Data Analysis on Blockchain Distributed File Systems: Systematic Literature Review

Miguel Rodrigues Baptista
Instituto Superior Técnico & INOV
Lisbon, Portugal *miguelbaptista@tecnico.ulisboa.pt*

Miguel Mira da Silva
Instituto Superior Técnico & INOV
Lisbon, Portugal *mms@tecnico.ulisboa.pt*

Paulo Rupino da Cunha
University of Coimbra DEI & CISUC
Coimbra, Portugal *rupino@dei.uc.pt*

Cláudia Antunes
Instituto Superior Técnico
Lisbon, Portugal *claudia.antunes@tecnico.ulisboa.pt*

Abstract

The interest on the discovery of information hidden in large amounts of data exploded in the last decade, bringing to light the need of efficient and effective tools to access all sources and kinds of data. On the other hand, the need to secure and share valuable data led to the development of new technologies, like blockchain, that warrant data integrity and transparency. Combining both is a natural demand, but several issues become clear, such as the lack of access efficiency and the need of data replication in common solutions. Indeed, the unique existing approach is by emulating queries, mostly through Smart Contracts, and applying traditional machine learning algorithms over the resulting data, stored externally for allowing multiple accesses. In this paper, we performed a systematic literature review that provides the above conclusions. Later, we discuss a new system architecture for the analysis of data stored in a blockchain, exploring the scalability and high-performance of data access in distributed file systems and the fast and up-to-date predictions of a streaming analysis approach.

Keywords: Blockchain; Information System Security; Data Analysis; Incremental Machine Learning; Distributed File System

1. Introduction

Blockchain and Data Analysis are topics of high interest, and are being integrated together in a multitude of applications [1]. However, research combining them does not provide neither guidelines on how to access data on a blockchain, nor how to analyse the collected data. Indeed, this process is not as straightforward as when mining traditional databases. There is no standard data structure for the data stored in the blockchain that makes the analysis efficient in time, like data cubes for data warehouses. In particular, blockchain does not have a built-in query system, so most solutions can be classified into one of two categories: emulating querying with smart contracts and custom search engines, or extracting the data to a traditional database and accessing it from there. However, both solutions have issues. Querying data through smart contracts has high costs and slow performance, and extracting data to an off-chain database looses the data integrity protections afforded by the blockchain, requiring additional storage. Lastly, with smart contracts and custom search engines, analysing data stored in a blockchain is

a time-consuming process and due to the nature of batch learning, it is a process that is repeated multiple times.

The objective of this paper is to answer our three research questions on thus topic and organize the current body of knowledge, with the goal of creating a starting point in the development of tools to analyse data stored in blockchains. To that end, in this paper discussion we propose a system architecture where data is stored using distributed file systems to reduce storage costs and, by saving the hash of the data in the blockchain, identify data tampering. From this, the data analysis is then made using a streaming approach, allowing for the incremental learning of information.

The paper is organized as follows. Section 2 describes the research design. Sections 3, 4, and 5 represent the systematic literature review phases of the topic and answer the three proposed research questions: (5.1) Which distributed file systems are used with blockchain? (5.2) How is data accessed on architectures using blockchain and distributed file systems? (5.3) Which are the current streaming data architectures used in blockchain? Section 6 presents the conclusions of the systematic literature review and the proposed architecture overview. Section 7 presents related work. Lastly, Section 8 closes with conclusion remarks and future work objectives.

2. Research Methodology

A systematic literature review (SLR) is defined as a "means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area or phenomenon of interest" [2].

The SLR conducted was based on Kitchenham 2004 study [2], and comprises three steps: planning, conducting and reporting. The planning phase is composed of the first three tasks; identify the review need, develop a review protocol and define the research questions. The conducting phase is divided in two parts; screen and select the target studies and analyze the studies data. Lastly, the reporting phase purpose is to summarize the information gathered in the studies.

3. Systematic Literature Review Planning

This section presents our three research questions. The three main topics these questions pretend to explore are blockchain, distributed file systems and data analysis, more specifically using streaming data techniques.

Research Question 1: Which distributed file systems are used with blockchain?

Research Question 2: How is data accessed and analysed on architectures using blockchain and distributed file systems?

Research Question 3: Which are the current streaming data architectures used in blockchain?

We used the search engine EBSCO Discovering Service [3] that includes the main research sources, such as Scopus, Academic Search and Clarivate Analytics (itself including Web of Science, Current Contents Connect, Derwent Innovations Index, MEDLINE e SciELO Citation Index, and other resources, such as Citation Reports and Essential Science Indicators).

