Introduction to Blockchain and its applications in managing SME

1st To Duc Anh
DSEB K61, NEU
DSLab
Hanoi, Vietnam
11196328@st.neu.edu.vn

2nd Le Tien bang
DSEB K61
Hanoi, Vietnam
11190703@st.neu.edu.vn

3rd Nguyen Minh Hue
DSEB K61
Hanoi, Vietnam
11192163@st.neu.edu.vn

4th Nguyen Thuy Linh
DSEB K61
Hanoi, Vietnam
11192981@st.neu.edu.vn

***Abstract*—This report is to introduce about the Blockchain technology, one of the building block of Web 3.**

***Index Terms*—Blockchain, Hashing, Business Rules.**

I. Report introduction

This report includes two big sections:

- Introduction to Blockchain and how it works.
- Coding demo with Blockchain applications in managing SME.

This report were made with introduction purposes, so all complicated mathematical have been cut-off for reader-friendly purposes. For detailed version, we recommend reading the book [1] and other references we have included in the end of the report. Reports and project contributions can be found in the end of the report as well

This report was written with the help of LaTeX.

II. Introduction to Blockchain

A. The existing problem

The current problems is that mainly everything is build on trust (or faith if you prefer). This problems make transactions harder than it should be. Since how can some person trust some other one when they dont know each other? So, how to quickly process transactions for everybody, with the information is that the transactions possibly involving multiple people, in an environment might not always honest, where people may not trust one another?

Our society has hundreds, or thousands of years been relying on the intermediaries to solve that problem. If we do not trust each other, we will do the transaction through a trusted middleman in this situation: a trusted-by-many-financial-institution, namely, the Bank! But the trust placed in intermediaries is an assumed trust; we assume that they will do what they are supposed to do in the ideal scenario, which is not the case in practice; humans make mistakes, machines fail all the time, hackers are always looking for ways to penetrate into the systems, and even in an ideal world where such errors or attacks do not occur, the conventional way of relying on a central authority to store information, process transactions, or manage systems is flawed. It is increasingly expensive in both money and time when there are more workload. [1].

B. Blockchain comes in!

Blockchain overcomes the weakness of the current centralized systems. It is comprehensively decentralized with absolutely no intermediaries involved. Blockchain approach the problem with four imperative aspects:

- Trust
- Security
- Privacy
- Transparency

C. The way Blockchain provides the solution for the unanswered problem.

Blockchain can provide the above desirable properties thanks to its decentralized network architecture, utilizing by many computers on the internet. These computers can continuously store and process transactions in a autonomous mechanism, but also, they are still able to achieve agreement in decision making and be guarded against hacker attacks, system malfunctions, users dishonesty, and self interests.

Using the internet, one may send data from one computer to another without having to worry about how the data will be delivered or if it will be lost. The internet and Blockchain system manage all of that, enabling individuals to focus on the important business work at hand. Similarly, when people transact on the blockchain, they don't have to worry about a slew of what-if scenarios, such as whether the other party will act as agreed, if money will be lost, or data will be illegally altered.

D. Why Blockchain named "Blockchain"

The term blockchain refers to how transactions in a digital ledger are stored in blocks. Each new block is attached to the one before it, forming a chain, hence the name blockchain. Two subsequent blocks in the Blockchain are mathematically linked in such a way that any change to an existing block contradicts the mathematics of the connection with the following block.

E. Economics contribution projection

According to 2020 Global Blockchain survey by Deloitte, interviewing roughly 1500 business managers from 14 countries with certain knowledge about this new technology shows that 39% business applied Blockchain into their business, 55% considered Blockchain lies into their top 5 piority, and 82 % would hire Blockchain specialists in the next year [2]. Another report by PwC have projected the contribution of Blockchain will add 1.76 trillion USD to the GDP globally, besides 40 million new jobs created, and be used to support 10-20% global business infrastructure [3].

III. What is Blockchain

For tech-illiterate, Blockchain is a computing technology that allows transaction recording and processing that is safe, transparent, and trust-less.

