Overview of Blockchain and Batabase

Cheng Hu [1,*], Yuanguang Ding[2], Chunyi Wang[1], Yihao Zhai[1]

[1] School of International Education, Nanjing Institute of Technology, Nanjing 211167, China;

[2] School of Automation, Nanjing Institute of Technology, Nanjing 211167, China.

* Corresponding Author

Abstract

In recent years, the application of blockchain in finance and other fields has increased a lot. The vigorous development of this technology shows its deep potential and research value. The purpose of blockchain design is to achieve safe and reliable data storage, thus sacrificing efficiency. At the same time, the disadvantage of less technology accumulation makes blockchain lack of good auxiliary technical support. These problems limit the further development of blockchain technology. As a data storage technology, database has developed for decades and its technical system is relatively mature. Both blockchain and database are widely used data storage technologies, and there are many similarities and differences between them. It is a valuable research direction to integrate the two technologies through system architecture design, make the two technologies complement each other, and realize a new generation of data storage technology. Through the elaboration of the design concept, technical characteristics and overall architecture of blockchain and database technology, and taking the integration paradigm of blockchain-based database and database-based blockchain as the benchmark, this paper summarizes the existing work of blockchain and database technology integration. On this basis, we further analyze the key issues and future development direction of the integration of blockchain and database technology from all dimensions of data storage technology design.

Keywords

Blockchain, Database, Distributed System.

1. Introduction

As the supporting technology of digital currency, blockchain technology has received wide attention in various fields since it was formally defined in the concept of Bitcoin in 2008, especially its application in the financial field, which can be regarded as a model of the integration of computer and financial technology. The essence of blockchain is a blockchain data structure that stores information. It is used to verify the validity of its information through the integration of distributed technology, encryption technology and other technologies to achieve true and reliable data storage. Generally speaking, a traditional blockchain, that is, a narrow blockchain, is composed of multiple blocks connected with each other according to the generated time sequence, in which each block stores relevant block information and data in a predefined format. Because blockchain replicas are stored on multiple distributed servers, also known as nodes, mutual verification and restriction between multiple nodes can ensure that the data in the entire blockchain is not tampered with, greatly improving the security and credibility of data storage in the entire system, and solving the trust problem between users in the system.

As a relatively mature and reliable technology, the database, which has been used for data storage for decades, is an important basis for building modern computer systems. Unlike the

ISSN: 2664-9640

blockchain that focuses on improving the credibility of data storage, the purpose of database design is more pure, that is, only data persistence. In order to cope with the increasing demand of users for data, modern database systems provide a complete system of organizing, storing and managing data. From hierarchical database to relational database to non-relational database (NoSQL), with the change of data structure used for storage, database technology also provides rich convenience design for different application scenarios.

Although blockchain and database are both data storage technologies, due to the different design purposes, their technical characteristics are also significantly different. Blockchain technology focuses on security and credibility, but its efficiency is poor, and lacks mature optimization technology and design, which can only support simple basic operations on data. Database technology not only provides basic data storage, but also considers the availability of data. It is not only efficient, but also provides rich and convenient data operations. With the continuous improvement of informatization in various industries, we are ushering in an era of rapid expansion of information. The traditional single data storage technology has been difficult to meet the growing demand for data storage. For example, the traceability of goods has high requirements for data security, and its time series data storage characteristics also naturally conform to the characteristics of blockchain data storage. However, the efficiency of blockchain data query is limited, which is difficult to meet the actual needs; In addition, in the multi-party trade scenario, the commonly used traditional distributed database storage is difficult to meet the requirements of data consistency, and economic disputes often occur due to data differences. Under these circumstances, a new data storage system combining the characteristics of blockchain and database data storage technologies can better meet the actual business needs.