To identify the relevant work, we used the following search expressions: (1) "AB (Blockchain) AND AB ("Distributed File System" OR "Decentralized File System" OR "Interplanetary File System")"; (2) "AB (Blockchain) AND AB ("Data Stream" OR "Data Streaming" OR "Data Flow" OR "Data Flows")".

The keyword AB indicates to the search engine we have used – EBSCO Discovery Service – that the search should be carried out in the title and the abstract. The papers were filtered automatically by the search engine according to Table 1.

The first search string resulted in 256 studies and the second in 111 studies. The merged results, after duplicates were removed, were 277 studies.

Table 1. Filtered Studies

Included	Excluded
Equivalent Subjects	Not Peer Reviewed
Full Text	Not Written in English
	Not Academic Journal or Conference Material

The studies abstracts were analysed and classified as out of scope according to our inclusion/exclusion criteria, presented in Table 2.

The purpose of this criteria was to analyse novel data analysis architectures, such as new data access processes; new or different architectures for distributed file systems and blockchain or new distributed file systems technologies that were not included before. Studies with data management components were included since these could identify technical problems or solutions in current real world applications of these technologies.

An objective of this study is to understand how data analysis is being conducted in blockchain based systems, supported by distributed file systems. As such, blockchain specific technical improvements or blockchain technology integration in an industry such as using blockchain for agriculture, was deemed as out of scope. Personal data applications were likewise excluded since these are not in the scope of the study.

Table 2. Scope Inclusion/Exclusion Criteria.

Inclusion Criteria	Exclusion Criteria
Data Management	General Security Improvements
Data Processes	Personal Data Applications
Data Access Architectures	Specific Integration of Blockchain in an Industry
Different Distributed File Systems	Performance Improvements by Consensus Algorithms
Technologies Not included Before	

4. Systematic Literature Review Conducting

The abstract of every paper was studied which resulted in excluding a total of 181 papers on this phase. In the following phase we analysed the introduction and conclusion of the remaining papers finishing this phase with a total of 30 papers. Figure 1 represents the process through a PRISMA flow diagram.

5. Systematic Literature Review Reporting

5.1. RQ1: Which distributed file systems are used with blockchain?

Table 3 presents the distributed file systems in use as well as the blockchain being used when mentioned.

InterPlanetary File System (IPFS) is the most used distributed file system with blockchain

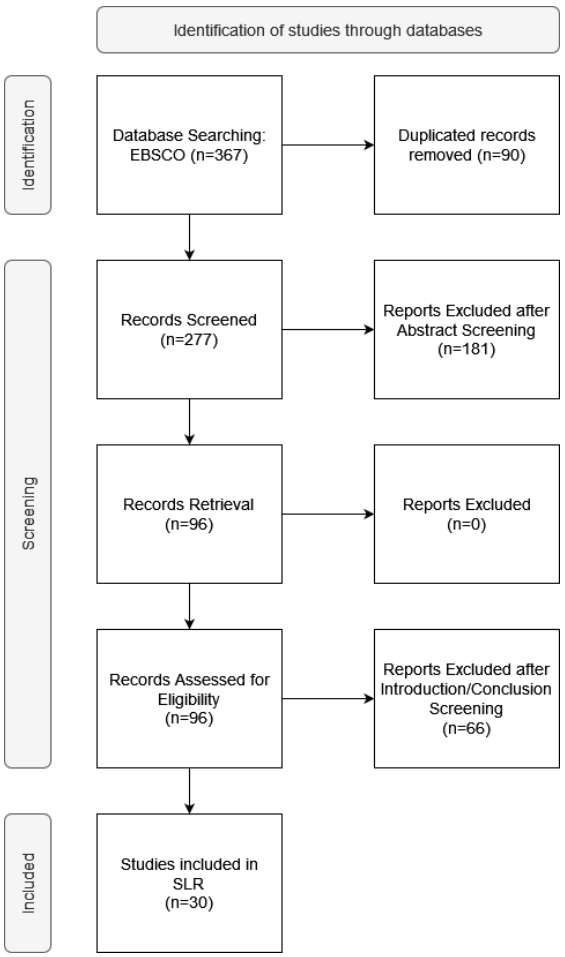

Fig. 1. PRISMA flow diagram for the Systematic Literature Review

in our sample. IPFS is a peer-to-peer hypermedia protocol where no nodes are privileged and a common computer system suffices as a node. The nodes store the IPFS objects in their local storage. Nodes then connect to each other and transfer objects. These objects represent the files and other data structures [4]. The object is chopped into smaller chunks of itself, hashed and given a unique Content Identifier (CID), which serves as a fingerprint. To access the object, the returned CID is necessary. IPFS "solve the shortage of blockchain in storing big files" [5] since "storing a document on the blockchain is expensive" [6].

