IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

DESIGNING SOFTWARE TOOL SET FOR DATA STORAGE IN DECENTRALIZED STORAGE SYSTEM

Prof. Prashant Mandale, Om Gosavi, Ritu Mahajan, Nikita Rathod, Shreyash Sharma
Department of IT,
International Institute of Information Technology, Pune, India.

Abstract : The goal of the project is to Decentralize data by creating a storage system that addresses issues with bandwidth throughput, data accessibility, security, and other factors. 3 entities will be in the model. Client, Guardian server, and Storage Nodes come first. In our scenario, the Guardian server has a significant impact. It deals with Network verification and analysis, Authentication and Authorization, key generation (Cryptography), and Storage information. Because storage nodes will run on commodity hardware, they will only be concerned with data routing, storage, and replication, giving clients the least amount of system management possible for security reasons. We upload our meta information to the Guardian server. The suggested approach implements access roles that the owner has written in smart contracts to ensure security and access control. Prior to submitting the requested amount as a gas fee for digital content, users must first authenticate themselves using our custom authentication system based on TCP connections. The user is invited to submit reviews of the data after its successful delivery. A study of gas consumption and actual cost necessary in terms of USD is conducted in the simulation results in order to appropriately determine the cost of executing the implemented scenario in a real-world context.

IndexTerms - **Authentication and Authorization, Block-chain, Ethereum, Fernet symmetric encryption, Meta-mask.**

I. INTRODUCTION

The internet is the most convenient means of sharing info in the modern world. Internet-based technology known as "cloud computing" is reliant on sizable storage providers. These storage companies manage vast amounts of data kept in the cloud in their capacity as unreliable third parties. This data can include private information that is the property of numerous people or businesses. Such models might be subject to security problems like privacy and integrity. In this project, we provide a working prototype of a multi-user system for data access control that secures shared data storage using blockchain technology.The owner of the data is permitted to upload the data via our web portal.The owner will designate to whom permission should be given. Only the user who is authorized to view a certain piece of server-stored data may do so. The aforementioned data operation will be documented on the Guardian server. Owner always has access to the logs to look for any unusual data operations. Consequently, using smart contracts, immutability, and the blockchain ledger, data privacy is guaranteed. We're going to put forward a decentralized storage system based on blockchain technology that performs data integrity at the Guardian server for the user and grants access to data stored at peer-to-peer servers after validating the user's identity. An overwhelming amount of data is kept on cloud servers, which are centralized

Distributed ledgers can be separately stored on various network nodes using a technique called "Decentralized storage." The processing and storage capacity limit of network nodes is the issue. There can never be a single point of failure. Although it differs from web3, it is comparable. It manages content addresses and functions in a way that is comparable to bit torrent. Data availability is ensured through Decentralised platform storage. On the other hand, when a user pays in ethers to view the content, the accuracy and integrity of the data cannot be compromised. There is just one suggestion for a remedy to the many problems that are present. In the hypothetical situation, the owner is in charge of sharing.

II. LITURATURE SURVEY

Dennis Trautwein[1] Digital Content Protection: A technique that ensures the authenticity and non-repudiation of digital data has been established in order to protect the confidentiality of traceable encryption in blockchain. The authors' method tackles the problem that the user's secret key does not contain the user's particular information when it is shared with other entities. It is

challenging to determine the origin of the secret key in the event that the shared key is corrupted or misused. Moreover, a bottleneck for current systems is the leakage of private data in access control.

A. Cloud Blockchain Because of its remarkable qualities, experts are working to implement blockchain in a wider area. A blockchain offers a safe, dependable, and trustworthy platform for data sharing. Nevertheless, because it is spread, networks' capacity for control is diminished. The size of data is significantly growing each day in the present digital era. Users' terminal storage is now a problem as a result. Using cloud services with these storage-constrained nodes could be the answer to this problem. Decentralized storage systems, though, like a file.
B. Access Control Using blockchain storage effectively is important, especially when network nodes need to maintain a lot of data. While they cannot store very large data sets, terminal nodes have a limited storage capacity. This issue raises a number of problems, including computational capacity and the high expense of handling enormous amounts of data. A decentralized storage solution employing IPFS and access control mechanisms was suggested taking into account the aforementioned difficulties. On each node, files are kept as chunks of data. But, unless the right rights are granted to users, a file cannot be accessed.

Dff Z. Ullah, B. Raza[2] a system for general-purpose Unix local area network (LAN) machines to monitor resources peer-to-peer. Despite being created to monitor system resources on a single LAN, PeerMon can be set up on numerous LANs provided resource sharing between LANs is permitted. The PeerMon data is used by three programmes, SmarterSSH, autoMPIgen, and a dynamic DNS binding system, to choose "excellent" nodes for task or process placement in a LAN. While deciding which machine to ssh into, SmarterSSH makes use of information gathered by the peer monitor process. As of right now, it is feasible to choose the "best" machines based on combinations of CPU load, memory load, and CPU core count. The second application, autoMPIgen, uses PeerMon data to generate MPI host files automatically based on the system's resource availability and consumption. Dynamic DNS binding, which takes into account the system's resource usage, is the third technology.PeerMon is a minimally intrusive solution that rapidly delivers precise information on the general purpose LAN systems' system-wide resource utilization. Its peer-to-peer architecture is fault resilient and grows effectively to big systems.

