International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)

ISSN: 2456-236X

Vol. 10 Issue 01 | 2025

Decentralised file storage system using block chain technology

Ms. Prajakta Narnaware[1]

[1] *Assistant Professor, CSE Department, Padm. Dr. V. B. Kolte College of Engineering, Malakpur*

DOI: 10.5281/zenodo.15751585

ABSTRACT

Centralized data storage systems have serious flaws: they're vulnerable to single points of failure, data breaches, and can't easily grow. Decentralized file storage systems built on blockchain technology offer a strong solution. This paper details a decentralized storage system that prioritizes data integrity, privacy, and accessibility without relying on a central authority. Blockchain provides a distributed ledger, securing and authenticating data transactions for transparency and immutability. Enhanced security and protection from unauthorized access are achieved through cryptographic hashing, data encryption, and consensus algorithms. Data files are fragmented and distributed across a network of nodes, with blockchain storing metadata and pointers for efficient retrieval. This decentralized setup improves fault tolerance and allows for cost-efficient scalability by using untapped storage across the network. Incentive mechanisms, like token-based rewards, encourage node participation, making the system even more reliable. This technology has many applications, including secure data sharing, archival storage, and managing sensitive information in industries such as healthcare, finance, and the Internet of Things (IoT). Our proposed system demonstrates that blockchain-based storage can overcome the weaknesses of traditional methods, offering a secure, scalable, and transparent platform for data storage and management.

Keywords: Ledger, Decentralized, Blockchain, Metadata, Platform

1. Introduction

With growing concerns over privacy, security, and user control, data sovereignty has become a critical issue for digital platforms. Traditionally, this meant data was governed by the laws of the country where it was collected or processed, often leading to data localization policies that required data to stay within specific geographic borders. However, as digital interactions increasingly cross national boundaries, this model's limitations are clear. Both individuals and organizations are now seeking more autonomy over their data. Similarly, centralized storage systems, where third parties manage data, are increasingly criticized for their vulnerability to breaches, censorship, and unauthorized access. These weaknesses highlight the urgent need for solutions that empower users to control their own data, giving rise to the concept of Data Self-Sovereignty (DSS).

Understanding Data Self-Sovereignty (DSS): DSS expands on traditional data sovereignty by giving individuals and organizations complete control over their data, regardless of where it's stored or processed. It emphasizes user autonomy, allowing individuals to decide how their data is stored, accessed, and shared without relying on central authorities or intermediaries. This aligns with the broader shift towards decentralized digital infrastructures, where trust is distributed across participants rather than concentrated in one entity.

The Role of Decentralized Storage and Blockchain: Decentralized storage systems, often powered by blockchain technology, are crucial for advancing DSS. Blockchain's decentralized nature, combined with its transparency, immutability, and cryptographic security, ensures data remains secure, tamper-resistant, and firmly under user control. A key blockchain feature supporting DSS is smart contracts. These are self-executing agreements embedded directly into the blockchain, automating and enforcing rules for data access, usage, and sharing without needing intermediaries. This ensures that ownership and control stay with the data's rightful owners. Unlike traditional centralized solutions, decentralized storage systems distribute data across multiple network nodes. This approach significantly reduces reliance on a single point of failure, enhancing overall security, privacy, reliability, and user autonomy.

2. Project Objectives:

- To ensure that data is securely stored and protected from unauthorized access and breaches.
- To Provide immutability through blockchain, ensuring that stored data cannot be tampered with or altered without detection.
- To Empower users with full control over their data, allowing them to decide how it is stored, accessed, and shared without reliance on centralized entities.
- To eliminate single points of failure and ensure consistent availability even if some nodes go offline.

- To safeguard sensitive information while preventing unauthorized entities from accessing the data.