For Pro-tech, Blockchain can be defined as a decentralized computing system is made up of five basic components:

- Decentralized networking
- Mathematical cryptography
- Distributed consensus mechanism
- A digital transaction ledger
- Smart contracts

The 5 fundamental components can be seen as below:

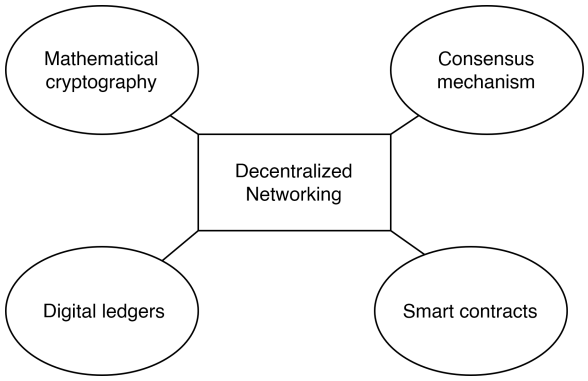


Fig. 1. 5 fundamental components of Blockchain

IV. Decentralized networking

Currently, there are two methods of storing blockchain data:

- Data are store in a centralized server (which might be unsafe, error prone).
- Blockchain Vision (decentralized) Advantages:
  - Many computers can join, automatically connect and coordinate to process transactions
  - Peer-to-peer model: avoid concentration of influence

A. Peer-to-peer (P2P) systems

Distributed system is decentralized without any sole player's control, where each node a peer (or have equal power, vote,...) to each other, being equivalent in functionality.

There are 2 types of P2P system:

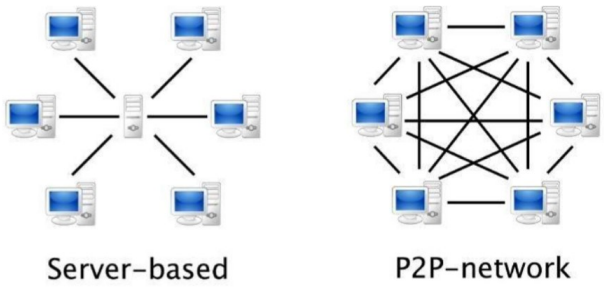


Fig. 2. Network types

- Unstructured P2P network: This network facilitates the users with an easy connection with other devices. In this kind of P2P network, all systems play their role, but clients suffer to find the rare content as there is a lack of structure. Moreover, this kind of network also has expensive communication cost.

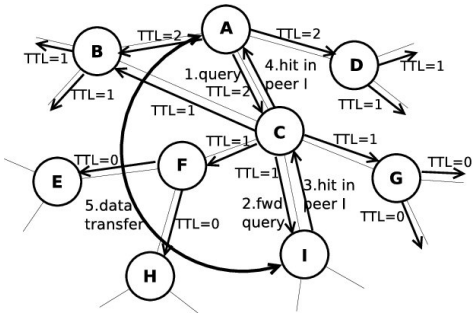


Fig. 3. Gnutella Unstructured P2P network

- Structured P2P network: efines a graph structure (topology) to connect the nodes and each node can only communicate with its neighbors. This network has some properties like:
  - Good network **traffic**: Less messages transmitted, avoid flooding
  - **Quick search**: Finding a node should take a short route
  - **Light** node overhead: Few neighbors to keep track of
  - **Fairness**: nodes should carry equal workload

B. Decentralized Networking Topology

*1) **Regular Graph**:* Every node has exactly the same constant number of neighbors.

Advantage: Very fair, and low overhead because the number of neighbors each node has is constant, independent of the graph size.

Disadvantage: Very long routing distance node-to-node routing path is $O(n)$ linear with the network size.

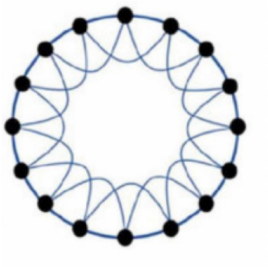


Fig. 4. Regular Graph

*2) **(Uniformly) Random Topology**:* The neighbor-relationship is equally random: probability of connecting two nodes is the same p for every edge.

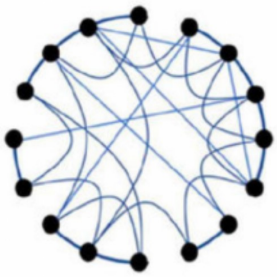


Fig. 5. Random Graph

Advantage: Short routing distance. Average node-to-node routing may take a path that is $O(log(n))$ nodes long.
Disadvantage: nodes have different degrees leads to unfair workloads.

