security and user access", *IEEE Veh. Technol. Mag.,* vol. 13, no. 1, 2018.

[32] N.O. Nawari, and S. Ravindran, "Blockchain and the built environment: Potentials and limitations", *J. Build. Eng.,* vol. 25, p. 100832, 2019.
http://dx.doi.org/10.1016/j.jobe.2019.100832

[33] B.F. Marr, "How much data do we create every day?", Available from: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#4146a89b60ba

[34] C.W. Tsai, C.F. Lai, H.C. Chao, and A.V. Vasilakos, "Big data analytics: A survey", *J. Big Data,* vol. 2, no. 1, p. 21, 2015.
http://dx.doi.org/10.1186/s40537-015-0030-3 PMID: 26191487

[35] M. Chen, S. Mao, and Y. Liu, "Big data: A survey", *Mob. Netw. Appl.,* vol. 19, no. 2, pp. 171-209, 2014.
http://dx.doi.org/10.1007/s11036-013-0489-0

[36] J Manyika, M Chui, B Brown, J Bughin, R Dobbs, C Roxburgh, and AH Byers, *Big data: the next frontier for innovation, competition, and productivity.* McKinsey Global Institute, 2011.

[37] D. Saidulu, and R. Sasikala, "Machine learning and statistical approaches for Big Data: Issues, challenges and research directions", *Int. J. Appl. Eng. Res.,* vol. 12, no. 21, pp. 11691-11699, 2017.

[38] R.L. Villars, C.W. Olofson, and M. Eastwood, "Big data: What it is and why you should care", *White Paper IDC,* vol. 14, pp. 1-14, 2011.

[39] D. Laney, "3D data management: Controlling data volume, velocity and variety", *META Group Res Note.,* vol. 6, no. 70, p. 1, 2001.

[40] A. Jain, *The 5 Vs of big data. IBM Watson Health Perspectives,* 2017. Available from: https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

[41] "IBM big data and analytics hub. Extracting Business Value from the 4 V's of Big Data", Available from: http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data

[42] D. Snow, "Dwaine Snow's thoughts on databases and data management", In: *Adding a 4th V to BIG Data - Veracity.* 2012.

[43] A. Gandomi, and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics", *Int. J. Inf. Manage.,* vol. 35, no. 2, pp. 137-144, 2015.
http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007

[44] N.R. Vajjhala, K.D. Strang, and Z. Sun, "Statistical modeling and visualizing open big data using a terrorism case study", *3rd international conference on future Internet of things and cloud (FiCloud), Aug 24-26, 2015,* Rome, Italy, pp. 489-96, 2015.
http://dx.doi.org/10.1109/FiCloud.2015.15

[45] B. Marr, "Really big data at Walmart: Real-time insights from their 40+ Petabyte data cloud", Available from: https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/#2a0c16916c10

[46] J. Pokorný, P. Škoda, I. Zelinka, D. Bednárek, F. Zavoral, M. Kruliš, and P. Šaloun, "Big data movement: A challenge in data processing", In: *Big Data in complex systems.* Springer: Cham, 2015, pp. 29-69.
http://dx.doi.org/10.1007/978-3-319-11056-1_2

[47] D. Court, "Getting big impact from big data", *McKinsey Q.,* vol. 1, pp. 52-60, 2015.

[48] M. Tang, M. Alazab, and Y. Luo, "Big data for cybersecurity: Vulnerability disclosure trends and dependencies", *IEEE Trans. Big Data,* vol. 5, no. 3, pp. 317-329, 2019.
http://dx.doi.org/10.1109/TBDATA.2017.2723570

[49] U. Bodkhe, S. Tanwar, K. Parekh, P. Khanpara, S. Tyagi, N. Kumar, and M. Alazab, "Blockchain for industry 4.0: A comprehensive review", *IEEE Access,* vol. 8, pp. 79764-79800, 2020.
http://dx.doi.org/10.1109/ACCESS.2020.2988579