Hadoop Distributed File System (HDFS) is the second most used distributed file system with blockchain in our studies sample. HDFS is an isolated master–slave data storage network composed of NameNodes and DataNodes. HDFS "is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets." [7]. HDFS is "mainly used for batch processing of data" [8]. HDFS is most suited when the nodes can be trusted.

Swarm is another distributed file system used with blockchain. Swarm is very similar to IPFS. Its biggest difference is that IPFS uses a Distributed Hash Table (DHT) and Swarm uses an immutable content address chunkstore to generate the content identifiers [9]. Swarm has a natural integration with Ethereum blockchain and an incentive system that benefits from smart contracts.

Merkle Tree based File System (MTFS) is a distributed file system that was integrated with blockchain. In MTFS a node consists of a "batch of servers with professional connection sitting in a data center" [10]. MTFS uses asymmetric cryptography including proxy re-encryption

(PRE), to ensure data privacy. Its peer-to-peer network broadcasts data like a tree having redundant nodes and connections in case of failure. This file system has less adoption and implementation examples when compared with the previously mentioned file systems.

When adding data to a distributed file system, most of the studies follow a similar process, which can be summarized as follows:

1. **Data Source:** Create Data Entry and send to API
2. **API:** Send (Encrypted) Data to Distributed File System
3. **API:** Upload Data and Generate Hash from Data
4. **DFS:** Send Data's Hash to API
5. **API:** Send Transaction to Blockchain with the Data's Hash
6. **Blockchain:** Send Confirmation of Success to API

Table 3. Distributed File Systems Used

Distributed File Systems	Support Literature
IPFS and Ethereum	[5] [11] [12] [13] [6] [14] [15] [16] [17] [18] [19] [20]
IPFS and HLF	[21] [22] [23]
IPFS and Multi-Chains/ Custom-Chain	[24] [25] [26]
IPFS	[27] [28] [29] [30]
HDFS and Ethereum	[31]
HDFS and HLF	[8]
HDFS	[32]
MTFS	[10]
Swarm and Ethereum/ Hyperledger Fabric	[33]

5.2. RQ2: How is data accessed for analysis on architectures using blockchain and distributed file systems?

Table 4 presents the data access architectures used by blockchain and distributed file systems found.

Smart Contracts, or Custom Search Engine Query, are the most common data accessing mechanism among the distributed file system and blockchain architectures, within the research studies. In these methods, after the data content identifier is obtained from the distributed file system, the identifier is saved in the blockchain ledger, along with relevant metadata, such as access authorization. In the case of custom search engines, it is also saved in a local or a cloud database. A smart contract or a traditional query in a local or a cloud database obtains the data content identifier by matching saved metadata such as a keyword. Using off-chain sources greatly improves access speed, however, since it is off-chain, it can be a target for malicious participants.

Hadoop Integration is the second most used accessing data mechanism identified. In these systems, the distributed file system used is HDFS where it is possible to use MapReduce that "is a pre-built framework in HDFS" [8]. In these cases, MapReduce can be used to analyse the data.

Share by Smart Contracts is another method used to access data from a distributed file system and a blockchain network where all the participants are trusted. The data content identifier is broadcast to all the participants through a smart contract. In this case every participant is able to directly access the saved file through the identifier in the distributed file system.

Table 4. Data accessed on Distributed File Systems and Blockchain

Data Access/ Analysis Found	Support Literature
Smart Contracts or Custom Search Engine Query	[5] [11] [12] [13] [6] [14] [27] [21] [31] [15] [16] [19] [17] [25] [23]
Hadoop Integration	[34] [8]
Shared by Smart Contract	[22]

5.3. RQ3: Which are the streaming data architectures used in blockchain?

Only one streaming data architecture in blockchain was found in the analysed studies - "ITrade: A Blockchain-based, Self-Sovereign, and Scalable Marketplace for IoT Data Streams" [35] (see Table 5). In this study, blockchain (Ethereum) and smart contracts are used for security, availability and trust purposes. Also, this system uses a pull-based message consumption model (Kafka) as the basis of its streaming architecture. This system's purpose is to give a data buyer the ability to subscribe to a data stream.

