

Retraction

Retracted: An Effective Big Data Sharing Prototype Based on Ethereum Blockchain

Scientific Programming

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Scientific Programming. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] S. Song, "An Effective Big Data Sharing Prototype Based on Ethereum Blockchain," *Scientific Programming*, vol. 2022, Article ID 6544031, 14 pages, 2022.

Research Article

An Effective Big Data Sharing Prototype Based on Ethereum Blockchain

Su Song 

Computer Science Teaching and Research Group, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu 224005, China

Correspondence should be addressed to Su Song; s1138599@s.eduhk.hk

Received 1 November 2021; Revised 15 December 2021; Accepted 18 December 2021; Published 19 March 2022

Academic Editor: Rahman Ali

Copyright © 2022 Su Song. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because of its many advantages, big data has been extending to various domains of science, health, education, and commerce. Despite its many applications, big data sharing typically suffers from some key issues, such as user control, lack of incentives, cost, and the right of data. This paper proposes a decentralized big data sharing prototype to improve the applications and services of big data. The method makes use of Ethereum blockchain and related technologies to systematically recommend the implementation guidelines. The research provides a detailed description of the design and implementation of each sublayer of a big data system. As the method is based on blockchain technology, the key technical points are properly addressed in each of the layers. For evaluation, relevant data were collected, and functional testing was performed. A comparison was performed about the sharing frequency and blockchain consensus performance of similar platforms. The dual mining node of the proposed prototype succeeded in processing 1366 blocks and 300 messages. A comparatively satisfactory file access time in the range of 10 m to 20 s and file transmission time between 100 m and 200 s were achieved. The results obtained show that this prototype can effectively verify the feasibility of the model, the layered architecture, and the related sharing mechanism. For the functional and performance testing, practical projects were implemented and evaluated. The promising results obtained testify that the research offers a theoretical background for innovative research in the domain and specialized guidelines for practical implementation.

1. Introduction

With the rapid surge in global data traffic, research interest in big data is increasing. Big data, as defined by [1], is a new generation of technology explored to analyze a colossal amount of data and to extract the key characteristics. Because of the rise of social networks [2], e-commerce [3], mobile communication [4], and the Internet of Things (IoT), human society has entered the era of big data [5]. In big data, the amount of data is measured in petabytes, whereas for accurate analysis, prediction, and decision making, the technology of distributed computing is used [6].

Smartphones and wearable devices trace every bit of change in our behavior and location. Moreover, physiological information is also recorded and analyzed for various purposes. Big data are stored and controlled by various entities, such as government agencies, business alliances,

research institutions, and even individuals, forming the phenomenon of *data silos*. Interconnecting and sharing the scattered nodes of big data for obtaining valuable information about products and services has become an unavoidable need in business, research, and public services [7].

The most valuable thing in the twenty-first century is not only talent but also data. A few of the typical applications of big data sharing in public services, business, and research are presented in the following lines: the *Centers for Disease Control and Prevention* (CDC) collaborates with Google to compare the 50 million most frequent search records of Americans. Using the CDC's seasonal flu data of the past five years, Google's massive search commands successfully predicted the H1N1 virus epidemic in the winter of 2009 weeks in advance. This greatly enhanced the U.S. epidemic combating strategy. The *Farecast* project integrated tens of trillions of price records from various airlines and successfully predicted domestic flight fares with an accuracy

rate of 75 percent. It helps in saving an average of more than \$50 per ticket, resulting in significant cost savings for travelers and flight optimization and utilization. It took scientists ten years from 2003 to 2013 to complete a study of 3 billion base pairs in the human body and perhaps another decade to sequence 3 billion base pairs in the human body. However, today we can do the same in just 15 minutes, thanks to the sharing of genetic decipherment data among organizations around the world. Even the closest life applications, such as Taobao, Weibo, and Youku, are the products of distributed big data sharing.

Despite the long list of advantages of big data, issues are there in every emerging technology, and big data is no exception. The key challenges of big data include security, privacy, management and interpretation of data, energy management, real-time processing, and intelligent interpretation. Some of the challenges have been addressed by recently carried out research works. For instance, the authors in [8] used machine theory and coalitional game to secure the social network data. For the smart grid system, Ren et al. [9] designed a security-aware algorithm based on reinforcement learning.

Akin to other countries, the open sharing of big data is a major demand for China's informatization development as well. The General Office of the State Council emphasized the importance of information data in the *Outline of National Informatization Development Strategy* [10] released in July 2016. In the "13th Five-Year" National Informatization Plan [11] jointly released with the General Office of the Central Committee of the Communist Party of China, it was clearly pointed out that it is necessary to properly manage and control the integration and sharing of information. Consequently, in May 2017, the State Council issued the implementation plan for the integration and sharing of Government Information Systems [12]. The plan suggested that the open sharing of governmental data can help promote the promotion and deployment of policies and is one of the important links in the top-level design of the country. President Xi Jinping sent a congratulatory letter to the China International Big Data Industry Expo, established in Guiyang City, Guizhou Province, on May 26, 2018. The official letter underlined that China attaches great importance to the development of big data and should adhere to the development concept of openness, sharing, innovation, and the theme of *Digitalization of Everything*. The theme of the expo was *"Integration of Intelligence"* to promote the innovative development of the big data industry. Although there are many application cases and policies to drive the sharing of big data, the phenomenon of "data silos" is still a serious concern. With a large amount of data in the hands of departments and individuals, the overall sharing degree is quite low. This hinders the transformation of data value and human social progress [13]. The root cause of this phenomenon is the lack of a transparent, open, and credible data-sharing environment for all parties involved in data sharing. This makes it difficult to jointly negotiate the distribution of data benefits, and hence, data security is not strongly guaranteed. This, in turn, also gives rise to three major problems:

difficulty in data connection, difficulty in data control, and lack of customization capability of cross-domain data services.

The emergence of blockchain technology brings new opportunities to solve the aforementioned problems. Blockchain is a distributed digital ledger system where hashed and encrypted authenticated data is stored. The ledger with the data is immutable, and any mistake or change is traced back to the source [14]. The key characteristics of blockchain include decentralization, tamper-evident, and data traceability. Thus, a blockchain system provides the possibility to build a transparent, open, secure, and trustworthy data sharing environment to connect big data in various fields. In this context, this paper proposes a blockchain-based big data-sharing model. The research work attempts to consider reliable, transparent, and efficient big data connection, data authority management, and data service customization mechanisms. The study finally implements a blockchain-based big data sharing prototype system using the cutting-edge technologies of Ethereum blockchain [15], smart contracts [16], IPFS [17], and Laravel. On the basis of the theoretical findings of the research, practical projects were designed to carry out functional and performance testing. The research not only offers significant theoretical findings but also leads a developer on how to practically implement such research works. The prototype is applicable in verifying the feasibility, the layered architecture, and the related sharing mechanism of a big data model.

The rest of the paper is organized into 5 sections. Related work is covered in Section 2. A detailed discussion about blockchain architecture is presented in Section 3. Section 4 is about big data sharing. The performance test and analysis of the prototype system are elaborated in Section 5. Finally, the conclusion is presented in Section 6.

2. Related Work

The section deals with the background and latest developments of big data and blockchain. For easy understanding, the literature is categorized into subsections.