[50] C. Xu, K. Wang, P. Li, S. Guo, J. Luo, B. Ye, and M. Guo, "Making big data open in edges: A resource-efficient blockchain-based approach", *IEEE Trans. Parallel Distrib. Syst.,* vol. 30, no. 4, pp. 870-882, 2019.
http://dx.doi.org/10.1109/TPDS.2018.2871449

[51] X. Fan, and Y. Huo, "Blockchain based dynamic spectrum access of non-real- time data in cyber-physical-social systems", *IEEE Access,* vol. 8, pp. 64486-64498, 2020.

[52] J. Sun, X. Yao, S. Wang, and Y. Wu, "Blockchain-based secure storage and access scheme for electronic medical records in IPFS", *IEEE Access,* vol. 8, pp. 59389-59401, 2020.
http://dx.doi.org/10.1109/ACCESS.2020.2982964

[53] T.R. Gadekallu, N. Kumar, S. Hakak, and S. Bhattacharya, "Blockchain based attack detection on machine learning algorithms for IoT based E-health applications", *IEEE Internet of Things Magazine,* vol. 4, pp. 30-33, 2020.

[54] L-A. Hîrtan, and C. Dobre, "Blockchain privacy-preservation in intelligent transportation systems", *2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Oct 29-31, 2018,* Bucharest, Romania, 2018.
http://dx.doi.org/10.1109/CSE.2018.00032

[55] J. Bernal Bernabe,.L. Canovas, J.L. Hernandez-Ramos, R. Torres Moreno, and A. Skarmeta, "Privacy-preserving solutions for blockchain: review and challenges", *IEEE Access,* vol. 7, pp. 164908-164940, 2019.
http://dx.doi.org/10.1109/ACCESS.2019.2950872

[56] M.A. Rahman, M.M. Rashid, M.S. Hossain, E. Hassanain, M.F. Alhamid, and M. Guizani, "Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city", *IEEE Access,* vol. 7, pp. 18611-18621, 2019.
http://dx.doi.org/10.1109/ACCESS.2019.2896065

[57] J. Liu, and Z. Liu, "A survey on security verification of blockchain smart contracts", *IEEE Access,* vol. 7, pp. 77894-77904, 2019.
http://dx.doi.org/10.1109/ACCESS.2019.2921624

[58] A. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy- preserving healthcare blockchain for IoT", *Sensors,* vol. 19, no. 2, p. 326, 2019.
http://dx.doi.org/10.3390/s19020326 PMID: 30650612

[59] T. McGhin, K.K.R. Choo, C.Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities", *J. Netw. Comput. Appl.,* vol. 135, pp. 62-75, 2019.
http://dx.doi.org/10.1016/j.jnca.2019.02.027

[60] J.D. Vyas, M. Han, L. Li, S. Pouriyeh, and J.S. He, "Integrating blockchain technology into healthcare", *Proceedings of the 2020 ACM Southeast Conference,* 2020 pp. 197-203.
http://dx.doi.org/10.1145/3374135.3385280

[61] L. Zhang, Y. Xie, Y. Zheng, W. Xue, X. Zheng, and X. Xu, "The challenges and countermeasures of blockchain in finance and economics", *Syst. Res. Behav. Sci.,* vol. 37, no. 4, pp. 691-698, 2020.
http://dx.doi.org/10.1002/sres.2710

[62] V. Astarita, V.P. Giofrè, G. Mirabelli, and V. Solina, "A review of blockchain-based systems in transportation", *Information,* vol. 11, no. 1, p. 21, 2019.
http://dx.doi.org/10.3390/info11010021

[63] A. Lei, H. Cruickshank, Y. Cao, P. Asuquo, C.P.A. Ogah, and Z. Sun, "Blockchain-based dynamic key management for heterogene-

ous intelligent transportation systems", *IEEE Internet Things J.,* vol. 4, no. 6, pp. 1832-1843, 2017. http://dx.doi.org/10.1109/JIOT.2017.2740569