Table 5. Streaming Architectures used in Blockchain

Data Streaming Architecture Found	Support Literature
Event-based Message Model	[35]

6. Discussion

Most blockchain architectures available in studies usually focus on adapting blockchain to an industry. In section 5.1, although different combinations of technologies are presented, (blockchains and DFS), the architecture between them is usually similar. Also, most of these architectures do not include or propose in their systems a mechanism or methodology for analysing the data stored in their systems.

In section 5.2 we can observe the solutions used to access data. Most of them could be more efficient or secure making the analysis process under-performing. The smart contracts query system does not scale well and such these implementations are introduced with custom built search engines. The problem with custom build solutions is the lack of comparability across different frameworks. Also, since these solutions are not on-chain, they can be subject to malicious participants and do not work on a public blockchain. HDFS is naturally compatible with MapReduce. However, like the previously mentioned case, it is not suited for public settings, since HDFS intended use is when its nodes can be trusted. Likewise, the last solution found is also not suited for public settings. These results motivate the proposal of a different architecture.

To improve the analysis process, we conceptualize an architecture that is divided into a data storage and collection layer composed by a distributed file system, integrated with blockchain based on the results analysed in the systematic literature review and a data stream pipeline.

In Figure 2, we present a Unified Modeling Language (UML) sequence diagram that showcases how new data is processed in the system. A user starts by sending data to the API through, for example, a website. The server's API can encrypt the data if needed and will send the data to a distributed file system to be saved. The distributed file system, after saving the data, will return the content identifier back to the API. The API will send a new transaction to the blockchain with the content identifier and if successful the confirmation of new data will be sent to both the API and then the user. After data is saved and the confirmation is sent to the user, the API will also send the new data to the data analysis pipeline for it to be readily available when an analysis request is submitted. When a user sends a data request through, for example, a website, a request is sent to the blockchain with the transaction identifier. Then, the blockchain returns the transaction data that contains the content identifier in the distributed file system. The content identifier is sent in a request to the distributed file system and the data is returned to the user by the API. The analysis request is sent to the data analysis pipeline and the requested analysis is returned from the data already analysed in the database.

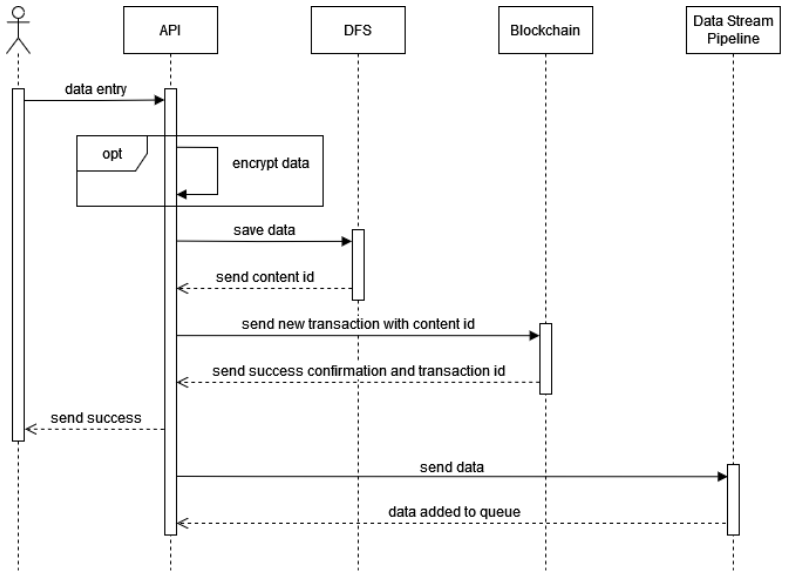

Fig. 2. Adding Data Process, UML Sequence Diagram

In Figure 3 we can observe the analysis process inside the data stream pipeline. The data stream pipeline has three components: the ingestion layer, event-based message bus system (based on the results of subsection 5.3); a stream processing application; an incremental learning module. The data stream messages ingestion system is responsible for managing the incoming data to be analysed from the API. The stream processing application requests the messages from the data stream messages ingestion system and processes the data and saves it, if necessary, in the results database. Lastly, the incremental learning algorithm pulls the data from the data stream messages ingestion system and the latest model from the database; then, it processes the new data and updates the incremental learning model with the latest data. The stream processing application results may be of interest to the incremental learning algorithm and it is possible to use it as part of the input for the model. With a stream processing application, efficient data pre-processing can be integrated.

