A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions

Preprint · September 2020

CITATIONS
0

READS
4,517

9 authors, including:

N. Deepa
VIT University
48 PUBLICATIONS **411** CITATIONS
SEE PROFILE

Quoc-Viet Pham
Pusan National University
137 PUBLICATIONS **2,023** CITATIONS
SEE PROFILE

Dinh C. Nguyen
Purdue University
53 PUBLICATIONS **989** CITATIONS
SEE PROFILE

Sweta Bhattacharya
VIT University
51 PUBLICATIONS **1,004** CITATIONS
SEE PROFILE

A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions

NATARAJAN DEEPA*, School of Information Technology and Engineering, VIT
QUOC-VIET PHAM†, Research Institute of Computer, Information and Communication, Pusan National University
DINH C. NGUYEN‡, School of Engineering, Deakin University
SWETA BHATTACHARYA*, School of Information Technology and Engineering, VIT
PRABADEVI BOOPATHY*, School of Information Technology and Engineering, VIT
THIPPA REDDY GADEKALLU*, School of Information Technology and Engineering, VIT
PRAVEEN KUMAR REDDY MADDIKUNTA*, School of Information Technology and Engineering, VIT
FANG FANG§, Department of Engineering, Durham University
PUBUDU N. PATHIRANA‡, School of Engineering, Deakin University

Big data has generated strong interest in various scientific and engineering domains over the last few years. Despite many advantages and applications, there are many challenges in big data to be tackled for better quality of service, e.g., big data analytics, big data management, and big data privacy and security. Blockchain with its decentralization and security nature has the great potential to improve big data services and applications. In this article, we provide a comprehensive survey on blockchain for big data, focusing on up-to-date approaches, opportunities, and future directions. First, we present a brief overview of blockchain and big data as well as the motivation behind their integration. Next, we survey various blockchain services for big data, including blockchain for secure big data acquisition, data storage, data analytics, and data privacy preservation. Then, we review the state-of-the-art studies on the use of blockchain for big data applications in different vertical domains such as smart city, smart healthcare, smart transportation, and smart grid. For a better understanding, some representative blockchain-big data projects are also presented and analyzed. Finally, challenges and future directions are discussed to further drive research in this promising area.

CCS Concepts: • **General and reference** → *Surveys and overviews*; • **Information systems** → *Distributed storage*; • **Computing methodologies** → *Decentralization*; • **Networks** → *Peer-to-peer architectures*; • **Security and privacy** → *Security services*.

Authors' addresses: Natarajan Deepa, deepa.rajesh@vit.ac.in, School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu, 632014; Quoc-Viet Pham, vietpq@pusan.ac.kr, Research Institute of Computer, Information and Communication, Pusan National University, Busan, Korea, 46241; Dinh C. Nguyen, cdnguyen@deakin.edu.au, School of Engineering, Deakin University, Waurn Ponds, Australia, 3216; Sweta Bhattacharya, sweta.b@vit.ac.in, School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu, 632014; Prabadevi Boopathy, prabadevi.b@vit.ac.in, School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu, 632014; Thippa Reddy Gadekallu, thippareddy.g@vit.ac.in, School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu, 632014; Praveen Kumar Reddy Maddikunta, praveenkumarreddy.g@vit.ac.in, School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu, 632014; Fang Fang, fang.fang@durham.ac.uk, Department of Engineering, Durham University, Durham DH1 3LE, UK; Pubudu N. Pathirana, pubudu.pathirana@deakin.edu.au, School of Engineering, Deakin University, Waurn Ponds, Australia, 3216.

0360-0300/2020/9-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Additional Key Words and Phrases: Blockchain, Big Data, Vertical Applications, Smart City, Smart Healthcare, Smart Transportation, Security.

ACM Reference Format:
Natarajan Deepa, Quoc-Viet Pham, Dinh C. Nguyen, Sweta Bhattacharya, Prabadevi Boopathy, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, Fang Fang, and Pubudu N. Pathirana. 2020. A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions. *ACM Comput. Surv.* 1, 1 (September 2020), 29 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The global data traffic has increased at an unprecedented rate over the last decade, thus the special interest in "big data". As reported in [1], the big data market shall reach 229.4 billion $ in 2025 and significantly reduce the expenditure for various vertical industries like healthcare, retail, transportation and logistics, manufacturing, media and entertainment. Despite the lack of a precise definition, attention to big data can be seen in many scientific and engineering areas, e.g., computer vision, Internet of Things (IoT) data analytics, operation management, and smart cities. Adding to the structural embodiment, [2] considered big data from three aspects, including attributive, comparative, and architectural. According to [3], big data can be identified as a new generation of technologies and architectures investigated to analyze a large amount of data and capture its main characteristics (e.g., high velocity, knowledge discovery, and analytics). The comparative aspect considers big data as the datasets, which has a very large size and dimensionality and cannot be stored, managed, analyzed, and captured by conventional database tools [4]. From the architectural viewpoint, big data is identified as the datasets, which have very large volume, velocity, and representation, and require significant horizontal scaling methods for efficient processing [5].

Nevertheless, there are various challenges and issues associated with big data techniques and applications, for example, data security and privacy, energy management, scalability of computing infrastructure, data management, data interpretation, real-time data processing, big data intelligence. Among these challenges, security and privacy have been considered as important issues since big data often involves different types of sensitive personal information, e.g., age, addresses, personal preference, banking details, etc. There have been various solutions and techniques investigated to preserve data confidentiality and private information. An example is [6], where matching theory and a coalitional game were jointly utilized to optimize a resource allocation problem so as to secure mobile social networks with big data. The use of reinforcement learning was investigated in [7] to design a security-aware algorithm for a smart grid system. Recently, blockchain as a ledger technology has emerged as attractive solutions for providing security and privacy in big data systems. For example, it was shown in [8] that blockchain can play a vital role in providing high-quality data and securing data sharing for industrial IoT applications. In [9], a blockchain-based mechanism was proposed for securing data collection in mobile ad hoc networks and incentivizing mobile nodes for efficient data collection. Furthermore, blockchain was also integrated with edge computing servers to enhance the data quality and process the compute-intensive tasks requested by IoT devices with security guarantees [10]. With its unique advantages, blockchain has the great potential to transform current big data systems by providing efficient security features and network management capabilities for enabling newly emerging big data services and applications. In this survey, we present a comprehensive review of blockchain for big data, ranging from approaches to opportunities and future directions.

1.1 State of the Arts and Our Contributions

Due to the importance of blockchain and big data, there have been a number of surveys published in related topics over the past few years. One of the earliest surveys on blockchain was carried out in [11]. Privacy and security issues of blockchain systems were reviewed in [12–15]. The survey in [16] presented applications (e.g., game for

mining management, game for security/privacy issues, and game for blockchain applications) of game theories for blockchain systems. Various surveys have been conducted to study applications of blockchain for other technologies. For example, the possibility of utilizing blockchain for IoT systems can be found in [14, 17–19]. The integration of blockchain with edge computing and 5G systems were studied in [20] and [21], respectively. The surveys in [22, 23] carried out reviews of applications and opportunities of blockchain for smart grid networks. Moreover, several surveys have been dedicated to reviewing the fundamentals and applications of big data analytics. A survey on techniques and technologies for big data management was presented in [24]. Recent studies in [25, 26] reviewed and discussed the roles and applications of big data for IoT systems and smart cities. Big data analytics have also found applications in smart grid and intelligent transportation systems, and representative surveys can be found in [27, 28]. The concept of mobile big data was reviewed in [29] and recently found many applications for next-generation wireless systems (e.g., 5G, beyond 5G, and 6G), from the physical and MAC layers to the application layer [30, 31].

In spite of many research efforts, we are not aware of any survey that comprehensively studies the applicability of blockchain for big data applications. Although the survey in [32] reviews blockchain for big data applications and challenges, it is very short and not updated since it has been published several years ago. The survey in [33] mainly reviews the use of blockchain to address security issues in edge computing-based IoT applications. Other surveys in [18, 21, 34] also mention the interplay between blockchain and big data, but they only provide brief introductions on this topic without an in-depth survey unlike our paper. Motivated by the above observations, we provide a comprehensive survey on blockchain for big data, which covers fundamental knowledge, up-to-date approaches, opportunities, research challenges, issues, and future directions. The key objective of this survey is to inspect the state-of-the-art studies and to carry out a review on the applicability of blockchain for big data applications. In summary, the contributions and features offered by this work can be stated as the following.

- Firstly, we present an overview of blockchain and big data as well as the motivations behind the use of blockchain for big data. We show that blockchain has the great potential for facilitating big data analytics such as control of dirty data, enhanced security and privacy, enhanced quality of data, and the management of data sharing.
- Secondly, we review four main blockchain services for big data, including blockchain for secure data acquisition, blockchain for secure data storage, blockchain for data analytics, and blockchain for data privacy preservation.
- Thirdly, we provide an extensive discussion of the use of blockchain in several popular big data applications, including smart healthcare, transportation and logistics, smart grid, and smart cities. Moreover, some popular blockchain-based big data projects are also introduced and analyzed.
- Finally, we discuss a number of research challenges that arose from the state-of-the-art survey on the use of blockchain for big data. We also highlight open research opportunities that provide a roadmap for future research.

1.2 The Survey Organization

The structure of this survey is organized as Fig. 1. An overview of blockchain and big data is presented in Section 2, along with a discussion of the motivations of their integration. The main parts of this survey are given in Sections 3 and 4, which respectively present 1) blockchain services for big data and 2) blockchain-big data applications and projects. Section 5 discusses and highlights a number of research challenges, issues, and future directions. Finally, Section 6 concludes the article.

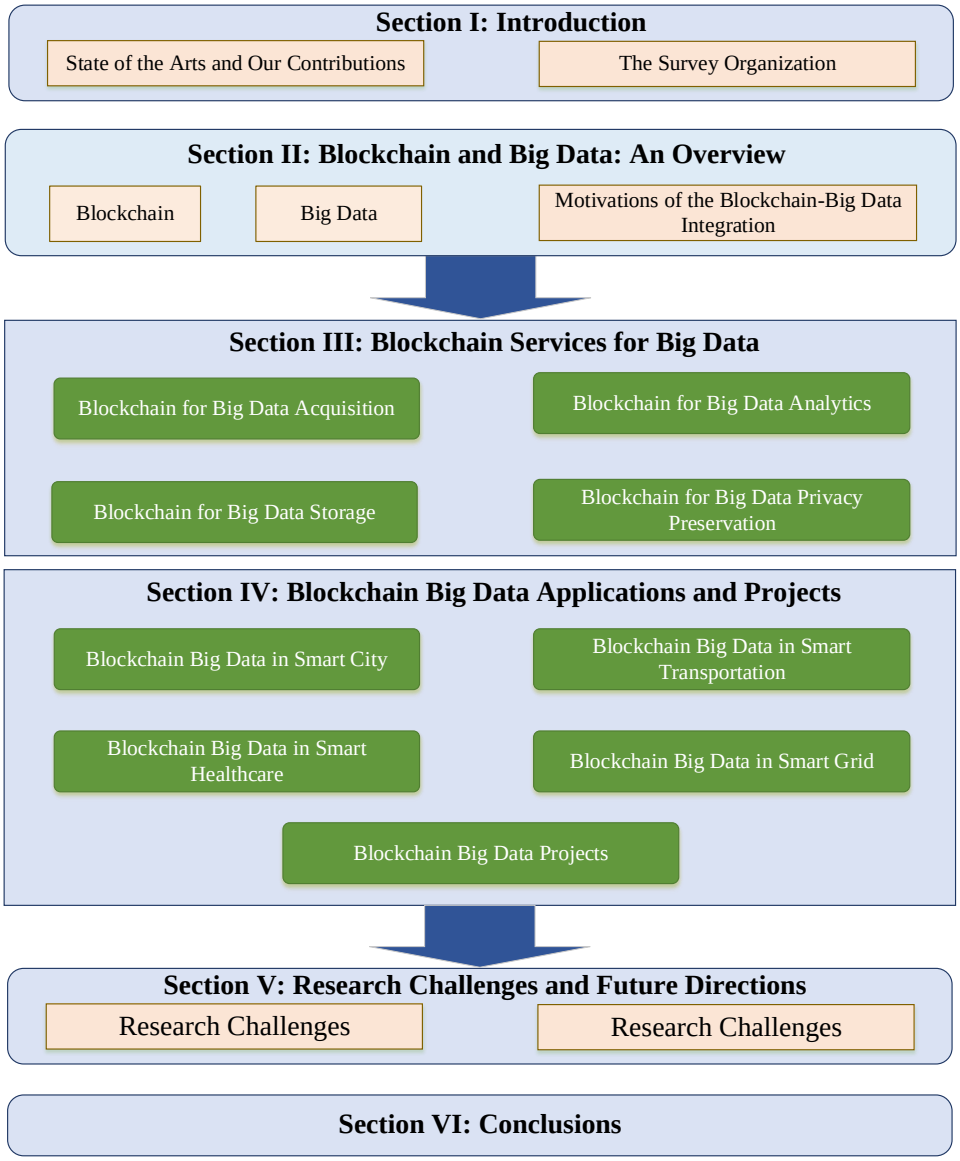

Fig. 1. Organization of this article.