3. Literature Review

Csirmaz et. al. states that synchronizing diverged copies of some data stored on a variety of devices and/or at different locations is an ubiquitous task. The last two decades saw a proliferation of practical and theoretical works addressing this problem. File synchronization is a feature usually included with backup software in order to make is easier to manage and recover data as and when required. File synchronization usually delivered through cloud services. Dedicated file synchronizing solutions frequently come with additional tools not just for managing the saved data, but also to allow for file sharing and collaboration with stored files and documents. These cloud storage services are easily accessible for the end-user because the service front-ends are very well integrated into web clients as well as desktop and mobile environments. Simple user interfaces hide the complex and sophisticated service back-ends. Collaboration services are frequently integrated into the “cloud storage” environment. For example, Google Docs is an application layer integrated into Google Drive storage, Office 365 is integrated with One Drive storage and Dropbox Paper service is an extension of [1]

Dürsch et. al. states that an application that enable digitally conducting QDA are grouped as Computer assisted qualitative data analysis software. One representative of Computer assisted qualitative data analysis software (CAQDAS) is called QDAcity1. QDAcity is a cloud-based web application, developed and operated by the Professorship for Open Source Software at the Friedrich-Alexander-Universität Erlangen- Nürnberg. QDAcity provides an environment for multiple analysts or researchers to collaboratively conduct QDA. Since QDA deals with big amounts of fuzzy and subjective data, enabling researchers to share and discuss different interpretations, ideas, and conclusions can be very beneficial for the process of QDA. The approach of enhancing a process by promoting close collaboration and "shrinking the feedback loop" can also be found in other fields. Agile approaches of software development like Extreme Programming (Beck, 2000) serve as examples of this. However, currently QDAcity only allows the simultaneous collaboration of multiple researchers in a shared project, but not on a more granular level in a shared document. Real-time collaborative editing of a shared document is a classic form of digitally enabled, close collaboration. [2]

Martins et. al. states that in recent years the cloud has become ubiquitous. Many apps and services with users spread across the world resort to these solutions. Cloud applications with global scale user base like social media tend to resort to distributed databases that prioritize lower latency over strong consistency. Such solutions don’t require coordination which would require reads and writes to contact a majority of replicas in a communication process that can cross continents, penalizing performance. This kind of of applications along with the database replicas usually run in multiple datacenters. Instances are usually geo-replicated to accommodate users from different parts of the globe with fast response time. When the amount of replicas grows it also becomes important to partition data in a way that doesn’t break the fault tolerance guarantees of replication, since having every piece of data in every replica of the database might not be necessary and can definitely become very expensive. In such a setting it is not enough to have database replicas close to the users. The coordination between replicas performing reads and writes also needs to be minimized in order to achieve the desired low latency. Consider that you have a local server close to a client. In order for a client database operation to complete, it would need to contact a majority of database instances. This would completely break the desired low latency. For this reason weak consistency models have been rising in popularity recently. [3]

4. Motivation

Centralized data storage systems suffer from critical weaknesses: they're susceptible to single points of failure, data breaches, censorship, and high operating expenses. Decentralized storage offers a robust, secure alternative to overcome these issues. Given rising global concerns about data privacy, ownership, and regulatory compliance, users need greater control over their data. Decentralized storage inherently supports data sovereignty by empowering user autonomy. Blockchain technology, with its immutability, transparency, and cryptographic security, can revolutionize data management when integrated into file storage systems. This integration ensures data integrity and trust without needing intermediaries. As cyber threats grow more sophisticated, highly secure and private storage systems are essential. Decentralized systems, by distributing data across multiple nodes, significantly increase network resilience against attacks. The advancements in peer-to-peer networks and blockchain provide a strong foundation for developing these decentralized storage solutions, and further research in this area can lead to groundbreaking innovations.

