BTP Report

Naman Jain, Pranshu Maheshwari, Saksham Tanwar

April 26, 2021

Contents

1 Introduction

1.1 Generic Application

A typical application can be broken down into three logical layers namely:-

- **Presentation Layer:** It is the layer responsible for display information and collect information from the user.
- **Business Logic Layer:** In this layer, information collected from presentation layer is used to perform the various calculations and operations needed to be performed by the application.
- **Data Layer:** This layer is responsible for storing and retrieving data to be used by other layers.

Data Layer is typically implemented using a database, which is an organized collection of data. A database can be of two types : Centralized and Distributed.

1.2 Disadvantages of a centralized database

Most applications require data in some form or another and need to store the said data for future use, this is where databases come into picture. A database is an organized collection of data, generally stored and accessed electronically from a computer system. Access to this data is usually provided by a "database management system" (DBMS) consisting of an integrated set of computer software that allows users to interact with one databases and provides access to all of the data contained in the database. DBMS's provide various functions that allow management of a database and its data such as Data definition which defines the database model i.e. the logical structure of the data model, Updating of actual data which includes insertion, modification, and deletion of the actual data, Retrieval of actual data which may or may not include data processing and Administration which includes allowing access to various users, enforcing data security and various other checks. [1] A database model is a type of data model that determines the logical structure of a database and fundamentally determines in which manner data can be stored, organized and manipulated. The most popular example of a database model is the relational model, which uses a table-based format. The ACID (Atomicity, Consistency, Isolation and Durability) database design is one of oldest and most important database design. A relational database that fails to follow the ACID model cannot be considered reliable.

Databases can categorized in many categories depending on how they store data, where it is stored and the additional functionality the DBMS can provide on the stored data. Broadly databases can categorized into two major categories centralized and distributed. They both have their advantages and disadvantages. A centralized database is a database that is located, stored, and

maintained in a single location and has a single DBMS but can be accessed by the users distributed in the network. A distributed database system consists of a collection of local databases, geographically located in different points (nodes of a network of computers) and logically related by functional relations so that they can be viewed globally as a single database [2]. There are two types of distributed databases homogeneous and heterogeneous. In a homogeneous distributed database all the locations store the database identically i.e. all locations have the same management system and schema. Whereas in heterogeneous distributed database different locations can have different management software, schemas. For a database management system to be distributed, it should be fully compliant with the twelve rules introduced by C.J. Date in 1987 [3]: local autonomy; the absence of a dependency from a central location; continuous operation; location independent; fragmentation independent; replication independent; distributed query processing; distributed transaction management; hardware independent; operating system independent; independent of communication infrastructure; independent of database management system.

If a centralized database is used to store the data then the whole system is prone to Single Point of Failure, if for some reason the database fails the whole system will fail. As all the data is stored in a single location, if there are no precautionary measures taken then a hardware failure can lead to complete data loss. The databases are also vulnerable to cyberattacks which can cause data loss, data leaks, data inconsistency and many other vulnerability. This makes centralized database dubious with very low reliability. But at the same time, centralized database ensures data consistency and easy management in compliance with ACID design [4].

A major advantage of distributed database is that by sharing a database across multiple nodes can obtain a storage space extension and also can benefit from multiple processing resources. A distributed database system is robust to failure to some extent. Hence, it is reliable when compared to a centralized database system. It is also more robust compared to a centralized database as it doesn't have a single point of failure, if a node fails another node or group of nodes can provide the necessary data. But to make this possible complex software's are required which incur additional costs and processing overheads. Marinating data integrity is difficult and hence minimum redundancy and ACID properties are more relaxed than compared to a centralized database [4]. Distributed environment also faces problems such as fragmentation and data replication. A data fragment constitutes some subset of the original database. A data replica constitutes some copy of the whole or part of the original database. The fragmentation and the replication can be combined: a relationship can be partitioned into several pieces and can have multiple replicas of each fragment [5].

1.3 Advantages of using Blockchain

Blockchain is a type of decentralized database where the data is stored in blocks and each block is linked to the previous block using its cryptographic hash, thus forming a chain. As the blockchain grows in size, it becomes more and more difficult to tamper the data in these blocks. Blockchain based databases are more trustworthy, reliable and secure than traditional databases.

A blockchain based system is decentralized. Hence, a centralized authority is not necessary. Decentralized systems are also resilient to single point of failure. A blockchain network comprises of various nodes, each maintaining its own copy of database. Hence, all the data is replicated, shared and synchronized across all the nodes in the network. This makes the data almost tamper proof as it will be very costly and thus highly unlikely for a malicious person to change the data across the whole network. No node can directly write data to the blockchain as it can make the data inconsistent, to avoid this the blockchain network uses a consensus algorithm which helps all the nodes in the network to come to a consensus and commit the data in the blockchain. This helps to keep all the nodes in the network to get synchronized with each other.