Both blockchain and database technologies belong to data storage technology, and each has its own advantages and disadvantages. Therefore, it is an effective idea to integrate the two technologies. In recent years, there has also been a lot of work focusing on the technical integration of blockchain and database, optimizing the design from the overall architecture, application functions, storage structure and other aspects, and realizing more comprehensive data storage by using the complementary advantages between the two technologies. However, due to completely different design concepts, the storage structure and function design of blockchain and database are very different, and there are many contradictions in their main characteristics. Simply overlaying or splicing the two types of data storage systems cannot achieve an effective new data storage system. Therefore, there are many difficulties in the integration of blockchain and database technologies. It is not only necessary to design a suitable data storage structure, but also to design a reasonable overall architecture to efficiently integrate the characteristics of the two types of technologies without too much damage to their own characteristics.

Starting from the design concept of blockchain and database, this paper introduces the technical characteristics of blockchain and database in detail, as well as their similarities and differences in infrastructure. Based on the existing work of blockchain database technology fusion design, and according to the key ideas of their respective design, this paper expounds in detail the relationship between their design and characteristics and makes a comprehensive analysis, and accordingly proposes a paradigm for the fusion of the two types of technologies, which provides a reference idea for future research in this direction.

2. Overview of Blockchain

2.1. Definition of Blockchain

Blockchain is a blockchain data structure used for data storage. It was first proposed by a person who called himself Satoshi Nakamoto in the article "Bitcoin: A Point-to-point Electronic

ISSN: 2664-9640

Cash System" on November 1, 2008, and was fully defined with the birth of Bitcoin. Because blockchain was born with Bitcoin, a digital currency. In early research and practice, blockchain often existed as the core supporting technology of digital currency. It is a block-type data structure that stores data and connects each other into a chain-type data structure in chronological order. Through encryption technology, data can be tampered with and forged, It is mainly used to store digital currency transactions, that is, blockchain in a narrow sense. With the continuous development of blockchain technology, new technologies have also been incorporated into the blockchain system, such as smart contracts, alliance chains, etc. While enriching the blockchain functions, it has also gradually changed the blockchain from an appendage of digital currency to an independent data storage technology, that is, the generalized blockchain. Broad blockchain technology, in addition to data storage based on blockchain data structure, focuses more on the management and use of data, such as the use of consensus algorithms in game theory to add data, the use of cryptography technology to maintain data security, and the provision of programmable smart contracts to achieve automated scripts.

Generally speaking, a blockchain system consists of data layer, network layer, consensus layer, incentive layer, contract layer and application layer. The data layer, as the lowest and most basic part, provides the specification for encapsulating data into blocks and basic algorithms such as asymmetric encryption and timestamp to ensure data security, and realizes data storage by connecting with the data storage engine. The network layer mainly realizes the communication between various nodes in the blockchain network. It relies on distributed technology to build a network for multiple nodes that maintain the blockchain, and provides stable and reliable data transmission and verification mechanism to maintain the stable operation of the entire blockchain network. The consensus layer is the core layer of the blockchain system to achieve information security. By encapsulating various consensus algorithms based on game theory, the network layer is used to communicate between nodes to achieve data tamper-proof and unforgeable. As an extension of the consensus layer, the incentive layer introduces economic factors into the blockchain system, provides the generation and distribution mechanism of economic incentives, provides the basis for the operation of consensus algorithms in the consensus layer, and ensures that the blockchain system can operate spontaneously, persistently and independently. The contract layer is the key to provide programmable modules and improve the scalability and convenience of the blockchain. It provides flexible and available functions for the blockchain system by encapsulating various automatic scripts, algorithm mechanisms and smart contracts. The application layer is separated from the technical level and focuses on various application scenarios. It provides different response mechanisms to different requirements by encapsulating the underlying functions and technologies.

