ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 5, May 2024

File Storage Using InterPlanetary File System and Blockchain

[1] Yash Agarwal, [2] Yash Shah, [3] Rajdeep Chaurasia, [4] Dhiraj Bari, [5] Santosh Kumar
[1] [2] [3] [4] Undergraduate (AI&DS), Vishwakarma Institute of Information Technology, Pune, Maharashtra, India,
[5] Associate HOD (AI&DS), Vishwakarma Institute of Information Technology, Pune, Maharashtra, India
Corresponding Author Email: [1] yashagarwal9765@gmail.com, [2] yashromilshah@gmail.com,
[3] rajdeepchaurasia14331433@gmail.com, [4] dhirajbari911@gmail.com, [5] santosh.kumar@viit.ac.in

Abstract*— Traditional centralized file storage system relies on a single entity to store and access data which leads to problems such as non-transparency and lack of proper data security. Data storage has evolved significantly over the years transitioning from local storage devices such as USB and hard drives to cloud storages. While this form of storage may be beneficial for the users, it poses some difficulties as well. This includes problems such as single point of failure and privacy and security reasons. In this paper, we have discussed the use of IPFS (Inter Planetary File System), a decentralized peer-to-peer file system to secure the way the data is distributed and stored on the web. Blockchain technology along with its decentralized, unchangeable and its unambiguous nature provides a safe and verifiable way to manage file ownership and transfership in a decentralized manner. The proposed system consists of an IPFS network and an Ethereum-based smart contract for managing file ownership and transfers.

Keywords: Blockchain Technology, DAGs, Ethereum, IPFS, PoS, PoW.

I. INTRODUCTION

The IPFS is built on several core principles such as content addressing, distributed web and Merkle DAGs (Directed Acyclic Graphs). The IPFS (InterPlanteray File System) employs content addressing instead of using traditional location based addressing such as URLs(Uniform Resource Locators). IPFS stores immutable data, remove duplication, and obtain address information for storage nodes to search for files in the network [1]. The content addressing involves the use cryptographic hash of its contents to identify the file. Distributed web means the use of peer-to-peer network rather than a centralized one and DAGs is used to showcase and link data that ensures efficient data verification and tamper-resistance. Secondly we have also made use of Ethereum for developing smart contracts which are self-executing programs that run blockchain. Uses such as PoW (Proof of Work), PoS (Proof of Stake) make Ethereum ideal to validate and maintain the integrity of Blockchain. The mechanism behind proof-of-work was a breakthrough in the space because it simultaneously solved two problems. First, it provided a simple and moderately effective consensus algorithm, allowing nodes in the network to collectively agree on a set of canonical updates to the state of the Bitcoin ledger. Second, it provided a mechanism for allowing free entry into the consensus process, solving the political problem of deciding who gets to influence the consensus, while simultaneously preventing sybil attacks [5].

II. KEY FEATURES

A. Storing Files

1. The user selects the file(s) they want to store on the decentralized system.
2. The client application performs client-side chunking and encryption (optional) of the file(s).
3. The client interacts with the IPFS network, adding the file chunks and obtaining a unique CID(Content Identifier) for each file.
4. The client submits a transaction to the Ethereum smart contract, calling the `addFile` function and providing the CID and owner's address.
5. The smart contract records the file metadata (CID, owner address) on the blockchain, establishing ownership.

B. Managing Ownership

1. The client application queries the smart contract to retrieve the ownership records for the user's files.
2. The application displays the list of owned files, along with relevant metadata (e.g., file name, size, CID).
3. Users can view and manage the ownership of their files through the application's user interface.

C. Transferring Ownership:

1. The user initiates an ownership transfer through the client application, specifying the file CID and the recipient's Ethereum address.
2. The client submits a transaction to the Ethereum smart contract, calling the `transferOwnership` function with the provided CID and recipient address.
3. The smart contract verifies that the caller is the current owner of the file and performs access control checks.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 5, May 2024

4. If the checks pass, the smart contract updates the ownership mapping, transferring ownership of the file to the new owner.
5. The new owner can now manage and access the file through the decentralized system.

III. METHODOLOGY

A. System Architecture

The suggested solution is composed of an Ethereum-based smart contract to handle file ownership and transfers, and an IPFS network to store and distribute files. Data integrity and deduplication are guaranteed by content addressing as files are divided into segments and dispersed around the IPFS network. Using the immutability and transparency of the Ethereum blockchain, smart contracts allow users to securely move files between each other and record file ownership.

The IPFS architecture consists of several components, including:

- Distributed Hash Table (DHT): A decentralized key-value store for locating and retrieving content across the network.
- BitSwap: A data exchange protocol that facilitates efficient transfer of data blocks between peers.
- IPLD (InterPlanetary Linked Data): A data model for representing and linking data structures across different protocols and platforms.