2 BLOCKCHAIN AND BIG DATA: AN OVERVIEW

This section presents the background and recent developments of blockchain, big data, and motivations of their integration.

2.1 Blockchain

Blockchain is presently one of the most prevalent disruptive technologies which is paving the way for emerging financial and industrial services [35]. Conceptually, it consists of a list of records commonly known as blocks wherein information stored is encrypted ensuring privacy and security. Also, unlike other technologies, blockchain is a decentralized network wherein the participating members have complete authority to monitor all transactions in the blockchain network in a peer-to-peer (P2P) manner [36, 37]. The blockchain technology is an amalgamation of varied multidisciplinary concepts such as software engineering, cryptography, distributed computing, creating an infrastructure emphasizing on digital assets related security. The combination of all these concepts is commonly termed as cryptoeconomics that create robust P2P networks for facilitating the use and transfer of assets among computers in digital markets [38]. Cryptocurrency can be considered as the present and future mode of financial transactions which support the aforementioned transparency and security aspects of blockchain technology. In the digital market, cryptocurrencies are existing in different forms such as Bitcoin, Ethereum, Litecoin, Stellar, Ripple, Z-cash, Dash, etc [39, 40].

Bitcoin as the most popular cryptocurrency platform was introduced by Sakato Nakamoto and since then almost 1600 cryptocurrencies have evolved using the Bitcoin concept [41]. In the case of Bitcoin, whenever a sender initiates a transaction, it is sent to the receiver through the transaction being performed on the public bitcoin network. User verification is conducted by miners in the network who also ensure that the sender has the necessary number of bitcoins to be sent to the receiver without affecting the basic sanity of the network. After approval and verification by the miner, the transaction is added to the block which eventually becomes a part of the blockchain network. Finally, the relevant transactions pertaining to the block get executed thereby updating the ledgers across all the nodes so that all participants share the same copy of transaction for ensuring transparency and security [42].

Blockchain platforms can be classified into three types; public, private and hybrid blockchains, based on their areas of application [43]. A public blockchain does not have any specific single owner and are visible to everyone in the network. Bitcoin is an example of public blockchain which is decentralized with its consensus process being available to all participants in the network. The private blockchain on the contrary is permissioned and controls the participation of network members to read from and write to the blockchain. In hybrid blockchains, the public access is given to only specific group. It is a partially decentralized framework where the consensus process is guided by rules agreed among all parties regarding the control and access over the blockchain [44]. Some of the most important features of blockchain are as follows:

(1) *Immutability*: Blockchain is almost impossible to corrupt due to a permanent and unalterable network. It works differently from the traditional banking system using collection of nodes and each node in the system has a copy of the digital ledger [45]. When any transaction is initiated, nodes check its validity and authenticate to add to the ledger. Hence, the success of any transaction depends on the consensus across all major nodes which makes the framework transparent and secure. It also eliminates the chances of corruption which is especially evident in a public blockchain that allows everyone to see the transactions but does not allow altering the data stored in the blockchain [46, 47].
(2) *Decentralization*: The network is not governed by a single authority but a group of nodes that are responsible for maintaining the network. This decentralized approach allows participants to access the blockchain from the web and store their replicated information using private keys [48].
(3) *Security*: Blockchain with its decentralized and immutable natures can provide high degrees of security [35]. The use of cryptography includes implementation of complex algorithms acting as firewalls against unauthorized attacks. Each information is hashed which hides its actual nature and also provides a unique identification for each data. In the chain, each block in the ledger holds its own hash and also the hash of its previous block which makes it immutable to tamper the data. Hashing also makes the framework

irreversible. Such that, it is impossible to have a public key and create a private key out of it and corrupting the network would basically mean changing each data stored on each node in the network [49].

(4) *Consensus*: The operation of the blockchain frameworks relies on associated consensus algorithms, which is responsible for deciding the group of active nodes on the network. This makes the validation process for a transaction faster and similar to a voting system [50].
(5) *Accelerated Financial Settlement*: The blockchain transactions are processed much faster in comparison to the traditional banking systems. This technology enables faster transfer of money to foreign workers and overseas travelers. Smart contracts running on the blockchain also help ensure faster settlement of contractual accounts [51].

2.2 Big Data

Big data is typically characterized by 4-V features, including volume, velocity, and variety, and veracity[52]. Here, we briefly describe these features of big data.

(1) *Volume*: Volume simply means the quantity of data, i.e., whether or not a dataset is considered as big data. Regarding big data processing, one usually faces several challenges, which may include the curse of modularity (i.e., not available to store/load the complete data in memory and hard disk), the curse of class imbalance (i.e., there may exist different data distributions), the curse of dimensionality (i.e., the dataset has many features and attributes) [53]. Moreover, data non-linearity, variance and bias, and computing availability are also considered as challenges associated with the *volume* feature of big data.
(2) *Variety*: Variety represents various types of data such as video, text, and audio, which are generally composed of structured data, semi-structured data, and unstructured data. The major challenges caused by variety may include data locality, data heterogeneity, dirty and noisy data [54]. Here, data locality expresses that the complete data cannot be stored in a data center and is typically distributed over a large number of physical locations. Data heterogeneity is referred to as various heterogeneous sources of data, thus having different data types, formats, models, and semantics. Dirty and noise data means that the data can contain noise and dirty, which would be caused by data collection methods, data sources, and generation time.
(3) *Velocity*: Velocity refers to the generation speed of data, i.e., how fast the data is generated to meet the demand. A massive number of mobile devices will be 13.1 billion in 2023, from 8.8 billion in 2018, which can generate an enormous amount of traffic [55]. Other good examples of the unprecedented growth of data are high-definition videos, video gaming, and streaming platforms (e.g., YouTube and IBM Cloud Video). In some literature, this feature is also considered as *variability*, that is, different applications may have different rates of data flow [26]. For example, a vehicular crowdsensing system may generate more data in peak hours due to the participant of a large number of vehicles on the road.
(4) *Veracity*: Veracity refers to the quality aspect since the data can be collected from multiple sources, which may include low-quality and noisy samples. It is reasonable since data can be generated by malfunctioning or uncalibrated IoT devices, untrusted devices, and can be transmitted to the data center via fading and dynamic wireless environments [56]. To improve the quality and analytical accuracy of big data, the challenges of data provenance, uncertainty, dirty and noisy data should be effectively tackled.

Big data analytics is about extracting useful information and patterns from the dataset, which are then used for different purposes and to create business and social values. In the literature, this is usually considered as the fifth feature of big data, namely *value*. Big data has found applications in many vertical domains such as smart grid, mobile and e-health, transportation and logistics, and wireless and communication networking. Besides great opportunities, we have a number of technological challenges and issues of big data, for example, big data management, data cleansing, imbalanced system capacities, imbalanced data, data analytics, and learning from data [53]. For more details, we refer the interested reader to the survey in [53] and the references therein.

2.3 Motivations of the Blockchain and Big Data Integration

Governments and private organizations are investing heavily in big data and blockchain technologies due to their great potential in solving many real-world problems. In modern life, the customers are more inclined to do the transactions online, and expanding amount of data is being generated every day. This exponential rise in the digital data generated creates new opportunities for industries to understand the customer needs, purchasing patterns and trends of the customers. Big data analytics, which uses data mining and statistical models to analyze massive datasets, is playing a major role in helping the industries to gain insights into the purchase patterns of the customers[57]. However, the tremendous growth in the big data presented its own challenges. Some of the key challenges of big data are security and privacy issues, dirty data, reliability of the data sources, sharing of the data, etc[58]. These challenges faced by the big data can be addressed by the unique properties of the blockchain like decentralized storage, immutability, transparency, and consensus mechanisms. The motivations of integrating blockchain with big data are discussed as follows.

- *Improving Big Data Security and Privacy*: As the number of devices connected to the Internet is growing day by day, the quantity of the data stored at third party locations like cloud is increasing rapidly. This brings new challenges like data breach or threats caused by curious third parties [59]. The traditional security solutions like firewalls cannot address this issue of big data since the organizations have no control over the data as it is not stored within the network perimeter of the organizations. The usage of blockchain to store the big data has the potential to address this issue. The encrypted and decentralized storage of the data in the blockchain network makes it very difficult for any unauthorized access to the data.
- *Improving Data Integrity*: There exists a likelihood of people tampering the records in big data to influence the prediction of big data analytics in their favor. The immutability property of the blockchain ensures that it is next to impossible to tamper with the data stored in the blockchain network. If someone wants to modify the data in the blockchain network they have to modify the data in at least 50% of the nodes in the blockchain network, which is nearly impossible in practice. Also, the immutability property of the blockchain ensures that data stored the blockchain network is reliable.
- *Fraud Prevention*: The existing big data solutions rely on the analysis of patterns in the historical data to detect fraudulent transactions. Hence big data cannot solve the problem of fraudulent transactions in the financial sector. The storage of the big data in blockchain enables the financial institutions to monitor each transaction in real time, hence allowing them to assess the potentially fraudulent transactions on the fly. As a result, the integration of blockchain in big data can help the financial institutions to prevent the frauds to protect their customers.
- *Real-Time Data Analytics*: Since the blockchain stores every transaction, it makes the real-time analytics of big data achievable. The banks and financial institutes can settle the cross-border transactions including large amounts in near real-time as the blockchain integrated big data analytics enables the financial institutes to settle the transactions quickly. Also, banks can monitor the changes in the data in real time, thus enabling them to make decisions like blocking of the transactions in real time.
- *Enhancement of Data Sharing*: The integration of blockchain with big data helps service providers to share the data to other stakeholders with minimal risk of data leakage. Also, if the big data generated from the different sources is stored in blockchain, the repetition of the analysis on the data can be eliminated as each experiment carried out is recorded in the blockchain.
- *Enhancement of the Quality of Big Data*: Data scientists spend most of their time on data integration as different sources follow different formats in data collection. By using blockchain for data storage, the quality of the data can be improved as it is structured and complete. Hence, data scientists can work on the quality data to come up more accurate predictions in real time.

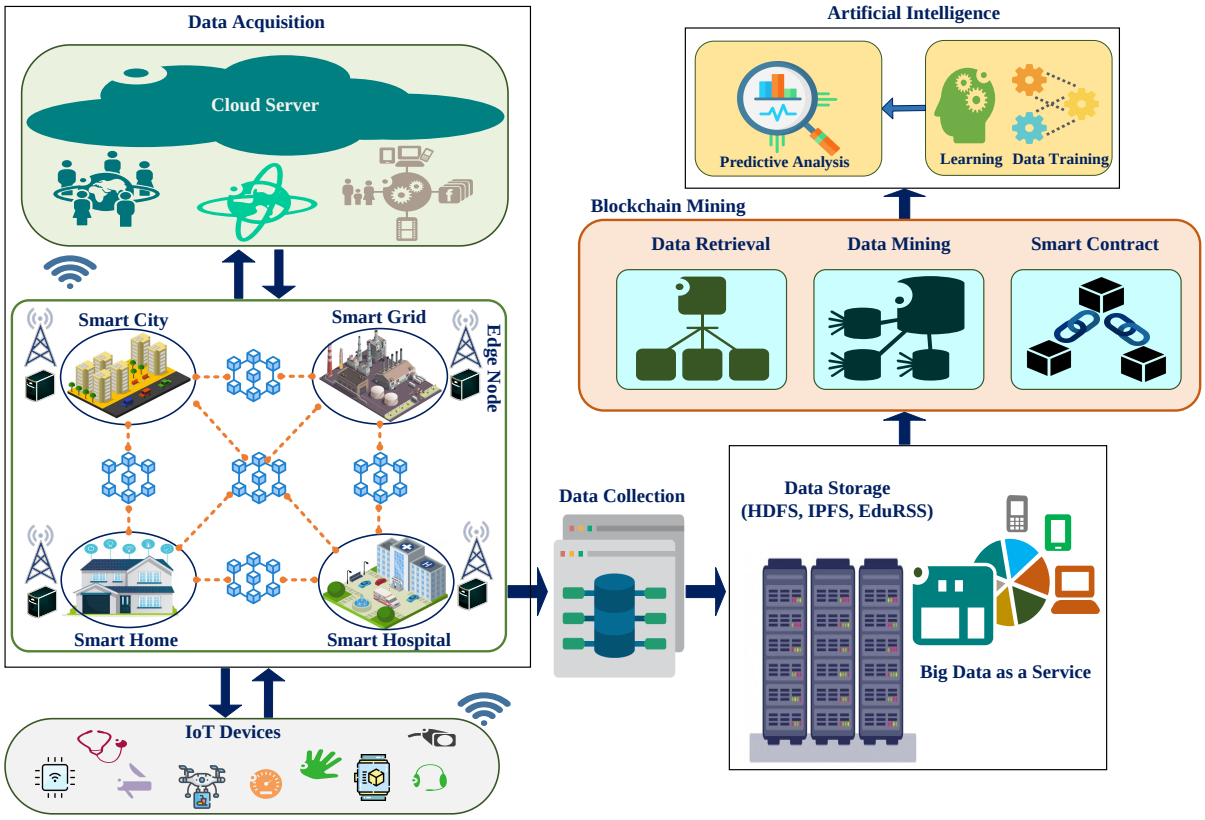

Fig. 2. An overview of blockchain services in big data environment.