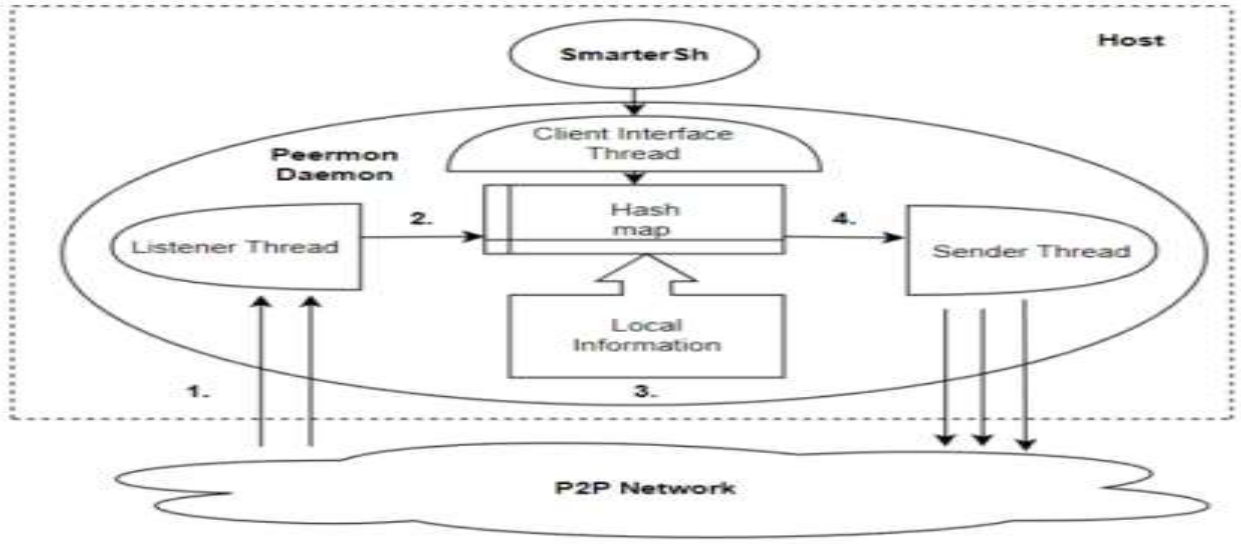


Fig 1: (Peer-to-Peer) Network Monitoring System

Athanere, Smita, and Ramesh Thakur[3] A distributed hash table is used to track files in the peer-to-peer Interplanetary File System (IPFS; Benet, 2014), a group of subprotocols (DHT).

a) System initialization and setting of system parameters fall under the purview of the administrator.
b) Domain Authority (DA): DA is responsible for giving IPFS users their credentials. Each DA consists of a number of users in a particular domain.

c) CSP: Data owners can utilize a cloud service provider (CSP) as a tool to store their data.
d) Data owner (DO): As CSP cannot be relied upon to submit files securely, DO uses encryption to limit access to data. DO establishes access controls, encrypts the data files, and transmits the key and encrypted data to IPFS separately before sending the data files to the cloud provider.
e) IPFS: The aforementioned organizations are all a part of an IPFS network. For each piece of uploaded data and key ciphertext, IPFS creates a unique hash code. IPFS stores access metadata and public parameters to protect its integrity and immutability. Additionally, it enables enterprises to perform partial trusted computing and gives multiple DA the ability to manage user credentials jointly.

All data users can download cipher text for free from the CSP (DUs). A user must ask for the IPFS's secret information in order to access encrypted data.Users from distinct domains have varying levels of decryption privileges. The successful decryption depends on the authentication scheme stored in the ciphertext matching the customer's credentials. As a result, the decryption engine is connected to permitted users.

Alizadeh, K. Andersson and O. Schelen[4] An instance of a self-certifying filesystem is used by IPFS to maintain data integrity while in transit (SFS). To do this, each data item d is given the address addr(d) = H, which is a distinct immutable address and the content's hash (d). By comparing the hash of the received data with the requested address, recipients can determine whether the data has been tampered with. An addr(d) is encoded as a "content identifier" on the IPFS (CID). A Merkle DAG4 is used to structure directories and files. This topology allows nodes to have more parents than Merkle Trees, and IPFS enables data storage on non-leaf nodes. For instance, a directory on IPFS is encoded as a node carrying both the hashes and metadata for each entry in the directory. After being broken into smaller data bits, big files are encrypted using multi-layered directed acyclic graphs (DAGs). The method also enables caching and deduplication of directory entries and file contents. Around 3.5 GB of compressed BitSwap traces were produced as a result of their ongoing fifteen-month collection of traces. Throughout the first two months of our observations, we kept just one monitoring node operational. Two nodes were used to scan the remaining time. Our monitors experienced a few outages as well as minor configuration and version modifications since the data gathering started. They are categorized as short (>5 s) and major (>1 h) outages, and the days they occurred are counted.After 450 days of operation, the US monitor experienced one or more large outages on 17 of those days and one or more minor outages on an additional 15 of those days. There were 18 days with one or more large outages and an additional 5 days with one or more minor outages during the 385 days the de monitor was in use in Germany. On a single monitor, we monitored about 806 million different CIDs over the course of fifteen months.