*3) **Small World Graph**:* A node has most neighbors nearby and few long-distance neighbors

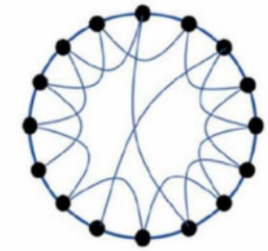


Fig. 6. Small World Graph

Advantage: Very short routing distance. Average node-to-node routing takes a path that is $O(loglog(n))$ nodes long.

Pay attention that:

- Most real-world networks are small-world
- Small-World is the combination of Regular and Random Graph
- Reduce node unfairness (better than Random)
- Keep short routing distance (better than Regular)

$\rightarrow$ Small-World seems like the best graph topology

*4) **Scale-Free Graph**:* Degree distribution follows a power law: the fraction P(k) of nodes having k neighbors is proportional to a power of k: $P(k) \approx \frac{1}{k^c}$

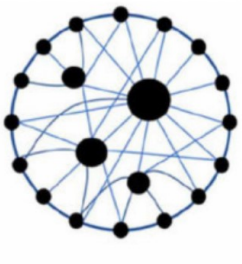


Fig. 7. Scale Free Graph

Advantage: This grapth has a very short routing node-to-node distance. Average traveling routing between node has the time complexity of $O(loglog(n))$ nodes long.

Disadvantage: The minority of node have the majority of neighbor. The majority of nodes have very few neighbors $\rightarrow$ this create extremely unfair workload situation among the nodes.

C. Chord P2P Network

Chord is a protocol and algorithm for a peer-to-peer distributed hash table. A distributed hash table stores key-value pairs by assigning keys to different computers (known as "nodes"); a node will store the values for all the keys for which it is responsible. [5]

1) The traditional query: The Chord protocol is used to find successors by inquiring about a key from a client (often a node) (k). If a node is unable to find the key locally, the query is sent to the node's successor. This gives an O(N) query time, where N is the number of computers in the ring.

D. The Chord way

Chord avoids the linear search mentioned above by requiring each node to keep a finger table with up to m entries, where m is the number of bits in the hash key. The i_th entry of node n will contain $((n + 2^{i-1}) mod 2^m)$ successors. The foremost entry of the table will be the node's direct successor. If a node wants to find some key k, it will pass that query to the closest predecessor or successor to check whether k is in its finger table until a node contains k or the immediate successor contains k or terminate if the whole network does not contains k.

With the help of finger table, the time complexity of finding the key k in an N-Node Network is $O(log(n))$ [5].

V. Mathematical Cryptography

A. Structure of the Chain

By default, the Blockchain ledger follows a chain topology, which is employed in all but a few disputed Blockchain

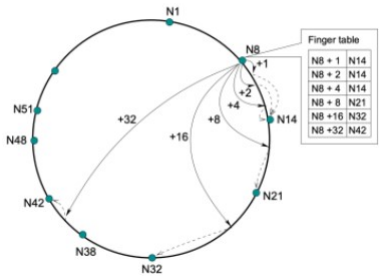


Fig. 8. Chord and Finger Table

designs.

The data is organized in a sequence of data blocks designated b1, b2, b3,... When new transactions need to be saved, they are placed in a new block that is added to the end of the existing chain.

Each block header has two important elements, which allow the Blockchain to store the transaction data, the chain state if applicable, and necessary:

- Block ID: this is set to the hash value of the block content b_i using a cryptographic hash function, this hash function is predefined and publicly known.
- Previous hash: This is the ID of the previous block b_{i-1}, which block b_i is appended to.