- *Streamlining the Data Data Access*: The use of blockchain would simplify the life cycle of big data analytics by online streamlining the data access. Indeed, by involving multiple departments in an organization in a common blockchain, authorized users can get access to the secure, trusted data without having to go through several checks.

3 BLOCKCHAIN SERVICES FOR BIG DATA

The big data technology has grown tremendously as large corporations and organizations use advanced analytical tools to store, visualize and analyze data. However, due to the enormous data utilization and data transmission, big data security is a major challenge. Cloud computing has been widely used for big data services despite some security concerns. Some third-party applications and intruders can easily perform malicious activities such as stealing sensitive data, crashing the server when proper security mechanisms are not used [60]. Big data faces challenges from a variety of perspectives, such as data collection, data sharing, data storage and data analysis. In this section, we survey the blockchain-based approaches and services for big data. An overview of blockchain services in big data environment such as big data acquisition, big data storage, big data analytics and big data privacy preservation is depicted in Fig. 2.

3.1 Blockchain for Big Data Acquisition

In general, big data applications have data sources from diversified sources in a different format (unstructured data). These data cannot be processed in the native form. Therefore, the data must be converted to a structured format from which various predictions on the application domain can be made. Blockchain, with its capability of handling vast data effectively, provides structured data for making predictions. Blockchain ensures data integrity through consensus algorithms thereby mitigating the data attacks. Moreover, smart contracts can be used for data trading, which allows users to share their unused data storage, and the reused block can be added to blockchain as a new block. Henceforth, AI models can be used for making valuable predictions. Here, we analyze two subdomains in blockchain services for big data acquisition, including blockchain for secure big data collection and blockchain for secure big data transmission/sharing.

3.1.1 Blockchain for Secure Big Data Collection. Nowadays, big data applications have gained popularity but faced major security issues and challenges. Data collection is a very important task in the life cycle of data processing. It is a vital step in various types of data applications which provides the outcome of data analysis. Suspicious sources of data and communication links allow the data collection to expose to various malicious attacks and threats. Therefore, secure data collection methodology is vital for various data applications. Several research works have been done so far to provide secure data collection. For example, a secure big data collection scheme based on blockchain is introduced in for mobile crowdsensing (MCS) [61]. Due to the rapid growth of portable smart mobile terminal devices such as mobile terminals (MT) and sensors, MCS has been efficiently applied for industrial Internet of thing (IIoT) environment. An MCS framework is developed with cloud servers and a set of MTs. The MCS servers publish some set of tasks related to sensing and choose MTs in the particular area to complete the tasks. The main challenge in performing data collection is limited energy resource in MT, the range of sensing devices and secure data sharing between MTs. A framework was proposed by [62] to overcome these challenges using blockchain and deep reinforcement learning (DRL). It provides energy efficient collection of data and security for data sharing in a distributed environment. The distributed blockchain based DRL approach for each MT provides extensive data collection and maximum range for sensing devices. An Ethereum blockchain platform is used to provide data reliability and security while MTs share the data. Ethereum maintains a secure ledger and shares with the cooperating MTs without a trusted third party. The proposed framework provides solutions for various attacks such as majority attack, device failure, eclipse attack, etc [62].

3.1.2 Blockchain for Secure Big Data Transmission/Sharing. Blockchain with its decentralized and immutable nature is able to provide secure big data transmissions and reliable data sharing from data sources to data analytics, aiming to solve security and privacy issues remained in traditional data transmission protocols. Blockchain can ensure big data training and prevent data theft to facilitate big data transmissions. That is, data can be recorded from ubiquitous sources such as data reports, data libraries, social media, or assistive gadgets. Then, they are added to the blockchain with signature and hash values before sharing with data analytic services in a fashion both data source owners and data analytic users can trace and monitor the data sharing flow over the network which in return provides high transparency and reliable data sharing. Such an example of a big data transmission model with blockchain is illustrated in Fig. 3. In the literature, there are also some research efforts devoted to use of blockchain for supporting big data transmissions and sharing. The emergence of edge computing has seen an increasingly vast amount of data on edge nodes, allowing end-users to optimize latency and the processing time. However, sharing sensitive information without proper authorization is a challenging task. The work in [63] introduces a blockchain model to share reliable data at the edge node. During this process, the authors pay attention to the reduction of the computational process at the edge nodes using proof-of-collaboration. Besides, to reduce response time and storage overhead, authors introduced a blockchain-based futile transaction filter algorithm that accesses data from the cache layer rather than the storage layer. Finally, to authenticate

asynchronous transactions, the authors proposed express transactions smart contact, and hollow block is formed, which tends to reduce resource duplication. The experimental results demonstrate that the proposed model decreases 90% of computing resources, 95% of storage resources and 27% of network resources. The immense growth of the cyber-physical system helps in providing faster information services, real-time sensing. Big data sends information to the cyber-physical system, which utilizes the radio spectrum. There is incredibly huge competitiveness in the spectrum auction and restricted license-free spectrum access. In [64], the authors propose a blockchain-based solution for license-free spectrum access using smart contracts, which facilitates transferring non-real-time data in a secure manner. The proposed framework using edge node aiming to reduce latency provides a blockchain-based protocol, which improves the transaction process in a safe mode where multiple channels are created for spectrum, and each channel is allocated the dedicated blockchain. During the process, two blocks are created, namely key block and micro block, where key block ensures to select an efficient spectrum license holder, and the micro block takes the responsibility of maintaining all the transaction details. Finally, the valid and authorized node gets the spectrum license from the key block, and the node maintains the license until the key block identifies the next holder. However, we found that the protocol uses PoS-after-PoW for generating the key block to select a user, which requires high computation cost and consumes more number of resources.

3.2 Blockchain for Big Data Storage

3.2.1 Blockchain for Secure File Systems. There are several cloud based services available to store and access files from anywhere on any machine. Users, particularly organizations are hesitant to store sensitive information on the system managed by a third party. Even though encryption of files before storing to the cloud storage is one of the solutions but still some challenges are faced by the cloud provider in terms of security. At present electronic information system are most popularly used in medical treatment. Volumes of data are produced every day such as medical images, medical records, diagnostic reports, etc. Electronic medical information can affect the treatment of the patients and the experience of the patient is shared with other medical academies. If the shared patient medical data is illegally misused, the privacy of the patient is compromised. Some mechanisms should be adopted to control the access of medical data. Blockchain integrated with interplanetary file system (IPFS) provides the solution for these kinds of security problems. IPFS is a decentralized storage platform developed to address the issue of file redundancy. It defines a unique hash value for the stored file and the user is allowed to find the file based on the hash address. Attribute based encryption method is applied to the medical data before it is stored in the cloud storage. The private key of the user is associated with their attributes and ciphertext with their policy. Any user can perform decryption on the ciphertext if the private key of the user satisfies the access policy available in the ciphertext. Also, blockchain is used to record the data storage and retrieval process. The hash value of medical storage data is stored in the blockchain to provide evidence for the authenticity of user verification. The decentralized blockchain framework helps to provide security for the file storage and avoids single point of failure [65].

3.2.2 Blockchain for Secure Database Management. Data which is stored in various types of database management system is vulnerable to attacks from internal and external sources. Database tampering detection methods were used to detect the malicious updates in the databases. It uses single-way cryptography hash functions along with digital watermarking to identify data misuse. But the method cannot be applicable to distributed databases. A blockchain based solution is applied to store the data on distributed databases and detect the malicious user transactions. Blockchain avoids data tampering by applying time stamping. Virtual shared ledger is incorporated to store the history of transactions. All the transactions are recorded in block and each block is interconnected with each other with cryptographic hash values. When the data available in a block is updated by a malicious attacker, the hash value of the block gets updated and the block becomes invalid. A blockchain based scheme, namely education records secure storage and sharing scheme, is proposed for privacy preserving and secure

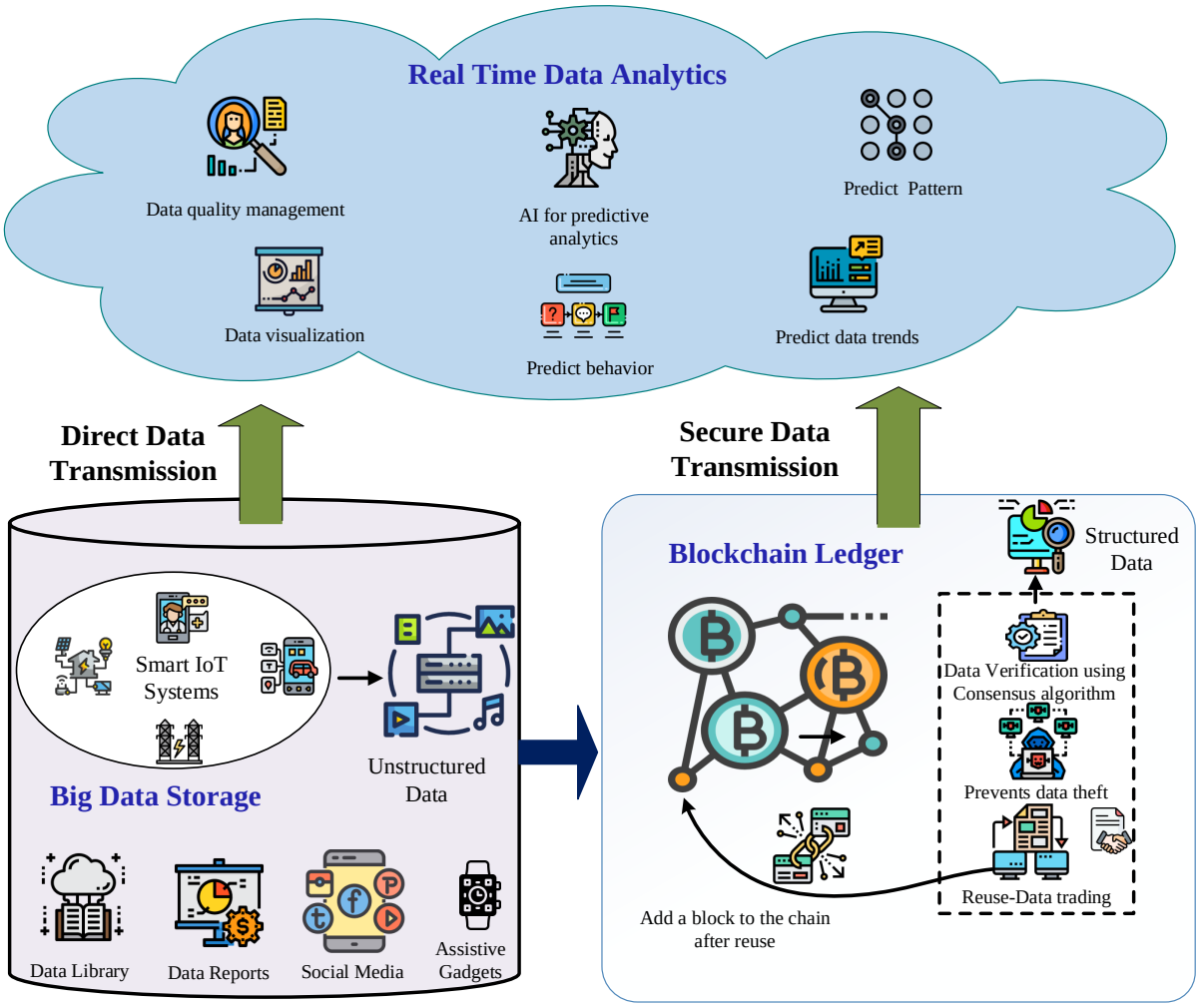

Fig. 3. Secured blockchain services for big data processing.

storage of education records. The scheme integrates storage servers, cryptographic algorithms and blockchain to develop a safe and reliable environment. Blockchain is integrated to provide reliability and security to the education database. The smart contracts in blockchain are applied to control the data sharing and storage process. The educational records are linked with the hash information stored in the blockchain to provide security to the stored data. Cryptographic algorithms and digital signatures are used to maintain the encryption of the records [66].

3.2.3 Blockchain for Big Data Storage Infrastructure. In the past few years, big data has grown into a new standard which provides huge amount of data and prospects to enhance the decision making applications in science and engineering. At the same time, it faces challenges in storing, processing and transmitting the data. Cloud computing offers basic support to address the issues with shared resources such as networking, storage

and computing. The increased readiness of data in AI provide opportunities in the healthcare industry. The present machine learning (ML) algorithms transform a person data to medical data for data analytics preventing the patients access to their medical data. Blockchain can provide security solutions to motivate the biomedical research and allow the patients to access and control their personal data along with the capability of monitoring their health records. Blockchain provides decentralization facility for a transparent and secure distributed personal data [67].