2.1. Existing Big Data Sharing Platform. The interoperability and sharing of big data can generate business value [18]. As an important resource, customer-related data is important for major enterprises [19]. By obtaining and analyzing multiparty customer data, we can judge the future trend of customers [20]. It helps enterprises optimize their business, effectively guide business behavior, and even determine key decisions. Typical business application fields based on big data sharing include education, advertising, film and television entertainment, real estate, and precision marketing. The main application fields include website analysis, financial applications, investment and financial management, and mobile applications. Major internet enterprises, such as bat and Sina, are the farthest away in the commercialization and sharing of big data. Baidu Index platform is a data analysis platform based on data, such as search, click, and access records of Baidu

netizens [21]. It is also one of the most important statistical analysis platforms of the domestic internet. A considerable number of enterprises even take the big data of the platform as the basis for marketing strategy formulation. Through this platform, we can clearly understand the changes in news and public opinion and the overall trend of the industry in a certain period of time by studying the search trend of single words. Tencent's WeChat team provides the WeChat index platform [22]. By searching keywords, one can see the heat and trend in a certain period of time. Such information is convenient for the government to carry out timely and effective public opinion analysis and monitoring and can also provide a guarantee for businesses to obtain user interest points. Sina Weibo's Micro index platform provides Weibo hot keyword index and real-time data [23]. Big data sharing can not only improve the breadth and depth of the research field but also effectively shorten the research cycle, reduce data storage, and management costs. It leads to other much greater contributions in scientific and technological progress. Typically, big data sharing applications for scientific research include gene sequencing, data publishing and citation, data reuse of mainframe computers, and scientific instruments. Under the call of the advocates of open data sharing culture and scientific researchers in various fields, relevant government departments and large scientific research organizations have issued appropriate data resource sharing policies. The aim behind this is to encourage and even force the project leader and paper author to store the supporting data related to the research results in a publicly accessible third-party database [24] for centralized storage and management. It has formed a centralized data-sharing platform for different disciplines and fields. Representative platforms include PhysioNet [25, 26], the physiological data sharing platform of National Science Foundation (NSF), digital coast, a network survey data platform [27], and Crawdad [28], a wireless data sharing platform of Dartmouth University [29], the national data-sharing system data.gov of the United States, and the National Earth System Science sharing service platform of the Chinese Academy of Sciences [27]. The data of these platforms have the characteristics of interdisciplinary, large quantity, and wide influence on other related domains and platforms as well.

In February 2018, Crawdad had 11,902 users from 124 countries and 2723 citations of academic papers. It not only realized data reuse but also contributed to the development of data format standardization in relevant fields and avoided academic fraud and experimental data forgery to a certain extent. As of May 2018, data.gov has opened 190,058 data sets, including 124,522 geospatial data and 65,536 other types of data. The data sources are mainly the environmental, agricultural, meteorological, and educational departments of the local governments. As of March 2018, China's National Earth System Science sharing service platform had a total of 150.08TB of data resources, providing 539.25TB of data services to the scientific and technological community and the public. It provides effective data services for 2384 major scientific research projects and topics

(including national 973 projects, National Science and technology support projects, National Natural Science Foundation projects, 35 major construction projects, and 34 livelihood projects) [30]. The platform under the model is mainly managed by the independent departments of the government and the Academy of Sciences. All the data providers offer corresponding development interfaces for data users. After being verified and authorized by the platform, data can be downloaded.

2.2. Research and Application of Data Sharing. Blockchain technology was originated in 2008. Nakamoto wrote a paper "bitcoin: a peer-to-peer e-cash system" in the cryptography e-mail group [31]. As the core technology is an encrypted digital bitcoin, it has been widely considered by the current scientific and industrial circles [32]. In the field of data sharing, blockchain research is mainly applied in three aspects: data privacy storage and protection, data authentication, and data management.

In terms of data privacy storage and protection, Zyskind et al. [33] proposed a blockchain-based distributed personal data management system to ensure that users own and control their private data. It was in view of the security vulnerabilities caused by third parties that may lead to the disclosure of users' privacy. The paper proposes the design of an automatic access control protocol to verify the data storage and query of the mobile terminal. It was to ensure that personal data was not accessed by the applicants. Ahmed and Ten Broek [34] aimed at the problem that all blockchain transactions are transparently reflected on the public network, resulting in the disclosure of transaction privacy. The study proposed to adopt the hawk protocol based on a smart contract. It encrypts the communication between both parties to ensure the absolute security of information. Ali et al. [35] proposed a blockchain lightweight application for the internet of things devices to solve the security and privacy issues involved in the internet of things. The method also claims for ensuring confidentiality, integrity, and availability in the smart home scenario. In [36], an encrypted smart contract is proposed to protect public and private files using public and private keys and to provide audit and tracking facilities.

In terms of data storage and authenticity assurance, Sivarajah et al. [37] made an in-depth discussion on the future impact of the audit industry based on big data analysis and blockchain technology. The author studied how to incorporate blockchain into future audit procedures based on the existing theoretical framework. Ali et al. [38] proposed the design of a secure global naming storage system using blockchain. The system includes a three-tier architecture of a naming system based on blockchain. The layers used are the data layer, routing layer, and blockchain layer. Finally, the system implemented on namecoin, migrated the original network to bitcoin network, and updated more than 33,000 items and 200,000 transactions. Currently, a PKI system is available for 55,000 users. Alammary et al. [39] proposed a method to ensure the authenticity of microbial sampling robot data based on the bitcoin blockchain. The robot will

not be disturbed by human factors in the process of collecting data, especially the malicious tampering of third-party regulators.

In terms of data management, the author [40] proposed blockchain-based solutions for the three types of trusted problems existing in the current IoT. The problems are data-sharing management, decentralized IoT data sharing architecture, and smart contract collection. Similarly, [33] constructed the IoT access control architecture to provide a more flexible management scheme. In [34], a resource access control method is proposed based on blockchain technology. The method is implemented in a rule-based reliable authority control mechanism by combining XACML and bitcoin. Da Xu and Viriyasitavat [41] introduced the application of blockchain as a software middleware in data value, sharing sensitive information, and other data management projects. In terms of medical data management, Deloitte white paper [42] proposes that blockchain is a new model for medical information interaction. The document describes in detail the coupling between blockchain technology and medical data sharing and the feasibility of building a medical blockchain. Azaria et al. [43] combined blockchain with the smart contract to realize access and authorization management of medical data. As stated in [44], Peterson proposed specific methods to build a medical blockchain from the aspects of the medical data storage structure and a consensus mechanism in view of the main problems existing in medical data sharing. The challenges in big data sharing are presented in the subsequent subsections.

2.2.1. Difficult Data Connection. To realize big data sharing, it is necessary to connect the mutually fragmented and decentralized data sources. However, because of the lack of detailed and transparent institutional standards, open policies, and pricing mechanisms among enterprises, research, and government institutions, it is difficult to achieve fairness and equality among the sharing parties. Data diversity, inconsistent storage structures, interaction standards, and the lack of a transparent communication environment are also hindering data interaction. These problems together make it difficult to connect data. Therefore, a transparent and open data connection is needed to record data information, access conditions, standards, and other relevant information before interconnection can be realized.

2.2.2. Data Control Is Difficult. Primarily, the problem behind the phenomenon of “data silos” is a matter of interest and security. As data is kept by large data platforms or other intermediaries, the owner loses control over the data. The government agencies and internet companies are sensitive to data, and a leak can have a significant social impact. The 18-March Facebook data privacy breach caused the company’s market value to evaporate by more than \$6 billion in one single day. Therefore, there is an urgent need for a “de-intermediated” or decentralized data control mechanism that records data ownership and access control information. In this way, no party will be able to alter the process of data permission and interaction.

2.2.3. Inability to Customize Cross-Domain Data Services. Data as a service is already a trend, such as Baidu’s DMP data marketing cloud platform, which uses Baidu AI brain to integrate search data and provide users with the most accurate placement strategies. However, because of the limitation of the first two problems, the current data-sharing platform is still oriented to narrow service areas. It is difficult to meet the multidimensional modeling and clustering analysis of cross-domain data, and it lacks transparent, efficient, and automated data service customization mechanisms. Therefore, under the premise of connecting, cross-domain big data sets are providing flexible and reliable data control functions and a scalable data service customization mechanism. It is predominantly needed to realize reliable and efficient automated data distribution.