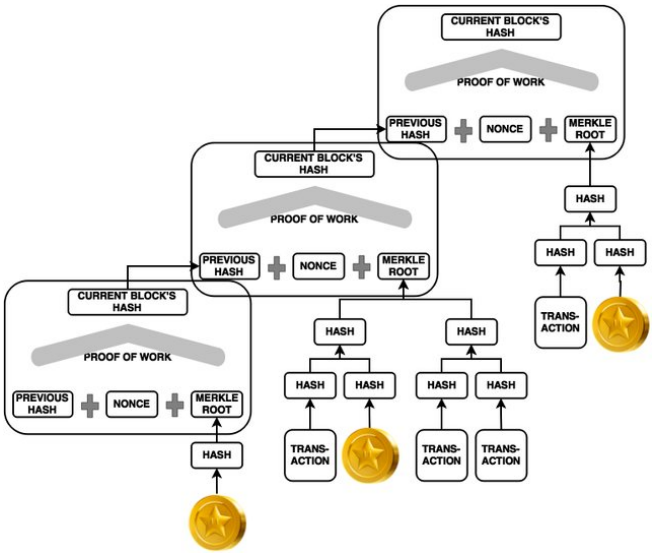


Fig. 9. Simple Blockchain structure information

It is noted that the Block ID may not necessarily be stored inside the block header since it can be re-calculated from the blocks content. The previous hash information is very important in maintaining the data integrity of the chain. Since if there are any part of any block is modified after it is recorded in the Blockchain, this will be detected. This is because for a new block to be added to the Blockchain it must pass a procedure called block validation. a new block b_{i+1} valid if and only if:

- The previous hash is consistent, or $Previous - hash(b_{i+1}) = Hash(b_i)$
- All the transactions in b_{i+1} are valid
- Previous block b_i is valid

VI. Use of Cryptography

It is now clear that the Blockchain's data immutability is achieved by the use of previous hash information linking subsequent blocks in the Blockchain. However, a hash function can have different input values and create the same hash output, which implies that block b_{i-1} can be changed while keeping its hash value, $Hash(b_{i-1})$, which equals the previous hash value b_{i-1} recorded in block b_i. If this happens, all the magical data immutability will be gone to ashes. So that, the hash function that are used in Blockchain must be a cryptographic hash function not any arbitrary hash function.

A cryptographic hash function is a hash function with three properties:

- **Collision resistant**: it is infeasible to find different input messages x and y such that $Hash(x) = Hash(y)$.
- **Hiding**: given the output $c = Hash(x)$, is infeasible to find an input x.
- **Puzzle friendly**: If we know the hash value $c = Hash(r||x)$ an input message made by concatenation of r and x and even if we know part of the input, x we cannot reconstruct the remaining input r in time complexity faster than 2^n where n is the binary length of output [1].

The solution above is called "Commitment Scheme" in cryptography [7]. The secret r generated by the node must come from a large number space. If the binary length n is small, it will only take a little time for the attacker to try all possible values of r and combine with x to determine which combination might find out r or x that satisfies $c = Hash(r||x)$. When such combination is found, attackers can deceive individuals by telling them that the outcome is the inverse value of x obtained when n is small. Despite the fact that x might has only a few possible values, the attackers will not be able to to reconstruct the secret x when r is sample from a large number space owing to Hash's puzzle-friendly nature as a cryptographic hash function.

The above is an example of how mathematics cryptography may assist creating a trust-less system.

VII. Distributed consensus mechanism

Consensus is a computing research subject that has been studied for more than SP years before bitcoin became famous. It began in the 1970s with the NASA-backed project, software integrated fault tolerance (SIFT), which attempted to construct a durable flight control system [8].The task was to replicate the system on numerous machines such that the entire system could withstand multimachine failures.

The work by Lamport et al. in 1982 has developed the term "Byzantine Fault" to describe a state in a distributed system in which certain nodes are untrustworthy and may look arbitrarily normal or malignant, and conspire with one another such that there is no consistent information for the other nodes to announce their dysfunction. It formulated this task as the "Byzantine Generals Problem" (BGP).

Byzantine Fault Tolerance (BFT) system must avoid complete failure and for that the nodes must agree on a concerted strategy and live by this consensus, knowing that some nodes may fail or act maliciously. BGP laid the foundation for research in distributed consensus. companies like Google and Facebook started adopting scientific results in BFT consensus for mission critical services such as Google wallet and Facebook credit.

But, the consensus system made by Google and Facebook is not entirely decentralized nor permissionless since it still has the nodes in the system belong to these corporations.