3.3 Blockchain for Big Data Analytics

3.3.1 Blockchain for Secure Data Training. The development of edge and cloud computing has increased the amount of data in various scenarios. Several ML and deep learning (DL) methods are applied for effective data analysis. Support vector machine (SVM) is one of the popular ML methods applied for its efficiency and accuracy. In vehicular social networks, data are gathered from various entities namely social network companies, vehicular manufacturers and vehicle management agencies. Data from various data sources normally differ in the attributes. When training with SVM classifier, the entities face problem of the data with inadequate attributes due to the diversity of sources. Therefore, various entities must share data to integrate the dataset with multiple attributes and train the classifier. Data privacy issue occurs due to sharing of data from various entities. A privacy preserving blockchain based SVM training method was proposed for vertically partitioned dataset from various data providers. In this method, a blockchain consortium and homomorphic cryptosystem were developed to implement a secure training platform without the need of a trusted third party. The training operations are performed over the original data locally and the interactions between the entities are secured by the homomorphic cryptosystem and blockchain consortium. Blockchain consortium helps to build a public and secure data sharing environment for effective communication between the entities when they share the attribute values [68].

3.3.2 Blockchain for Secure Data Learning in AI Algorithms. The extensive construction and generation of data from sensors, social media, web and IoT devices resulted in the growth of Artificial Intelligence (AI) techniques. The data can be applied to ML and DL algorithms for the purpose of data analytics. These methods depend on centralized server for training purpose and it leads to tampering of data. Thus the decisions obtained from AI are erroneous and risky. Therefore the decentralized AI came into existence to solve this problem and it is integration of blockchain and AI. Several limitations of blockchain and AI are solved by integrating these two technologies. AI techniques depend on data to learn, gather and provide decisions. These techniques perform better when the input data are gathered from various secure, reliable and trusted data repositories. Blockchain provides secure environments through distributed ledger in which data can be recorded and transacted. Here, data are stored with high resiliency and integrity in blockchain and cannot be tampered. When smart contracts are used for learning purpose in AI algorithms to obtain decisions and analytics, the results can be undisputed and trusted. Therefore, the integration of blockchain with AI can provide immutable, decentralized and secure environment for learning the highly sensitive data. This integrated framework provides substantial development in various domains such as banking, medical, financial, personal and trading [69].

3.4 Blockchain for Big Data Privacy Preservation

3.4.1 Blockchain for Privacy Preservation in Big Data Processing. Due to the rapid increase in generation of data, privacy preserving has become a main concern nowadays. In this era of big data, the data is regularly being gathered and examined which leads to commercial and innovation growth. Big organizations and companies utilize the collected data to provide better customer services, optimize the decision process and forecast the future developments. Thus data has become a valuable asset in recent days. Big data is widely applied in smart city environment for extensive monitoring of city traffic and maintenance, ensuring quality of air and water etc. A blockchain based model is proposed [70] for privacy preserving in intelligent transportation system (ITS) for

in-car navigation system in smart city environment. The model applies offline blockchain based storage in which all the sensitive information from the users are stored in a secure manner. The sensitive data is encrypted with the help of shared key associated with a group of cars. Users can use various security features such as sharing details about speed enable and disable, location enable and disable, etc [70].

3.4.2 Blockchain for Privacy Preservation in Big Data Storage. The big data era is threatening the user privacy in various digital scenarios. Third party organizations are benefited in the management of user data by gathering, analysing and managing the huge amount of user personal information. These services provided by the third parties are prone to security breaches and data misuse without the knowledge of the users. Blockchain provides various solutions to the challenges faced by the user data. User transactions in blockchain do not face privacy concerns and users are provided with options to control their personal information. The details about when, by whom, which and what personal information is revealed in each transaction. Privacy preserving solutions are emerging for blockchain built on crypto-privacy methods to allow the users to become unidentified and gain control over their personal information during their digital transaction in ledger [71]. The various services provided by cloud environment for big data, challenges and blockchain based solutions are tabulated in Table 1.

4 BLOCKCHAIN BIG DATA APPLICATIONS AND PROJECTS

Blockchain technologies have gained immense momentum with its varied applications in various spheres of life. The technology is still going through its phase of infancy and is being experimented for providing solutions to various challenges pertinent to security, data ownership, decision support systems, identity verification and decentralization. Our present generation is traversing through an era of overwhelming volume of digital data, being generated by man and machines. Hence there emerges a desperate need to store, organize, process and analyze this big data where the use of blockchain technologies has a potentially significant role to play [72]. As an example, maintaining data ownership, data transparency and management of access control has always been a major challenge. Blockchain technology resolves this issue by storing access policies to personal data in the blockchain framework. By using the blockchain technology, a decentralized personal data management system is created by implementing a protocol allowing users to own and manage their data. The dependency on third party is completely eliminated allowing organizations to focus more on data utilization rather than security management and compartmentalization [73]. The application of blockchain in combination with big data is visible in two segments - data management and data analytics. The various blockchain based big data applications are summarized in Table 2. In case of data management, blockchain technologies being sure and distributed, are implemented to store important data. It can also evaluate data authenticity and stop tampering of sensitive data. In the applications of data analytics, blockchain is used to analyze trading trends, prediction of potential customers, diseases or business partners [74].

4.1 Blockchain Big Data in Smart City

The rapid urbanization have led to the development of smart cities which requires efficient and intelligent solutions for its transportation, administration, environment and energy optimization. The integration of IoT, big data and energy efficient Internet technologies has the capability to provide such infrastructural solutions required for the smart city life. But there are numerous problems related to inferior security, reliability, maintenance, adaptability and costs. The blockchain technology caters to such needs having transparency, energy efficiency, space, recover-ability and maintenance of the IoT devices. The study in [19] discusses the use of hash, asymmetric encryption, consensus algorithm, a blockchain structure and a Merkle tree in ensuring a tamper free transaction. This framework has blocks interlocked with one another within the block itself with the help of a Merkle tree which makes it even more secured for performing seamless transactions. The recent years have also witnessed a surge in the development of big data based auditing systems termed as third party auditors (TPAs). The TPAs

Table 1. Services provided by cloud environment for big data, challenges and blockchain based solutions.

Ref	Cloud based services	Challenges faced by Big data	Solutions provided by Blockchain
[62]	Data collection.	Data collection is exposed to various malicious attacks and threats.	Blockchain provides energy efficient data collection and secure data sharing environment using Ethereum.
[63]	Data transmission/sharing.	Lack of authorization for data sharing in edge nodes and response time is more.	Blockchain based futile transaction filter algorithm helps to access data from cache layer instead of storage layer and helps to reduce response time and storage overhead. Smart contracts are used for authorization.
[65]	File storage system.	Unauthorized access to the electronic file system. Privacy, security and redundancy problems.	Blockchain integrated with IPFS provides the solution by implementing decentralized platforms to solve file redundancy problems and provides security to the file storage system. Hash value of data is stored in blockchain to provide authenticity to the users and an attribute based encryption method is applied before data storage in cloud.
[66]	Database management system.	Data stored in distributed database is exposed to internal and external attacks.	Blockchain overcomes data tampering using time stamping method. Virtual shared ledger is applied to store the transaction history. Database transactions are recorded in block and each block is interconnected with each other using cryptographic hash value. Blockchain based solution integrates storage servers, cryptographic algorithms for a reliable database access.
[68]	Data training/learning process.	Various entities share data to integrate the dataset with various attributes and train the ML classifier. Data privacy issue occurs while sharing data from various entities.	Blockchain consortium and homomorphic cryptosystem provide a secure training platform without the intervention of a trusted third party. Blockchain provides a secure environment for communication between the entities.
[71]	Data privacy preservation.	User privacy is an issue in digital scenarios in big data era. Services provided by third parties are exposed to security breaches and data misuse.	Blockchain provides immutable, verifiable and decentralized ledger to record the transactions in digital scenarios.It provides facilities to the user to control their personal data. Crypto-privacy methods are applied to solve privacy preserving problems.

are centralized frameworks which are subjected to security issues within the cloud environment. Blockchain technologies have been used to create decentralized TPAs for smart cities with enhanced security and reliability. This framework is named as Data Auditing Blockchain (DAB), the entire audit history is traced and also allows owners to audit their files at any point in time. It also includes the feature of batch verification of various auditing proofs ensuring security and prevention of privacy [75]. In [76], a blockchain based infrastructure is presented that provides secured spatio-temporal smart contract services. The framework provides sustainable IoT based shared economy in smart mega cities. The huge generation of big data has created the need to collect, analyze and utilize the same for autonomously predicting any risky or exceptional events from occurring. The framework consists of device-to-device (D2D) communication systems and fog nodes installed onsite to enable the blockchain and other offline operations. A three-tier architecture is used for supporting shared economy services in the blockchain based smart city environment. The client tier includes the smart applications, IoT and associated infrastructures. The client tier communicates with the mobile edge tower through WiFi, ZigBee, 5G and other related technologies. The MEC tower hosts the blockchain nodes, data storage client, related databases and cloudlet applications, thereby manages the load efficiently. The data from the blockchain, IoT and social

network are finally fed into the AI engine for performing sophisticated analysis such as digital forensics, emotion extraction and various others.

4.2 Blockchain Big Data in Smart Healthcare

Recent advances in the healthcare sector have led to a drastic rise in medical data generation. These data are extremely important for diagnosis, predictions and treatment purposes. Healthcare professionals have recently started focusing on the use of IoT and related wearable technologies wherein sensors, devices are vehicles are connected through the Internet providing services for the benefit of mankind. As an example the remote patient monitoring system is a common device for treating elderly patients in particular. Although these technologies have enormous benefits but have aforementioned security issues while transferring and logging of data transaction information. But these issues have possibilities of extreme violation of data security and privacy. The use of blockchain is a potential solution that would provide security and efficiency in analysing data but it is costly and lags energy optimization. The study in [77] proposes a framework that resolves such issues using public key, private key, light weight cryptographic techniques in integration with blockchain technology. The framework thus provides an access control of medical records for patients with improved privacy and security. In [78] a secured smart health care system is proposed using blockchain. The various private data, public data and related sensitive information are captured using sensors and then encrypted using blockchain technologies. These types of information are further stored in a distributed format rather than centralized cloud storage systems, which can be accessed only by authorized individuals having approvals from patients. Similarly, the healthcare professionals seeking to access the patient records need to send request to the patient and once real time notification is processed, information is available to them. All the entities such as IoT devices, Electronic Health Records (EHRs), Encryption/decryption system, blockchain mechanisms in this framework remain connected through wireless sensor networks (WSN) to conduct seamless yet secured communication. In [79], a private blockchain framework is proposed using Ethereum protocol wherein the sensors communicate with the smart devices. These smart devices call smart contracts which keep records of all events on the blockchain. Thus, these smart contract systems help in monitoring patients in real-time and also send notifications to healthcare professional when medical interventions are required. The saved records are secured, due to the connectivity in the blockchain which provides authentication and eliminates possibilities of data tampering of EHRs.

4.3 Blockchain Big Data in Smart Transportation

Transportation helps to move human beings and goods from one location to another. Although the application of blockchain has the immense potential towards benefiting the transportation sector, but individuals in this sector are not well informed about this emerging technology. Various other technologies namely Mobility as a Service (MaaS), IoT, AI and DL have converged with blockchain technologies to revamp the traditional approach involved in transportation. The automotive sector has also used blockchain technologies for developing intelligent transportation systems and offer services like remote software based vehicle operation system, automated insurance services, smart charging and cab sharing services [80].