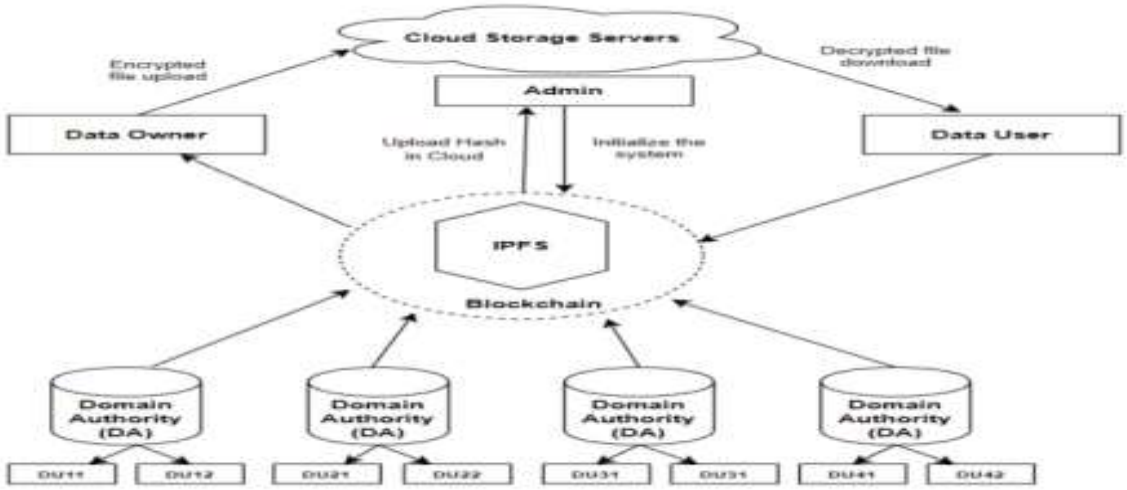


Fig 2: Blockchain based hierarchical semi-decentralized approach using IPFS

III. TECHNOLOGY USED

1. Blockchain Technology

A sophisticated database system called Blockchain Technology allows for open information sharing inside a company network. In a blockchain database, data is kept in blocks that are linked together in a chain. The data remains constant over time since the chain cannot be destroyed or modified without network approval.You may utilize blockchain technology to build a permanent ledger that can be used to track orders, payments, accounts, and other transactions. It is possible to view these transactions collectively thanks to the system's built-in features, which also stop unapproved transaction inputs.

2. Smart Contract

Smart contracts allow businesses to manage contracts on their own, doing away with the requirement for a third party to mediate disputes. These are scripts that are kept in the blockchain system and are launched automatically in response to certain circumstances. To finish transactions with assurance, if-then checks are performed.

3. Ethereum

Anyone who wants to create open-source, decentralized applications for the blockchain can do so using the Ethereum platform. For corporate use cases, Ethereum Enterprise was created.

4. Hardhat Network

For the purpose of creating and testing Ethereum applications, the Hardhat Network is a local Ethereum network. It is a component of the Hardhat development environment, a well-liked tool used by programmers to create decentralized apps (dApps) and smart contracts on the Ethereum blockchain. In addition to being simple to install and operate, Hardhat Network has a number of characteristics that make it ideal for local development and testing. They comprise:

1) Quick and reliable block times are a result of Hardhat Network's deterministic mining algorithm, which produces predictable and consistent block times. This makes testing dApps and smart contracts that depend on specific block times simpler.
2) Control over gas prices and limits is available through Hardhat Network, which is helpful for testing gas-related functionality.
3) Network forking: Hardhat Network enables network forking, which allows developers to build a clone of the Ethereum network at a specified block. This can be helpful for testing dApps and smart contracts in various network setups.
4) Built-in accounts: Hardhat Network includes a selection of accounts that have been pre-funded and may be used to test smart contracts and decentralized applications (dApps) that demand account balances.
5) Integration with Hardhat: The Hardhat development environment, which offers a number of tools and functionalities for creating Ethereum applications, is intimately connected with the Hardhat Network.

5. TCP Socket

Network sockets known as TCP (Transmission Control Protocol) sockets provide error-checked, dependable data transfer between programmes running on various hosts throughout a network. TCP, along with IP, is one of the foundational protocols of the Internet Protocol (IP) family and is in charge of delivering dependable data packet transit across a network.
TCP sockets exchange data in a stream of bytes between two endpoints, referred to as a client and a server, after connecting to each other. The client is the one who first requests a connection with the server, starting the connection. A three-way handshake is carried out to synchronize the sequence numbers needed to identify each data packet if the server accepts the connection.
After the connection is made, reliable and organized data transfer between the two destinations is possible. TCP ensures that data is not corrupted during transmission, that packets are delivered in the proper order, and that missing packets are retransmitted.
TCP sockets are commonly used in client-server applications including file transfers, email, and web browsing that demand dependable data delivery. They are also utilized in numerous other applications, such as online gaming, streaming videos, and database access, that call for a dependable, well-organized stream of data.

