BD - 10

Team: ForToZero

Know Your Customer (KYC) Implementation on a Decentralized Architecture with Hyperledger Fabric and IPFS

22 Oct, 2022

Table of Contents

Abstract:

The traditional KYC process has many shortcomings. It is a very inefficient and time-consuming process which leads to financial loss both for the client and the organization. If clients want to open accounts in multiple banks, then the sufferings know no bounds. To overcome the shortcomings of traditional KYC process this proposed model uses modern technologies like blockchain and peer to peer file sharing system to ensure better efficiency and user privacy.

In the proposed model, integration of Hyperledger Fabric (a private blockchain) and Interplanetary File System (IPFS) will be used to build a decentralized application for Banks to upload files containing KYC details. The proposed model will be designed using the private blockchain for its' striking features like enhanced privacy and transparency. The trust issue for KYC file storage will have been solved by the Interplanetary File System (IPFS) which is a decentralized platform. The success of this project would create a user-friendly system for KYC verification. This system would not only ensure data privacy but also make the whole procedure more efficient and faster. The customers would not need to go through separate KYC verification for each new bank. This would encourage more financial activities in different banks from the customers. This would also create a more secure system for storing sensitive personal information of the customers.

Introduction:

KYC verification can include ID card verification, biometrics verification, and document verification (bank statements, utility bills, and more). Banks have to comply with KYC regulations and anti-money laundering regulations to detect and eliminate fraud. Complying with KYC regulations is a responsibility financial institutes have to uphold. Non-compliance with KYC and AML regulations can lead to heavy fines imposed by regulatory bodies. As the world moves forward, these regulations will experience more changes. Criminal activities like money laundering and terrorist financing have increased; therefore, all regulatory authorities have made some changes to the KYC/AML regulations. After the COVID-19 pandemic, banks facing significant challenges and more rigid laws. The traditional system isn't enough to keep up with these issues.

Blockchain has many advantages over the traditional database system. One of the major advantages is, it stores data cryptographically encrypted via distributed and decentralized way in peers. This solves the availability of data within banks and revokes the access to the outside world. For more security the proposed model is built with private blockchain so data breaching is nearly impossible.

Hyperledger Fabric (private blockchain) and Interplanetary File System (IPFS) is used to solve these problems as much as possible. Through the public KYC smart contract, a user registers and uploads their KYC information to the network and it is stored in IPFS with a unique ID so that it can be retrieved easily when needed. Hyperledger Fabric ensures enterprise level security and distribution of user data in a decentralized manner. IPFS is used mainly to store user data in a most secure way and share them among peers without any hassle. With the combination of these two technologies the presented system introduces an enterprise KYC solution. Also, the data storing and sharing process is much more transparent with IPFS and smart contracts of Hyperledger Fabric, so the trust issues among peers are eliminated. A central authority can always monitor the whole system and take necessary steps if any kind of fraud occurs.

Enterprise blockchain solutions attempt to solve the crucial matter of user privacy, albeit that blockchain was initially directed towards full transparency. In the context of Know Your Customer (KYC) standardization, a decentralized schema that enables user privacy protection on enterprise blockchains is proposed with two types of developed smart contracts. Through the public KYC smart contract, a user registers and uploads their KYC information to the exploited IPFS storage, actions interpreted in blockchain transactions on the permissioned blockchain. Furthermore, through the public KYC smart contract, an admin user approves or rejects the validity and expiration date of the initial user's KYC documents. Inside the private KYC smart contract, CRUD (Create, read, update and delete) operations for the KYC file repository occur. The presented system introduces effectiveness and time efficiency of operations through its schema simplicity and smart integration of the different technology modules and components. This developed scheme focuses on blockchain technology as the most important and critical part of the architecture and tends to accomplish an optimal schema clarity.

1. Opportunity

From the business perspective, the problem exists due to insufficient trust and the effort to build that trust is an insurmountable obstacle without the use of some technology to help coordinate and reconcile. From the technology perspective, the solution shows that blockchain can address the problem better than other technologies.

1.1 Problem

- **Duplication:** If a client wants to open a bank account, he has to submit different types of documents like NID, Driving License etc. Once the documents have been verified the bank will open an account. But when the same client wants to open an account in another bank he again has to go through the same process. Clients are frustrated with the number of times they are asked to provide the same personal information to different banks over and over again in order to KYC verification. As such, complying with KYC regulations also takes

longer for corporates and companies are now spending an average of 26 days on the process, up from 23 days in 2016. However, corporate customers claim that, on average, they are spending 32 days on KYC compliance.