Consider a broadcast system of nodes in which a sender node must broadcast a message (value, for example, "Hello World") to all nodes in a peer-to-peer way. a The sender gets an input value m at the start of the broadcast protocol, and each node i must output a value m_i at the conclusion. The sender and receivers may be honest or dishonest, and this protocol achieves BFT if it meets two conditions.

- **Consistency**: All honest nodes i and j must output the same value:Z $m_i = m_j$.
- **Validity**: If the sender is honest, all honest nodes o must output value $m_i = m$.

A system might be consistent but not valid, and otherwise. All the honest node might output the similar value but that output value might be completely different from the original message. The sender might be dishonest and honest nodes might output different value from the dishonest node original message. Due to the above reasons, both of the **requirements must be satisfied**.

Blockchain is a type of BFT system. To overcome discrepancies caused by blockchain nodes' autonomous and independent operation, the traditional method is for each node to agree on the consensus that the longest blockchain copy, the one with the most blocks, is the globally valid version. Because blockchain copies that are shorter than the correct blockchain are not used, nodes want to keep their copies as current as possible to avoid wasting time adding blocks to the erroneous blockchain.

Toward these different consensus mechanisms have been used for Blockchain. Major among them are the methods of practical Byzantine Fault Tolerance (PBFT) [11]. Namely Proof of Work (PoW) [10], Proof of Stake (PoS) [12].

VIII. A DIGITAL TRANSACTION LEDGER

A distributed ledger is a database that is consensually shared and synced by numerous persons and locations. It enables transactions to have public "witnesses." Each network participant can view and own an identical copy of the recordings shared throughout the network. Any modifications or additions to the ledger are reflected and duplicated to all participants in a couple of seconds or minutes. [14]

A. Understanding Distributed Ledgers

Since ancient times, ledgers have been at the heart of economic activities, recording contracts, payments, buy-sell transactions, and the transfer of goods or property. The journey that began with recording on clay tablets or papyrus took a big leap ahead with the invention of paper.

Computers have greatly facilitated and accelerated the process of record-keeping and ledger maintenance during the previous few decades. With the advancement of technology, information saved on computers is becoming cryptographically secure, quick, and decentralized. Companies may use this technology in a variety of ways, one of which is through distributed ledgers.

The ledger's information is safely and precisely recorded using cryptography and may be retrieved via keys and cryptographic signatures. When information is saved, it becomes an immutable database governed by network rules.

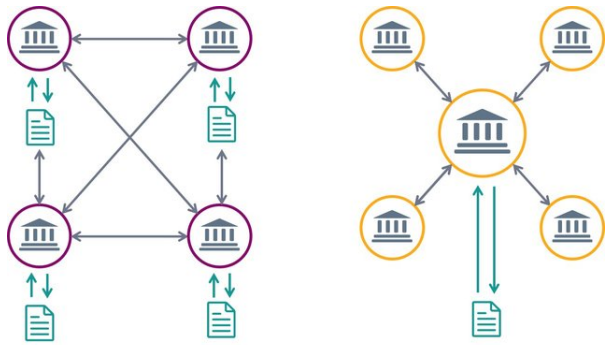
B. Advantages of Distributed Ledgers

While centralized ledgers are vulnerable to cyber attacks, distributed ledgers are inherently more difficult to attack since an attack must target all distributed versions at the same time to be successful. These records are also resistant to malicious changes performed by a single individual. Because distributed ledgers are difficult to manipulate and attack, they provide high transparency.

Distributed ledgers also reduce operational inefficiencies, shorten the time it takes to complete a transaction, and are automated, so they work around the clock, cutting total costs for firms that use them.

Distributed ledgers also allow for a smooth flow of information, making an audit trail easier to follow for accountants doing financial statement checks. This reduces the potential of fraud developing on a company's financial accounts. The decrease in paper consumption is also beneficial to the environment.

A distributed ledger is a ledger of any transactions or contracts that is maintained dispersed among multiple places and people, removing the need for a central authority to keep a check on manipulation. A central authority is not required to authorize or authenticate any transactions in this manner.



(a) (b)

Fig. 10. Distributed Ledgers (a) [15]

IX. Smart contracts

Smart contracts are programs that store on a certain Blockchain to run a particular jobs when a predetermined conditions are met. They are used to automate an agreement execution, and the outcomes can be delivered immediately to all participants in the Blockchain, without any time loss. They can also automate workflow or triggering some actions when conditions are met.