Blockchain technologies have seen a rapid growth due to its potential to revolutionize intelligent transportation systems (ITS). Such developments can be used to create secured, reliable and autonomous ITS ecosystems with optimized usage of relevant infrastructure and resources. As an example, the study by [81] presented a seven layer conceptual model for ITS that would help in characterizing the architecture and major components in a blockchain based system. The physical layer holds the different vehicles, devices and assets relevant to ITS. The main aspect of this layer includes the use of IoT for providing enhanced security and privacy for the blockchain based transportation systems. The data layer provides the data blocks and associated encryption algorithms, hashing algorithms and Merkle trees. The network layer defines the process involved in distributed networking,

data forwarding and authentication. The packaging of the consensus algorithms is done by the consensus layer followed by the incentive layer which specifies the mechanisms for issuance and allocation of coins to nodes in the blockchain network. The contract layer constitutes algorithms and smart contracts that activates the process of data storage in the blockchain. Finally, the application layer encompasses the scenarios and use cases of blockchain based ITS. Security is often a major concern in vehicle communication systems. The study in [82] presented a secure key management framework for accomplishing network security. The study utilizes the role of security managers who capture vehicle departure data and encapsulate the blocks to transport keys and later implement rekeying to the vehicles within the secured domain. The framework proposes an efficient key management system for key transfers among the security managers in a heterogeneous vehicle communication network architecture. In addition to security provision, blockchain technologies play a significant role in privacy protection of ITS especially in car navigation. The framework proposed in [83] is based on an offline blockchain storage system wherein all sensitive data extracted from the users are stored and later shared using specific encryption keys relevant to a particular car cluster. The system uses two major applications namely the client application installed in the users smart phone and the main application installed at the server side. It is assumed that the smart phone and server are configured securely and security policies both simple and complex are used depending on user types for data sharing. All clients in the network are grouped into clusters depending on the location to optimize the use of computational resources, reduce network delays and overheads. The system allows users to define the privacy policies and they are later automatically implemented at the client application that provides accurate transportation routes. Blockchain technologies can potentially solve various problems relevant to car insurance. The insurers with the help of this technology will be able to track their claims seamlessly by searching the trusted ledger. The study by [84] presented a prototype framework for fine-grained transportation insurance services where the premium was calculated based on vehicle usage and behavior of the driver. These information were collected by streaming IoT data collected using mobile sensors. This unique framework initiated transparent insurance and also motivated drivers to drive safely in-order to achieve insurance incentives. The mobile GPS sensors in this framework were strategically placed in vehicles for continuous monitoring of their GPS location. The GPS trajectory data were further uploaded to the public cloud or data center using the IoT suite and later saved in the GIS database. The IoT messages trigger spatio-temporal data analytic function to extract driver behavior and vehicle usage data. These data get saved in the distributed ledger system on the blockchain ensuring transparency, trace-ability and safety. In case of Ethereum based framework the premium evaluation is done based on driver behavior and vehicle usage which is tokenized according to varying risk levels through a fine grained process.

4.4 Blockchain Big Data in Smart Grid

Blockchain technologies can contribute significantly to improve the efficiency of practices and processes in the energy sector. Blockchain integrated with big data has the ability to accelerate the speed of development of IoT platforms and digital applications thereby innovating the P2P energy trading and decentralization services. The present energy systems are experiencing radical transformations due to the advancement of distributed energy resources and use of information and communication technologies. The blockchain architectures have the capability to solve issues relevant to controlling and managing of decentralized energy systems and micro-grids [85]. Smart grid is a technology that makes electrical power grids more efficient, robust and less pollutant. The advanced metering infrastructure (AMI) is one of the major components in smart grid architecture that ensures two way communication between users and the utility device, by installing a smart meter at the user end. Key management plays a major role in this process and most of the traditional architectures depend on a single entity to distribute the keys and maintenance. The study by [86] proposes a distributed key management system to maintain optimum security in the smart grid system. A key agreement protocol is proposed between the utility

and the smart meter followed by the use of a distributed multi-case key management scheme which allows group members to effectively manage their group communication. The blockchain architecture enables distributed entities to interact with each other in the distributed P2P network ensuring security, scalability and efficiency. In [87], a data integration and regulation system is proposed based on consortium blockchain. A signcryption algorithm is implemented to multidimensional data acquisition and the receivers in the blockchain framework. As part of the regulation process, the control center, the grid operator and the grid supplier receive fixed blocks from the blockchain and later obtain plaintext from the decryption process. At the outset, multidimensional data are analyzed by the relevant receivers. This results in creation of control pieces. These control pieces takes care of the security and data integrity aspect thereby reducing communication costs. In [88], a blockchain based demand response management system is proposed. This system is termed as GUARDIAN and is capable of taking trading decisions pertaining to the energy sector. The system is extremely secured and also contributes significantly towards load management in the residential, industrial and commercial sector. The minor nodes in this framework termed as block verifiers, are selected using their specific power consumption and processing power capability. These nodes help in the authentication of energy transactions in the smart grid network. The transaction process in the proposed framework is initiated by the end user which creates the block of transaction for energy trading. The miner nodes validate these blocks, add them to the blockchain and they become eligible to be part of the energy trading. This helps in achieving security and eliminates unauthorized entries in energy trading.

4.5 Blockchain Big Data Projects

Blockchain is a technology that empowers cryptocurrencies such as bitcoin and ethereum. On the other hand, big data is an advanced concept of data science which involves larger dataset with great variety, size and velocity. These datasets are analyzed to reveal interesting patterns, association and trends. Interestingly, blockchain is a type of distributed ledger that records transactions in a way that cannot be altered. There is an immense trust factor associated with blockchain that eliminates the need of third parties to regulate transactions ensuring the data is immutable. Blockchains have many applications in data science where data integrity is maintained while performing data analysis and data sharing [89]. The benefits of the application of blockchain are applied in three areas namely for decentralized data storage, performing blockchain enabled data analysis and finally in maintaining blockchain enabled data security as shown in Fig. 4. Some of the blockchain projects for big data applications are discussed below.

4.5.1 Storj. Storj is an end-to-end decentralized storage project which utilizes the excess hardware and bandwidth capacity, enabling peer to peer authentication of storage contracts between the providers and the users [90]. The process involves encryption of the files at the client side which are the split into pieces termed as "shards". These shards are later stored three times to maintain backups at the farmer side. The client only has access to the data which provides additional security than the traditional centralized cloud services. The Storj cryptocurrency allows renters to check on the farmers files and also pay for the maintenance of this storage system. The renters pay only for the space used without any additional fees pertaining to user requirements of setup costs.

4.5.2 Omnilytics. Omnilytics is a blockchain platform for big data analytics that provides insights for sales, marketing and merchandising industry [91]. It uses blockchain, big data analytics, ML, AI and various other technologies to integrate data from different industries. The platform provides data analytics and related services for competitor benchmarking, trend analysis and pricing analysis for the clients. Blockchain is used to empower smart contracts, distributed data finger printing, data exchange and other services to track the trend of data, provide incentives through micropayments.

4.5.3 Rubix. Rubix blockchain [92] uses the concept of decentralization to integrate the cryptocurrency traders in a common trading platform to authenticate their credibility and predictions. The protocol is based on the

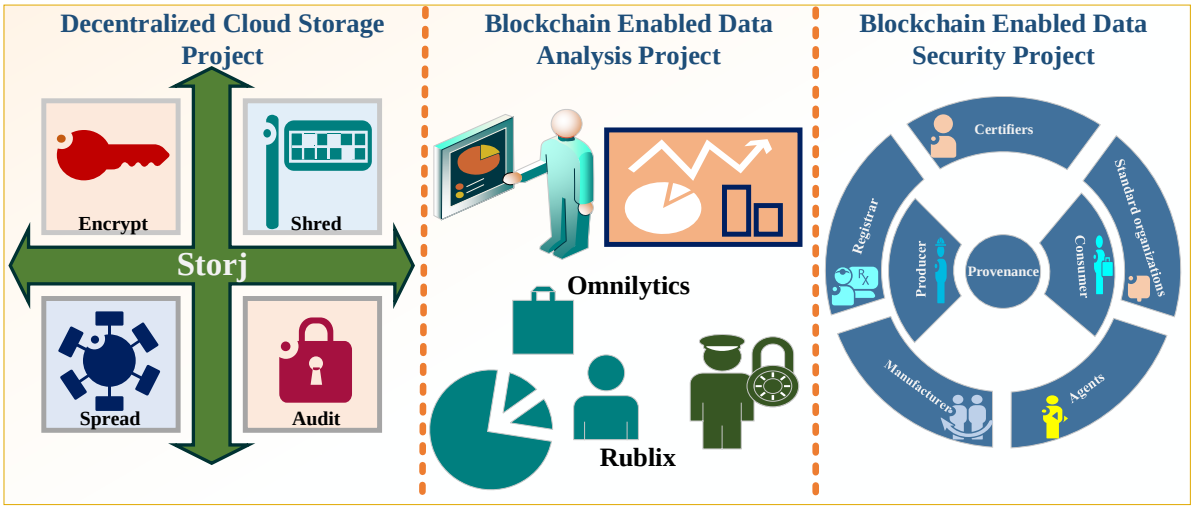

Fig. 4. Popular blockchain big data projects.

transparency and immutability attribute of blockchain in combination with investment data analytics to generate more accurate trading predictions. The traders as a result are ranked based on the accuracy of their predictions wherein the blockchain verifies the traders and incentivizes them based on the content quality.

4.5.4 Provenance. Provenance is a blockchain platform mainly used in supply chain management that helps to gather important product information and shares the same in a trusted, secure and accessible manner [93]. The blockchain architecture used encompasses of six participants namely - the producer, the manufacturer, registrar, standard organizations, agents like certifiers or auditors and finally the customers. The protocol provides access to information to its consumers on origin of the products, its journey along various points in the supply chain, product quality and its impact on the environment.

4.5.5 FileCoin. FileCoin intends to create a decentralized storage network that would allow traders to buy and sell storage in an open market. The FileCoin allows users to rent storage on devices having excess storage spaces using the filecoin cryptocurrency. The clients spend cryptocurrencies for sharing or retrieving of the data and miners earn the filecoins through storage and services of data. When the miners mine a particular block, they need to submit a proof-of-space-time (PoST) to the network, which validates if a storage provider is performing the required responsibilities for storing outsourced data for the stipulated time frame. The filecoin consists of blockchain, retrieval nodes, storage nodes and a native filecoin token. The storage nodes store sealed copies of data and the transactions are recorded by the blockchain. The retrieval nodes fetches and delivers the files to the users abiding to the PoST [94].

4.5.6 Datum. Datum (DAT) is a decentralized, distributed, high performance and NOSQL platform supported by Ethereum, Bigchain DB and IPFS. It basically enables users to store data anonymously and securely from social network, IoT devices and wearable technologies. The platform also acts as a marketplace for sharing and selling of data by providing Datum users with a unique Datum ID which are managed by the Datum mobile application available for Android and iOS services [95]. The summary of big data blockchain applications is described in Table 2.

Table 2. Summary of blockchain big data applications.

Ref.	Application	Description	Benefits	Limitations
[19]	Smart City	Use of hash, asymmetric encryption, consensus algorithm, blockchain and merkle tree.	-Tamper free transactions in IoT devices. -Development of decentralized, secure and auditable environment for IoT devices.	-Maintaining balance between privacy and accountability. -Implement Blockchain for crowd sensing.
[75]		Secured big data auditing scheme using DAB.	-Elimination of centralised third party auditors. -Improvement in reliability and stability.	NA.
[76]		Use of fog nodes and D2D to enable blockchain.	Prediction of risk or any exceptional event in smart contract.	Large scale testing in shared economy scenarios could be implemented.
[77]	Smart Healthcare	Access control of patient records using cryptographic techniques integrated with blockchain.	Provides security and privacy for IoT based health monitoring systems.	-Resource constraints of IoT acts as a major challenge. -Commercialization in collaboration with industry partners.
[78]		-Sensitive information collected using sensors are encrypted and distributed using blockchain. -Implemented in IoT based EHRs systems connected through WSN.	Ensures identity verification and fraud detection.	Implementation on large scale healthcare data.
[79]		Implements Ethereum smart contract where sensors communicate with the smart devices.	-Monitors patients in real-time. -Sends alerts for medical interventions.	-Large scale adaptation. -Resource utilization.
[80]	Smart Transportation	Implementation of Smart contract, track & trace, fast payment and supply chain finance using blockchain.	Data authentication, decentralization to provide knowledge in shipping, logistics, transportation .	-Varied successful applications considering transportation engineering still not predominant.
[82]		Implements a key management system for key transport in VC systems.	Captured vehicle departure data, ensures secured key transport and re-keying of vehicles.	-Maintaining of balance between security and privacy. -Pseudonym management can be included.
[83]		Enables users to travel between locations, make insurance and finance decisions on blockchain based disruptive technology.	Defines privacy policies and resolves issues to transportation route finding, car insurance and tracking of claims.	Use of GPS positioning of the users not included.
[84]		Implements a fine grained transportation insurance service based on vehicle usage, driver behavior using hyperledger and cryptocurrency.	Promotes safe driving and unbiased insurance claims.	Implementation in large scale in cities and other similar applications.
[85]	Smart Grid	Implementation of blockchain, distributed consensus algorithms in energy industry.	Innovations in P2P energy trading and decentralized energy generation.	Achieving market penetration and commercial viability.
[86]		Distributed key management for AMI in Smart grids.	Use of multi-case key management scheme for security of smart grids.	-Scalability to process transaction. -Difficulty in prediction of price due to volatility in supply and demand.
[87]		Implementation of signcryption algorithm for data integration and regulation system inblockchain framework.	-Security, of multidimensional data. -Reduction of communication costs.	Real-time analysis could be included.
[88]		Implements GUARDIAN, blockchain secured demand response management.	-Enables accelerated decision making for energy trading. -Load management in residential, industrial & commercial sector.	-Testing and deployment of the scheme on larger dataset. -Optimization of the scheme to reduce latency and increase network throughput.