Different from traditional distributed databases, blockchain technology pays more attention to the security of data, and ensures the data can not be tampered with and forged through complex encryption and consensus algorithms and communication between nodes. Due to its blockchain data structure based on time stamps and consensus algorithm based on economic incentives, its operation efficiency is greatly reduced compared with traditional distributed databases, which is also an important factor limiting the development of blockchain. On the whole, blockchain technology has rich application scenarios in finance, foreign trade and other fields due to its advantages of security and credibility, and is gradually replacing the status of traditional distributed database, but the efficiency and convenience problems it brings to improve security have greatly limited its development in other fields.

ISSN: 2664-9640

2.2. Characteristics of Blockchain

2.2.1. Decentralization

Decentralization is the most prominent feature of blockchain compared with traditional distributed databases. Traditional distributed databases often distribute permissions among multiple storage nodes based on master-slave structure. There are one or more central nodes with higher permissions in the whole system. Although the existence of these central nodes improves the efficiency of data management in the database system, it introduces the risk that data may be tampered with and forged, and brings data security problems. In the blockchain system, each node can retain a copy of the blockchain, and verify the data based on it without querying other nodes, thus realizing the data storage decentralization. Compared with centralized database, blockchain technology uses P2P (Point to Point) to communicate, avoiding trust problems caused by multi-hop communication to the greatest extent, and realizing decentralized communication between nodes. At the same time, the blockchain relies on the consensus algorithm and incentive mechanism to achieve decentralized or multi-centric distributed applications through the game constraints between all nodes in the entire blockchain network.

2.2.2. Security

The design concept of blockchain is data security. The blockchain is based on the blocks that store data, and is connected into a chain-like data structure in one direction through the time sequence of block generation. In the block, the stored data is encrypted using asymmetric encryption technology to ensure the security of the data, and the Merkle tree (https://github.com/ethereum/wiki/wiki/Patricia-Tree)The data is stored in the form of Merkle tree. The root of Merkle tree is stored to verify the data conveniently. Once the stored data has errors, it can be quickly judged based on it. After each block is generated, it is also necessary to add relevant block element information, including the hash value, timestamp, random number, etc. of the previous block, and broadcast it to all nodes in the blockchain network. After verification, it can be connected to the existing blockchain, further ensuring data security. In addition, in the blockchain system, the existence of the consensus algorithm between nodes also makes the economic cost of tampering with forged data expensive. Because the blocks of the blockchain are connected in chronological order and the latter block contains the information of the previous block, tampering with forged data requires modifying all blocks after the block, and also requires huge computational power or capital to pass the consensus algorithm, Such extreme conditions also guarantee the security of blockchain data storage. Unlike the storage form of adding, deleting, changing and checking data in the database, the blockchain system stores data in the way of adding only. This way can ensure that all the historical states of data are saved on the blockchain, realizing the data can not be tampered with and forged. Compared with the database system where the log and data records can be modified, the data security and credibility of the blockchain system are more significant.

2.2.3. Traceability

Another key feature that blockchain can be widely used in the financial field is traceability. Although there are corresponding records for data addition, deletion, modification and query in the database system, that is, log mechanism, these records are difficult to persist and store, and it is difficult to obtain their historical status after multiple modifications. Existing mainstream databases cannot access the data of the old version, and technologies such as snapshot isolation through multi-version concurrency control can only solve this problem to a limited extent. Database systems are difficult to trace from both design and application perspectives. On the contrary, data modification is not allowed in the blockchain system, but can only be implemented in incremental form by adding data records. This incremental form can effectively retain all data storage records, and avoid the problem that the historical status

ISSN: 2664-9640

of data cannot be obtained after multiple modifications. At the same time, the chain structure of the blockchain organized in chronological order also makes it easier to trace data. The data on the entire blockchain can be traced by traversing the data in block order. However, due to the performance problems of most blockchain systems at present, large-scale traceability of data is expensive and still difficult to apply in practice.

2.2.4. Anonymity

The anonymity of the blockchain system is not completely anonymous. It is mainly achieved through the encryption of accounts, transactions and other information. In the database system, the data stored in clear text is completely anonymous, and the data stored in encryption is also greatly damaged due to the existence of centralized nodes. In contrast, user information and data information in the blockchain system are encrypted using encryption algorithms, and the decentralized nature avoids the destruction of data encryption by the central node, which can achieve anonymity to a considerable extent.