3. Blockchain Architecture

Swan [45] proposed the method to divide the blockchain architecture from different development stages, including the subsequent three main versions.

3.1. Blockchain 1.0 Architecture. Blockchain 1.0 refers to the underlying technology of early virtual currency. It mainly provides the common functions of digital currency, such as transfer and payment. The most famous application is bitcoin. Bitcoin and cryptocurrency technologies speculate that the bitcoin white paper appeared before the bitcoin system, and the continuous module optimization was carried out. Blockchain is based on P2P architecture, which is different from the traditional well-known three-tier architecture of C/s, B/s, or MVC. Although it has a client, it does not have the traditional so-called server in the background. As all clients have equal status, the only one called the server is the JSON RPC server for blockchain 1.0 clients. This component is only used to provide HTTP and JSON RPC interfaces for blockchain interaction rather than interfere with the network. Blockchain 1.0 architecture is shown in Figure 1.

3.2. Blockchain 2.0 Architecture. Although the 1.0 architecture makes this kind of blockchain flexible by supporting transaction notes, it is limited to support scenarios other than digital currency. In recent years, the IT industry has paid more attention to how to integrate blockchain with different fields and solve practical problems. Therefore, the concept of blockchain 2.0 came into being. Its core idea is to endow the blockchain with programmable characteristics and introduce smart contracts. It not only takes the blockchain as a decentralized digital cryptocurrency payment platform but also provides more multidimensional applications by adding extensible functions on the chain, such as real estate contract, equity certificate, intellectual property protection, automobile, and authentication of high-end works of art. The most representative one is the Ethereum blockchain. The architecture is shown in Figure 2.

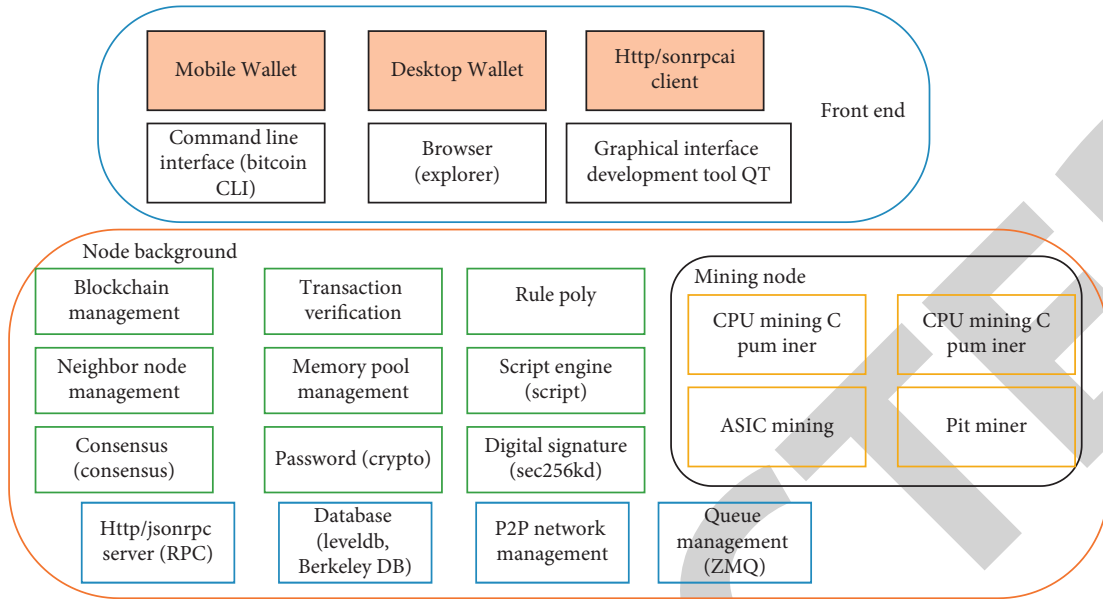


FIGURE 1: The blockchain 1.0 architecture.

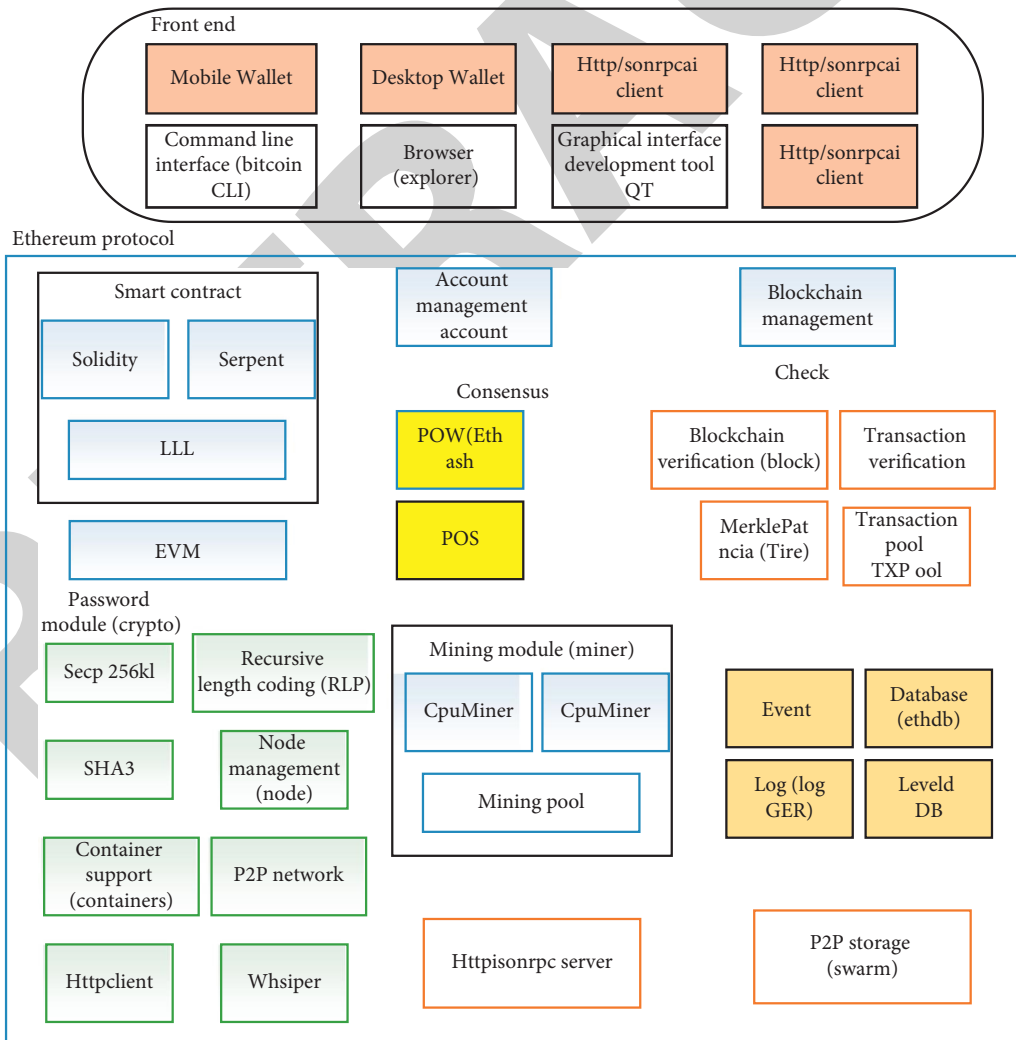


FIGURE 2: The blockchain 2.0 architecture.