An identified challenge of typical distributed data stream processing frameworks is "how to accurately ingest and integrate data streams from various sources and locations into an ana-

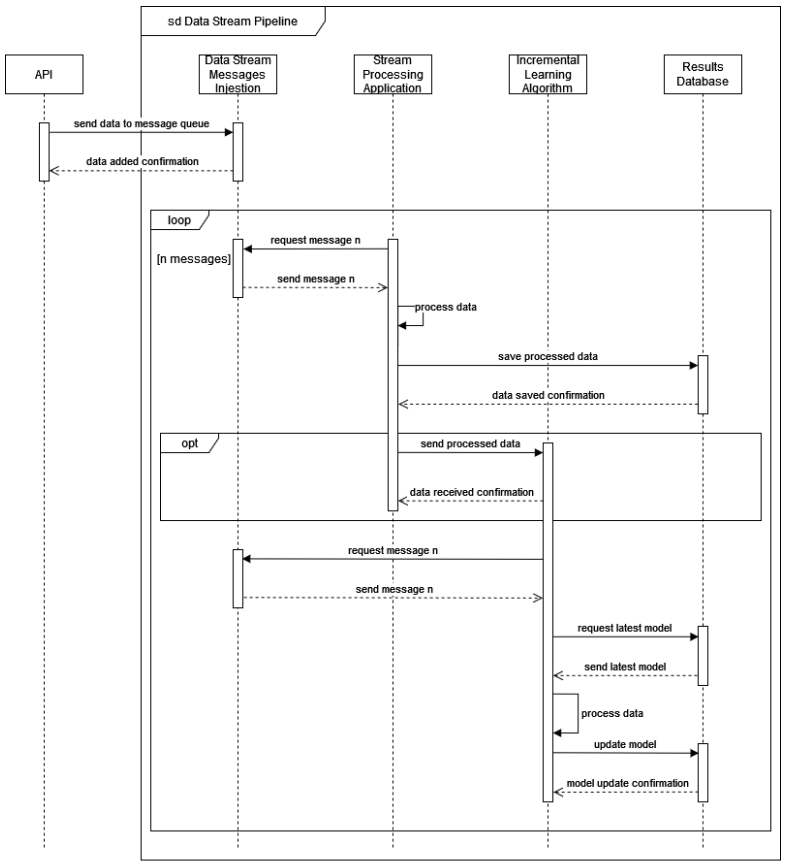

Fig. 3. Analyzing Data, UML Sequence Diagram

lytics platform" [36]. Our proposed architecture solves this issue since it aggregates multiple data sources into a single one through blockchain. It is also compatible with different types of blockchains (public or private), resulting in an architecture that is not bound to a single application. Another issue solved by our proposed architecture is adding data analysis functionalities to an existing blockchain based system. For example: applying the proposed pipeline and expanding the system's API on a blockchain already in use, would add data analysis functionalities. One of the benefits of this new architecture is that it is modeled like microservices. Since its modules are loosely-coupled, with small changes to the overall architecture, features can be added or removed (for example, encryption, access control, or data pre-processing).

7. Related Work

There are various SLRs about blockchain and its applications in industries in general or in specific applications such as healthcare [37], supply chains [38], energy [39], Internet of Things [40] and smart cities [41], finance [42], government [43], education [44], agriculture [45]. There are also various reviews in the research field of blockchain and different technology improvements to blockchain security [46] and privacy [47].

Huang [48], summarizes the current state of blockchain and DFS, demonstrating challenges and open issues. However, this study is not an SLR. Deepa [49], presents a survey about the state of big data and blockchain, including the data analysis topic; however, it only briefly mentions distributed file systems. As such, while researching this field, no SLR was found that addresses data analysis in blockchain (and DFS) within the same scope. Hussain [50] study is not an SLR and does not consider every phase of the analysis process.

8. Conclusion

Given the recent increase in blockchain systems popularity, there are several proposals that explore this technology applications in multiple fields. Blockchain technology stores data in architectures that provide high data integrity and provenance, as well as a platform where different participants can share data with a high degree of trust. However, this data only has value if it can be accessed and analysed in an efficient way transforming data in information. With the use of DFS the amount, speed and type of data is improved. Also, streaming data technologies allow for a higher data flow from the moment data is accessed to the analysis.

Through this work, we have answered three research questions and organized the current body of knowledge identified. The SLR was performed to identify which technologies were used with blockchain, the methodologies used to access data in these architectures and which streaming data architectures were being used.