6. Fernet Symmetric Encryption

A symmetric encryption technique called Fernet is created to offer a quick and safe way to encrypt data. It is a component of Python's cryptography library, which offers applications cryptographic services.
Data is encrypted and decoded using a single key in Fernet encryption. This indicates that the encryption and decryption processes use the same key. Because the same key is used for both processes, these types of encryption are known as symmetric encryption.
The Fernet algorithm combines symmetric encryption and message authentication to safeguard the data's confidentiality.The plaintext data is initially encrypted using a key that is created during the encryption procedure. Moreover, a special message authentication code (MAC) is generated, which guarantees the data's integrity by confirming that the data wasn't altered during transmission or storage.

7. GUI Technology

Three of the main technologies used to create websites and online apps are HTML, CSS, and JavaScript. Each language has a particular function in the creation of websites:
The industry standard markup language, HTML (HyperText Markup Language), is used to construct the structure, design, and content of web pages. HTML is used to create text, pictures, and other media elements as well as to define the structure of a web page. It makes use of tags to specify things like headings, paragraphs, and links.
Cascading Style Sheets, or CSS, are used to specify how web pages are presented, including their layout, colors, and fonts. It is simpler to update and maintain a website's look and feel thanks to CSS, which enables developers to isolate a web page's presentation from its content.
JavaScript is a scripting language that is used to build dynamic and interactive web sites. JavaScript enables web designers to include features like form validation, animations, and user interaction. Moreover, it may be used to build web apps and communicate with web APIs.
Modern web development is built on the basis of HTML, CSS, and JavaScript. To build dynamic and responsive online applications, they are often used in conjunction with web development frameworks and tools.

8. MetaMask

Interacting with the Ethereum network requires the usage of the MetaMask cryptocurrency wallet. Users can access their Ethereum wallet through a browser extension or mobile app to interact with decentralized applications. [1] [2] By focusing on Ethereum-based infrastructure and tools, ConsenSys Software Inc., a blockchain software company, developed MetaMask.

IV. RESEARCH METHODOLOGY

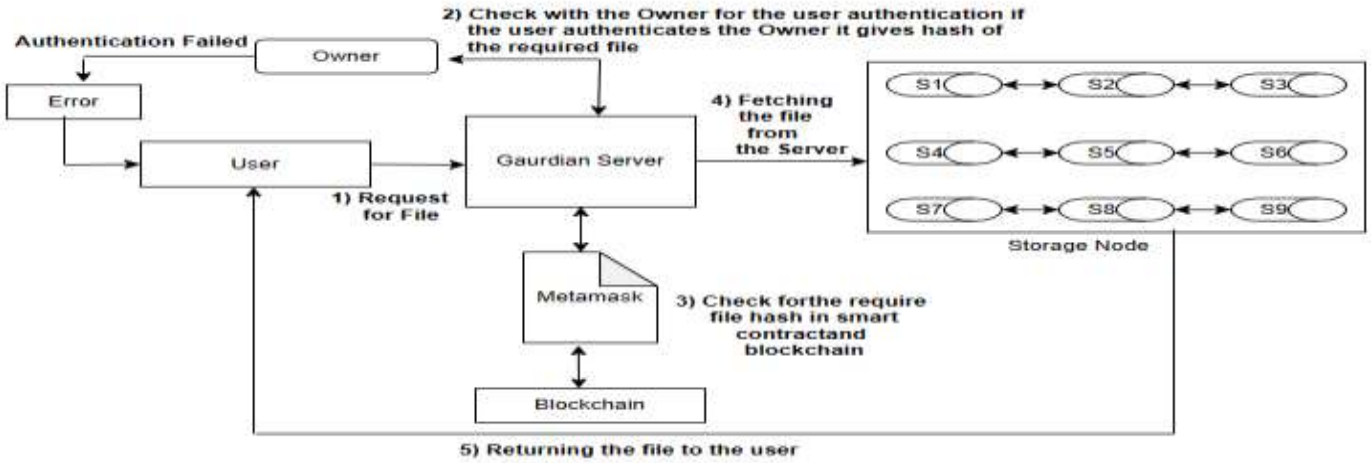


Fig 3 Data Flow Diagram

The goal of this project is to decentralize data by creating a storage system that addresses issues with bandwidth, throughput, data availability, security, and other factors. The project's goal is to provide a decentralized storage system in which all data transactions are recorded in a blockchain and utilize cryptography. The market is dominated by centralized systems, but as data grows, most businesses are switching to decentralized systems because old technologies are inefficient for managing data. The permanent record of agreements between data owners and users is maintained via a blockchain. A user contracts with the owner and requests access to the data by There are regulations that apply to the blockchain. The suggested method makes use of blockchain to give research data transparency, traceability, and fine-grained granularity. An exclusive Ethereum address has been given to each entity (owner and user), which may be used to conduct transactions on the Hardhat network. A decentralized platform to keep the author's files is not thought upon, yet.