- **Regulation:** There is no global standard for know your customer (KYC) regulations, so different banks will have varying documentary requirements for their corporate customers. The report states: "Survey respondents reported an average of 10 global banking relationships — each one placing different demands on corporates, resulting in frustration, rising costs and wasted time." Corporates were also frustrated with having to deal with many different people within the bank. Thus, the entire process is very inefficient.
- **Redundancy:** Most large files use similar data and processes to verify an equivalent client. The solution benefit is to eliminate the redundancy documentations that got to be verified only once before the approval information is shared.
- **Security concerns:** Concerns over security and who is viewing documents is also an important issue for corporates, partly because the documents required for KYC include the personal documents (such as passports) of company directors. KYC processes enable regulated institutions to verify the identity and intentions of individuals or businesses to whom they provide services. That's essential for complying with fraud and anti-money laundering regulations.
  The worrying fact is that some of the customers' most sensitive personal documents end up stored in various locations by multiple third parties. There is now ample evidence of the vulnerability of even the world's largest data referencing companies. The Equifax data breach in 2017, for example, exposed sensitive personal information belonging to 143 million Americans.
- **Costly:** The traditional KYC process is slow and expensive. Companies need to hire a team of competent staff, who charge high salaries. Alternatively, companies are spending unprecedented sums on training existing and new staff. The average cost of annual compliance training is $45/hour per member of staff, with an average of 5 hours needed. For firms with 1000+ employees, compliance training can cost in excess of $225,000 a year. If these AML mechanisms aren't up to par, companies face unimaginable fines. Financial institutions have paid 46.4 billion dollars from 2008 to 2020 related to AML KYC data privacy and KYC regulations.

1.2 Solution

Blockchain technology offers several opportunities for streamlining KYC verification and disrupting traditional identity management solutions:

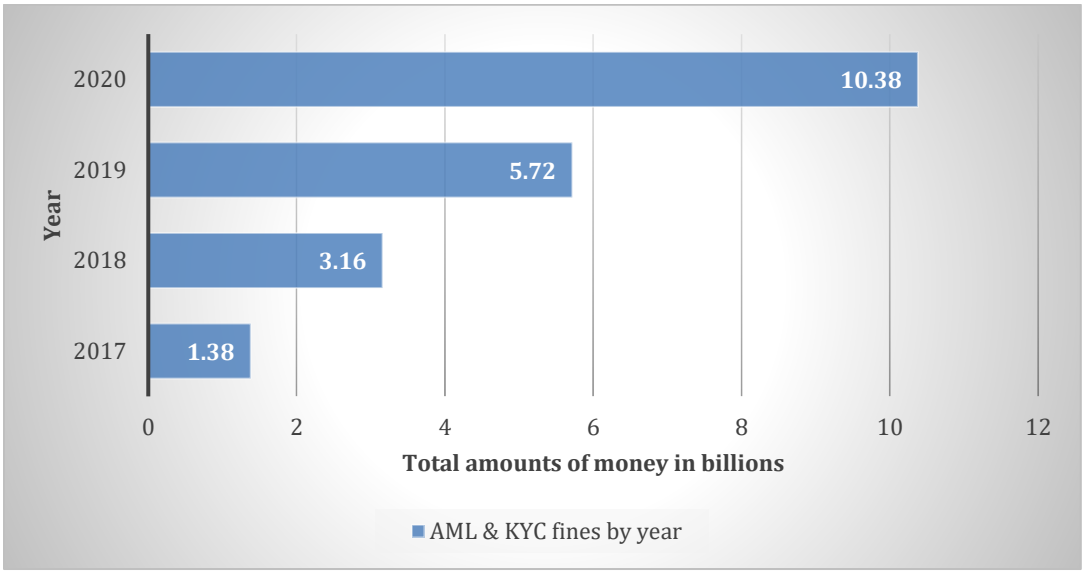
- **Remove duplication:** A significant benefit of streamlining the KYC verification process on a permissioned blockchain, like Hyperledger Fabric, is that only authorized and permitted members of the chain can add to or exchange customer data. This differs dramatically from a permissionless and public blockchain, such as Ethereum.
- **Added security:** Storing customer identity documents as encrypted files on the chain is more secure than traditional methods. Cryptographic hashes enable the blockchain ledger to remain as a permanent record that cannot be altered or reversed. With blockchain, the

financial institution only validates the user and does not have access to sensitive information included in the credential document

- **Increased transparency:** Verification information may be shared securely between financial institutions which makes the KYC regulation process convenient.
- Efficient: Blockchain technology will be very efficient in the KYC process. It will save a lot of time in every step of the KYC process.
- **Reduced cost:** Once a customer opens an account and provides credentials to one financial institution, a smart contract archives and hashes the document containing the credentials. Other institutions can then access that smart contract in order to verify the customer's identification.

