WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Data Sharing Over Blockchain

Kajal Mahajan[1], Jay Wanjari[2], Aayushi Ghag[3]

[1]Information Technology, Vidyavardhini's College of Engineering and Technology, Vasai Road, India kajal.191994201@vcet.edu.in
[2]Information Technology, Vidyavardhini's College of Engineering and Technology, Vasai Road, India Jaywanjari1010@gmail.com
[3]Information Technology, Vidyavardhini's College of Engineering and Technology, Vasai Road, India aayushi.191854201@vcet.edu.in

ABSTRACT—

The project is titled as "Data Sharing Over Blockchain" is a peer-to-peer network and each node provides the storage service to the user's data. We are proposing a new user modeling platform using blockchain, permitting users to share data without leaving behind control and ownership. Our new platform solves three important problems: ensuring user privacy and control, and encouraging sharing. It verifiably tracks who shared what, with whom, when, how and for what purpose. Files shared via blockchain so that other nodes can process and consume the data efficiently. The smart contract verifies and enforces the agreed data usage terms and transfers digital tokens to users as a reward. The developed system is concerned with storing parts of a single hashed file, each part being stored on different nodes such that only the user can recover the parts to alter the original file. The motive of the project is to demonstrate and experiment a functioning Secure File Storage System for credential files of small sizes, based upon the blockchain technology. The peer-to-peer network is designed in such a way that permits each node to connect and leave the network anytime, acknowledge it's nearest nodes, and look for any node. The results show that nodes respond quickly with appropriate transaction co

***Index Terms*—Text analytics, Natural Language Processing (NLP), Sentiment analysis, Rule Based Model; Word Cloud**

I. INTRODUCTION

A. Overview

With increase in the demands of fields of Information Technology, Internet Of Things and Digitization of every profession, organizational projects and work, Information has become the substantial and beneficial asset for everyone. Data has become the most prominent thing in today's world. With the ample amount of data and it's ever developing nature, it's evenly dominant to reserve it in a formulated way such that it's easily approachable and safe. For this motive databases are being used as a repository to store data. Database play a pivotal role for any individual as well as any institution and business to reserve its data. Acquiring the significance of data and inadequacy of storage, databases are copied, chunked and backed up in distinct ways. Users stores data in the cloud offered by distinct private companies. Organizations construct their data centers at distinct part of the world to store its data. For the safety and bandwidth, data are distributed and copied to distinct servers at different sites. This seems to produce a good solution for the management of speedily growing data and also guarantees data security. In future, the rate of enlargement of data is sure to reach at its peak. To confront with it, the present database system needs to be more dependable, secure and accessible anytime. 1

B. Objectives

To establish a project over Ethereum Blockchain which can reserve the user's data in a decentralized database scattered across the P2P network. The certain objectives are as follows:

- To study about blockchain, peer-to-peer network and web technologies.
- To add in the active investigation on decentralized web applications.
- To build a distributed storage system using IPFS.

C. Problem Statement

Distinct cloud service organizations and dispersed data centre of organization guarantees the data accessibility and security. However,many of them have terms that permits the company to update, modify, allow, remove, view and examine your content. Reserving delicate data only on local machine can sometimes be regreting because once they are taken, lost or demolished by any other means, individual cannot make a retrieval. However, many of the private accounts of cloud storage also do not provide the security of data, take authority in case of data loss due to calamitous failure as well as guarantees data accessbilties anytime. Storing individual's sensitive data to cloud is not treated as a good option when it comes to the high potential value of that data.

D. Organization and contribution of the report

For the above mentioned problem statement, the solution proposed here is to develop a decentralized storage system which will reserve data in P2P network such that there will be no central system with the right to utilize and modify user's data. The data will be divided in to multiple groups, hashed using distinct cryptographic algorithms, reserved at different nodes. None of the node will know what data and whose data they are reserving even if any hacker hacks into any node and pulls data, it'll only retrieve part of the data which is hashed. So, it's difficult to grab absolute data in decrypted form by any hacker. Therefore, this is more safe than Cloud. For the decentralized storage system, the network will provide the storage of the nodes and the user will pay for the service.