3.3. Blockchain 3.0 Architecture. Blockchain 3.0 mainly refers to other applications beyond financial transactions, especially in scientific research, industry, agriculture, the internet of things, and other fields. Generally speaking, such blockchains exist in the form of enterprise alliance chains or private chains. At present, there is no mature blockchain 3.0 platform in the industry; however, it can be summarized according to several nascent platform architectures that meet the definition of blockchain 3.0, as shown in Figure 3.

4. Big Data Sharing System

Based on the blockchain platform architecture and big data sharing mechanism, this chapter designs and implements a big data sharing prototype system based on the Ethereum blockchain. Firstly, the main functions and module design of the system are briefly introduced, and then, the implementation is discussed in depth. Finally, to verify the function of the system, the system is deployed in the LAN of our school, and the related big data sharing interactive experiments and the related performance tests are carried out on it.

4.1. Big Data-Sharing Prototype System Based on Ethereum. Based on the aforementioned sharing mechanism, this paper constructs a blockchain-based big data-sharing prototype system, including five modules: account system, data management, data service, data quality evaluation, and background management, as shown in Figure 4.

4.1.1. Account System Module. The account system module is composed of account registration, account information, and points system. It provides user access and secure interaction functions. It adopts blockchain account key technology to ensure account security, distributed file system to store account details, and smart contract to customize a configurable integration system.

The account management contract records the relationship between the registered user's Ethereum account master address (public key), all user-related information, and the update function. The registration management function is responsible for account creation, account audit, user information, and update function. Taking the registration function as an example, the steps, which are shown in Figure 5, are as follows.

- (1) Users register blockchain accounts to obtain public and private keys.
- (2) As per the system audit specification, the personal information is put into IPFs in the form of a file to obtain the hash address of the file.
- (3) Send the public key and hash address of the blockchain account to the e-mail with qualification for review by e-mail or other means.
- (4) Auditors obtain personal information using the hash address. It is verified whether the information meets the standard and health of the data source. It is also gauged whether the data source is accessible or not. If the approval fails, it needs to be filled in again. After

passing the registration, the registration contract confirms the user's registration qualification.

- (5) The user binds the username, which means that after the contract is verified, it is written into the blockchain in the form of a transaction. Finally, the blockchain network receives the transaction to form a consensus and persists it to each node of the point-to-point network.

4.1.2. Data Management Module. The data management module includes data source management, data release, data query, data request, and permission management functions to realize the separation of data storage and control. It is also to ensure the absolute control of data by the data owner. The specific management process is as follows: the data publisher, firstly, stores the relevant description and accesses the method of data in the distributed file system in the form of documents. The next step is to call the data publishing application interface to fill in the data type, basic information, and the hash address of the data description document and call the contract to store it in the blockchain. The data demander queries the required data set and writes it into the distributed file system in the form of documents according to the requirements of the data uploader (such as signing the use agreement and paid payment).

In the system, the file-hash address is used as a parameter to make a request to the platform. After receiving the request, the data publisher uses the authority-based access control mechanism of the platform to authorize it. After obtaining the authorization, the data requester obtains the authority token to make a request to the data source. After passing the verification, the data set is successfully downloaded. The interaction process of each layer of the system is shown in Figure 6.

The specific contract set includes a data management master contract, data retrieval contract, data permission contract, type information contract, and data information contract. The data information contract is used to store the data name, IPFs file address, hash, and permission contract. The type information contract uses the key value method to store the data category and its attributes. The data retrieval contract uses the hash table to store the corresponding relationship between each data and its category to ensure convenient category retrieval. The permission contract stores the username and approval status of the requested dataset, and it is responsible for verifying the permission. The data management master contract records the published DST list and RT list of each user and provides functional interfaces for data publishing, retrieval, request, and permission control.

4.1.3. Data Service Module. Data service includes three functions: service customization, service authorization, and service notification. Service customization refers to the service demander customizing relevant data services according to the relevant data types and keywords involved in the service. The service authorization function can

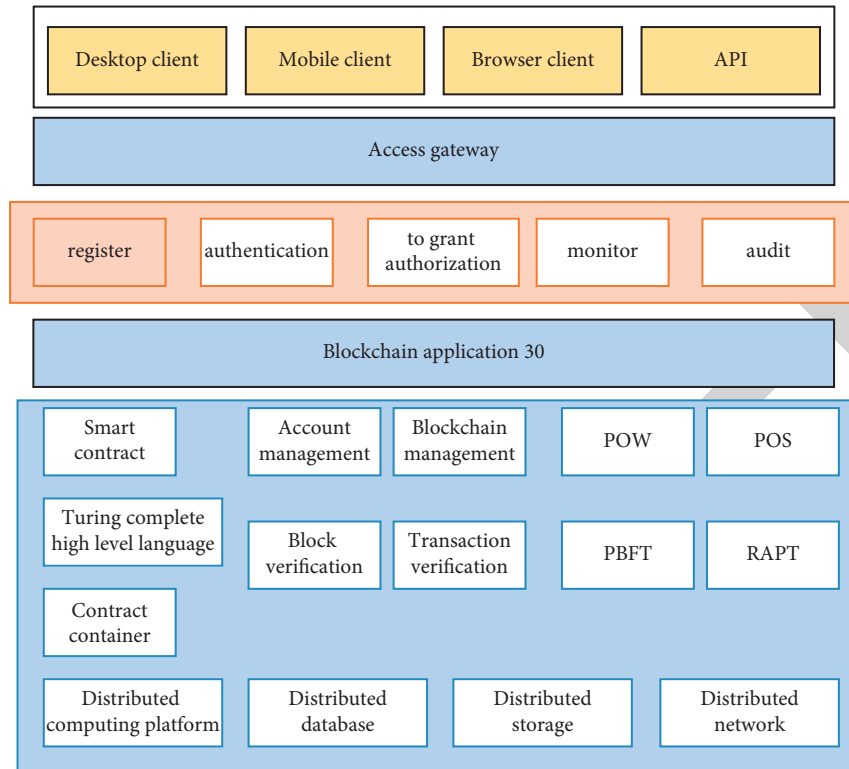


FIGURE 3: The blockchain 3.0 architecture.

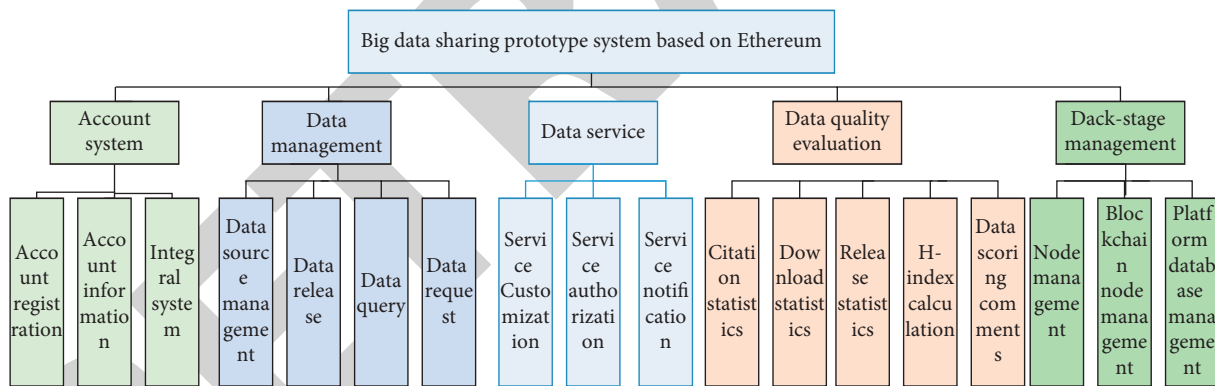


FIGURE 4: Functional architecture of the big data-sharing prototype system based on Ethereum.

automatically send a request to the data owner and handle relevant authorization operations by the system’s service publishing and subscription contract. After all authorizations involved in the service are completed, the service notification function sends a completion notification to the service customization. The module flow is shown in Figure 7.