2. Market

Financial institutions have reported spending $60 million annually, based on research conducted by Consult Hyperion in 2017. Some are spending up to $500 million each year on KYC. Global spend on Anti-Money Laundering (AML)/ Know-Your-Customer (KYC) data & services is projected to total a record $1.35 billion in 2021.



From the graph we can see the that AML and KYC fines have been rising for the past few years. It was simply 1.38 billion dollars in 2017. But in 2020 we can see a massive increase of 9 billion from 2017 fines.

As such, there is a huge market for this project to tap-into despite the pre-existing competitors and alternative solutions besides Hyperledger Fabric.

3. Partners

The organizations of the same private network will be called partners. In this KYC model, the partners are -

- Central Bank, State-Owned Commercial Banks, Private Commercial Banks and Foreign Commercial Banks
- Credit unions
- Wealth management firms and broker-dealers
- Private lenders and lending platform

Domain operator and regulator will be the central bank. The incentive will be given by the Central Bank. Central bank and all the partners bank will be in the same private blockchain network. Central Bank is the moderator in charge of adding new partners to the network.

4. Competition

There are many e-KYC registration services in the market like Identity Mind, Shufti Pro, Beam Solutions etc. They verify their customer's identity and address electronically through documents authentication.

In the current e-KYC process, all the data is at the same location. If multiple users try to access the database simultaneously it creates a problem. This may reduce the efficiency of the system. In IPFS, same data is fetched from different peers. As a result, there’s no latency or inefficiency in the system. If there are no database recovery measures in place in the centralized database system and a system failure occurs, then all the data in the database will be destroyed. And also, the database recovery measures are very costly. In IPFS the data will be distributed in all the peers. If any peer’s database destroys, the data can be retrieved from other peers.

Data security is always an issue in centralized database. If an attacker somehow manages the access of the database, he will have the full control of the database. So current e-KYC model is not very trustable. In IPFS, every piece of data is cryptographically hashed, resulting in a safe, unique content identifier: CID. So, it is cryptographically hard for any attacker to access the data.

The data is mutable in this current e-KYC process that means anyone can alter the information for personal gain. In blockchain, the data is immutable that means no one can alter the blockchain's distributed ledger about all the committed blocks. If the data has been manipulated, replaced, or falsified by a partner or its employees, everyone in the network can see what information is altered by the partner or its employees in the block.

The central bank can't monitor the whole system in the current e-KYC process. As a result, if any bank or organization try to misuse the data, they can surely do it. In the proposed model, the central bank can monitor the full network. It can add any node to the network and also has the power to remove a node from the network for the misuse of data.

Nowadays there are many enterprises which are giving KYC solutions with blockchain technology like Antier, KYC-chain etc. They are giving particular organizations or companies blockchain-based KYC solutions. In our proposed model, we are trying to solve the whole KYC-related problems of an entire country. Also, companies like Antier are using Ethereum, which is a public blockchain. Because of the Ethereum blockchain, transaction data is available for everyone. No banks want to share their transaction information publicly. So Hyperledger Fabric is a good solution for developing these types of KYC systems.

5. Risks

Blockchain technology is currently new to many people especially private blockchain. So, it is somewhat difficult to create an enterprise solution with this technology. But now-a-days people are more interested in learning new technology than before so they are embracing blockchain very quickly. Not only the developer but also the users need some time to adopt this KYC solution. People who have a high level of technological expertise are needed to develop such a network. Many Clients will face some difficulties in the initial phases.

The main security threats of Hyperledger fabric are - Denial of Service, Consensus Manipulation, MSP Compromise, Smart Contract Exploitation. To mitigate this risk, smart contracts should be designed with security in mind at the onset by following a secure software development life cycle framework. Before deploying, smart contract security should be assessed with smart contract analysis tools like the Hyperledger Lab Chaincode Analyzer.

Besides there are already much customer data, so it is a huge responsibility for the developer to move all that data into blockchain database maintaining privacy. Also, the cost would be high for doing so. But this is a one-time cost and after the work is done properly every bank can get benefit from it.