Steps:

- The user is requesting the file from the guardian server.
- In order to verify the user's authentication, Guardian Server will speak with the owner.
- The user will see the ERROR message if they have not been authorized.
- If the user successfully authenticates, the file's hash will be returned to him.
- The file's hash will be verified on the servers utilizing blockchain technology and smart contracts.
- The user can verify the data's immutability with the aid of smart contracts.
- The addresses of the data from the server will be stored on the blockchain.
- After obtaining the necessary file, the programme will.
- By checking the file's hash on the server, users can verify the file's integrity.

V. RELATED WORK

The three entities are required for this particular custom authentication system based on Kerberos to successfully finish the authentication process. The guardian server will assign a secret key to each client and node in our system, which will enable us to authenticate connections and encrypt sensitive data. The client can now immediately connect to the storage nodes and upload/download data after the verification is complete.

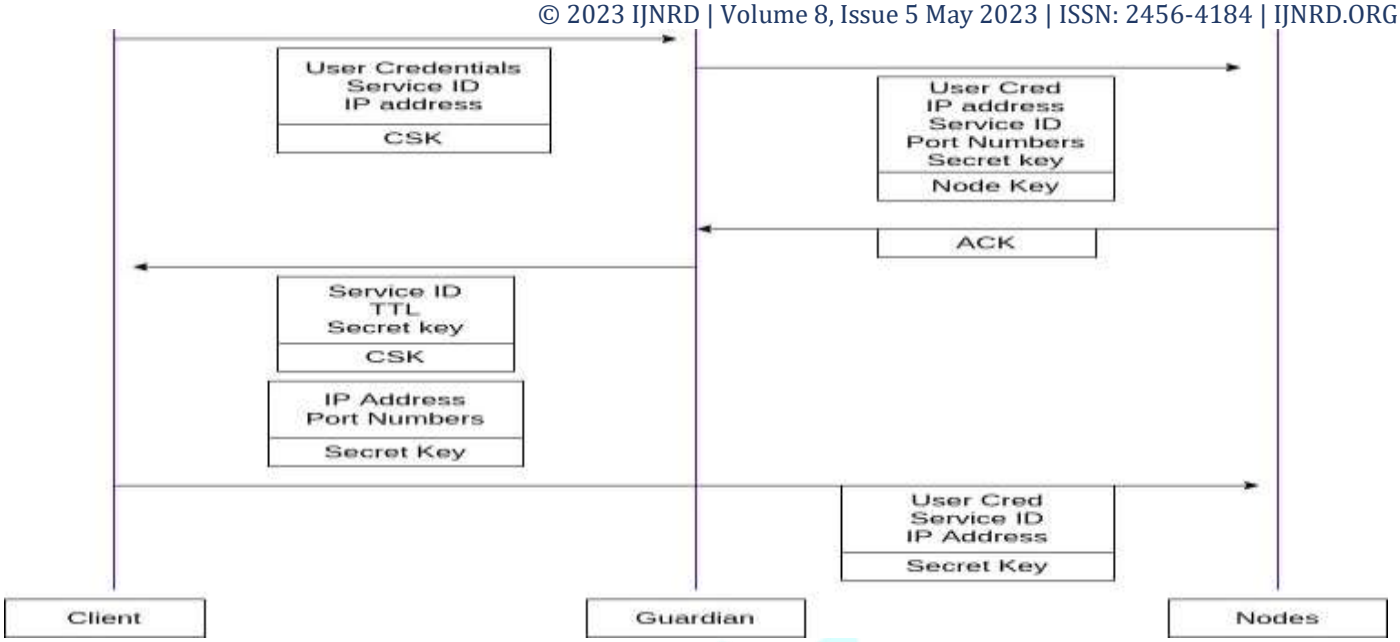
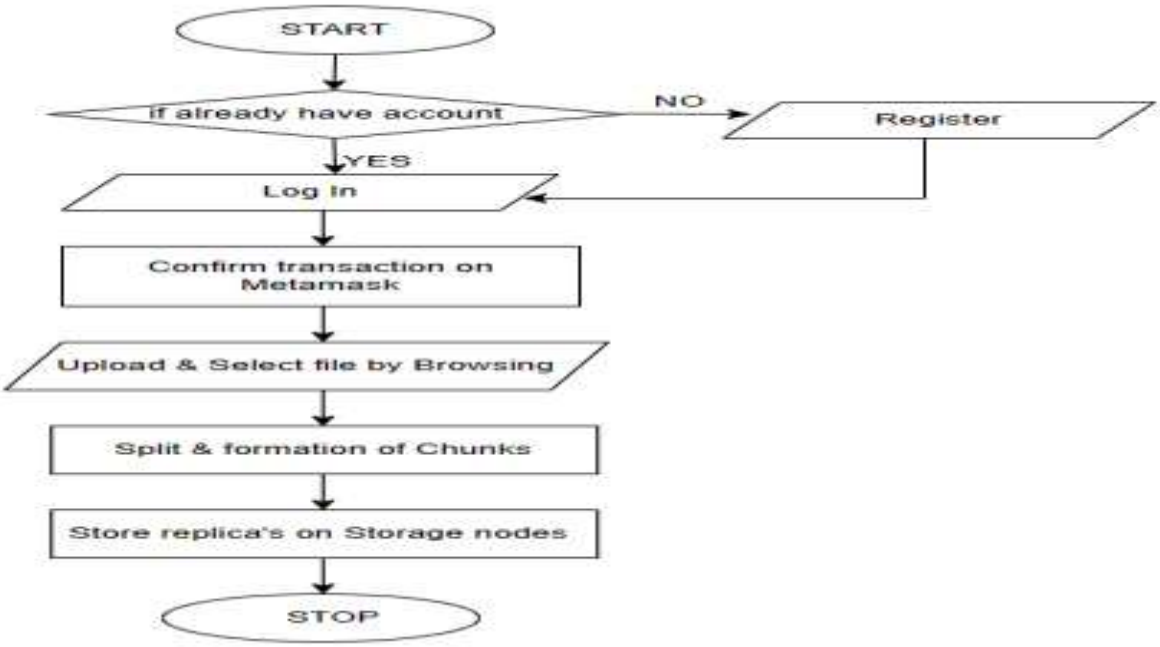