5 RESEARCH CHALLENGES AND FUTURE DIRECTIONS

The robust technologies blockchain and big data are evolving in almost all domains. When these powerful technologies are integrated, the integration opens up new research opportunities due to massive data accumulations in today's data centers. Big data is evolving in business organizations today gaining higher profits. Similarly, data held up in the blockchains worth more through its sensitive nature. Blockchain validates the data ensuring quality in data management, whereas big data analytics makes better predictions on a large quantity of data. These technologies have specific challenging issues to be addressed when used individually and in combination with their adoption. The most prominent challenges are massive data silos in the big data environment that should be secured, ensuring integrity and repudiation in data transactions. Therefore, blockchain with its decentralized

Table 3. Research challenges in integration of blockchain and big data.

Ref.	Challenges	Application	Description	Benefits
[97]	Security and privacy enhancement in big data	Big data and cryptocurrency.	Integration of big data and cryptocurrency for decentralized data management.	Secured data sharing and decentralized data access.
[98]		Blockchain smart contracts for big data.	Vulnerability scan and programming correctness for security and correctness in smart contracts' operations.	Secured data sharing and privacy.
[33]		IoT big data, blockchain and fog computing.	Big data security in fog enabled IoT using blockchain.	Secured data transactions with low latency response.
[99]	Security and privacy in big data exchange	E-crime detection and bitcoin price predictions.	Interpretation of data stored in public blockchain.	Secured blockchain data transactions.
[100]		Smart toy assisted with MEC and blockchain.	Smart contracts are used for validating various data exchanges authorized using blockchain.	Secured and low-latency response in smart toy business.
[96]		Big data exchange and smart contract.	A fair way for protecting user data copyright and ensures privacy using SC.	Privacy in decentralized big data sharing.
[101]	Blockchain standardization	Blockchain and DLT.	High-level functional architecture for blockchain and DLT.	Standards for various functionalities in blockchain.
[102]		Early standardization for blockchain immutability.	Describes different levels of standardization and their importance.	Participatory standard for blockchain immutability.
[103]	Complexity in big data	Crime big data.	Mining of data using various state-of-art data mining techniques.	Efficient mining of data for crime departments.
[104]		Big data in mobile network optimization.	Explorers the features of big data from the perspective of users and network operators.	Effective mobile data management.
[105]	Computational overhead in blockchain	Cyber physical social systems.	A lightweight blockchain for big data.	Privacy Preserving data transaction with low-latency.
[106]		Blockchain in distributed databases.	Blockchain on distributed databases allows 1-million writes for a second.	Scalability and faster querying with sub-second latency.
[107]		Blockchain in cloud based healthcare big data.	Suggests off chain computation of healthcare data with control in the blockchain.	Secured and immutable medical transactions.
[108]		Blockchain for supply chains with 5G MEC.	One-way hash and bitwise rotation make the system light.	Low-latency response.
[109]	Network Virtualization	SDN, big data, blockchain and 5G MEC.	SDN and big data are integrated with faster 5G and immutable blockchain.	Faster query processing and secured data transactions.

framework and secured immutable nature will be an optimal choice. Indeed, blockchain possesses some challenges to be addressed on its deployment. The goal of integration is to store the massive data on the decentralized ledgers instead of centralized servers with authorized data access [96] and allowing the users to share their unused storage on the exchange of cryptocurrencies like bitcoins [97]. This section presents the key challenges and future directions related to blockchain big data research. The summary of research challenges upon the integration of blockchain and big data is described in Table 3.

5.1 Research Challenges

5.1.1 Security in Blockchain. The blockchain, a valid ledger keeps track of various digital transactions across diversified domains such as IoT applications (includes data transaction from heterogeneous devices), in fifth-generation (5G) network, healthcare and financial services. Some of the notable services of the blockchain through decentralization are data security and privacy demanding more computational power (about 50%) for the malicious users trying to deceive the block information [110]. This type of attack is called 51% attack. Though 51% of the computational resources are required for any user to deceive information from the blocks, the double-spending

attack is still possible. Blockchain smart contracts reinforce the environment to avoid the double-spend attacks [98]. Blockchains' distributed nature (which shares each transaction network-wide) will induce greater complexity for fraudulent block transactions. As the blockchain stores history of all transaction in the same state as it was performed makes it an essential candidate for big data application. Data-intensive applications like the healthcare industry where big data is employed for managing the voluminous data from medical practitioners, patients, clinicians, laboratory and pharmaceutical requires privacy-preserving data sharing. The researchers suggest that the secured blockchain framework can be employed for controlled access to the voluminous data using its decentralized data management [12].

Moreover, the permissionless blockchains which allow any user to join the chain without permission are secured by the hyper ledger [111]. Hyperledger strengthens the permissionless blockchain by allowing the users involved in the transaction to join through permissioned blockchain guaranteeing data provenance. One of the possible network threats to these kinds of a public blockchain is a Sybil attack, which enables a node in a blockchain to add enormous malicious users under its control [112]. But, the PoW consensus algorithms will mitigate these attacks by allowing the malicious node to devote more of its computational resources to accomplish the attack. An analysis in [33] suggests that resource-constrained IoT environments where more data is accumulated from varied sources cannot be secured with traditional cryptographic infrastructure. Also, the authors recommend that the secured, distributed, and anonymized nature of blockchain is essential challenger for such environments. Furthermore, lightweight blockchain should be preferred for optimal computational resource utilization in the resource-constrained IoT environment with big data services.

The distributed ledger technology ensures trusted data transactions with immutability and transparency via peer-to-peer networking services. Blockchain assures better scalability than centralized architectures [113]. But, as the chain grows longer and longer, the entries in the blockchain will be more and computational load in processing the data will increase tremendously. In blockchain applications like IoT, the nodes are simple and resource-constrained. Still, the security capability using cryptographic functions in blockchain consumes more computational resource for key exchange, encryption, decryption and digital signatures. The miners (the node that performs mining) which are responsible for creating new blocks and linking it with the existing chain requires higher computational load.

Therefore, the cryptographic techniques or security measures used for enhanced security in the blockchain environment should not impose more computational resources. The application of blockchain is increasing every day with an increase in the complexity of the data stored in the blockchain. Henceforth, blockchain data analytics must be explored to ensure better performance of blockchain with varied complexity of the data [99]. Also, before integrating the blockchain with other technologies like big data, the type of blockchain public or private, security measures adopted, data processing capabilities should be considered for network safety and better performance.

5.1.2 Standardization. Blockchain was initially developed as the solution to the problem of digital cash (the cryptocurrency named bitcoin). It facilitates secure transaction of digital assets over different banks. Blockchain automized the global payment over the Internet irrespective of any topographical constraints within hours. Whereas the traditional financial system takes many days to perform any financial transactions worldwide. Nevertheless, the scope of adoption of blockchain has been hindered by its interoperability challenges. These challenges not only include the differences among different cryptocurrencies but also consists of the differences in the diversified transaction. Therefore, it is tedious for the blockchains to interoperate and integrate compatibly with the legacy systems. This, in turn, may hinder the regulatory acceptance of blockchains. One possible solution for this type of open systems is standardization to provide common technical guidelines for any industry.

An analysis of blockchain terminologies and various initiatives taken by non-profit organizations for standardization for blockchain was carried out in [101]. For standardizing distributed ledger technology (DLT) and blockchain, the international organization for standardization (ISO) which develops and publishes standards, has

formed an ISO/TC 307 technical committee led by standards Australia for standardizing DLT and blockchain. The primary motive of this committee is to publish the standards related to blockchain privacy, taxonomy, smart contract, security (for users and data), privacy, interoperability, governance, and various use cases of blockchain. The different workgroups and their activities under ISO/TC 307 are summarized in [114]. International telecommunication union (ITU), the working group under ISO, focuses on identifying and standardizing the DLT application, its services, best practise to be adopted for its implementation and further research on related standards. The world wide web consortium (W3C) which implements web standards has initiated standards for developing blockchain message formats (ISO20022), guidelines on blockchain storage (public, private and side-chain) and approving the use-cases. IEEE has developed a standard framework for blockchain use in IoT and a handbook on blockchain asset exchange. The Internet engineering task force (IETF), an open group that develops interoperability standards for network communications has a greater impact on blockchain standardization.

Furthermore, a framework in [102] was designed for the implementation of blockchain immutability. Also, they have discussed the effects of early standardization in blockchain immutability. They suggest that three different types of standards, namely, anticipatory, participatory, and responsive standards. The anticipatory standards are developed before the acceptance of a new service or technology. The participatory standard is developed and adopted during the implementation of the technology to test the conformance specification. And the responsive standards are adopted after technology adoption (or during its evolution). A framework for participatory standards to deploy the immutability concept of blockchain and its operation is discussed.

Blockchain for big data allows the data sharing in cross-domain environment irrespective of the risk factors concerned with accumulating data from various data silos. Therefore, while adopting blockchain, proper standards and guidelines should adhere to the smooth functioning of the technology.

5.1.3 Complexity of Big Data. The emergence of cloud computing, smart IoT applications have led to the massive accumulation of data. Along with the enormous growth of data in this information age, data management issues like inaccessible data, dirty or unclean data, and data privacy have also increased [115]. With the advent of big data, data quality management is more challenging. Furthermore, while handling more significant and complex datasets, the companies should ensure the authenticity of data source, cleanliness, and data breach. Because of this, the complete digital transformation of entire legacy data is still a challenging issue. The security perspective of the data management can be assured by blockchain, but yet, the complexity in big data management should be considered on its integration.

The prominent challenges of big data are due to the nature of the data, conventional analysis models and inefficient data processing systems. The big data is inherently complex, making it challenging to represent and interpret, thereby increasing the computational complexity. The big data has heterogenic sources that exhibit different patterns and behaviors. Some of the essential characteristics are complex data type, its structure, more intricate relationships and wide-ranging quality. The big data mining activities such as data retrieval, analyzing the topic, text mining (sentiments and semantics extraction) will be challenging than the traditional data [103]. The lack of knowledge of these characteristics and domain-specific data processing techniques will result in inefficient computational models. A clear understanding of the attributes of inherently complex big data is mandatory for designing the computational models with the highest level of abstraction. Apart from diversified sources and massive volume, the critical feature of big data is its dynamically changing data (real-time information) [116].

The big data processing systems are complex enough in handling the inherent complexity of the big data. These systems were built with high processing capability with more computational resource requirement. The system complexity includes the elaborate architecture, different processing modes and computing requirements. Basic knowledge of the system complexity will directly impact the performance of the big data systems. Also, the parameters affecting the energy utilization of big data processing systems must be considered while designing a

robust framework. Some of them are system throughput, energy consumption, resource utilization, distributed data storage, parallel computation and accuracy in job calculation.

Furthermore, big data offers more chances for mobile networks to improve their service quality. The study in [104] has explored the integration of big data with mobile network optimization, with a focus on investigating the characteristics of big data from the perspective of the mobile network operator and users. The user-specific data obtained from user equipment include profile (location, communication pattern), behavior and other application data. The data of network operators include data from the core network, radio access network and Internet service providers. The core network provides data related to network performance, call details and application usage index. Information sourced from radio access network includes cell configuration, mobility, handover details, resource utilization (source details and link utilization), interference details, signal measurements and notification signal messages among different components in the mobile network. The effectiveness of the service laid by mobile network depends on how effectively network operators process this information and make valuable decisions. Efficient data analytic mechanisms are essential for better network optimization.

Therefore, the complex nature of the big data must be ensured while integrating it with blockchain as it will improve the way how the data is handled in big data processing models. Also, the mapping between complexity vs computation, energy consumption vs efficiency should be evaluated for laying out effective means of data sharing, trusted transactions, data access, intruder detection and enhanced security through decentralized blockchains.

5.2 Future Directions

Big data which is proprietary for variability, volume, veracity, value and complexity, requires the data processing systems with higher computational capability. Also, the decentralized distributed ledger blockchain offers immutable, secured, and transparent data transactions that require more computational power for effective services. Upon integration of these complicated big data with blockchain, incur an unexpected computational complexity leading to the poor performance of the system. Therefore adaptive blockchain designs should be preferred, thereby alleviating the computational resource utilization for blockchain and 5G network communication can be utilized for faster services. This section presents the future directions for the integration of blockchain with big data.

5.2.1 Adaptive Blockchain Design for Big Data. The adaptive blockchain reduces the computational power required for processing the blocks even if the chain grows exponentially. The most preferred adaptive blockchain designs are lightweight blockchain for real-time big data and scalable blockchain for large scale big data.The framework in [105] was designed for large scale and real-time big data application cyber-physical social systems (which integrates cyber, physical and social systems) uses blockchain for access control. The framework uses fog computing at the edge nodes for processing the local data dynamically. A lightweight symmetric algorithm is used for encryption for privacy-preserving data transactions. The cyber-physical social system big data is accessed using the account address of blockchain node. The access control details such as authentication, authorization are stored and managed in a blockchain. Experiments proved that the system is feasible and efficient, but it incurs more time to ensure privacy as all the authorizations are performed in the blockchain. Also, the retrieval mechanisms must be strengthened for better performance.