Banks should be highly co-operative with each other because of the data flow. Let's say, if someone open an account in a particular bank the address of his document stored in the private IPFS will be stored in the blockchain as that bank's private collection. So, when that person approaches to another bank, the bank needs to send request to previous bank for that person's personal data. In that case these two banks are sharing valuable information with each other and that sharing process should not take much time. If the banks aren't co-operative with each other, it would delay the process and affect the customer satisfaction.

6. Architecture

The blockchain technology that will be used in the project is called 'Hyperledger fabric'. There are many reasons for choosing Hyperledger Fabric for this project.

- **Identity management:** Hyperledger Fabric's Membership Service Provider (MSP) provides support for identity management
- Privacy and confidentiality: Hyperledger Fabric offers private channels that are out of bounds from peers or members without permission. This enables secure, private data sharing among channel members.
- **Efficiency:** Hyperledger Fabric's concurrency and parallelism feature improves its performance by allowing transactions to be threaded and processed faster.
- **Chaincode execution:** Chaincodes are the business logic on the Fabric network. They are responsible for executing transactions on-chain. Chaincode logic is written in general-purpose languages, making it possible for developers to build Chaincodes in the language with which they are most comfortable.
- **Governance:** Hyperledger Fabric makes use of governance policies to identify peers who can deploy a Chaincode or add an MSP to a channel
- **Modular architecture:** Hyperledger Fabric's modular architecture makes it flexible to changes. The Hyperledger Fabric blockchain can be configured with multiple Channels, and multiple banks or partners can join a single Channel or join different Channels for data sharing. Network administrators create their own CA in the blockchain network and then apply for a public-private key and a digital certificate using the X.509 standard from the CA to provide signatures for transactions and to endorse the results of transactions. The digital certificate contains basic information, e.g., version number, serial number, business registration number, public key, enterprise tax number, and valid time.

The name of the decentralized storage is IPFS. IPFS is a distributed system for storing and accessing files, websites, applications, and data built by Protocol Labs. This system is a service that relies on a distributed network of computers that delivers information based on its content, which is stored on many nodes or computers around the world. It uses "content-based addressing" which serves up files based on its content. The system is a decentralized, peer-to-peer file-sharing

network and open-source Web3 service, designed to overcome centralized points of failure and censorship efforts, to ensure that the web is freely accessible to all. Users in a local network can communicate with each other, even if the Wide Area network is blocked for some reason. Since, no servers are required, creators can distribute their work without any cost. Data load faster as it has higher bandwidth.