5. Proposed Methodology:

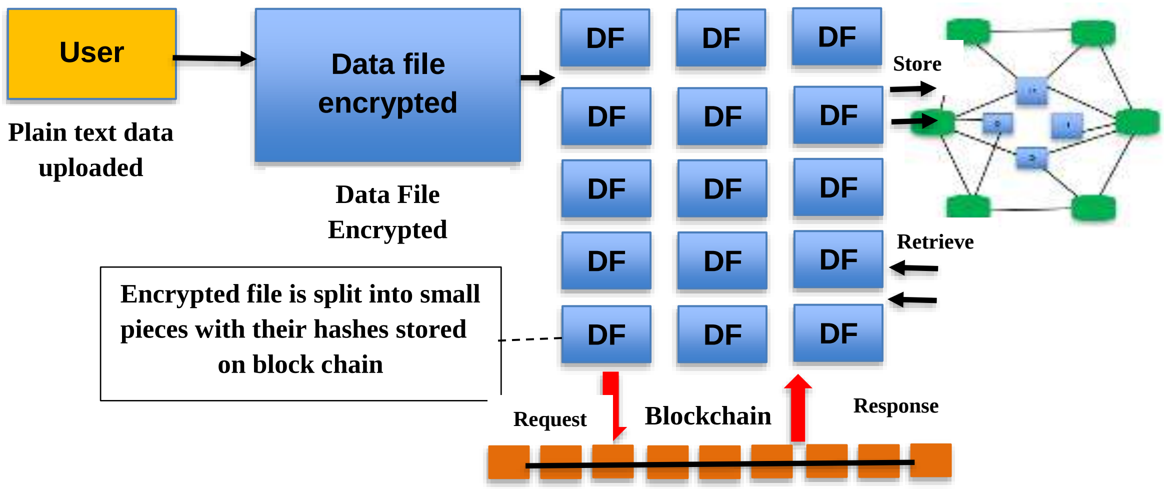

Fig. 1 Architecture of Block Chain

Decentralized storage systems function on a peer-to-peer (P2P) network, where participants contribute their unused storage space. In return, they receive incentives such as tokens or cryptocurrencies. Blockchain technology is crucial to this model, as it facilitates the creation and management of these digital rewards for network contributors. This incentivized approach fosters active participation, which in turn ensures the storage ecosystem remains sustainable, scalable, and efficient. By aligning the interests of participants with the overall health of the system, decentralized storage maintains consistent data availability, reliability, and security.

A typical blockchain-based decentralized storage system functioning on a P2P network involves four key steps:

1. Data Uploading: Users upload plaintext files to the decentralized storage system.
2. Data Encryption: Uploaded files are encrypted using advanced cryptographic algorithms. This process converts plaintext into ciphertext, ensuring data privacy and confidentiality. Only users with the correct decryption keys can access the original data.
3. Data Fragmentation (Sharding): Encrypted data files are divided into smaller fragments, or shards. This process enhances system scalability and performance by enabling secure distribution of data fragments. Sharding also improves retrieval speed, as individual fragments can be accessed independently.
4. Data Chunk Distribution: Encrypted fragments are distributed across multiple nodes in the P2P network. Each node, contributing storage and participating in data operations, ensures redundancy and high availability. Even if some nodes fail or go offline, the system maintains data integrity and accessibility by storing fragments on other nodes.

This architecture leverages blockchain's transparency and cryptographic security to create a resilient, efficient, and user-driven storage environment.

6. Benefits

Blockchain technology offers a wide range of benefits, making it a transformative tool across various industries. Here are the key advantages:

Decentralization

- Eliminates the need for a central authority by distributing data across a peer-to-peer network.
- Reduces the risk of centralized points of failure, ensuring higher system resilience.

Transparency

- Transactions are recorded on a public ledger that is visible to all participants in the network.
- Enhances trust as all parties can verify transactions independently.

Immutability

- Data once recorded on the blockchain cannot be altered or deleted, ensuring data integrity.
- Provides a tamper-proof record of transactions, reducing fraud and errors.

Security

- Utilizes advanced cryptographic techniques to secure data and transactions.
- Ensures that only authorized parties can access or modify information through encryption and private keys.

Cost Efficiency

- Reduces the need for intermediaries or third-party verifications, lowering transaction costs.
- Automates processes through smart contracts, saving time and resources.

Traceability

- Tracks the provenance of goods or transactions, making it ideal for supply chain management.
- Provides an auditable trail of records, enhancing accountability and reducing counterfeit risks.