A distributed ledger differs from a centralized ledger, which is the type of ledger used by the majority of businesses. Because it provides a single point of failure, a centralized ledger is more vulnerable to cyber assaults and fraud. [13]

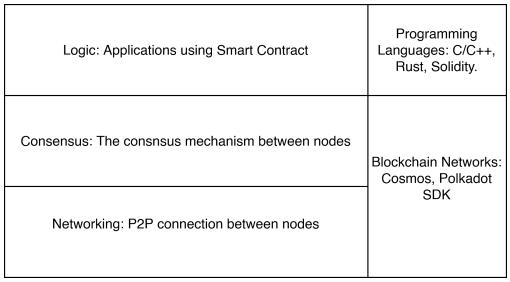


Fig. 11. Blockchain High-level Structure

Smart contracts are written using a high level programming language eg: Solidity, Viper, Flint, Bamboo... Solidity is the most popular language for smart contract networks. It is Turing complete, meaning that it can simulate any computation. [1]

A. How Smart Contracts work

Smart contracts operate by executing basic conditional assertions encoded into Blockchain code. When preset circumstances are met and validated, a network of computers conducts the activities. These activities might include transferring payments to the proper parties, registering a vehicle, providing alerts, or issuing a ticket. When the transaction is completed, the Blockchain is updated. This implies that the transaction cannot be modified, and the results are only visible to persons who have been granted permission.

A smart contract can have as many specifications as necessary to reassure the participants that the work will be executed correctly. Participants must identify how transactions and associated data are represented on the Blockchain, agree on the conditions of the conditional statement rules that govern those transactions, investigate all conceivable exceptions, and design a framework for resolving disputes in order to set the terms.

After agreement have been reached, developers can then build the smart contract. Companies that use Blockchain for business are increasingly giving templates, web interfaces, and other online tools to help with smart contract creation, so that creating rules or conditions are now more accessible.

B. Benefits of Smart Contracts

Smart Contracts offer some benefits as follow:

- Speed, efficiency and accuracy
- Trust and transparency
- Security
- Savings

1) ***Speed, efficiency and accuracy****:* Once a condition is met, the contract is executed immediately. Because smart contracts are digital and automated, there's no paperwork to process and no time spent reconciling errors that often result from manually filling in documents [13].

2) ***Trust and transparency****:* Because there's no third party involved, and because encrypted records of transactions are shared across participants, there's no need to question whether information has been altered for personal benefit [13].

3) ***Security****:* Blockchain transaction records are encrypted, which makes them very hard to be hacked. Moreover, since each record is connected to the previous and subsequent records on a distributed ledger, hackers would have to alter the entire chain to change a single record. [13].

4) ***Savings****:* Smart contracts remove the need for intermediaries to handle transactions and, by extension, their associated time delays and fees. [13].

X. Blockchain applications

A. Blockchain applications in economy

With several remarkable achievements along the development process, it is difficult to dispute that the advent of Blockchain technology has triggered a revolution in many areas throughout the world. When Blockchain technology is used, the economic sector, in particular, benefits the most from innovation. Because of its decentralized character, this technology has done what previous databases have not: it has solved the transaction trust problem. All information in transactions using Blockchain technology will not be controlled by a single party, removing the need for a third party and allowing transactions between strangers to take place smoothly under the absolute trust of both parties without the need for a reputable company or organization to act as an intermediary. As a result

of this feature, financial and banking services are substantially simplified, as is convenience. These sectors are infamous for being unnecessarily difficult and time-consuming, since we will have to go through a variety of legacy systems to validate payments. Blockchain has simplified complex processes by allowing anybody to access financial and banking services regardless of geography, nation, or internet connection. Blockchain-based cryptocurrencies will allow you to move money around the world almost immediately and at a minimal cost, removing the previously existing problem of time and expensive remittance fees. When all transactions were done on paper and there was the risk of them being altered from white to black, the economy faced several challenges with the equitable allocation of goods. The immutability of Blockchain information has altered the face of this problem, allowing us to democratize the process of splitting assets, guaranteeing that they are properly benefited, contracts, or papers. Legal papers will be maintained on Blockchain to prevent arbitrary changes or theft, resulting in economic stability.