BigChainDB, a scalable blockchain for distributed big data was proposed in [106] by integrating the traditional blockchain with distributed databases to attain scalability through faster querying mechanisms. Blockchain network allows a user to join the chain based on the consensus. PoW and PoS are basic consensuses approach evolved to allow anyone to enter the network based on hash rate and stake(digital coins) respectively. BigChainDB can let 1-million writes for a second with sub-second latency (less than a second) and petabyte capacity. Also, it enforces permissioning system that can be employed for both public and private blockchains. Similarly, HBasechainDB framework in [117], a scalable blockchain for big data store deployed on the Hadoop ecosystem.

The immutable and decentralization nature of the blockchain is built on the Hbase database of Hadoop. Instead of scaling the chain, the blockchain is implemented on a distributed database. Results proved that the system scales up linearly with sub-second latency and higher transaction throughputs. But, this framework is well suited for blockchain adoption to the organizations running on the Hadoop ecosystem. Therefore, a standard adaptive framework supporting different types of blockchain for the big data system should be explored.

5.2.2 Blockchain for 5G Big Data. The diverse requirements change and an exponential increase in the number of mobile devices make it difficult for the 4G to meet future demands. Though 5G has been integrated with technologies like SDN, Cloud, ML, network virtualization, it is still hard to meet the diverse requirements change. Also, these technologies procure different types of challenges in terms of decentralization, security and privacy, transparency, interoperability and immutability. Therefore, the blockchain with its feature will be an essential aspirant for massive computation with 5G big data applications [118]. In the 5G era, big data can rely on some key technologies to build its platforms such as cloud computing, mobile edge computing, and software defined networking. In this context, blockchain can come as viable solution for realizing 5G technology-based big data services.

Blockchain for cloud-based big data: A framework was proposed in [119] to enhance the current building information modeling (BIM) by integrating tamper-resistant blockchain and mobile cloud through big data sharing. BIM collects a huge amount of data throughout the project and needs to access the historical data for making certain decisions. The framework uses a private blockchain authorized by a trusted center for audit and data provenance in BIM as a cloud service (outsourced storage and computation). The system is evaluated based on the block size, security, hashing time and packaging period. The results show that blockchain can effectively resolve security issues and data quality with BIM and promotes it further. Likewise, a blockchain framework for cloud-based healthcare data handling was proposed in [107]. Though blockchain ensures immutability and secured transaction of medical records stored in the cloud, the authors suggest off-chain storage for healthcare data due to protection of privacy laws and dynamic changing behavior of healthcare data (old record might be unusable). But, the immutable hashes can be stored on the chain, whereas off-chain data can be modified saved as a distributed database.

Blockchain for mobile edge computing based big data: The integration of mobile edge computing (MEC) and blockchain is a right solution to achieve low latency response which shrinks the computing power by local data processing. A distributed blockchain framework was proposed in [120] for privacy protection with heterogeneous MEC in 5G and beyond networks. The consensus in multidomain collaborative routing is achieved without exposing the network topology. Blockchain is used for multilevel mutual trust and collaborative routing. The privacy of heterogeneous MEC collaboration in 5G is highly improved with blockchain adoption.

A prototype for a smart toy, an IoT device based on edge computing and blockchain (hyper ledger fabric) was proposed in [100] for secured data exchange. Smart toy environment has many different types of data exchanges among single supplier, single demander, multiple supplier and multiple demanders. For each data exchange, the ids are stored and validated through smart contracts (handles accounting). Chaincode, a blockchain-based smart contract handles the complex accounting in the smart toy business. The edge computing is used for local client computations ensuring low- latency response. The framework ensures a secured, flexible, scalable and confidential data exchange among smart toy business participants. Similarly, a lightweight blockchain with RFID enabled authentication system for supply chains with 5G MEC was proposed in [108]. The one-way hash function and bitwise rotation make the system light with less computational power compared to existing protocols.

Blockchain for software defined networking-based big data: Software defined networking (SDN) makes the network agile and flexible by improving the network control with faster response for changing requirements. Upon integration of big data and SDN, both the technologies are benefited seamlessly [121]. SDN can assist big data in solving most of its issues like data delivery, data processing in the cloud, transmission, data scheduling and

resource optimization. Likewise, big data can assist SDN in handing traffic data, security vulnerabilities and inter and intra data center network communications. Though they serve each other effectively, there are some open issues to be resolved. The overflooded queries in the SDN controller may degrade its performance, in turn, making it insufficient for accommodating more extensive big data entries, thereby imposing the scalability challenge and central point of failure. Resource-constrained and unintelligent network switches can flood the SDN controller with raw data packets imposing a higher computational overhead. Also there is no high-level programming environment available for the development of big data applications. Above all, security vulnerabilities are more in SDN. Therefore the SDN and big data can serve each other better when integrated with faster 5G and the scalable, immutable and decentralized blockchain. The 5G enables faster processing speed, whereas blockchain can resolve scalability and security issues in SDN [109]. The fusion of SDN with blockchain and MEC [122] will solve the majority of issues in network virtualization, making it fit for processing big data applications.

Therefore, big data will be promoted when the blockchain technology meets MEC, SDN, Cloud with robust 5G communications. All these technologies together ensure security and privacy, scalability, parallel computing, transparency and trust when individual challenges are met for compatible functioning of the underlying technologies.

6 CONCLUSIONS

Blockchain is a disruptive ledger technology that has sparked a significant interest to support big data systems with high security and efficient network management. In this article, we have conducted a state-of-the-art review on the application of blockchain for big data. We have first discussed the recent advances in blockchain and big data and explained the motivation behind the integration of these two technologies. Particularly, we have provided an extensive survey on the use of blockchain in a number of key big data services, including big data acquisition, big data storage, big data analytics, and big data privacy preservation. Then, we have explored the opportunities brought by blockchain in important big data applications, such as smart city, smart healthcare, smart transportation, and smart grid. The emerging blockchain-big data platforms and projects have been also highlighted and analyzed. From the extensive literature review on blockchain-big data services and applications, we have identified some key technical challenges and pointed out possible future directions to spur further research in this promising area.

REFERENCES

[1] "Big data market worth $229.4 billion by 2025," 2020. [Online]. Available: https://www.marketsandmarkets.com/PressReleases/big-data.asp

[2] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *IEEE access*, vol. 2, pp. 652–687, 2014.

[3] J. Gantz and D. Reinsel, "Extracting value from chaos," *IDC iview*, vol. 1142, no. 2011, pp. 1–12, Jun. 2011.

[4] J. Manyika, "Big data: The next frontier for innovation, competition, and productivity," 2011. [Online]. Available: http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

[5] S. Pouyanfar, Y. Yang, S.-C. Chen, M.-L. Shyu, and S. Iyengar, "Multimedia big data analytics: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 1, pp. 1–34, 2018.

[6] Z. Su and Q. Xu, "Security-aware resource allocation for mobile social big data: A matching-coalitional game solution," *IEEE Transactions on Big Data*, 2017, in press.

[7] J. Wu, K. Ota, M. Dong, J. Li, and H. Wang, "Big data analysis-based security situational awareness for smart grid," *IEEE Transactions on Big Data*, vol. 4, no. 3, pp. 408–417, Sep. 2018.

[8] C. H. Liu, Q. Lin, and S. Wen, "Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3516–3526, Jun. 2019.

[9] G. Liu, H. Dong, Z. Yan, X. Zhou, and S. Shimizu, "B4SDC: A blockchain system for security data collection in MANETs," *IEEE Transactions on Big Data*, 2020, in press.

[10] 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 Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4187–4195, Jun. 2020.

[11] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *International Journal of Web and Grid Services*, vol. 14, no. 4, pp. 352–375, Oct. 2018.
[12] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, vol. 107, pp. 841–853, Jun. 2020.
[13] Q. Feng, D. He, S. Zeadally, M. K. Khan, and N. Kumar, "A survey on privacy protection in blockchain system," *Journal of Network and Computer Applications*, vol. 126, pp. 45–58, Jan. 2019.
[14] J. Sengupta, S. Ruj, and S. D. Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT," *Journal of Network and Computer Applications*, vol. 149, p. 102481, Jan. 2020.
[15] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. H. Nyang, and D. Mohaisen, "Exploring the attack surface of blockchain: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, 2020, in press.
[16] 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. 47 615–47 643, 2019.
[17] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, "On blockchain and its integration with IoT. challenges and opportunities," *Future generation computer systems*, vol. 88, pp. 173–190, Nov. 2018.
[18] H.-N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for Internet of things: A survey," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8076–8094, Oct. 2019.
[19] 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 Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1676–1717, Secondquarter 2018.
[20] 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 Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1508–1532, Secondquarter 2019.
[21] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for 5G and beyond networks: A state of the art survey," *Journal of Network and Computer Applications*, p. 102693, Sep. 2020.
[22] P. Zhuang, T. Zamir, and H. Liang, "Blockchain for Cyber Security in Smart Grid: A Comprehensive Survey," *IEEE Transactions on Industrial Informatics*, 2020, in press.
[23] M. B. Mollah, J. Zhao, D. Niyato, K.-Y. Lam, X. Zhang, A. M. Ghias, L. H. Koh, and L. Yang, "Blockchain for future smart grid: A comprehensive survey," *IEEE Internet of Things Journal*, 2020, in press.
[24] 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," *Journal of Network and Computer Applications*, vol. 71, pp. 151–166, Aug. 2016.
[25] M. Ge, H. Bangui, and B. Buhnova, "Big data for Internet of things: A survey," *Future generation computer systems*, vol. 87, pp. 601–614, Oct. 2018.
[26] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2923–2960, Fourthquarter 2018.
[27] M. Ghorbanian, S. H. Dolatabadi, and P. Siano, "Big data issues in smart grids: A survey," *IEEE Systems Journal*, vol. 13, no. 4, pp. 4158–4168, Dec. 2019.
[28] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, Jan. 2019.
[29] X. Cheng, L. Fang, L. Yang, and S. Cui, "Mobile big data: The fuel for data-driven wireless," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1489–1516, Oct. 2017.
[30] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, Thirdquarter 2019.
[31] M. A. Al-Garadi, A. Mohamed, A. Al-Ali, X. Du, I. Ali, and M. Guizani, "A survey of machine and deep learning methods for Internet of things (IoT) security," *IEEE Communications Surveys & Tutorials*, 2020, in press.
[32] E. Karafiloski and A. Mishev, "Blockchain solutions for big data challenges: A literature review," in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*, Ohrid, Macedonia, 2017, pp. 763–768.
[33] N. Tariq, M. Asim, F. Al-Obeidat, M. Zubair Farooqi, T. Baker, M. Hammoudeh, and I. Ghafir, "The security of big data in fog-enabled IoT applications including blockchain: a survey," *Sensors*, vol. 19, no. 8, p. 1788, Apr. 2019.
[34] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Integration of blockchain and cloud of things: Architecture, applications and challenges," *IEEE Communications Surveys & Tutorials*, 2020.
[35] Q. Zhou, H. Huang, Z. Zheng, and J. Bian, "Solutions to scalability of blockchain: A survey," *IEEE Access*, vol. 8, pp. 16 440–16 455, 2020.
[36] W. Viriyasitavat and D. Hoonsopon, "Blockchain characteristics and consensus in modern business processes," *Journal of Industrial Information Integration*, vol. 13, pp. 32–39, Mar. 2019.
[37] L. Da Xu and W. Viriyasitavat, "Application of blockchain in collaborative internet-of-things services," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1295–1305, 2019.
[38] C. Berg, S. Davidson, and J. Potts, *Understanding the blockchain economy: An introduction to institutional cryptoeconomics.* Edward Elgar Publishing, 2019.