B. System Flow:

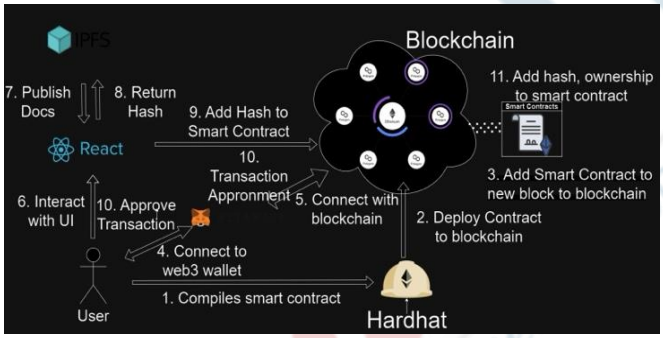

Fig .1 System Architecture

The proposed decentralised file storage system uses IPFS for storage of files and distribution along with Ethereum blockchain for managing the ownership and transfers through smart contracts. The diagram shown above can be explained through the following steps:

a. File Upload Process:-

1. After the user uploads a file, the client application chunks the file on the client side and, if desired, encrypts it.
2. The client communicates with the IPFS network to add file chunks and acquire the file's distinct Content Identifier (CID).
3. The file metadata (owner address, CID) is recorded on the blockchain by the client when they submit a transaction to the Ethereum smart contract.

b. File Distribution and Retrieval:

1. The CID is used for content addressing when distributing and storing the file chunks among several IPFS network nodes.
2. A user asks the IPFS network with the CID to download a file, and the network finds and retrieves the necessary pieces from nodes that are involved.
3. To recover the original file, the recoverable chunks are put back together and, if necessary, decrypted.

IV. POTENTIAL USE CASES

Reference [3] used the IPFS as their data sharing infrastructure and blockchain for transferring data in a peer-to-peer network to transport pre-trained deep learning models to others.[4] combined blockchain with cloud technology to mitigate data security, privacy, availability, and resource utilization.

A. Management Documents

For safe and open document management, the suggested decentralised file storage system can be used, especially in situations where ownership, auditability, and data integrity are important considerations. As examples, consider:

- Legal contracts and agreements: Ensuring ownership records and transfer histories are clear and unchangeable, as well as storing and transferring legal documents in a way that makes them auditable and impenetrable.
- Intellectual property management: Keeping track of and overseeing ownership of assets like trademarks, patents, and copyrights in a secure manner while allowing for easy transferability of ownership.
- Corporate records and compliance: To lower the risk of tampering or unauthorised access, maintain and manage corporate records, financial statements, and compliance papers in a decentralised and auditable manner.

B. Data Backup and Archiving

- Redundant storage: IPFS distributes file storage over a number of network nodes. Thanks of this redundancy, your data is always available and accessible from the remaining nodes, even in the event that some nodes go offline, offering reliable backup.
- Immutable data: A file's content cannot be altered or removed once it has been added to IPFS. Because of its immutability, IPFS is a good choice for archiving data that must not change over time, such scientific, legal, or historical records.
- Versioning: Versioning is supported by IPFS and lets you monitor file changes over time. This functionality comes in very handy when preserving documents or datasets that change over time and need to keep a history of modifications.
- Permanent persistence: IPFS aims to provide persistent data storage by encouraging nodes to continuously host

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 5, May 2024

popular content. This persistence makes it suitable for long-term archiving of valuable information.

V. FUTURE WORK AND CHALLENGES

Blockchain data storage is expensive and not suitable for storing large amounts of data. This is a key barrier to the large amount of user data generated by DApp, and the risk of data loss, tampering and so on if the data is stored in a centralized service organization [2]. Although the suggested method provides improved security and transparency, the following possible privacy issues need to be addressed:

- Metadata visibility: Since the file metadata, including ownership records and CIDs, is kept on the public Ethereum blockchain, any sensitive information in the metadata may give rise to privacy problems.
- File content visibility: The contents of the files are dispersed throughout the IPFS network and may be viewed by any network node.

The following techniques can be used by the system to lessen these privacy concerns:

- Encryption: To guarantee that only authorised users with the decryption key may access the contents of a file, files can be encrypted prior to uploading to IPFS.
- Privacy-preserving approaches: To enable privacy-preserving activities on the blockchain, such as confirming file ownership without disclosing the file information or content, techniques like zero-knowledge proofs or secure multi-party computation could be investigated.
- Access control rules: Based on predetermined policies or permissions, the smart contract may apply extra access control rules to limit access to file metadata or content.

VI. CONCLUSION

The proposed system not only challenges the existing centralized models but also empowers individuals and companies to take control of their assets. Secure and auditable ownership management and transfers inside decentralised storage solutions are made possible in large part by blockchain technology. Blockchains' transparency and immutability make ownership records tamper-proof and verifiable, which promotes systemic trust. Blockchain-based smart contracts automate and uphold ownership transfer regulations, doing away with the need for middlemen and encouraging peer-to-peer exchanges.

People and organisations may take back control of their data by adopting decentralised storage solutions based on blockchain technology. This will lessen their dependency on centralised service providers and reduce the dangers related to censorship, loss of access, and data breaches. This paradigm change has the ability to upend established data storage practices and provide users more control, transparency, and resilience over their priceless digital assets.

REFERENCES

[1] Q. Xu, Z. Song, R. S. Mong Goh and Y. Li, "Building an Ethereum and IPFS-Based Decentralized Social Network System," 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), Singapore, 2018, pp. 1-6, doi: 10.1109/PADSW.2018.8645058.

[2] Tang, X., Guo, H., Li, H., Yuan, Y., Wang, J., & Cheng, J. (2021, January 1). A DAPP Business Data Storage Model Based on Blockchain and IPFS. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-030-78612-0_18

[3] A. ul Haque, M. S. Ghani, and T. Mahmood, “Decentralized transfer learning using blockchain & IPFS for deep learning,” in 2020 International Conference on Information Networking (ICOIN), pp. 170–177,IEEE, 2020.

[4] M. Shah, M. Shaikh, V. Mishra, and G. Tuscano, “Decentralized cloud storage using blockchain,” in 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), pp. 384–389,IEEE, 2020.

[5] Ethereum Whitepaper | ethereum.org. (n.d.). ethereum.org. https://ethereum.org/en/whitepaper/