II. METHODOLOGY

A. Existing Data Storage Systems

Currently, users use offline storage devices and other sec- ondary storage for data backup and protection. Most of us often use cloud services provided by Amazon, Google, Drop- box, Microsoft, etc. A huge amount of user data is stored in the cloud. The cloud is actually someone's computer or storage device. Such organizations have full authority over user data. In last few years, there has been a quick increase in the tendency for users to utilize this data and gain more advantage from it without acknowledging and obtaining permission from these companies.

B. Necessity of blockchain for a distributed storage system

In 2008, after the global financial crisis, a new paper brought about by person or a group named Satoshi Nakamoto introduced a concept of a peer-to-peer distributed currency with a paper titled "Bitcoin: A peer-to-peer Electronic Cash System" [1]. With there usually being a central authority bearing a higher power in a system may it be in economics, finance, technological or general, there exists a risk of the central authority misusing its power for one's gain while other's loss. Hence, Blockchain with its social perceptions related to Bitcoin not only solves the problem of "Double Spending" but provides us a solution for a problem with this specific criteria:

- Possibility of Fraud.
- Intermediaries or a middle man.
- Throughputs (Number of transactions/sec).
- Stable data.

Hence, called the FITS model which defines a environment where there is a possibility of Fraud, a middle man for transactions and exchange of other resources, transactions occurring in many number per seconds and a stable data i.e. data that is not constantly changing. However, taking this step further, Blockchain allows for what is known as a Distributed Autonomous Organization where it is an organization self- sustained by members without a central authority and trust built around what is known as "Smart contracts". [2] Hence, using blockchain technology we can build a system far exceed- ing the typical client server system in terms of application and scope but with the same amount of trust we put in a centralized server. Like a cloud platform where you are storing your data. You will trust the server of the given cloud service provider to hold the data without tampering and betrayal of trust and make it available for use any time you want. This trust is built by the company itself with continuous service and a good policy. However, in a distributed service this is hard to do as there is no central authority However, with blockchain we can implement a distributed system with a trust system. The same kind of trust people put in data storage services like a dedicated server or even simple cloud storage applications like Google Drive. Hence, for an implementation of a distributed storage network maintained by the people and for the people to use without an authority that provides trust but have trust built into the system, technology like blockchain is crucial. [3]

C. Existing Decentralized Database System using Blockchain

Although blockchain technology has been developed for more than 2 years, its practical application in various technical fields is only just beginning to bear fruit. There are few examples of distributed databases using blockchain. Some of the most popular blockchain applications in distributed database management are storj.io, bigchaindb, etc. We are talking here about BigchainDB, Storj. io, Sia and MaidSafe as they are more relevant to our project.

BigchainDB: A scalable blockchain database that uses blockchain technology to store user data across multiple nodes. BigchainDB inherits the characteristics of modern distributed databases: throughput and capacity scale linearly with the number of nodes, full NoSQL query language, queries and user permissions. Because it is built on an existing distributed database, much of its code base also inherits hard company code. [4] Storj is intended to be a cloud storage platform that cannot be audited and monitored without crashing. It is the first decentralized, end-to-end encrypted cloud storage that uses blockchain technology and cryptography to protect user files. Store is a decentralized platform, cryptocurrency and suite of applications that allows you to store data in a secure and decentralized manner. Files are encrypted, shredded into small pieces called "fragments" and stored times on a decentralized network of computers around the world. [5]