4.1.4. Data Quality Evaluation Module. Data quality evaluation includes reference, download, and release statistics, along with the h-index calculation and other functions. The focus is to provide reference-based data evaluation and data feedback mechanisms. Data quality evaluation is divided into data set quality evaluation and data publisher influence evaluation. Data citation and download mean users’ recognition of data and its

publishers. Data comment and feedback play an important role in data management and sharing. Among them, the h-index measures the amount of published data and the number of cited references. The two basic measures are about the number of articles whose cited frequency is greater than or equal to H. If the h-index of an institution/individual is 30, it indicates that the subject has at least 30 data sets whose citation frequency is greater than or equal to 30.

To ensure the uniqueness of the data set, the naming format is required to follow the BibTeX standard to facilitate the citation of article authors, retrieval, and verification of publishers and other scholars. Each dataset name is unique, such as BJUT/WSN/forest_Temp refers to the organization, collection method, and data name for collecting data.

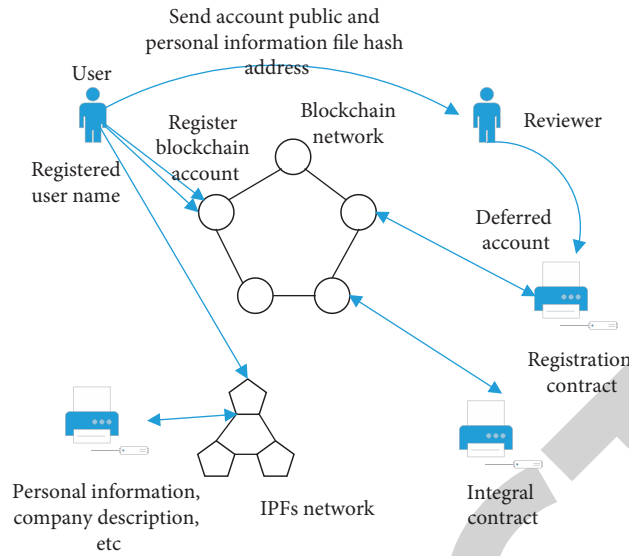


FIGURE 5: The user registration process.

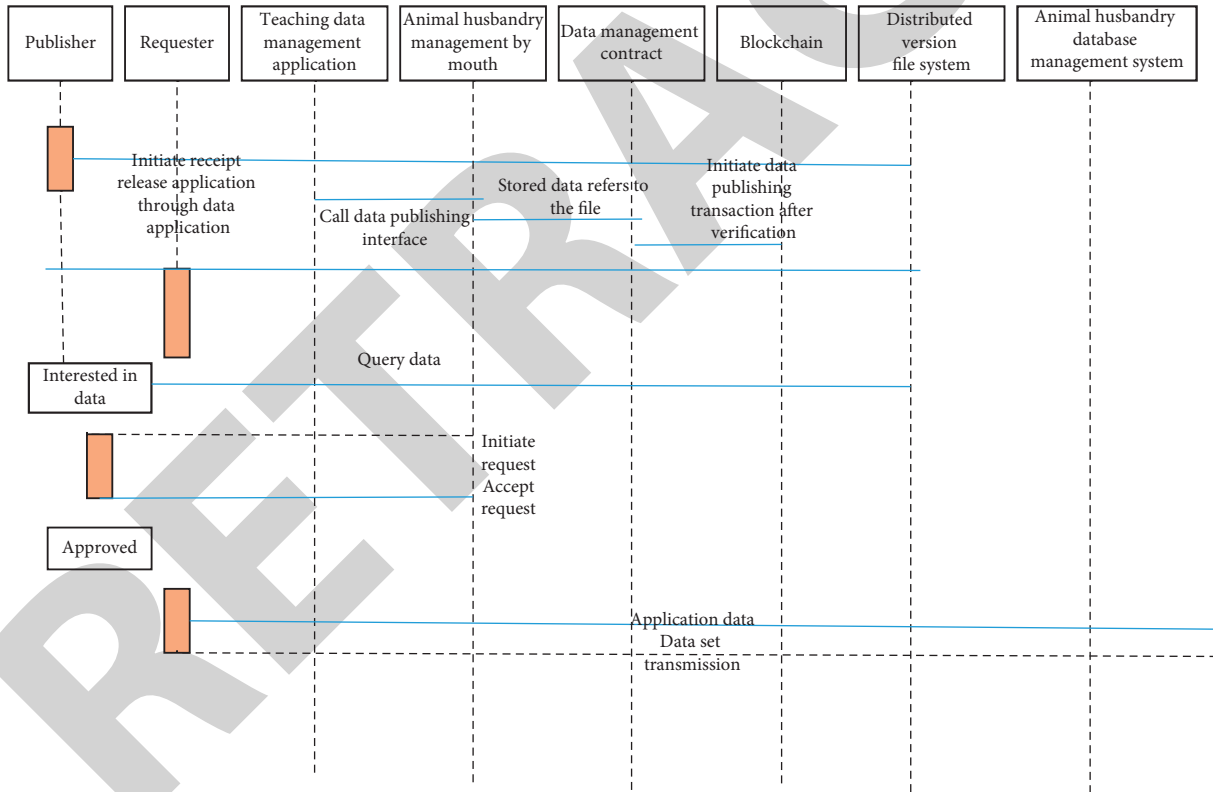


FIGURE 6: Schematic of the data management sequence.

4.2. *Prototype System Implementation.* The corresponding relationship between the technology adopted by the prototype system and each layer of the architecture is shown in Figure 8.

The application layer adopts bootstrap, HTML5, and other front-end technologies to be executed on the web browser and mobile terminal. It ensures the display of different services in a visual fashion. The service layer uses Laravel and web3.js to package various contract functions

into different service interfaces. The contract layer develops smart contracts containing the business logic of each module based on the solid language, and the blockchain adopts Ethereum go Ethereum nodes. The IPF interstellar file system is used in the routing layer, and MYSQL, MongoDB, and FTP-based database server are used in the data storage layer. The stored data sets come from various sensor data collected by sd-wsn and Intel Edison board development kit.

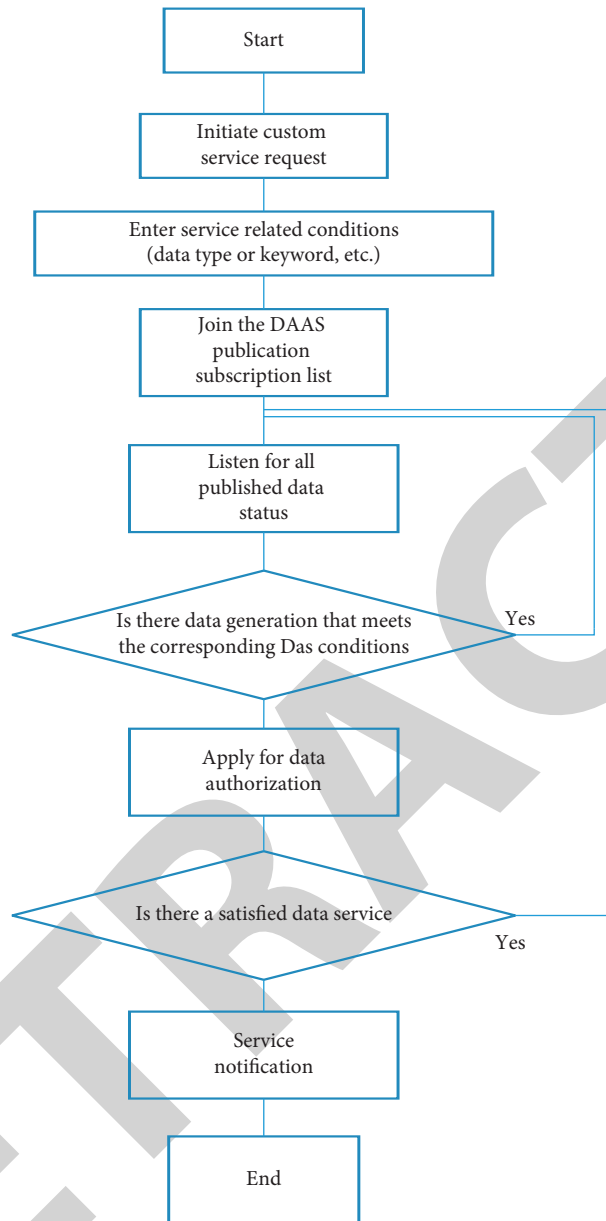


FIGURE 7: Flow chart of the data service module.