Fig 4:Activity Diagram

VI. Flowcharts:



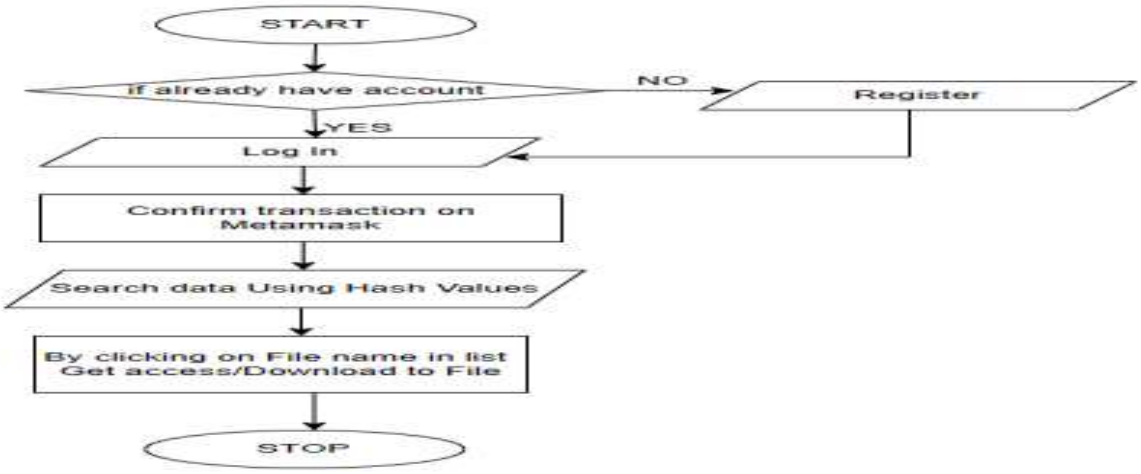


Fig 5: Upload data on Server

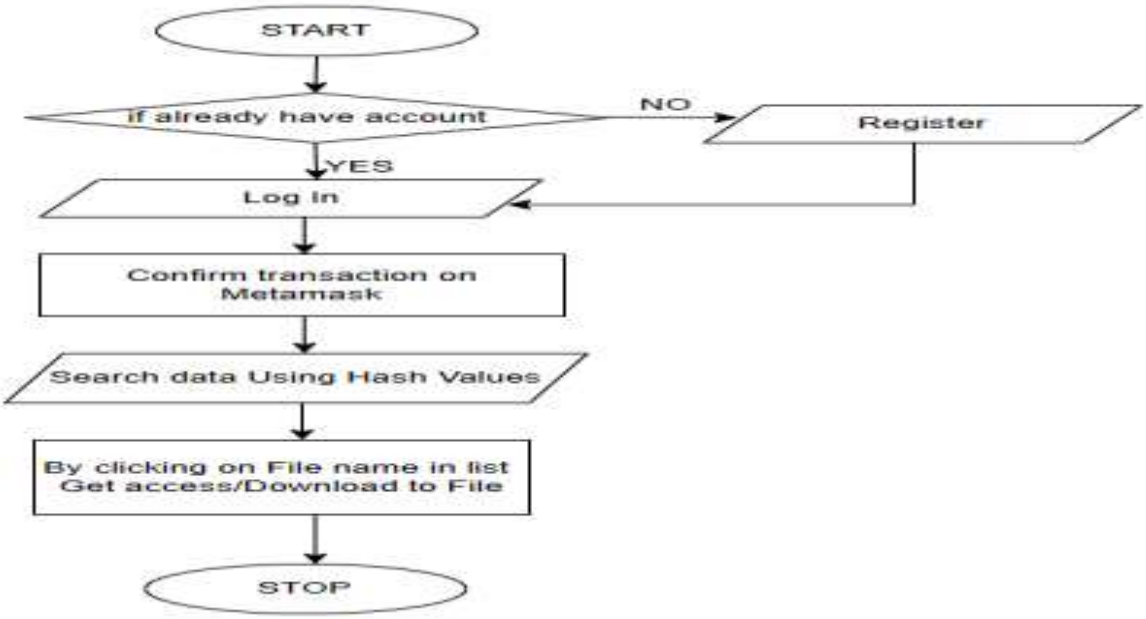


Fig 6: Get User's own uploaded Data

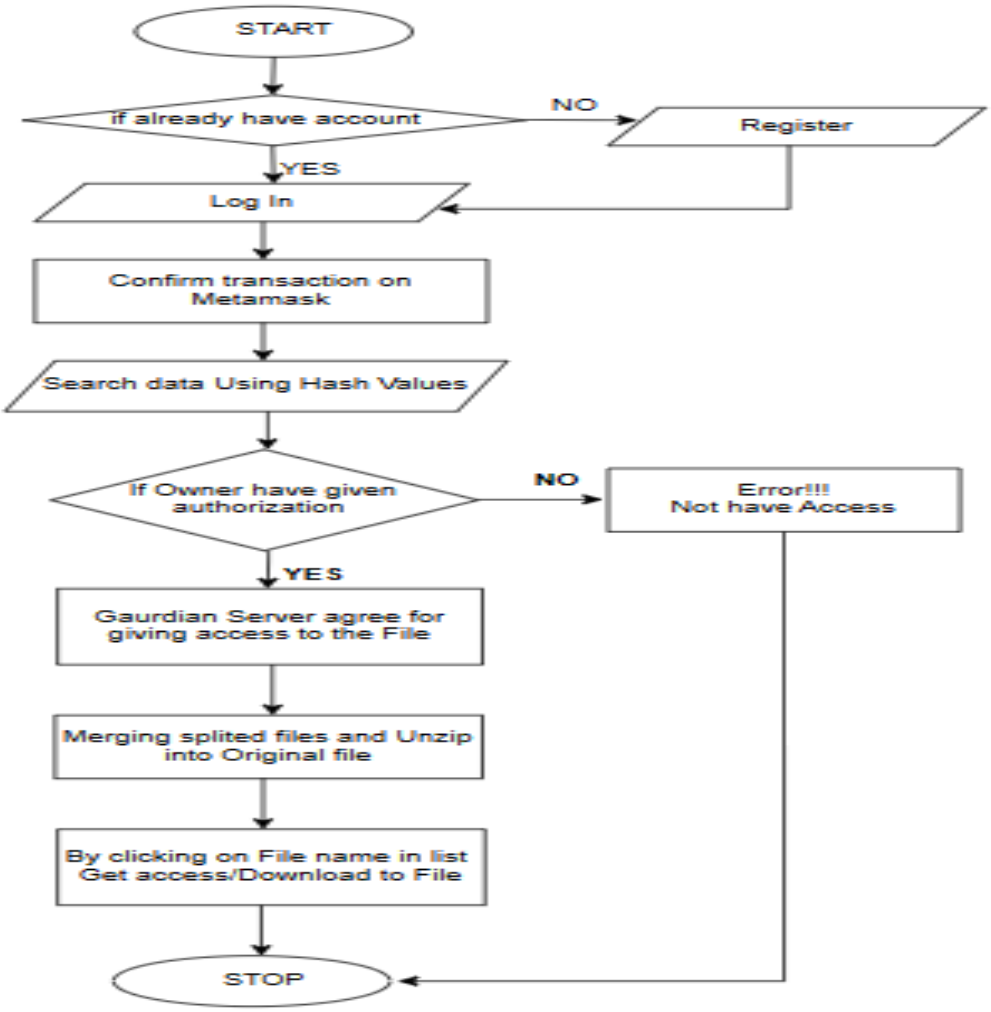


Fig7: Data access Flow

VII. Comparison Table:

	Research Papers				
Features	**A Secure Data Sharing Platform Using Blockchain and Interplanatay File System**	**PeerMon: Peer-to-Peer Network Monitoring System**	**Blockchain based hierarchical semi-decentralized approach using IPFS for secure and efficient data sharing**	**Monitoring Data Requests in Decentralized Data Storage System**	**Designing a Software tool set for Storing Data in Decentralized Storage System**
Access Control Scheme	Yes	Yes	Yes	Yes	Yes
IPFS	Yes	No	Yes	No	No
Overhead of Processing on the user side	No	Yes	No	No	Yes
Storage Workload division	No	Yes	No	Yes	Yes

VIII. Conclusion

Blockchain is a revolutionary technology with decentralized and peer-to-peer features that can revolutionize a variety of businesses. One of the most important and contentious issues of our time is how to access and store data in the cloud. Traditional storage systems have a number of drawbacks and problems that can be solved by blockchain-based storage solutions. In this survey, a novel data storage strategy is used with the goal of enhancing security and privacy.The fundamental objective of the suggested scenario is to provide clients with reliable and high-quality information while also providing owners with a strong financial foundation. The solution to the owner's end bloating issue is decentralized storage. Data hashes sent by the Guardian server are encrypted with Fernet Symmetric Encryption to prevent access by users who have not purchased digital content. The owner is protected from any hash leakage to unauthorized clients by doing this. By adopting a review-based system where users can leave comments and ratings on the data, data authenticity is maintained. In this approach, prospective clients can access the data's quality and save money.