Following the research, we discuss an architecture based on the review results. The architecture is composed of blockchain technology, for trust, security, traceability, data integrity, data sharing and provenance purposes. A DFS is included in the architecture, for storage scalability and to store different data types such as files or images. Lastly, we included a data stream pipeline as a data analysis solution (with stream processing capabilities for data transformation on the go and/or incremental learning model(s) to analyse said data).

This research is based on scientific literature only. However, the distributed file system, the blockchain and the data analysis topics also have developments described in gray literature. A multivocal literature review could be used to include that data, but that was not in the scope of our study.

In future research we will focus on producing a prototype based on the proposed architecture as well as evaluate said prototype against a simple solution that extracts data from a blockchain to a database and analyzes it using batch processing in terms of speed, accuracy and cost.

9. Acknowledgements

This research was funded by Guest Intelligence Chain: ALG-01-0247-FEDER-047399 and FCT - Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit - UIDB/00326/2020 or project code UIDP/00326/2020.

References

[1] Joe Abou Jaoude and Raafat George Saade. "Blockchain applications - Usage in different domains". In: *IEEE Access* 7 (2019), pp. 45360–45381. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2902501.

[2] Barbara Kitchenham. "Procedures for performing systematic reviews". In: *Keele, UK, Keele University* 33.2004 (2004), pp. 1–26.

[3] EBSCO. *EBSCO*. https://eds.s.ebscohost.com/eds/search/advanced. [Online; accessed 20-June-2022].

[4] Juan Benet. "Ipfs-content addressed, versioned, p2p file system". In: *arXiv preprint arXiv:1407.3561* (2014).

[5] Wenren Huang. "A Blockchain-Based Framework for Secure Log Storage". In: *2019 IEEE 2nd International Conference on Computer and Communication Engineering Technology-CCET* (2019).

[6] Kosala Yapa Bandara and John Breslin. "BaaS Architecture for DApps and Application for Veterinary Medicine Case Study in Ireland". In: *The 2021 International Symposium on Networks, Computers and Communications (ISNCC 2021) - Dubai, UAE* (2021).

[7] Dhruba Borthakur et al. “HDFS architecture guide”. In: *Hadoop apache project* 53.1-13 (2008), p. 2.

[8] Viraaji Mothukuri et al. “BlockHDFS: Blockchain-integrated Hadoop distributed file system for secure provenance traceability”. In: *Blockchain: Research and Applications* 2 (4 Dec. 2021), p. 100032. ISSN: 20967209. DOI: 10.1016/j.bcra.2021.100032.

[9] Swarm Team. *Swarm; Storage and communication infrastructure for a self-sovereign digital society*. https://www.ethswarm.org/swarm-whitepaper.pdf. [Online; accessed 27-June-2022].

[10] Jia Kan and Kyeong Soo Kim. “MTFS: Merkle-Tree-Based File System”. In: *ICBC 2019 - IEEE International Conference on Blockchain and Cryptocurrency* (May 2019), pp. 43–47. DOI: 10.1109/BLOC.2019.8751389.

[11] Dragos Daniel Taralunga and Bogdan Cristian Florea. “A blockchain-enabled framework for mhealth systems”. In: *Sensors* 21 (8 Apr. 2021). ISSN: 14248220. DOI: 10.3390/s21082828.

[12] Muqaddas Naz et al. “A Secure Data Sharing Platform Using Blockchain and Interplanetary File System”. In: *Sustainability (Switzerland)* 11 (24 Dec. 2019). ISSN: 20711050. DOI: 10.3390/su11247054.

[13] Aparna Kumari and Sudeep Tanwar. “A Reinforcement-Learning-Based Secure Demand Response Scheme for Smart Grid System”. In: *IEEE Internet of Things Journal* 9 (3 Feb. 2022), pp. 2180–2191. ISSN: 23274662. DOI: 10.1109/JIOT.2021.3090305.

[14] Yustus Eko Oktian, Sang Gon Lee, and Byung Gook Lee. “Blockchain-based continued integrity service for IoT big data management: A comprehensive design”. In: *Electronics (Switzerland)* 9 (9 Sept. 2020), pp. 1–36. ISSN: 20799292. DOI: 10.3390/electronics9091434.

[15] Zhili Zhou et al. “Reliable and Sustainable Product Evaluation Management System Based on Blockchain”. In: *IEEE Transactions on Engineering Management* (2021). ISSN: 15580040. DOI: 10.1109/TEM.2021.3131583.