[64] M. Andoni, V. Robu, D. Flynn, S. Abram, D. Geach, D. Jenkins, P. McCallum, and A. Peacock, "Blockchain technology in the energy sector: A systematic review of challenges and opportunities", *Renew. Sustain. Energy Rev.,* vol. 100, pp. 143-174, 2019. http://dx.doi.org/10.1016/j.rser.2018.10.014

[65] M. Fan, and X. Zhang, "Consortium blockchain based data aggregation and regulation mechanism for smart grid", *IEEE Access,* vol. 7, pp. 35929-35940, 2019. http://dx.doi.org/10.1109/ACCESS.2019.2905298

[66] X. Zhang, J. Grannis, I. Baggili, and N.L. Beebe, "Frameup: An incriminatory attack on Storj: A peer to peer blockchain enabled distributed storage system", *Digit. Invest.,* vol. 29, pp. 28-42, 2019. http://dx.doi.org/10.1016/j.diin.2019.02.003

[67] J. Moreno, E.B. Fernandez, E. Fernandez-Medina, M.A. Serrano, and B.D. Block, "a security pattern to incorporate blockchain in big data ecosystems", *Proceedings of the 24th European Conference on Pattern Languages of Programs,* 2019, pp. 1-8.

[68] H.M. Kim, and M. Laskowski, "Toward an ontology-driven blockchain design for supply-chain provenance", *Int. J. Intell. Syst. Account. Finance Manage.,* vol. 25, no. 1, pp. 18-27, 2018. http://dx.doi.org/10.1002/isaf.1424

[69] V. Gramoli, and M. Staples, "Blockchain standard: Can we reach consen-s us?", *IEEE Communications Standards Magazine,* vol. 2, no. 3, pp. 16-21, 2018. http://dx.doi.org/10.1109/MCOMSTD.2018.1800022

[70] M. Feng, J. Zheng, J. Ren, A. Hussain, X. Li, Y. Xi, and Q. Liu, "Big data analytics and mining for effective visualization and trends forecasting of crime data", *IEEE Access,* vol. 7, pp. 106111-106123, 2019. http://dx.doi.org/10.1109/ACCESS.2019.2930410

[71] L. Tan, N. Shi, C. Yang, and K. Yu, "A blockchain-based access control frame- work for cyber-physical-social system big data", *IEEE Access,* vol. 8, pp. 77215-77226, 2020. http://dx.doi.org/10.1109/ACCESS.2020.2988951

[72] S. Jangirala, A.K. Das, and A.V. Vasilakos, "Designing secure lightweight blockchain-enabled RFID-based authentication protocol for supply chains in 5G mobile edge computing environment", *IEEE Trans. Industr. Inform.,* vol. 16, no. 11, pp. 7081-7093, 2020. http://dx.doi.org/10.1109/TII.2019.2942389

[73] C. Chen, J. Wang, F. Qiu, and D. Zhao, "Resilient distribution system by microgrids formation after natural disasters", *IEEE Trans. Smart Grid,* vol. 7, no. 2, pp. 958-966, 2016. http://dx.doi.org/10.1109/TSG.2015.2429653

[74] G. Fanti, L. Kogan, S. Oh, K. Ruan, P. Viswanath, and G. Wang, "Compounding of wealth in proof-of-stake cryptocurrencies", *International Conference on Financial Cryptography and Data Security,* 2019, pp. 42-61. http://dx.doi.org/10.1007/978-3-030-32101-7_3

[75] P.M. Dhulavvagol, V.H. Bhajantri, and S.G. Totad, "Performance analysis of distributed processing system using shard selection techniques on elasticsearch", *Procedia Comput. Sci.,* vol. 167, pp. 1626-1635, 2020. http://dx.doi.org/10.1016/j.procs.2020.03.373