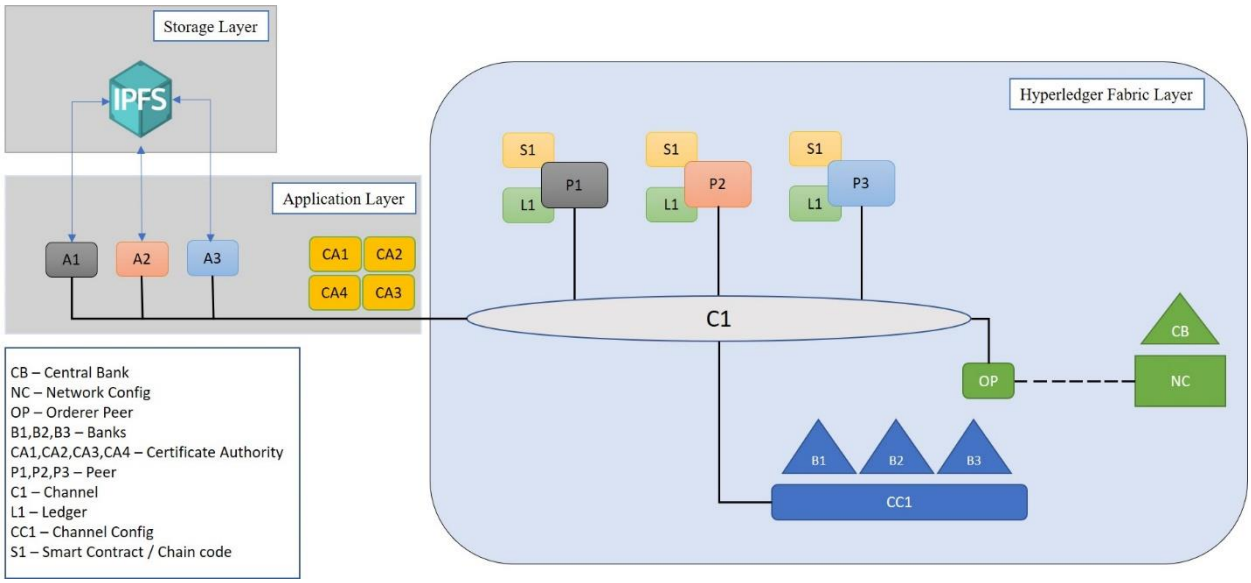


Fig: Architecture of the Hyperledger-based KYC process

The framework is divided into three layers -

1. **Hyperledger Network Layer:** This includes Peers, Orderer peer, Channels, and Certificate Authority (CA). The CA is responsible for issuing public and private keys and digital certificates. Administrators and Peers must be authenticated by the CA to become part of the blockchain network. The Channel is a private blockchain built based on data isolation and confidentiality. The data in the channel (e.g., Ledger information and member information) is known only to the members in the channel, and the data cannot be shared between different channels, and the channel mechanism ensures data sharing between different partners while protecting privacy. The Orderer peer only sorts and packs the transactions received in the channel and does not verify the legitimacy of the transactions, and then broadcasts the packaged transactions to all Peers in the channel. Peers are a network entity that maintains the ledger and runs the Chaincode to do read and write operations on the ledger.

2. **Application Layer:** The partners or the peers are connected to the blockchain network through the application, which uses the SDK (Software Development Kit) to interact with the blockchain network and can access the ledger through Peers using the Chaincode, and the administrator which is the central bank needs to register through CA to participate in transactions in the system.
3. **Storage Layer:** Peers that join the same channel will also join the channel's IPFS network, which is a distributed file system for storing and sharing data, and generating a hash address for storing data, which is a key component.

7. Governance

The whole process can be described in two major steps. The first one is when someone approaches to a bank to create a bank account who didn't open an account in any bank before, second is when someone with a previous bank account approaches to a new bank for another account. Let's discuss these steps below:

1. Let's assume someone wants to create a bank account for the first time in bank 'A'. First, he/she needs to send the necessary documents and picture needed for KYC process through web interface. Then bank A will verify the given data. The verification process will be defined by the central bank. After verification the bank will store these documents in its own database. Each bank needs to be authorized by the central bank to use the blockchain based KYC solution. The central bank has the right to approve or reject the authorization request in case any bank break the regulations. So now bank A will send request to central bank for authorization if it isn't authorized already. After approving bank A's request the central bank will store the bank's identity information in their own database. Now the bank will store customer's data in private IPFS and this process will return a unique content identifier known as CID (a level used to point material in IPFS). After that the bank will generate a unique KYC id for that customer which will be used to query the blockchain database in future. Now the bank will create a file where KYC ID, Customer Name, Account Type, NID no. will be stored which will be called KYC docs. This KYC docs will be stored in the blockchain through a smart contract and the CID of that customer will be stored in that bank's private collection. The private collection is a special feature of Hyperledger Fabric which is used to manage data privacy, so other banks in the same channel won't be able to access the CID unless bank A gives access to them. This ensures an extra layer of privacy to customer's data and only the bank someone wants to use the service of, will be able to access his/her personal details. With the upload of KYC docs in the blockchain the KYC process is finished and the bank will give the previously generated KYC id to the customer which will be used in the second process.