A lot of nations are presently researching and developing Blockchain technology. It is considered as a solution to minimize inter-banks transaction costs and clearing times while simultaneously establishing a more secure system. Many financial organizations have created partnerships to commercialize Blockchain technology, such as the R3 alliance of the three major Australian banks, including Westpac, Commonwealth, and NAB, as well as 40 banks and many more financial institutions globally. In June, CNBC reported that IBM is developing Blockchain technology specifically for the seven top European banks (including HSBC, KBC, Deutsche Bank, Natixis, Unicredit Societe Generale, and Rabobank) to boost transaction efficiency. Small and medium-sized firms can benefit from international trading. Blockchain technology development provides the framework for the growth of cryptocurrencies... These cryptocurrencies are seen as a danger to fiat money since they diminish our dependency on it. Despite the fact that many governments have not actively accepted this currency, many nations are eager to accept cryptocurrencies, and the UK central bank has stated that cryptocurrencies are a part of the financial revolution. Wire transfers have spread over all fields due to their ease. Furthermore, virtual currency exchanges allow anybody to invest through them, hence enabling the development of new sectors based on virtual currency investment.

B. Blockchain applications in Vietnam

According to Infinity Blockchain Lab data, Blockchain technology will be used mostly in financial services (more than 83 percent), supply chain (40 percent), public services (30 percent), energy (30 percent), education (30 percent), and many other industries in Vietnam by the end of 2021.Blockchain is being utilized in the supply chain, and it is particularly effective in the sector of food traceability (smart contracts) and is yielding beneficial results, such as in the My Xuong cooperative or the An An enterprise. Although the results were not reproduced on a wide scale, they serve as a foundation for using Blockchain to enhance Vietnamese agricultural goods. The most common use of Blockchain is in banking, and our nation has an edge in gaining successes in this field since many significant financial organizations have been functioning in Vietnam for many years. The include Big 4 auditing firms, banks such as Standard Chartered, Sumitomo Mitsui Banking, and MUFG Bank, among others. These enterprises, with the assistance of their foreign parent firms, are prepared to deliver Blockchain-integrated services in Vietnam as soon as this technology demonstrates its capabilities throughout the world. Furthermore, Vietcombank has begun to use Blockchain in digital banking. Blockchain, together with AI and the Internet of Things, can be considered the foundation of future technological industries. Viettel has created a Blockchain application for personal health management records, which contains a comprehensive information storage history of people's medical tests and treatments.

Vietnam has recently emerged the Blockchain and NFT phenomenon from the world. Many Vietnamese-founded and meticulously implemented Blockchain initiatives and those companies have become worldwide phenomenon, namely Kyber Network, Tomochain, Coin98, and Axie Infinity. These projects contribute to Vietnam's increased prominence on the global Blockchain map. Businesses, however, have yet to make major expenditures in the development of this technology. As a result, the implementation of this new technology will rely significantly on Blockchain start-ups and international firms with investment money in our country.

To keep up with the global development trend, Vietnam established the Blockchain Association on April 27, 2022. The formation of the association will create conditions for businesses and individuals interested in blockchain technology to use it more widely and efficiently. Through research and the application of blockchain technology solutions in the economy, the goal is to help promote the digital economy and soon bring Vietnam to an international level in the digital economy.

C. Benefits of using Blockchain in Business

Because of the remarkable benefits of immutability, transparency, and security, global businesses have aggressively implemented Blockchain into their processes. According to a Deloitte survey of 1,386 organizations in 12 countries, the percentage of businesses that have a good attitude and strongly support the use of Blockchain technology in their operations is quite high [18]. It is also proving its worth and significance in the industry. This article will highlight some of the benefits of adopting Blockchain in business.

1) ***In supply chain management****:* Supply chain management, notably the lack of transparency between suppliers and

brokers, is a big challenge for organizations today. With the emergence of the Blockchain distributed ledger, this problem is readily overcome. It allows numerous users to view the same database at the same time, boosting openness across the supply chain. [19]. The data in the Blockchain ledger cannot be modified, and all legitimate transactions are timestamped, making it much easier to identify and prevent theft. The application of Blockchain in supply chain management assists organizations in improving regulatory compliance, reducing "paperwork," and considerably lowering expenses. Food Trust, a Blockchain-based application that applies Blockchain to the supply chain, was recently released by IBM. Before releasing this product, IBM cooperated with Walmart to regulate and trace the flow of food from the field to the store. Blockchain will also help customers authenticate product information, allowing them to evaluate whether or not the information written on the package is correct. Blockchain technology can also be employed in the pharmaceutical business, where customers are particularly worried about medicine quality and provenance.