3. Overview of Database

3.1. Definition of Database

Unlike blockchain, which was born in 2008, database technology has a history of decades, is an important part of modern computer systems, and plays an important cornerstone role in the computer field. Database technology was born out of the demand for data persistence, and gradually evolved into the function of data organization and management. Looking back on the decades of database development, with the continuous change of users' data storage requirements, the database architecture is also constantly evolving. From hierarchical database to mesh database, to relational database, and even the latest NoSQL (NotOnlySQL, non-relational database) and NewSQL (new database) have played a significant role in promoting the vigorous development of the computer field at that time.

After decades of development, database technology has covered all aspects of technology and formed a relatively mature and complete architecture. Taking the current mainstream relational database as an example, it uses a relational model to organize data in the form of tables and store data in the form of rows and columns. In order to make the data storage more standardized, and also facilitate the optimization of query and storage space, relational databases require that the meta information of each field be defined before the data storage. At the same time, the defined data table cannot be easily changed, and the data can only be added vertically.In addition, the design of relational databases emphasizes ACID rules, namely atomicity, consistency, isolation and durability. Atomicity refers to the controllability of the database to meet the atomic granularity of transaction operations to ensure that the database can roll back transactions when necessary; Consistency requires that the database can only switch between consistency states to avoid the problem of data inconsistency caused by the interruption of the database during the execution of transactions; Isolation means that when dealing with concurrent access, the database needs to isolate the transactions of different users from each other, so as to avoid the impact on the results caused by mutual interference between transactions; Persistence means that each state of the database needs to be stable and persistent, which is the key to maintain the stable operation of the relational database system. However, due to the limitation of tabular data structure and strict ACID definition specification, relational databases are difficult to cope with the growing demand for data volume and complex data management requirements. At the same time, the current data and computing power are distributed on multiple nodes to ensure high availability and scalability. Therefore, it is necessary to design a new database architecture more suitable for distributed storage.

In the context of distributed architecture design, there are two development directions of database technology, namely NoSQL and NewSQL. NoSQL, or NotOnlySQL, refers to the

ISSN: 2664-9640

database type of non-determinant storage. Its design concept is to use the distributed architecture to process the massive stream data with complex types that do not meet the data integrity requirements in the era of big data. NoSQL has a wide variety of data structures, such as columnar database, graph database, etc. It can design corresponding data structures for different business scenarios to meet different business needs. Unlike standard relational databases, NoSQL often trades for better distributed performance and horizontal scalability at the expense of ACID features, and its flexible design also allows NoSQL to perform specific optimization to obtain higher performance for specific business data. Although NoSQL can meet the storage and management requirements for massive data and provide users with strong scalability, NoSQL does not support the ACID feature, lacks a standard query language and general and convenient data management capabilities. To solve these problems, NewSQL takes into account the specification and convenience of relational databases and the high performance and scalability of NoSQL. On the basis of traditional relational databases, it optimizes the distributed architecture to meet the storage and management requirements of massive data.

NoSQL and NewSQL, as the representatives of distributed databases in the new era, are quite different from the design concept of blockchain. The core of blockchain is data security. Its distributed architecture is not to improve concurrency performance, but to use the game between distributed nodes to achieve decentralization under the consensus mechanism, so as to ensure the data is tamper-proof. The core of distributed database is high-performance, highly scalable, and highly robust data storage with high concurrency and multiple copies. Compared with the blockchain that sacrifices performance for security, distributed database pays more attention to performance and convenience.