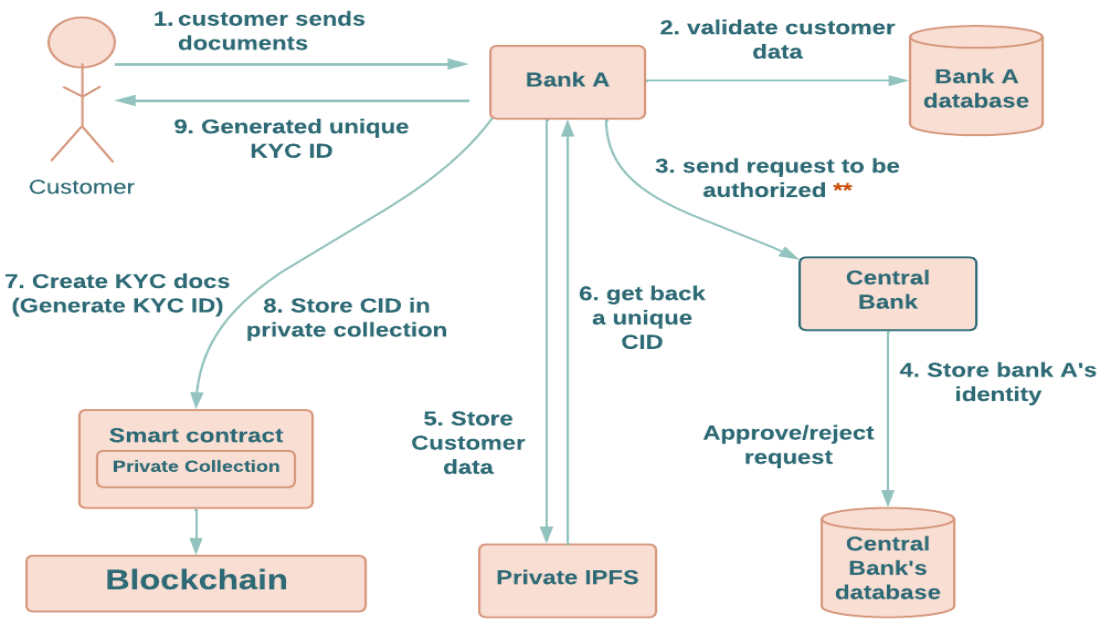


Fig1: Customer approaching the first bank

2. In this step it's assumed that someone wants to create a bank account who already has an account in another bank. So, the customer approaches the bank with his unique KYC id instead of the whole KYC document because it is already stored in the private blockchain and IPFS storage. The bank then sends a request to the central bank for permission to use the blockchain-based KYC solution if not already authorized. After finishing the authorization process the bank can now query and access the data of that customer from the private blockchain database. But the customer's CID is in the private collection of the previous bank, in which the customer already has an account. So that bank now needs to share its private collection with the new bank. After getting the customer's CID the bank retrieves his/her KYC documents from private IPFS storage and finishes the process of creating an account with this data.