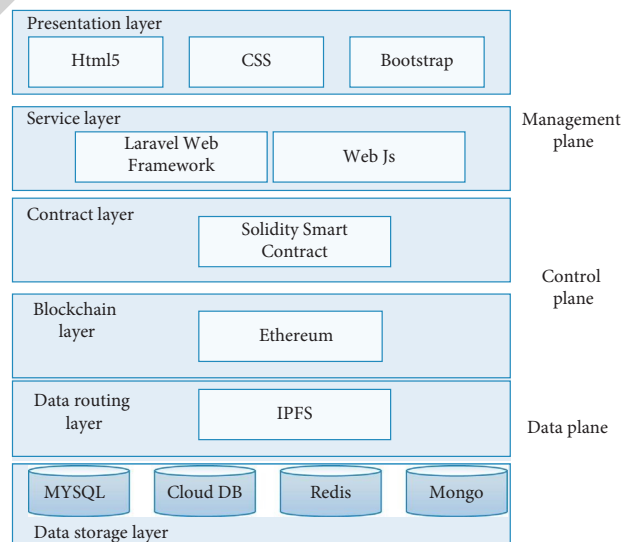


FIGURE 8: The technical architecture of big data-sharing prototype based on Ethereum.

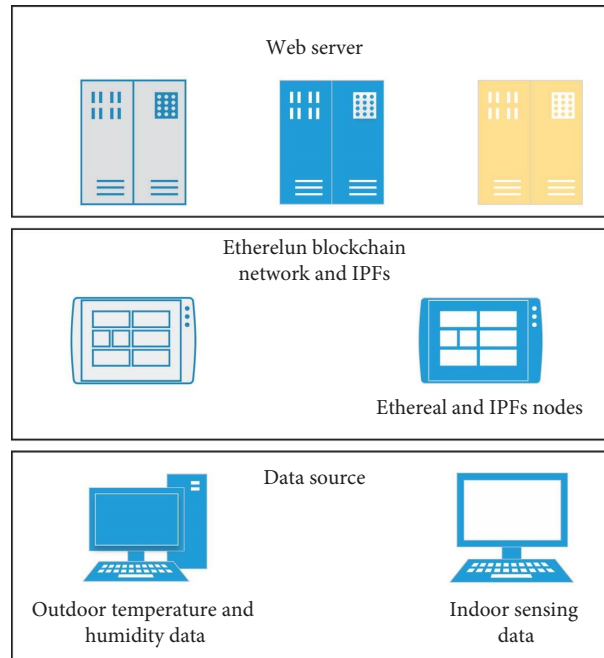


FIGURE 9: Prototype deployment diagram.

4.2.1. Ethereum and IPFs Network Construction. This section mainly introduces the construction of the Ethereum private chain and smart contract development environment and the application of IPFs. As a decentralized blockchain platform that can run smart contracts, Ethereum provides blockchain node clients based on various languages. In this paper, the Ethereum client node based on the go language (get for short) is adopted.

4.2.2. Smart Contract Set. In this paper, the contract is implemented in solid language. The conventions are as follows: the upper half of the contract is a member variable, the lower half is a function, + indicates a public variable (visible to all accounts), - indicates a private variable (visible to individuals), # and # indicate that only those with specific permissions can call and view it. The main contracts are introduced as follows.

4.3. Deployment Environment. In this paper, IPFs and Ethereum blockchain networks are deployed on the 20 hosts of Alibaba Cloud and Tencent Cloud server and laboratory. The mining consensus main network is composed of four desktop computers with Windows 7 system and 4G memory in the laboratory, which are deployed on the Ubuntu system with a 64-bit memory of 1g, as shown in Figure 9.

5. Performance Test and Analysis of the Prototype System

The Ethereum node can automatically synchronize the existing transaction data, participate in the maintenance of network security and stability, and carry a satisfactory level of scalability and reliability, as shown in Figure 10. In order

to verify the system concurrency and storage efficiency, the performance test results of Ethereum private chain mining speed and IPFs network transmission speed are shown in Figures 11 and 12.

The dual mining nodes of the system can reach 1366 block consensus within one hour. Block 5231011 of the Ethereum public chain contains 309 transactions, which can process approximately 300 messages concurrently and give feedback at the rate of 20 times/minute. On average, CrowdAd's total citations per month are about 40, and the most frequently downloaded data sets are 12.59 times per month. Dava.gov has the highest average monthly downloads of 9001, as shown in Table 1. Therefore, using the Ethereum blockchain to record the core interaction process can fully meet the sharing needs. The file access time of IPFs less than 10 m can be stabilized within 20 s, and the file transmission time of less than 100 m can be stabilized within 200 s. The average size of word per page is approximately 20 K, audio is approximately 10 m, and video is approximately 100 m. Therefore, it is completely feasible to use IPFs for nonbusiness core file storage.

The test results show that the flat prototype system can meet the sharing requirements in performance and improve and perfect in reliability, scalability, and so on.

Figure 13 shows the data storage distribution probability. A random variable x is given as the length of time, where the unit of time is taken as minutes. Over time, the data variables share distribution or have the same density function (or similar functions change some constants). The storage method focuses on the penetration of result-oriented data storage, presenting a student-centered tour guide business class with ideological and political elements. Data storage can be in the form of group discussion, group debate, role

```

INFO [05-11|06:45:00] Starting mining operation
null
> INFO [05-11|06:45:00] Commit new mining work          number=13 txs=3 uncles=0 elapsed=10.297ms
> miner.stop()INFO [05-11|06:45:06] Successfully sealed new block          number=13 hash=1b4309_c56265
INFO [05-11|06:45:06] mined potential block          number=13 hash=1b4309_c56265
INFO [05-11|06:45:06] Commit new mining work          number=14 txs=0 uncles=0 elapsed=1.212ms
> miner.stop()INFO [05-11|06:45:07] Successfully sealed new block          number=14 hash=5ae1e8_b55e70
INFO [05-11|06:45:07] mined potential block          number=14 hash=5ae1e8_b55e70
INFO [05-11|06:45:07] Commit new mining work          number=15 txs=0 uncles=0 elapsed=189µs
INFO [05-11|06:45:07] Successfully sealed new block          number=15 hash=84262e_f3ac00
INFO [05-11|06:45:07] mined potential block          number=15 hash=84262e_f3ac00
INFO [05-11|06:45:07] Commit new mining work          number=16 txs=0 uncles=0 elapsed=177µs

true
> INFO [05-11|06:49:45] Submitted transaction          fullhash=0x918e3b17cdef656d7538a8f6e9cd524d820f343f42
> miner.start()
INFO [05-11|06:49:52] Updated mining threads          threads=0
INFO [05-11|06:49:52] Starting mining operation
null
> INFO [05-11|06:49:52] Commit new mining work          number=16 txs=1 uncles=0 elapsed=2.632ms
> miner.stop()INFO [05-11|06:49:53] Successfully sealed new block          number=16 hash=733fd2_06acc9
INFO [05-11|06:49:53] mined potential block          number=16 hash=733fd2_06acc9
INFO [05-11|06:49:53] Commit new mining work          number=17 txs=0 uncles=0 elapsed=420µs
INFO [05-11|06:49:54] Successfully sealed new block          number=17 hash=3d5623_06fd95
INFO [05-11|06:49:54] mined potential block          number=17 hash=3d5623_06fd95
INFO [05-11|06:49:54] Commit new mining work          number=18 txs=0 uncles=0 elapsed=309µs
    
```

FIGURE 10: Automatic data synchronization of Ethereum node.