Increased Speed

- Processes transactions faster compared to traditional methods, especially cross-border payments.
- Eliminates bottlenecks caused by manual verifications and intermediaries.

Tokenization

- Allows physical and digital assets to be represented as tokens on the blockchain.
- Simplifies asset management and opens up new opportunities for fractional ownership.

Interoperability

- Supports seamless integration with other technologies and platforms.
- Facilitates the creation of multi-system networks for improved collaboration.

7. Conclusion:

This paper examines decentralized storage systems, focusing on their features, performance, and how they contribute to sustainable data self-sovereignty (DSS). Decentralized solutions, especially those using blockchain technology, are promising for implementing DSS by giving users greater control, privacy, and security over their data. The study highlights that these systems must align with user needs for optimal platform selection. Key findings show that while blockchain-based storage excels in security and data integrity, they vary significantly in complexity, cost-efficiency, and overall performance. Therefore, users must carefully evaluate these aspects to choose the best decentralized storage option for their specific requirements. Ultimately, these systems are vital for enabling self-sovereign data management by providing secure, resilient, and user-centric solutions. With increasing global emphasis on data ownership, privacy, and security, decentralized storage platforms are becoming crucial, offering a strong foundation for achieving data sovereignty in the digital age. This study serves as a valuable resource for users, developers, and researchers, helping them make informed decisions when selecting and deploying decentralized storage systems that best meet their sovereignty and operational needs.

References:

[1] Elod P. Csirmaz and Laszlo Csirmaz, “Synchronizing Many Filesystems in Near Linear Time”, arXiv:2302.09666v2 [cs.IT] 17 May 2023

[2] Martin Dürsch, “Scaling Real-time Collaborative Editing in a Cloud-based Web App”, Erlangen, 19 April 2023

[3] João Gonçalves Martins, “Query Processing in Cloud Databases with Partial Replication” NOVA University Lisbon March, 2023

[4] Masoumeh Hajvali, Sahar Adabi, Ali Rezaee and Mehdi Hosseinzadeh, “Decentralized and scalable hybrid scheduling-clustering method for real-time applications in volatile and dynamic Fog-Cloud Environments” (2023) 12:66, https://doi.org/10.1186/s13677-023-00428-4, Journal of Cloud Computing: Advances, Systems and Applications

[5] Elod P. Csirmaz 1,_ and Laszlo Csirmaz, “Data Synchronization: A Complete Theoretical Solution for Filesystems” Future Internet 2022, 14, 344. https://doi.org/10.3390/fi14110344

[6] Novak Boˇskov, Ari Trachtenberg, and David Starobinski. Enabling costbenefit analysis of data sync protocols, 2023.

[7] Elod P. Csirmaz and Laszlo Csirmaz. Data synchronization: A complete theoretical solution for filesystems. *Future Internet*, 14(11), 2022.

[8] Elod Pal Csirmaz. Algebraic file synchronization: Adequacy and completeness. *CoRR*, abs/1601.01736, 2016.

[9] John Day-Richter. What’s different about the new Google Docs: Making collaboration fast, 2010.

[10] Shimon Even. *Graph Algorithms*. Cambridge University Press, USA, 2nd edition, 2011.

[11] JiuLing Feng, XiuQuan Qiao, and Yong Li. The research of synchronization and consistency of data in mobile environment. In *2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, volume 02, pages 869–874, 2012.

[12] Rusty Klophaus. Riak core: Building distributed applications without shared state. In *ACM SIGPLAN Commercial Users of Functional Programming*, CUFP ’10, New York, NY, USA, 2010. Association for Computing Machinery.

[13] Zhenhua Li, Christo Wilson, Zhefu Jiang, Yao Liu, Ben Y. Zhao, Cheng Jin, Zhi-Li Zhang, and Yafei Dai. Efficient batched synchronization in dropbox-like cloud storage services. In David Eyers and Karsten Schwan, editors, *Middleware 2013*, pages 307–327, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.