[39] Y. Yuan and F.-Y. Wang, "Blockchain and cryptocurrencies: Model, techniques, and applications," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1421–1428, 2018.
[40] J. Abou Jaoude and R. G. Saade, "Blockchain applications–usage in different domains," *IEEE Access*, vol. 7, pp. 45 360–45 381, 2019.
[41] S. Shalini and H. Santhi, "A survey on various attacks in bitcoin and cryptocurrency," in *International Conference on Communication and Signal Processing (ICCSP)*, 2019, pp. 0220–0224.
[42] J. Parkin, "The senatorial governance of bitcoin: making (de) centralized money," *Economy and society*, vol. 48, no. 4, pp. 463–487, 2019.
[43] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: current status, classification and open issues," *Telematics and Informatics*, vol. 36, pp. 55–81, Mar. 2019.
[44] J. Zhang, S. Zhong, T. Wang, H.-C. Chao, and J. Wang, "Blockchain-based systems and applications: A survey," *Journal of Internet Technology*, vol. 21, no. 1, pp. 1–14, Jan. 2020.
[45] H. Huang, W. Kong, S. Zhou, Z. Zheng, and S. Guo, "A survey of state-of-the-art on blockchains: Theories, modelings, and tools," *arXiv preprint arXiv:2007.03520*, 2020.
[46] A. K. Dey, C. G. Akcora, Y. R. Gel, and M. Kantarcioglu, "On the role of local blockchain network features in cryptocurrency price formation," *Canadian Journal of Statistics*, vol. 48, no. 3, pp. 561–581, Sep. 2020.
[47] J. Angelis and E. R. da Silva, "Blockchain adoption: A value driver perspective," *Business Horizons*, vol. 62, no. 3, pp. 307–314, 2019.
[48] Y. Chen and C. Bellavitis, "Blockchain disruption and decentralized finance: The rise of decentralized business models," *Journal of Business Venturing Insights*, vol. 13, p. e00151, Jun. 2020.
[49] R. Zhang, R. Xue, and L. Liu, "Security and privacy on blockchain," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–34, Jul. 2019.
[50] S. Leonardos, D. Reijsbergen, and G. Piliouras, "PREStO: A systematic framework for blockchain consensus protocols," *IEEE Transactions on Engineering Management*, 2020, in press.
[51] L. Zhang, Y. Xie, Y. Zheng, W. Xue, X. Zheng, and X. Xu, "The challenges and countermeasures of blockchain in finance and economics," *Systems Research and Behavioral Science*, vol. 37, no. 4, pp. 691–698, Jul. 2020.
[52] A. Jindal, N. Kumar, and M. Singh, "A unified framework for big data acquisition, storage, and analytics for demand response management in smart cities," *Future Generation Computer Systems*, vol. 108, pp. 921–934, Jul. 2020.
[53] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big data technologies: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, Oct. 2018.
[54] H. V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi, "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, Jul. 2014.
[55] "Cisco annual internet report (2018-2023)," 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html
[56] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE network*, vol. 30, no. 3, pp. 22–29, May-Jun. 2016.
[57] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54 776–54 788, 2020.
[58] D. B. Rawat, R. Doku, and M. Garuba, "Cybersecurity in big data era: From securing big data to data-driven security," *IEEE Transactions on Services Computing*, 2019, in press.
[59] M. Tang, M. Alazab, and Y. Luo, "Big data for cybersecurity: Vulnerability disclosure trends and dependencies," *IEEE Transactions on Big Data*, vol. 5, no. 3, pp. 317–329, Sep. 2019.
[60] P. Sharma, R. Jindal, and M. D. Borah, "Blockchain technology for cloud storage: A systematic literature review," *ACM Computing Surveys (CSUR)*, vol. 53, no. 4, pp. 1–32, 2020.
[61] 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. 79 764–79 800, 2020.
[62] C. H. Liu, Q. Lin, and S. Wen, "Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3516–3526, Jun. 2019.
[63] 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 Transactions on Parallel and Distributed Systems*, vol. 30, no. 4, pp. 870–882, Sep. 2018.
[64] 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. 64 486–64 498, Aril 2020.
[65] 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. 59 389–59 401, 2020.
[66] H. Li and D. Han, "EduRSS: a blockchain-based educational records secure storage and sharing scheme," *IEEE Access*, vol. 7, pp. 179 273–179 289, 2019.
[67] C. Yang, Q. Huang, Z. Li, K. Liu, and F. Hu, "Big data and cloud computing: innovation opportunities and challenges," *International Journal of Digital Earth*, vol. 10, no. 1, pp. 13–53, 2017.

[68] M. Shen, J. Zhang, L. Zhu, K. Xu, and X. Tang, "Secure SVM training over vertically-partitioned datasets using consortium blockchain for vehicular social networks," *IEEE Transactions on Vehicular Technology*, 2019, in press.
[69] K. Salah, M. H. U. Rehman, N. Nizamuddin, and A. Al-Fuqaha, "Blockchain for AI: Review and open research challenges," *IEEE Access*, vol. 7, pp. 10 127–10 149, 2019.
[70] L.-A. Hîrtan and C. Dobre, "Blockchain privacy-preservation in intelligent transportation systems," in *2018 IEEE International Conference on Computational Science and Engineering (CSE)*. IEEE, 2018, pp. 177–184.
[71] J. B. Bernabe, J. L. Canovas, J. L. Hernandez-Ramos, R. T. Moreno, and A. Skarmeta, "Privacy-preserving solutions for blockchain: review and challenges," *IEEE Access*, vol. 7, pp. 164 908–164 940, 2019.
[72] S. Kim and G. C. Deka, *Advanced applications of blockchain technology*. Springer, 2020, in press.
[73] J. Chen, Z. Lv, and H. Song, "Design of personnel big data management system based on blockchain," *Future Generation Computer Systems*, vol. 101, pp. 1122–1129, Dec. 2019.
[74] A. A. Monrat, O. Schelén, and K. Andersson, "A survey of blockchain from the perspectives of applications, challenges, and opportunities," *IEEE Access*, vol. 7, pp. 117 134–117 151, 2019.
[75] H. Yu, Z. Yang, and R. O. Sinnott, "Decentralized big data auditing for smart city environments leveraging blockchain technology," *IEEE Access*, vol. 7, pp. 6288–6296, 2018.
[76] 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. 18 611–18 621, 2019.
[77] A. D. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors*, vol. 19, no. 2, p. 326, 2019.
[78] T. McGhin, K.-K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *Journal of Network and Computer Applications*, vol. 135, pp. 62–75, Jun. 2019.
[79] J. D. Vyas, M. Han, L. Li, S. Pouriyeh, and J. S. He, "Integrating blockchain technology into healthcare," in *Proceedings of the 2020 ACM Southeast Conference*, 2020, pp. 197–203.
[80] S. Wang and X. Qu, "Blockchain applications in shipping, transportation, logistics, and supply chain," in *Smart Transportation Systems 2019*. Springer, 2019, pp. 225–231.
[81] 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, 2020.
[82] A. Lei, H. Cruickshank, Y. Cao, P. Asuquo, C. P. A. Ogah, and Z. Sun, "Blockchain-based dynamic key management for heterogeneous intelligent transportation systems," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1832–1843, Dec. 2017.
[83] L.-A. Hîrţan, C. Dobre, and H. González-Vélez, "Blockchain-based reputation for intelligent transportation systems," *Sensors*, vol. 20, no. 3, p. 791, 2020.
[84] Z. Li, Z. Xiao, Q. Xu, E. Sotthiwat, R. S. M. Goh, and X. Liang, "Blockchain and iot data analytics for fine-grained transportation insurance," in *IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, 2018, pp. 1022–1027.
[85] 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," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 143–174, Feb. 2019.
[86] M. Baza, M. M. Fouda, M. Nabil, A. T. Eldien, H. Mansour, and M. Mahmoud, "Blockchain-based distributed key management approach tailored for smart grid," in *Combating Security Challenges in the Age of Big Data*. Springer, 2020, pp. 237–263.
[87] M. Fan and X. Zhang, "Consortium blockchain based data aggregation and regulation mechanism for smart grid," *IEEE Access*, vol. 7, pp. 35 929–35 940, 2019.
[88] A. Jindal, G. S. Aujla, N. Kumar, and M. Villari, "GUARDIAN: Blockchain-based secure demand response management in smart grid system," *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 613–624, Jul.-Aug. 2020.
[89] L. Lao, Z. Li, S. Hou, B. Xiao, S. Guo, and Y. Yang, "A survey of IoT applications in blockchain systems: Architecture, consensus, and traffic modeling," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, pp. 1–32, 2020.
[90] 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," *Digital Investigation*, vol. 29, pp. 28–42, Jun. 2019.
[91] J. Moreno, E. B. Fernandez, E. Fernandez-Medina, and M. A. Serrano, "BlockBD: a security pattern to incorporate blockchain in big data ecosystems," in *Proceedings of the 24th European Conference on Pattern Languages of Programs*, 2019, pp. 1–8.
[92] J. Kokina, R. Mancha, and D. Pachamanova, "Blockchain: Emergent industry adoption and implications for accounting," *Journal of Emerging Technologies in Accounting*, vol. 14, no. 2, pp. 91–100, 2017.
[93] H. M. Kim and M. Laskowski, "Toward an ontology-driven blockchain design for supply-chain provenance," *Intelligent Systems in Accounting, Finance and Management*, vol. 25, no. 1, pp. 18–27, 2018.
[94] J. Benet, *Filecoin (CoinDesk)*, April 2020(Accessed on July 10, 2020), https://www.coindesk.com/crypto/filecoin-fil.
[95] R. Haenni, *Datum (blockchain)*, November 30, 2017(Accessed on June 5, 2020), https://golden.com/wiki/Datum_(blockchain)-PBWGX9G.

[96] Y. Chen, J. Guo, C. Li, and W. Ren, "FaDe: a blockchain-based fair data exchange scheme for big data sharing," *Future Internet*, vol. 11, no. 11, p. 225, 2019.
[97] H. Hassani, X. Huang, and E. Silva, "Big-Crypto: big data, blockchain and cryptocurrency," *Big Data and Cognitive Computing*, vol. 2, no. 4, p. 34, 2018.
[98] J. Liu and Z. Liu, "A survey on security verification of blockchain smart contracts," *IEEE Access*, vol. 7, pp. 77 894–77 904, 2019.
[99] C. G. Akcora, M. F. Dixon, Y. R. Gel, and M. Kantarcioglu, "Blockchain data analytics," *Journal of IEEE Intelligent Informatics*, p. 4, 2018.
[100] J. Yang, Z. Lu, and J. Wu, "Smart-toy-edge-computing-oriented data exchange based on blockchain," *Journal of Systems Architecture*, vol. 87, pp. 36–48, Jun. 2018.
[101] V. Gramoli and M. Staples, "Blockchain standard: Can we reach consensus?" *IEEE Communications Standards Magazine*, vol. 2, no. 3, pp. 16–21, Sep. 2018.
[102] F. Hofmann, S. Wurster, E. Ron, and M. Böhmecke-Schwafert, "The immutability concept of blockchains and benefits of early standardization," in *2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K)*. IEEE, 2017, pp. 1–8.
[103] 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. 106 111–106 123, 2019.
[104] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE network*, vol. 30, no. 1, pp. 44–51, Jan.-Feb. 2016.
[105] L. Tan, N. Shi, C. Yang, and K. Yu, "A blockchain-based access control framework for cyber-physical-social system big data," *IEEE Access*, vol. 8, pp. 77 215–77 226, 2020.
[106] T. McConaghy, R. Marques, A. Müller, D. De Jonghe, T. McConaghy, G. McMullen, R. Henderson, S. Bellemare, and A. Granzotto, "BigchainDB: a scalable blockchain database," *white paper, BigChainDB*, 2016.
[107] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K. R. Choo, "Blockchain: A panacea for healthcare cloud-based data security and privacy?" *IEEE Cloud Computing*, vol. 5, no. 1, pp. 31–37, Jan 2018.
[108] 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 Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7081–7093, 2020.
[109] G. S. Aujla, M. Singh, A. Bose, N. Kumar, G. Han, and R. Buyya, "BlockSDN: blockchain-as-a-service for software defined networking in smart city applications," *IEEE Network*, vol. 34, no. 2, pp. 83–91, Mar.-Apr. 2020.
[110] C. Chen, J. Wang, F. Qiu, and D. Zhao, "Resilient distribution system by microgrids formation after natural disasters," *IEEE Transactions on smart grid*, vol. 7, no. 2, pp. 958–966, Mar. 2016.
[111] S. Namasudra, G. C. Deka, P. Johri, M. Hosseinpour, and A. H. Gandomi, "The revolution of blockchain: State-of-the-art and research challenges." *Archives of Computational Methods in Engineering*, 2020.
[112] H. Halpin and M. Piekarska, "Introduction to security and privacy on the blockchain," in *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2017, pp. 1–3.
[113] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka, "Security services using blockchains: A state of the art survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 858–880, Firstquarter 2019.
[114] H. Cha, W. Lee, Y. Choi, J. Lee, and K. Lee, "International standardization on blockchain," *Electronics and Telecommunications Trends*, vol. 34, no. 2, pp. 110–120, 2019.
[115] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and challenges of big data research," *Big Data Research*, vol. 2, no. 2, pp. 59–64, 2015.
[116] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 995–1004.
[117] M. S. Sahoo and P. K. Baruah, "HBasechainDB-a scalable blockchain framework on hadoop ecosystem," in *Asian Conference on Supercomputing Frontiers*. Springer, 2018, pp. 18–29.
[118] M. Tahir, M. H. Habaebi, M. Dabbagh, A. Mughees, A. Ahad, and K. I. Ahmed, "A review on application of blockchain in 5G and beyond networks: Taxonomy, field-trials, challenges and opportunities," *IEEE Access*, vol. 8, pp. 115 876–115 904, 2020.
[119] 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," *Mathematical Problems in Engineering*, vol. 2019, pp. 31–37, Mar. 2019.
[120] 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 Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7094–7104, Nov. 2020.
[121] 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 Network*, vol. 30, no. 1, pp. 58–65, Jan.-Feb. 2016.
[122] D. B. Rawat, "Fusion of software defined networking, edge computing, and blockchain technology for wireless network virtualization," *IEEE Communications Magazine*, vol. 57, no. 10, pp. 50–55, Oct. 2019.