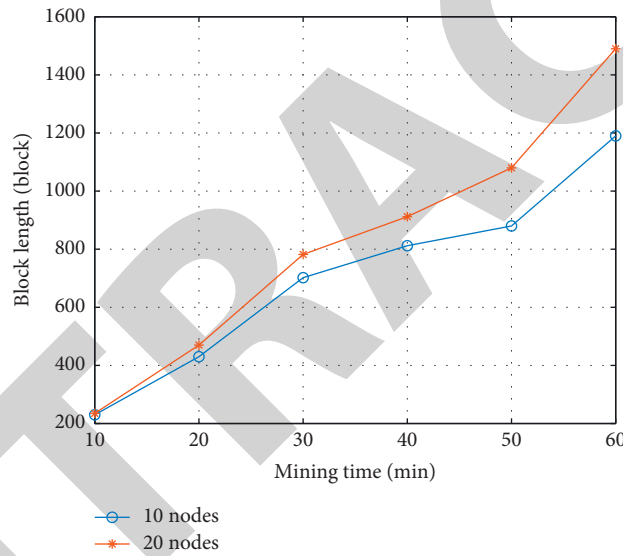


FIGURE 11: The node mining speed (in min).

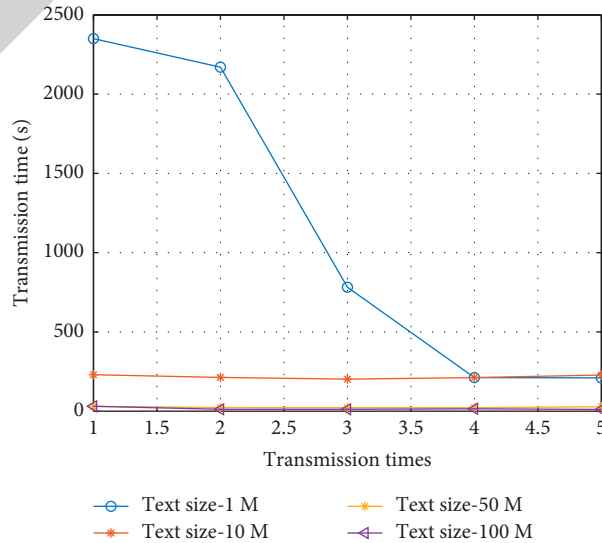


FIGURE 12: The IPFs transmission speed.

TABLE 1: Top five institutions of data upload and total upload of dava.gov in 2018.

Data sources	2018.1	2018.2	2018.3	2018.4	2018.5	Total over the years
Ministry of commerce	1	9.001	18	4	0	67.298
Department of homeland security	0	81	0	0	0	8.877
Ministry of energy	19	27	93	15	0	2.499
Ministry of health	3	28	2	7	0	1.417
Ministry of agriculture	1	401	299	0	0	1.417

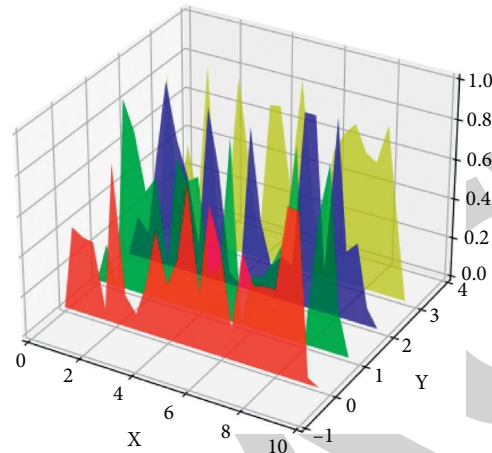


FIGURE 13: The data storage distribution.

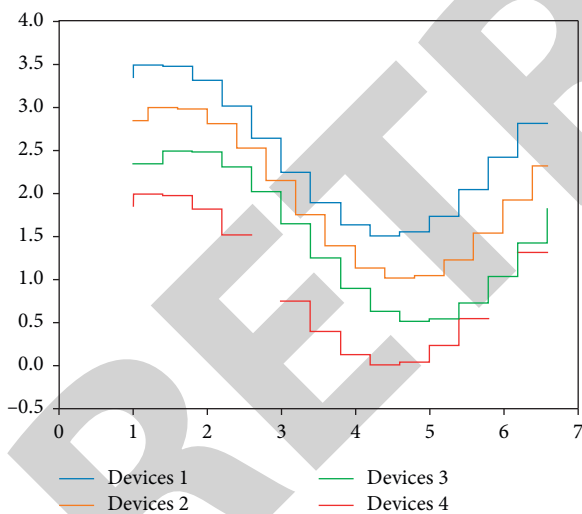


FIGURE 14: The data storage endurance of different equipment.

play, group PK, brainstorming, blue ink voting, and other forms, so that the students can truly participate in the classroom and actively think and reflect. For example, for a debate on the theme “*the rise of online tour guides, are you willing to try?*” the positive and negative sides not only cultivate the spirit of full cooperation but also shape their correct values by a wonderful defense.

In Figure 14, different devices used for data storage are depicted in distinct colors. The data saved included the word, picture, audio, video, and other types of files. The data can be converted into a format occupying less space

so that more data can be stored. In fact, there are many ways to save data, such as optical disk, hard disk, mechanical hard disk, and network disk. Therefore, if one has certain funds, it is recommended that one chooses a safer storage medium.