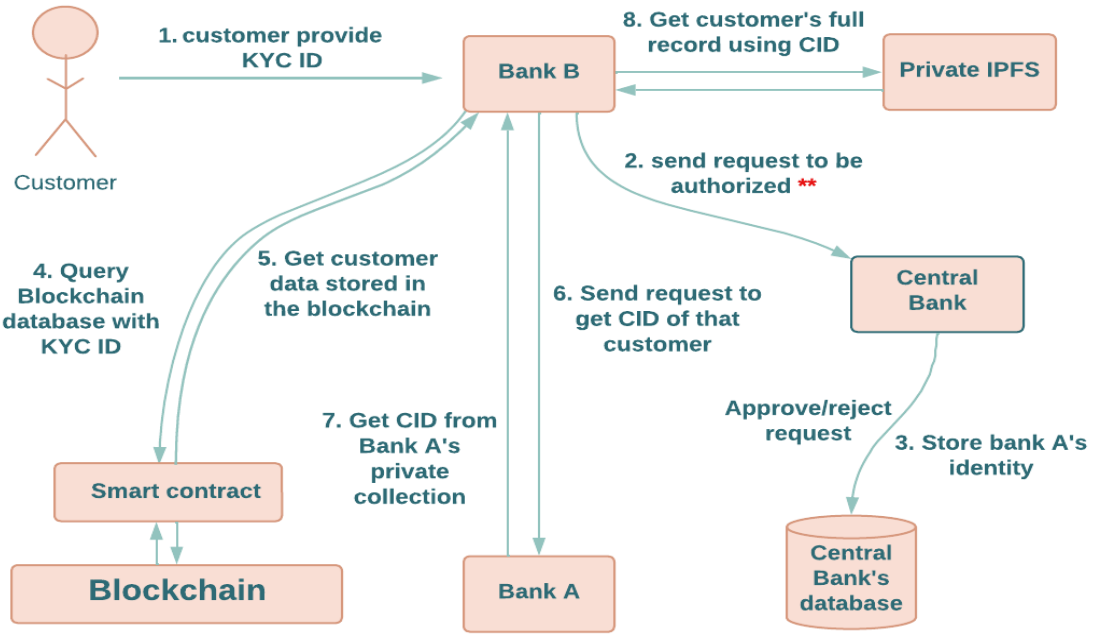


Fig2: Previous client approaching to a new bank

8. Value Proposition

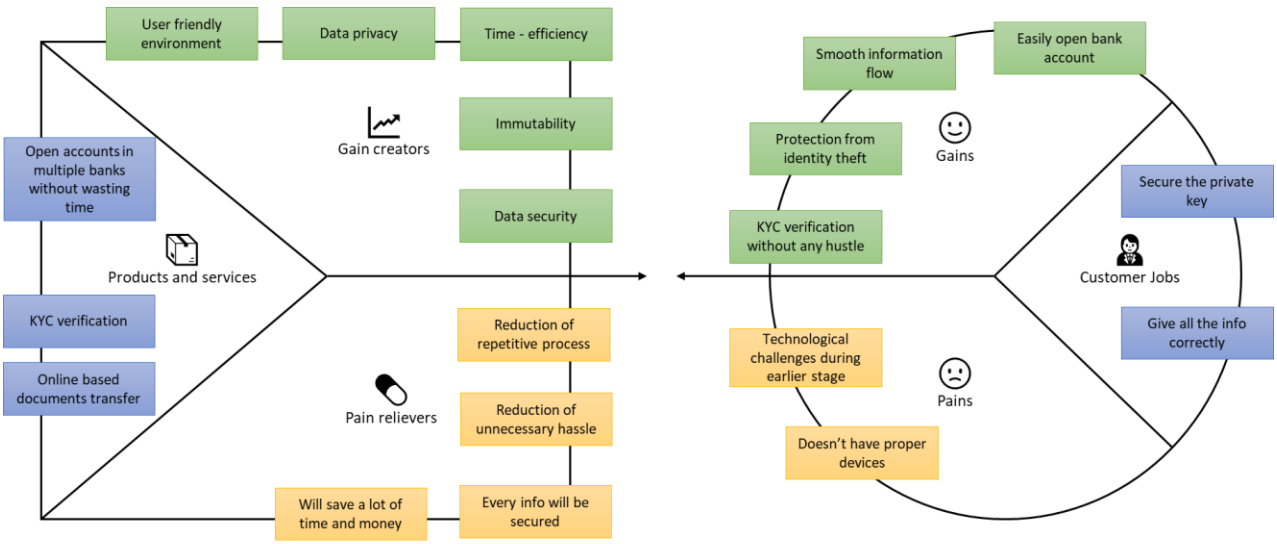


Fig: Value proposition canvas

9. SWOT Analysis

The Swot analysis canvas shows that this whole blockchain based KYC solution has more strength and opportunities than some minor weaknesses and threats. More research and development initiatives can make this project almost flawless.

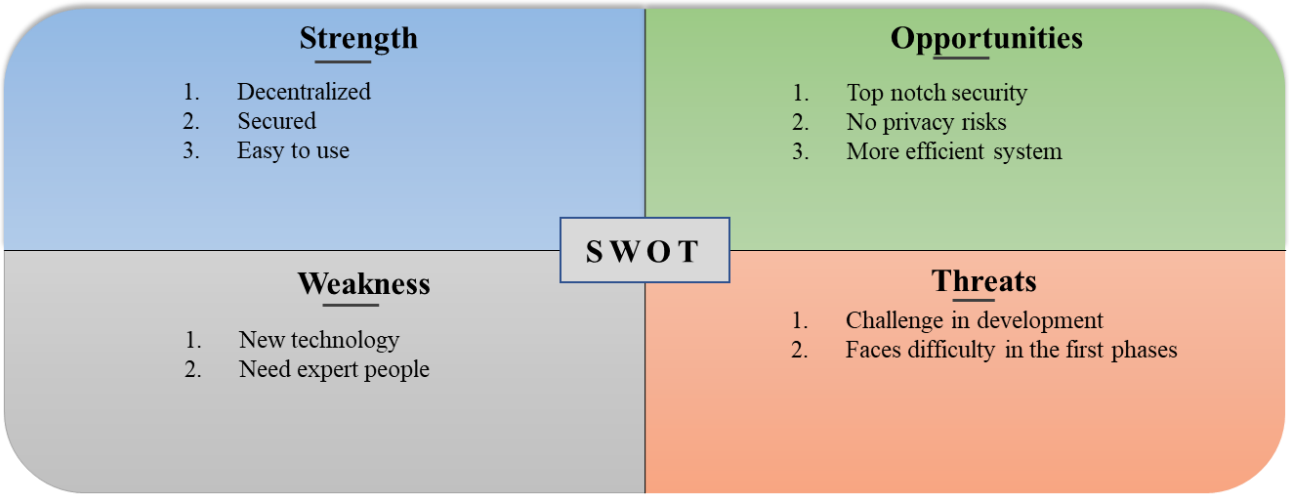


Fig: SWOT analysis

10. Cost Distribution

The cost for this initiative can be divided into two sections: the initial cost to set up the system and the maintenance cost to ensure smooth operation.