IX. Future Scope

Decentralized architecture is the foundation of the cloud storage of the future, offering Web3 performance, privacy, resilience, and sustainability for less money than you currently spend. Decentralized storage system utilized in decentralized apps that run on the blockchain as opposed to a server platform owned or under the control of one central authority. transitioning to decentralized storage for the following reasons:

- Data Breaches: No organization wants the feared instances of data breaches and angry customers. Thus the required safety measures are performed.
- Data Outages: DDoS assaults may prevent you and your clients from accessing their data. This could halt all services, including those that are essential.
- Skyrocketing Storage Costs: The necessity for data security, rising transmission costs, and evolving hacker techniques are the main causes for concern.
- Censorship and Monitoring: Many people have been lenient towards censorship and surveillance. Decentralized storage will become the standard, and by 2023, the market is expected to be worth more than $101 billion. Decentralized storage will effectively address all of the aforementioned issues in the future so that customers receive better service.

X. References

[1] Dennis Trautwein, Aravindh Raman, Gareth Tyson, Ignacio Castro, Will Scott, Moritz Schubotz,Bela Gipp, and Yiannis Psaras. 2022. Design and Evaluation of IPFS: A Storage Layer for the Decentralized Web. In ACM SIGCOMM 2022 Conference (SIGCOMM '22),August 22–26, 2022, Amsterdam, Netherlands. ACM, New York, NY, USA. https://doi.org/10.1145/ 3544216.3544232.

[2] TiaNewhall, JanisLıbeks, and JeffKnerr RossGreenwood. "PeerMon: A Peer-to-Peer Network Monitoring System."*Proceedings of LISA'10: 24th Large Installation System Administration Conference*. 2010.

[3] Athanere, Smita, and Ramesh Thakur. "Blockchain based hierarchical semi-decentralized approach using IPFS for secure and efficient data sharing. "Journal of King Saud University-Computer and Information Sciences 34.4 (2022): 1523-1534.:Blockchain based hierarchical semi-decentralized approach using IPFS for secure and efficient data sharing.

[4] Alizadeh, K. Andersson and O. Schelen, "Efficient Decentralized Data Storage BasedonPublic Blockchain and IPFS," 2020 IEEE Asia-Pacific Conference on Computer Science Data Engineering (CSDE),2020, pp. 1-8, doi: 10.1109/CSDE50874.2020.9411599

[5] Shrestha, A.K.; Vassileva, J. Blockchain-Based Research Data Sharing Framework for Incentivizing the Data Owners.

Blockchain; Springer: Cham, Switzerland, 2018; pp. 259–266

[6] S. Wang, Y. Zhang and Y. Zhang, "A Blockchain-Based Framework for Data Sharing With Fine-Grained Access Control in Decentralized Storage Systems," in IEEE Access, vol. 6, pp. 38437-38450, 2018, doi: 10.1109/ACCESS.2018.2851611.

[7] S. Ali, G. Wang, B. White and R. L. Cottrell, "A Blockchain-Based Decentralized Data Storage and Access Framework for PingER," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2018, pp. 1303-1308, doi: 10.1109/TrustCom/BigDataSE.2018.00179.

[8] Z. Dong, F. Luo and G. Liang, "Blockchain: a secure, decentralized, trusted cyber infrastructure solution for future energy systems," in Journal of Modern Power Systems and Clean Energy, vol. 6, no. 5, pp. 958-967, September 2018, doi: 10.1007/s40565-018-0418-0.

[9] https://www.researchgate.net/publication/358286608_A_Blockchain-based_Access_Control_for_Big_Data

[10] Alizadeh, Morteza, Karl Andersson, and Olov Schelén. "Efficient decentralized data storage based on public blockchain and ipfs. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)." (2020): 1-8.

[11] Dff Z. Ullah, B. Raza, H. Shah, S. Khan and A. Waheed, "Towards Blockchain-BasedSecureStorage and Trusted Data Sharing Scheme for IoT Environment," in IEEE Access, vol. 10, pp.36978-36994, 2022,doi: 10.1109/ACCESS.2022.3164081.

[12] Yeon-sup Lim ISSN : 1976-7277 Cost-Efficient Framework for Mobile Video Streaming using Multi-Path TCP KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS VOL. 16, NO. 4, April 2022

[13] Sandvine, "The Mobile Internet Phenomena Report," 1H2020.Available:https://www.sandvine.com/download-report-mobile-internet-phenomena-report-2020-sandvine