6. Conclusions

This paper implements a decentralized big data-sharing prototype system based on the Ethereum blockchain and related technologies. The study describes in detail the design and implementation of each layer of the system and collects relevant data for functional testing. The results obtained show that the prototype effectively verifies the feasibility of the model besides elaborating the hierarchical architecture and the relevant sharing mechanism. The study provides a platform for the data sharing of actual projects and can be followed for theoretical innovation and practical implementation. As our future strategy, we are to enhance the method to effectively control and manage big data, particularly educational research, using machine learning techniques.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] J. Gantz and D. Reinsel, "Extracting value from chaos," *IDC Iview*, vol. 1142, pp. 1–12, 2011.
- [2] Y. Ren, F. Zhu, P. K. Sharma et al., "Data query mechanism based on hash computing power of blockchain in internet of things," *Sensors*, vol. 20, no. 1, p. 207, 2020.
- [3] P. K. Sharma, J. H. Park, and K. Cho, "Blockchain and federated learning-based distributed computing defence framework for sustainable society," *Sustainable Cities and Society*, vol. 59, Article ID 102220, 2020.
- [4] D. Li, W. Liu, L. Deng, and B. Qin, "Design of multimedia blockchain privacy protection system based on distributed trusted communication," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 2, Article ID e3938, 2021.
- [5] O. Said and A. Tolba, "DORS: a data overhead reduction scheme for hybrid networks in smart cities," *International Journal of Communication Systems*, vol. 33, no. 12, Article ID e4435, 2020.
- [6] C. de Villiers, S. Kuruppu, and D. Dissanayake, "A (new) role for business - promoting the United Nations' Sustainable Development Goals through the internet-of-things and blockchain technology," *Journal of Business Research*, vol. 131, pp. 598–609, 2021.
- [7] Z. Su and Q. Xu, "Security-aware resource allocation for mobile social big data: a matching-coalitional game solution," *IEEE Transactions on Big Data*, vol. 7, no. 4, pp. 632–642.
- [8] 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, 2016.
- [9] Y. Ren, F. Zhu, and K. Zhu, "Blockchain-based trust establishment mechanism in the internet of multimedia things," *Multimedia Tools and Applications*, vol. 80, no. 20, Article ID 30653, 2021.
- [10] H. Si, C. Sun, Y. Li, H. Qiao, and L. Shi, "IoT information sharing security mechanism based on blockchain technology," *Future Generation Computer Systems*, vol. 101, pp. 1028–1040, 2019.
- [11] T. Xie, C. Zhang, Z. Zhang, and K. Yang, "Utilizing active sensor nodes in smart environments for optimal communication coverage," *IEEE Access*, vol. 7, Article ID 11338, 2018.
- [12] Z. Lv, L. Qiao, M. S. Hossain, and B. J. Choi, "Analysis of using blockchain to protect the privacy of drone big data," *IEEE Network*, vol. 35, no. 1, pp. 44–49, 2021.
- [13] L. Fan, J. R. Gil-Garcia, Y. Song, F. Cronemberger, G. Hua, and D. Werthmuller, G. Burke, J. Costello, B. R. Meyers, and X. Hong, "Sharing big data using blockchain technologies in local governments: some technical, organizational and policy considerations," *Information Polity*, vol. 24, no. 4, pp. 419–435, 2019.
- [14] A. K. Shrestha, J. Vassileva, and R. Deters, "A blockchain platform for user data sharing ensuring user control and incentives," *Frontiers in Blockchain*, vol. 3, p. 48, 2020.
- [15] Z. Zhang and L. Zhao, "A Design of Digital Rights Management Mechanism Based on Blockchain technology," in *Proceedings of the International Conference on Blockchain*, pp. 32–46, Springer, Seattle, WA, USA, June 2018.
- [16] L. Guo, H. Xie, and Y. Li, "Data encryption based blockchain and privacy preserving mechanisms towards big data," *Journal of Visual Communication and Image Representation*, vol. 70, Article ID 102741, 2020.
- [17] 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, 2019.
- [18] A. Z. Faroukhi, I. El Alaoui, Y. Gahi, and A. Amine, "Big data monetization throughout big data value chain: a comprehensive review," *Journal of Big Data*, vol. 7, no. 1, pp. 1–22, 2020.
- [19] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, "Customer relationship management and big data enabled: personalization & customization of services," *Applied Computing and Informatics*, vol. 15, no. 2, pp. 94–101, 2019.
- [20] L. Satish and N. Yusof, "A review: big data analytics for enhanced customer experiences with crowd sourcing," *Procedia Computer Science*, vol. 116, pp. 274–283, 2017.
- [21] T. Wang, Q. Xia, X. Chen, and X. Jin, "Use of Baidu index to track Chinese online behavior and interest in kidney stones," *Risk Management and Healthcare Policy*, vol. 13, p. 705, 2020.
- [22] C. Montag, B. Becker, and C. Gan, "The multipurpose application WeChat: a review on recent research," *Frontiers in Psychology*, vol. 9, p. 2247, 2018.
- [23] J. Wang, "Analysis of Young Chinese Users of Sina Weibo Based on Uses and Gratifications Theory," Thesis, Jönköping University, Jönköping, Sweden, 2018.
- [24] F. Bravo-Marquez, S. Reeves, and M. Ugarte, "Proof-of-learning: a blockchain consensus mechanism based on machine learning competitions," in *Proceedings of the 2019 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPCON)*, pp. 119–124, Newark, CA, USA, April 2019.
- [25] X. Liu, Z. Wang, C. Jin, F. Li, and G. Li, "A blockchain-based medical data sharing and protection scheme," *IEEE Access*, vol. 7, Article ID 118943, 2019.
- [26] L. Wang, C. Zhang, Q. Chen et al., "A Communication Strategy of Proactive Nodes Based on Loop Theorem in Wireless Sensor Networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, IEEE, Wanzhou, China, November 2018.
- [27] E. Sifah, Q. Xia, K. O. B. O. Agyekum et al., "Chain-based big data access control infrastructure," *The Journal of Supercomputing*, vol. 74, no. 10, pp. 4945–4964, 2018.
- [28] U. U. Uchibeke, K. A. Schneider, S. H. Kassani, and R. Deters, "Blockchain Access Control Ecosystem for Big Data security," in *Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 1373–1378, IEEE, Halifax, NS, Canada, August 2018.
- [29] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.
- [30] T. T. Thwin and S. Vasupongayya, "Blockchain Based Secret-Data Sharing Model for Personal Health Record system," in *Proceedings of the 2018 Fifth International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, pp. 196–201, IEEE, Krabi, Thailand, August 2018.
- [31] R. J. Garratt and M. R. Van Oordt, "Privacy as a public good: a case for electronic cash," *Journal of Political Economy*, vol. 129, no. 7, pp. 2157–2180, Article ID 21260, 2021.
- [32] Y. Sun, J. Zhang, Y. Xiong, and G. Zhu, "Data security and privacy in cloud computing," *International Journal of Distributed Sensor Networks*, vol. 10, no. 7, Article ID 190903, 2014.

- [33] G. Zyskind, O. Nathan, and A. Pentland, "Decentralizing privacy: using blockchain to protect personal data," in *Proceedings of the 2015 IEEE Security and Privacy Workshops*, pp. 180–184, San Jose, CA, USA, May 2015.
- [34] S. Ahmed and N. Ten Broek, "Blockchain could boost food security," *Nature*, vol. 550, no. 7674, p. 43, 2017.
- [35] D. Ali, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "LSB: a lightweight scalable blockchain for IoT security and privacy," 2019, <https://arxiv.org/abs/1712.02969> arXiv preprint.
- [36] R. Yuan, Y. B. Xia, H. B. Chen, B. Y. Zang, and J. Xie, "Shadoweth: private smart contract on public blockchain," *Journal of Computer Science and Technology*, vol. 33, no. 3, pp. 542–556, 2018.
- [37] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of big data challenges and analytical methods," *Journal of Business Research*, vol. 70, pp. 263–286, 2017.
- [38] M. Ali, J. Nelson, R. Shea, and M. J. Freedman, "Blockstack: a global naming and storage system secured by blockchains," in *Proceedings of the USENIX Annual Technical Conference (USENIX ATC 16)*, pp. 181–194, Denver, Colorado, June 2016.
- [39] A. Alammary, S. Alhazmi, M. Almasri, and S. Gillani, "Blockchain-based applications in education: a systematic review," *Applied Sciences*, vol. 9, no. 12, p. 2400, 2019.
- [40] W. Li, J. Wu, J. Cao, N. Chen, Q. Zhang, and R. Buyya, "Blockchain-based trust management in cloud computing systems: a taxonomy, review and future directions," *Journal of Cloud Computing*, vol. 10, no. 1, pp. 1–34, 2021.
- [41] 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.
- [42] A. Farouk, A. Alahmadi, S. Ghose, and A. Mashatan, "Blockchain platform for industrial healthcare: vision and future opportunities," *Computer Communications*, vol. 154, pp. 223–235.
- [43] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using Blockchain for Medical Data Access and Permission management," in *Proceedings of the 2016 Second International Conference on Open and Big Data(OBD)*, pp. 25–30, Vienna, Austria, August 2016.
- [44] H. A. Lee, H. H. Kung, J. G. Udayasankaran et al., "An architecture and management platform for blockchain-based personal health record exchange: development and usability study," *Journal of Medical Internet Research*, vol. 22, no. 6, Article ID e16748, 2020.
- [45] M. Swan, *Blockchain: Blueprint for a New Economy*, O'Reilly Media, Inc., California, CA, USA, 2015.