[16] Raja Guru R and Praveen Kumar S. “Self-restrained energy grid with data analysis and blockchain techniques”. In: *Energy Sources, Part A: Recovery, Utilization and Environmental Effects* (2020). ISSN: 15567230. DOI: 10.1080/15567036.2020.1852341.

[17] Norah Alrebdi et al. “SVBE: searchable and verifiable blockchain-based electronic medical records system”. In: *Scientific Reports* 12 (1 Dec. 2022). ISSN: 20452322. DOI: 10.1038/s41598-021-04124-8.

[18] Haya R. Hasan et al. “Trustworthy IoT Data Streaming Using Blockchain and IPFS”. In: *IEEE Access* 10 (2022), pp. 17707–17721. ISSN: 21693536. DOI: 10.1109/ACCESS.2022.3149312.

[19] Shunrong Jiang et al. “Verifiable Search Meets Blockchain: A Privacy-Preserving Framework for Outsourced Encrypted Data”. In: *ICC 2019-2019 IEEE International Conference on Communications (ICC)* (2019).

[20] Anisha Desai, Payal Shah, and Dayanand D. Ambawade. “VerifyB - Students’ record management and verification system”. In: *Proceedings - International Conference on Communication, Information and Computing Technology, ICCICT 2021* (2021). DOI: 10.1109/ICCICT50803.2021.9510144.

[21] Emmanuel Nyaletey et al. “BlockIPFS - Blockchain-enabled interplanetary file system for forensic and trusted data traceability”. In: *Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019* (July 2019), pp. 18–25. DOI: 10.1109/Blockchain.2019.00012.

[22] Xingyu Tao et al. “Distributed common data environment using blockchain and Interplanetary File System for secure BIM-based collaborative design”. In: *Automation in Construction* 130 (Oct. 2021). ISSN: 09265805. DOI: 10.1016/j.autcon.2021.103851.

[23] Chinling Chen et al. “Enterprise Data Sharing with Privacy-Preserved Based on Hyperledger Fabric Blockchain in IIOT’s Application”. In: *Sensors* 22 (3 Feb. 2022). ISSN: 14248220. DOI: 10.3390/s22031146.

[24] Chao Peng et al. “Research on Cross-chain Communication Based on Decentralized Identifier”. In: *HotICN 2021 - 2021 4th International Conference on Hot Information-Centric Networking* (2021), pp. 7–12. DOI: 10.1109/HotICN53262.2021.9680822.

[25] Amrendra Singh Yadav, Nikita Singh, and Dharmender Singh Kushwaha. “Sidechain: storage land registry data using blockchain improve performance of search records”. In: *Cluster Computing* 25 (2 Apr. 2022), pp. 1475–1495. ISSN: 15737543. DOI: 10.1007/s10586-022-03535-0.

[26] Justin S Gazsi et al. “VAULT: A Scalable Blockchain-Based Protocol for Secure Data Access and Collaboration”. In: *Proceedings - 2021 IEEE International Conference on Blockchain, Blockchain 2021* (2021), pp. 376–381. DOI: 10.1109/Blockchain53845.2021.00059.

[27] Mengji Chen et al. “Blockchain-Enabled healthcare system for detection of diabetes”. In: *Journal of Information Security and Applications* 58 (May 2021). ISSN: 22142126. DOI: 10.1016/j.jisa.2021.102771.

[28] Peter Altmann et al. “Creating a Traceable Product Story in Manufacturing Supply Chains Using IPFS”. In: *2020 IEEE 19th International Symposium on Network Computing and Applications, NCA 2020* (Nov. 2020). DOI: 10.1109/NCA51143.2020.9306719.

[29] Pearl Alisha Lobo and V. Sarasvathi. “Distributed File Storage Model using IPFS and Blockchain”. In: *2021 2nd Global Conference for Advancement in Technology, GCAT 2021* (Oct. 2021). DOI: 10.1109/GCAT52182.2021.9587537.

[30] Erik Daniel and Florian Tschorsch. “IPFS and Friends: A Qualitative Comparison of Next Generation Peer-to-Peer Data Networks”. In: *IEEE Communications Surveys and Tutorials* 24 (1 2022), pp. 31–52. ISSN: 1553877X. DOI: 10.1109/COMST.2022.3143147.

[31] Thomas Renner, Johannes Muller, and Odej Kao. “Endolith: A Blockchain-Based Framework to Enhance Data Retention in Cloud Storages”. In: *Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018* (June 2018), pp. 627–634. DOI: 10.1109/PDP2018.2018.00105.