[76] H. Baniata, and A. Kertesz, "A survey on blockchain-fog integration approaches", *IEEE Access,* vol. 8, pp. 102657-102668, 2020. http://dx.doi.org/10.1109/ACCESS.2020.2999213

[77] O. Bouachir, M. Aloqaily, L. Tseng, and A. Boukerche, "Blockchain and fog computing for cyberphysical systems: The case of smart industry", *Computer,* vol. 53, no. 9, pp. 36-45, 2020. http://dx.doi.org/10.1109/MC.2020.2996212

[78] R. Zheng, J. Jiang, X. Hao, W. Ren, F. Xiong, and Y. Ren, "bcBIM: A blockchain-based big data model for BIM modification audit and provenance in mobile cloud", *Math. Probl. Eng.,* vol. 2019, pp. 1-13, 2019. http://dx.doi.org/10.1155/2019/5349538

[79] L. Cui, F.R. Yu, and Q. Yan, "When big data meets software-defined networking: SDN for big data and big data for SDN", *IEEE Netw.,* vol. 30, no. 1, pp. 58-65, 2016. http://dx.doi.org/10.1109/MNET.2016.7389832

[80] P.K.R. Maddikunta, Q-V. Pham, B. Prabadevi, N. Deepa, K. Dev, T.R. Gadekallu, R. Ruby, and M. Liyanage, "Industry 5.0: A survey on enabling technologies and potential applications", *J. Ind. Inf. Integr.,* vol. 26, p. 100257, 2021.

[81] R. Qin, Y. Yuan, and F.Y. Wang, "Blockchain-based knowledge automation for CPSS-oriented parallel management", *IEEE Trans. Comput. Soc. Syst.,* vol. 7, no. 5, pp. 1180-1188, 2020. http://dx.doi.org/10.1109/TCSS.2020.3023046

[82] T. Li, A.K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions", *IEEE Signal Process. Mag.,* vol. 37, no. 3, pp. 50-60, 2020. http://dx.doi.org/10.1109/MSP.2020.2975749

[83] Y. Zhao, J. Zhao, L. Jiang, R. Tan, D. Niyato, Z. Li, L. Lyu, and Y. Liu, "Privacy-preserving blockchain-based federated learning for IoT devices", *IEEE Internet Things J.,* vol. 8, no. 3, pp. 1817-1829, 2021. http://dx.doi.org/10.1109/JIOT.2020.3017377

[84] A. Jyoti, and R.K. Chauhan, "A blockchain and smart contract-based data provenance collection and storing in cloud environment", *Wirel. Netw.,* vol. 28, no. 4, pp. 1541-1562, 2022. http://dx.doi.org/10.1007/s11276-022-02924-y

[85] Z. Shanshan, L. Shancang, and Y. Yufeng, *Blockchain Enabled Industrial Internet of Things Technology,* 2019. Available from: https://uwe-repository.worktribe.com

[86] H. Yang, Y. Liang, J. Yuan, Q. Yao, A. Yu, and J. Zhang, "Distributed blockchain-based trusted multidomain collaboration for mobile edge computing in 5G and beyond", *IEEE Trans. Industr. Inform.,* vol. 16, no. 11, pp. 7094-7104, 2020. http://dx.doi.org/10.1109/TII.2020.2964563

[87] H. Yang, J. Yuan, H. Yao, Q. Yao, A. Yu, and J. Zhang, "Blockchain-based hierarchical trust networking for jointcloud", *IEEE Internet Things J.,* vol. 7, no. 3, pp. 1667-1677, 2020. http://dx.doi.org/10.1109/JIOT.2019.2961187

DISCLAIMER: The above article has been published, as is, ahead-of-print, to provide early visibility but is not the final version. Major publication processes like copyediting, proofing, typesetting and further review are still to be done and may lead to changes in the final published version, if it is eventually published. All legal disclaimers that apply to the final published article also apply to this ahead-of-print version.