3.2. Characteristics of Database

3.2.1. High Efficiency

In addition to providing the function of data storage, the database system also needs to meet the requirements of rich data management. Therefore, the efficiency of storage and query is an important factor in database design. Compared with the blockchain data structure of blockchain, the data structure adopted by the distributed database is better for concurrency support and more efficient for data storage and management. In addition, the consensus and incentive mechanism of blockchain also greatly limits the efficiency of blockchain, while distributed databases do not need to sacrifice performance on these additional mechanisms. On the whole, the restriction of blockchain replacing database for large-scale application is largely due to the performance gap.

3.2.2. Concurrency

Similar to efficiency, concurrency is one of the most important features of distributed databases. In the distributed database, each data storage or operation request can be realized by any node in the distributed network. The distributed data and computing power make each node relatively independent and can achieve efficient concurrent operation through the distributed architecture, significantly improving the data flow. At the same time, this concurrency can ensure data storage and computational redundancy between nodes, and can improve the availability and scalability of the entire distributed network. On the contrary, although the blockchain also adopts a distributed architecture, due to the particularity of its data structure, the data can only be operated in a sequential manner during the actual operation, and it is difficult to achieve concurrency, which also limits the performance of the blockchain.

3.2.3. Convenience

Convenience refers to the convenient mechanisms and functions provided in the database for data organization and management, as well as the optimization of convenient user use designed for different types of business scenarios. After decades of development, database technology includes many technologies that focus on improving user efficiency, such as sharding, replication, logs, snapshots, etc. Most of these mechanisms and functions are encapsulated as pluggable modules and exist independently of the database system. However, due to their important role in various practical application scenarios, they have gradually become an indispensable part of the database system. In contrast, blockchain is difficult to meet complex data organization and management requirements due to its highly standardized data structure and strict security mechanism, and its convenience in practical application is greatly reduced.

3.2.4. Scalability

In the distributed database, data is stored in the form of data tables or data blocks. At the same time, there are multiple nodes in the distributed network that provide data and computing power. Both the data level and the node level meet high scalability. The operation of one of the data blocks or nodes will not have too much impact on the overall system. The data storage structure organized by the blockchain in chronological order makes each data block highly correlated. At the same time, the introduction and exit of nodes will have a huge impact on the consensus and incentive mechanism of the whole system, and the scalability is poor compared with the distributed database.

4. Conclusion

By analyzing the design concepts and characteristics of blockchain and database, as well as the similarities and differences of architecture design, this paper summarizes the existing work on the integration of blockchain and database technology from the perspective of the paradigm of the integration of the two blockchain and database technologies, and expounds the advantages and disadvantages of the existing work in detail. At the same time, the technology integration paradigm of blockchain+database is described in detail, and the future trend and significance of blockchain+database technology integration are explained.

References

[1] Blockchain Life will host the 10th Global Blockchain and Crypto Forum in Dubai[J]. M2 Presswire,2023.

[2] Ortega Julio César Úbeda, RodríguezMolina Jesús, MartínezNúñez Margarita, Garbajosa Juan. A Proposal for Decentralized and Secured Data Collection from Unmanned Aerial Vehicles in Livestock Monitoring with Blockchain and IPFS[J]. Applied Sciences,2022,13(1).

[3] The Africa Blockchain Center (The ABC) graduates 105 blockchain programmers in its first cohort of Blockchain Programming Training - one of the first formal blockchain training programs in Africa[J]. M2 Presswire,2022.

[4] ARES Blockchain GmbH launches the most advanced ARM Carrier Mainboard ever made[J]. M2 Presswire,2022.

[5] VRRB, a US based blockchain startup which is focused on building a lightweight blockchain consensus algorithm.[J]. M2 Presswire,2022.

[6] Biswas Debajyoti, Jalali Hamed, Ansaripoor Amir H, De Giovanni Pietro. Traceability vs. sustainability in supply chains: The implications of blockchain[J]. European Journal of Operational Research,2023,305(1).

[7] OKLink has launched Aptos Blockchain explorer, the 20th Chain Supported by OKLink Blockchain Explorer[J]. M2 Presswire,2022.