The data that is committed to the blockchain is immutable. After some data is inserted into the blockchain, it is very difficult to delete or alter it. This is realized by including, in each block, the cryptographic hash of its previous block. Because of this immutability, a blockchain database only supports create and read operations in contrast to traditional databases which supports create, read, update and delete (CRUD) operations. To update the database state, a subsequent write is required. Old writes persist in the blockchain forever, making it convenient to trace how the blockchain database state is updated over time. This makes it possible to trace the actions that resulted in a particular database state.

Another benefit of blockchain is that it provides transparency to the data stored in it. This is achieved by replicating the blockchain data among multiple nodes present in the blockchain network. Nodes in the blockchain network not only can view the data already committed to the blockchain but can also participate in the process of validating the data to be committed in the blockchain. The transparency also helps users to cross verify some data against the blockchain.

A blockchain ledger is logically a single public ledger. Since every node updates its copy of blockchain ledger in sync with each other and after consensus is reached, the whole network logically act as one single public ledger. This removes the complications present in traditional systems where different participants or organizations maintains multiple ledgers which needs to be reconcile and synchronized time to time.

Systems that are deployed on public blockchain network are more cost efficient as compared to the traditional systems. This is because the system utilizes the processing power and resources of a large number of nodes that are already connected to the blockchain network, thus significantly reducing the cost needed for setting up and maintaining centralized servers present in traditional systems.

Blockchain based systems are more secure and resilient to cyber attacks, which can cause data loss, data leaks, data inconsistency and many other vulnerability, than traditional sytems. Whenever a new block is introduced in the blockchain, its hash value is calculated which also includes the hash of its previous block. If a malicious user fraudulently tries to tamper the data inside a block, not only its hash value changes, but the hashes of the following blocks get changed too. This, along with the fact, that the data is stored in multiple nodes present in the blockchain network, makes it very difficult to tamper the data present in the blockchain.

1.4 IPFS

The Interplanetary File System (IPFS) [6] is a peer to peer distributed file system for sharing and storing data. IPFS provides a high throughput storage system where each file is uniquely identified by a cryptographic hash. In IPFS, multiple users store a piece of each others' data and actively participate in making it available in the network. In contrast to HTTP which uses location based addressing, IPFS uses content based addressing. In content based addressing, user request data using its content identifier instead of the location where it is stored.

IPFS uses Distributed Hash Table (DHT) [7] to enable routing and discovery of peers and content on the network. The hash table is split across all the participating nodes in the IPFS. These nodes coordinate with each other to enable efficient lookup and retrieval of data on the network. IPFS stores data using a data structure called Merkle DAGS [6]. Each node in a Merkle DAG stores the cryptographic hash of each of its children. Thus, altering a node's data changes the hash value of the node as well as all its ancestors in the DAG. This makes the data stored on the IPFS immutable.

1.5 Motivation

In India, currently if person is asked to prove if he/she is the owner of a piece of land then all they can show for it is a sale deed which just proves that the person was the owner at a particular time, but that person cannot prove that he/she is still the owner of that land. For proving the same he/she has to go to various government offices and collect various document showing that no sale

deed has been registered for that land after the one which the person has. When a person wishes to buy land in India, they have to be very careful and perform various checks such as

- Check if the deed title is in the name of the seller and if he/she has the full right to sell it
- Procure a Encumbrance Certificate from sub-registrar's office where the deed is registered which declares that the land is free of any legal hassle and unpaid dues
- Check if the Property tax and other bills are paid in full and the seller has the respective original receipts
- Check if the loan on the land has been completely repaid

If a buyer is lethargic in verifying these documents he/she can be easily duped into buying a disputed land parcel or it may even happen that the land in question did not even legally belong to the said seller, that is the seller was not the genuine owner of that land provided fake documents. As these documents are mostly maintained offline as hard copies these cases occur often.

Getting all the information such as the sale deeds and ownership history for a land parcel is very cumbersome, a simple solution to this problem is a smartphone application with which users/potential buyers can get land ownership history and related documents for a particular land parcel. These documents and data can be stored in a database by the government and will be updated every time a new sale deed is registered i.e. owner for a land parcel changes. This makes the process of acquiring the documents in question very easy and hassle free.