- **Initial cost:** The initial cost includes cost of making the whole blockchain and IPFS storage network and deployment server cost. This cost will be borne by the central bank. Other partners will have to pay a premium to the central bank in order to be added to the private blockchain network.
- **Maintenance cost:** As for the maintenance cost, it includes the cost of maintaining the server. Some developers and system admins will be in charge of maintaining it. The central bank will bear a certain percentage and the rest would be divided amongst the other partners. These partners will pay an annual fee according to the amount of resources allocated towards their needs and the number of their clientele in regards to the maintenance cost.

References:

[1] Uddin, Mueen & Memon, M & Memon, Irfana & Halepoto, Imtiaz & Memon, Jamshed & Abdelhaq, Maha & Alsaqour, Raed. (2021). Hyperledger Fabric Blockchain: Secure and Efficient Solution for Electronic Health Records. Computers, Materials and Continua. 68. 2377-2397.10.32604/cmc.2021.0 15354.

[2] Kapsoulis, Nikolaos & Psychas, Alexandros & Palaiokrassas, Georgios & Marinakis, Achilleas & Litke, Antonios & Varvarigou, Theodora. (2020). Know Your Customer (KYC) Implementation with Smart Contracts on a Privacy-Oriented Decentralized Architecture. Future Internet. 12. 41. 10.3390/fi12020041.

[3] Mamun, Abdullah & Kaiser, M. Shamim & Yousuf, Mohammad. (2020). Secure and Transparent KYC for Banking System Using IPFS and Blockchain Technology. 10.1109/TENSYMP50017.2020.9230987.

[4] Drgon, Matus & Georgiou, Lamprini & Kiayias, Aggelos. (2020). Robust KYC via Distributed Ledger Technology. 10.6084/m9.figshare.13301363.

[5] Thomson Reuters 2016 Know Your Customer Surveys Reveal Escalating Costs and Complexity. Available: https://www.thomsonreuters.com/en/press-releases/2016/may/thomson-reuters-2016-know-your-customer-surveys.html

[6] Global AML/KYC Spending Projected to Total $1.4 Billion in 2021, Rising 26% as Governments Continue to Focus on Fighting Financial Crime - New Burton-Taylor Report. Available: https://finance.yahoo.com/news/global-aml-kyc-spending-projected-185700931.html

[7] What is KYC and why does it matter? Available: https://plaid.com/resources/banking/what-is-kyc/

[8] Definitions and terminology related to crypto-economics, blockchain and distributed ledger technology. Available: https://smithandcrown.com/glossary/

[9] The problems with KYC: too long, inconsistent, security, no standard. Available: https://ctmfile.com/story/the-problems-with-kyc-too-long-inconsistent-security-no-standard

[10] Private IPFS network. Available: https://github.com/ipfs/go-ipfs/blob/master/docs/experimental-features.md#private-networks

[11] Hanafi, Jefrul & Prayudi, Yudi & Luthfi, Ahmad. (2021). IPFSChain: Interplanetary File System and Hyperledger Fabric Collaboration for Chain of Custody and Digital Evidence Management. International Journal of Computer Applications. 183. 24-31. 10.5120/ijca2021921808.

[12] Hegadekatti, Kartik & Yatish, Yatish. (2016). Roadmap for a Controlled Block Chain Architecture. SSRN Electronic Journal. 10.2139/ssrn.2822667.

[13] Bamidele, Awotunde & Ogundokun, Roseline & Misra, Sanjay & Adeniyi, Emmanuel & Sharma, Mayank Mohan. (2021). Blockchain-Based Framework for Secure Transaction in Mobile Banking Platform. 10.1007/978-3-030-73050-5_53.

[14] Amesar, Yash & Nerkar, Yash & Mali, Nitesh & Nitnaware, Ashwin & Prashant, Yawalkar. (2020). Decentralized Banking Application using Block chain Technology.

[15] KYC Blockchain Solutions Providers | Blockchain Enabled KYC Solution | Blockchain for Know Your Customer. (2021, October 12). Antier Solutions. Retrieved October 22, 2022, from https://www.antiersolutions.com/kyc-blockchain-solution/

[16] KYC-Chain - Blockchain & Banking KYC / AML Compliance Solution. (2019, September 11). KYC-Chain. Retrieved October 22, 2022, from https://kyc-chain.com/

[17] Hyperledger Fabric Security Threats: What to Look For. (2021, December 6). Hyperledger Foundation. Retrieved October 22, 2022, from https://www.hyperledger.org/blog/2021/11/18/hyperledger-fabric-security-threats-what-to-look-for