Filecoin and Sia are two other decentralized database sys- tems, both of which support smart contracts on the blockchain to set rules and storage requirements, and Store does not pay Store users who use them. This particular payment model means that if users disappear, hosts will no longer be paid for lending their space, which is a potential problem for those who will rent out their storage space. [6] MaidSafe also intends to do more than trade storage on its network; it positions itself as a "crowdsourced internet" on which not only data is stored, but decentralized applications run. Miners lease

their idle computing resources to the SAFE network, including hard drive space, processing power, and data connections, and are paid in native Safecoin. The SAFE network also supports Safecoin to access the market, and part of the payment will go to the app developer. Miners can also sell the coins they earn for other digital currencies, and these transactions can take place on the network or directly between individuals. [6] Storj, Sia, MaidSafe, and Filecoin all create local storage marketplaces where users and hosts can buy and sell storage space. All use mining to provide computing power to the network. In Filecoin, miners are not only rewarded with tokens for hosting files, but they must prove that they are constantly replicating files for more secure storage . They are also rewarded with for distributing content quickly, the miner who completes the content the fastest ends up with tokens. In Maidsafe's network called SAFE, users are paid in Safecoin when data is retrieved; however, this is done in a lottery system that randomly rewards miners. The amount of Safecoin someone can earn is directly related to the resources they provide and how often their computer is turned on.

III. PROPOSED SYSTEM

A. Requirement Analysis

The functional and non-functional requirements of this project are mentioned below:

1. Functional Requirements:

- The project will encrypt the data of user.
- The project will divide and merge the file.
- The project will built peer to peer network and permit clients to connect.[4]

2. Non-Functional Requirements:

- The project must ensure data retrieval of users.
- The project must permit client to store all kinds of files.
- The project should be dynamic enough to efficiently adapt with growing number of data

B. Software Requirements

1) Blockchain: A blockchain is a data structure used to create a public or private distributed digital ledger of trans- actions, not kept with a single vendor, but shared among a distributed network of computers. For example, distributed ledgers can be used to store key assets in supply chains to track their ownership and status changes. Data is stored in the blockchain in a distributed and redundant manner, with each node verifying each transaction. Therefore, it is difficult for malicious parties to attack and manipulate data to their advantage. As such, blockchains do not require a central server or trusted entity, often referred to as "trustless". The basic software model of blockchain technology was introduced in the source code of the Bitcoin digital payment system, but it is not limited to cryptocurrencies. The blockchain software model represents a digital ledger - a database, which contains an immutable record of every transaction. Each block collects a batch of timestamped transactions to place on the chain. Each block references a block's previous A signature, and the chain goes back to the first block (genesis) created in the chain.

The key idea is that there is no central authority to determine what is right or wrong; instead, several distributed parties come to a consensus, which is recorded in the register, which is then accessible to all. It is impossible for anyone (or less than the consensus of a majority) to go back and change history, because the blockchain represents a chronological chain of events. Since the data is stored in a distributed and redundant manner, every transaction is verified by every node, it is difficult for malicious nodes (corrupt parties) to attack and manipulate the data.

2) Ethereum: Ethereum is an open-source, public, per- missionless blockchain for building decentralized applications (dapps) in which users interact with online services in a distributed, peer-to-peer fashion, as evidenced by the review on Developers can use any of the popular programming languages and tools to design the interface and business logic.

Ethereum uses "ether (ETH)" as virtual currency, which can be used to pay transaction fees and provide the primary liquidity layer for trading digital assets. Ethereum also offers a technology called "smart contracts", which are software that automatically executes contracts. Unlike Bitcoin transactions, which can only be created externally, "messages" in Ethereum can be created by external entities or internal contracts. Ethereum messages also have an explicit option to include data and the Ethereum message recipient returns a response. Ethereum also has "transactions" - signed data packets that store messages to be sent from accounts belonging to third parties.

3) Solidity: Solidity is an object-oriented, high-level pro- gramming language for executing smart contracts. Smart con- tracts are programs which control the actions of accounts within the Ethereum state.

Solidity is a curly-bracket language created to mark the EVM. It is similar to C++, Python and JavaScript.

Solidity is statically typed, supports inheritance, libraries and complex user-defined datatypes among other characteris- tics.

4) Smart Contracts: A smart contract is an example of a contract deployed on the Ethereum blockchain , although the term was originally coined earlier in the context of electronic transaction agreements between strangers on the Internet.