Indian government launched Digital India Land Record Modernization Programme in 2008. Main aim of this programme is to create a system of updated land records, automated and automatic mutation, integration between textual and spatial records, inter-connectivity between revenue and registration, to replace the present deeds registration and presumptive title system with that of conclusive titling with title guarantee [8]. In addition to this, Indian government also plans on storing the digital records on a private blockchain and provide a system where citizens can view the ownership details and complete history of the property before going in for the purchase [9].

1.6 Proposed Solution

To takle this problem of accessing land records, a smartphone based solution is proposed. Users can request land records for a particular land parcel by visiting the land parcel or marking it on the map. Land records are stored on a private blockchain which provides methods to access the data. For help in migrating

old records to the new system, scanned documents can be uploaded. These documents are stored in a private IPFS cluster with their Content Identifier (CID) is stored on the blockchain. A verification portal is also part of the solution where users while accessing records will be given verification code which users can use to generate a digitally signed PDF for verifying the land records.

1.7 Organisation

Chapter two provides an introduction to the various offerings of the blockchain technology that is makes it useful for the proposed solution. Chapter two also provides a summary of related works. Chapter three describes the proposed solution's architecture including the various actors/users of the solution, various submodules as part of the solution and how they interact with one another. Chapter four provides details regarding the deployment of the proposed solution. Chapter five concludes with the main contributions of this work and proposes direction for future work.

2 Chapter Two //TODO Title

2.1 Related Works

Republic of Georgia, in 2016, undertook a phased project to create a Blockchain-based land titling system. The project has significantly improved the government's efficiency and ensures security and immutability of the data [10].

Sweden's land registry authority Lantmäteriet tested a blockchain based solution to store land records and allow users to buy and sell land by using a smartphone application. The authenticity of the process and digital signatures are secured using the blockchain [11].

[12] lists the various problems in the current land record management system in India and suggests a blockchain based solution and gives a detailed flow of the land registration process in India and how blockchain technology improves the registration system.

[13] suggests a permissioned blockchain based system for managing land records using Hyperledger Composer for permissioned blockchain implementation and IPFS for storing various documents. Implements the front end as an Angular based web application and uses Passport for authentication.

[14] Proposes a three phase project for adoption of Blockchain based solution for land titling problem in Bangladesh and provides an implementation of a prototype model using Ethereum. Phase One proposes adoption of public blockchain such as Etherium to store timestamps of land transactions on the blockchain. Phase Two proposes a hybrid blockchain solution consisting of

both public and private blockchain networks, where key information is stored on public blockchain and all data related to land is stored on private blockchain network. Phase Three proposes using IPFS to store digital documents in the hybrid blockchain solution.

References

[1] Wikipedia.

[2] M.T. Ozsu and P. Valduriez. *Principles of Distributed Database Systems.* Springer, New York, 3rd edition, 2011.

[3] C. Date. *An Introduction to Database Systems. Vols. I and II.* Addison-Wesley Publishing Co., 4th edition, 1987.

[4] Mirela Liliana Moise Nicoleta Magdalena Iacob (Ciobanu). Centralized vs. distributed databases. case study. *Academic Journal of Economic Studies*, 4:119–130, 2015.

[5] Korth H.F. Silberschatz, A. and S. Sudarshan. *Database System Concepts.* McGraw-Hill, 6th edition, 2010.

[6] Juan Benet. Ipfs - content addressed, versioned, p2p file system. 07 2014.

[7] Klaus Wehrle, Stefan Götz, and Simon Rieche. 7. distributed hash tables. volume 3485, pages 79–93, 01 2005.

[8] Department of land resources.

[9] Centre of excellence in blockchain technology.

[10] Qiuyun Shang and Allison Price. A Blockchain-Based Land Titling Project in the Republic of Georgia: Rebuilding Public Trust and Lessons for Future Pilot Projects. *Innovations: Technology, Governance, Globalization*, 12(3-4):72–78, 01 2019.

[11] Joon Ian Wong. Sweden's blockchain-powered land registry is inching towards reality. 2017.

[12] Vinay Thakur, M.N. Doja, Yogesh K. Dwivedi, Tanvir Ahmad, and Ganesh Khadanga. Land records on blockchain for implementation of land titling in india. *International Journal of Information Management*, 52:101940, 2020.

[13] H. Mukne, P. Pai, S. Raut, and D. Ambawade. Land record management using hyperledger fabric and ipfs. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–8, 2019.

[14] Kazi Masudul Alam, J.M. Ashfiqur Rahman, Anisha Tasnim, and Aysha Akther. A blockchain-based land title management system for bangladesh. *Journal of King Saud University - Computer and Information Sciences*, 2020.