[32] Fredrick Ishengoma. “NFC-Blockchain Based COVID-19 Immunity Certificate: Proposed System and Emerging Issues”. In: *Information Technology and Management Science* 24 (Dec. 2021), pp. 26–32. ISSN: 22559086. DOI: 10.7250/itms-2021-0004.

[33] Patrick Sylim et al. “Blockchain technology for detecting falsified and substandard drugs in distribution: Pharmaceutical supply chain intervention”. In: *JMIR Research Protocols* 7 (9 Sept. 2018). ISSN: 19290748. DOI: 10.2196/10163.

[34] M. Hena and N. Jeyanthi. “A Three-Tier Authentication Scheme for Kerberized Hadoop Environment”. In: *Cybernetics and Information Technologies* 21 (4 Dec. 2021), pp. 119–136. ISSN: 13144081. DOI: 10.2478/cait-2021-0046.

[35] Sina Rafati Niya, Danijel Dordevic, and Burkhard Stiller. "ITrade: A Blockchain-based, Self-Sovereign, and Scalable Marketplace for IoT Data Streams". In: *Proceedings of the IM 2021 : 2021 IFIP/IEEE International Symposium on Integrated Network Management : 17-21 May 2021, Bordeaux, France, virtual conference* (2021).

[36] Haruna Isah et al. "A survey of distributed data stream processing frameworks". In: *IEEE Access* 7 (2019), pp. 154300–154316.

[37] Anushree Tandon et al. "Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda". In: *Computers in Industry* 122 (2020), p. 103290. ISSN: 0166-3615. DOI: https://doi.org/10.1016/j.compind.2020.103290. URL: https://www.sciencedirect.com/science/article/pii/S0166361520305248.

[38] Yingli Wang, Jeong Hugh Han, and Paul Beynon-Davies. "Understanding blockchain technology for future supply chains: a systematic literature review and research agenda". In: *Supply Chain Management: An International Journal* (2018).

[39] Hamzah Khan and Tariq Masood. "Impact of blockchain technology on smart grids-A systematic literature review". In: *Available at SSRN 4003063* (2021).

[40] Sin Kuang Lo et al. "Analysis of blockchain solutions for IoT: A systematic literature review". In: *IEEE Access* 7 (2019), pp. 58822–58835.

[41] Charles Shen and Feniosky Pena-Mora. "Blockchain for cities—a systematic literature review". In: *Ieee Access* 6 (2018), pp. 76787–76819.

[42] Omar Ali, Mustafa Ally, Yogesh Dwivedi, et al. "The state of play of blockchain technology in the financial services sector: A systematic literature review". In: *International Journal of Information Management* 54 (2020), p. 102199.

[43] F Rizal Batubara, Jolien Ubacht, and Marijn Janssen. "Challenges of blockchain technology adoption for e-government: a systematic literature review". In: *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age* (2018), pp. 1–9.

[44] Faiza Loukil, Mourad Abed, and Khouloud Boukadi. "Blockchain adoption in education: a systematic literature review". In: *Education and Information Technologies* 26.5 (2021), pp. 5779–5797.

[45] Oscar Bermeo-Almeida et al. "Blockchain in agriculture: A systematic literature review". In: *International Conference on Technologies and Innovation* (2018), pp. 44–56.

[46] Paul J. Taylor et al. "A systematic literature review of blockchain cyber security". In: *Digital Communications and Networks* 6 (2 May 2020), pp. 147–156. ISSN: 23528648. DOI: 10.1016/j.dcan.2019.01.005.

[47] Francisco José de Haro-Olmo, Ángel Jesús Varela-Vaca, and José Antonio Álvarez-Bermejo. "Blockchain from the perspective of privacy and anonymisation: A systematic literature review". In: *Sensors (Switzerland)* 20 (24 Dec. 2020), pp. 1–21. ISSN: 14248220. DOI: 10.3390/s20247171.

[48] Huawei Huang et al. "When blockchain meets distributed file systems: An overview, challenges, and open issues". In: *IEEE Access* 8 (2020), pp. 50574–50586.

[49] Natarajan Deepa et al. "A survey on blockchain for big data: approaches, opportunities, and future directions". In: *Future Generation Computer Systems* (2022).

[50] Adedoyin A. Hussain and Fadi Al-Turjman. "Artificial intelligence and blockchain: A review". In: *Transactions on Emerging Telecommunications Technologies* 32 (9 Sept. 2021). ISSN: 21613915. DOI: 10.1002/ett.4268.