When triggered by an authorized or agreed event, a smart contract can execute an agreed stored procedure. All contract transactions are stored in chronological order for future access and a full audit trail of events. If one party attempts to modify a contract or transaction on the blockchain, all other parties can detect and block it. If either part fails, the system will continue to operate without loss of data or integrity. Thus, it creates a single huge secure logical computing arrangement without the costs, risks, and trust issues of a centralized model. We write smart contracts using the Solidity programming language because it only allows basic operations on its primi- tive types, allowing for lightweight code. The contract uses EVM (Ethereum Virtual Machine) code , which is made up of bytes, and each byte represents an operation. For the implementation and testing of smart contracts on the Ethereum blockchain, Truffle Framework can be used as it integrates the composition, coupling, deployment and binary management of smart contracts. Right now, our smart contracts run on the EVM. However, with the future implementation of Ethereum Web Assembly (EWASM), the development of smart contracts can be done using any programming language other than Solidity, and it will also speed up function calls. between Web Assembly and JavaScript (JS).

C. Software Development Approach

It takes more time and effort to create large and dynamic systems using traditional approaches like the waterfall model of software development. Therefore, in order to more flexibly and timely meet the requirements of the system, we chose Scrum methodology in the way of agile development. Rather than providing a complete and detailed description of how to do everything in a project, much of it is left to the project development team. Scrum teams are self-organizing. This is because there is no higher level team leader to decide who does what task or how the problem is solved. Each time each team member tried to solve her problems and find a solution. In Scrum, teams are cross-functional. In other words, it takes people each to move a feature from idea to implementation. Project progress was represented in a series of sprints.

IV. RESULT AND ANALYSIS

Based on our observations of the project, we have come to the following observations and evaluations.

- File storage services work well at all stages of func- tionality: select, encrypt, split, upload, list, download, merge, decrypt. Metadata for each record is stored in a smart contract. If a copy of every transaction is kept in a local database and reading from the contract for validation checks is only done in race conditions, the task is much smoother. This makes the user experience smoother.
- File chunking merging and encryption/decryption are highly dependent on two factors: file size and processor speed. The time for each file operation is linear depending on the file size. So uploading a very large file puts a lot of load on the Android thread, and in some cases it is killed by the system, causing a system crash.

V. CONCLUSION AND FUTURE WORK

The project that is designed so far will help us in concluding that this project comprises of a Peer-to-Peer network where a node can connect to the network and supply the storage ser- vices for the users. It has an edge over the traditional database method. It reserves the data cryptographically encrypted in a decentralized manner where peers can be resembled as computing machine.

This project requires a future implementation as mentioned:

- UI for the project is required.
- There needs to be a size limit for the files and as to what kind of files can be uploaded.
- Security can be improved as loopholes in security were discovered.

REFERENCES

[1]. S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008.

[2]. coldfustion, “Why blockchain matters more than you think!” Sep 2017. [Online]. Available: https://www.youtube.com/watch?v=sDNN0uH2Z3o

[3]. “The great chain of being sure about things,” Oct 2015. [Online]. Available: https://www.economist.com/news/briefing/21677228-technology-behind-bitcoinlets-people- who-do-not-know- or-trust-each-other-build-dependable

[4]. T. McConaghy, R. Marques, A. M uller, D. De Jonghe, T. McConaghy, G. Mc- Mullen, R. Henderson, S. Bellemare, and A. Granzotto, “Bigchaindb: a scalable blockchain database,” white paper, BigChainDB, 2016.

[5]. S. Wilkinson, T. Boshevski, J. Brandoff, and V. Buterin, “Storj a peer-to-peer cloud storage network,” 2014. S. Jung, “Filecoin v. sia, storj maidsafe: The crowded push for decentralized storage,” Aug 2017. [Online]. Available: https://medium.com/tokenreport/filecoin- vsia-storj-maidsafe- the-crowded-push