*2) **Customer information security**:* One of the most urgent concerns affecting the internet community today is the theft and impersonation of online users, especially considering that technical developments have not alleviate the problem. Users may use Blockchain technology to securely store personal information with an immutable feature without worry of it being stolen or modified. Because of the decentralized network, the data on the Blockchain is practically resistant to cyber assaults. Admin IDs can make it easier for users to access their data. Civic's secure identification platform supports multi-factor bio-metric authentication of usernames and passwords. The user must go through an identity verification procedure before an ID can be produced, and this ID can be used to confirm transactions by governments or banks. Furthermore, this ID can be Furthermore, this ID may be used to store user data on the Blockchain system, such as social media account information or medical records. These digital IDs not only let consumers authenticate online transactions safely and promptly, but also prevent persons and organizations from unlawfully gathering and benefitting from personal information.

*3) **Cost-effective and convenient solution for data sharing**:* For millennia, business data has been stored on paper. Data is housed in distinct, independent data silos, restricting the ability to share data in real time. Documents are securely stored on a distributed ledger on blockchain, making them easy to access when needed, but only authorized stakeholders have access to them. In the insurance business, for example, Blockchain is devoted to the confidentiality of people' health information. Organizations can use timestamps to check people's medical history. This data may then be securely shared across doctors, hospitals, pharmacies, and health insurance, giving individuals control over their information.

*4) **Management of candidate profiles in enterprises**:* Business administrators are interested in blockchain technology and its application to human resource management. In this situation, human resource recruiting is viewed as a commercial activity that affects the quality of products and services provided to consumers. Many studies, however, suggest that application fraud is commonly employed to offer applicants an edge [20]. Because of the characteristics of Blockchain, using this technology to verify candidate profile information can brings many benefits to businesses, namely:

- With the abilities of Blockchain technology, employers may obtain reliable information about candidates. A Blockchain is made up of connected blocks that are decentralized stored. To add a new block to the Blockchain, all network objects must agree. As a consequence, when the candidate wishes to include related information, all participants in the system must agree on and verify it. Furthermore, the Blockchain's cryptographic mechanisms restrict system users from modifying the information on the blocks. Any changes to the chain are documented and immediately accessible. This assures the correctness, integrity, and transparency of the documents. [21].
- The usage of Blockchain technology to validate applicant profile information decreases the HR department's job. The usage of Blockchain technology will allow organizations's human resource departments to focus on other operations that demand thought, such as establishing recruiting strategy, assessing prospects, and so on.
- Additionally, this technology helps employers make smarter judgments. One of the most significant methods for employers to make educated judgments is to have access to reliable candidate information. Accurate information also assists firms in placing human resources in appropriate places and organizing appropriate skill training for human resources.

XI. The downsides of Blockchain

Scalability is one of Blockchain's main flaws, owing to the fixed size of the block for storing information. Since sometimes, the block size is only 1 MB, it can only carry a few transactions on a single block.

Another disadvantage of Blockchain is that it is only a couple of years old technology, so people do not have much confidence in it and are not ready to invest in it. Despite the fact that several applications of Blockchain are doing well in various industries, it still needs to gain the trust of even more people in order to be recognized for its full utilization.

Besides, Blockchain might not completely secure [22]:

- **51% attack**: In the 51 percent attack, if an entity controls 51% or more of the network nodes, it can gain control of the network. They can change the data in the ledger and double-spend as a result. This is conceivable on networks where miners or nodes may be controlled. This indicates that private networks are more likely to be immune to 51% attack, but public networks are more vulnerable..
- **Double-spending**: Another issue with current Blockchain technology is double-spending. To avoid double-spending, the Blockchain network employs several consensus methods like as Proof-of-Stake and Proof-of-

Work. Only networks vulnerable to the 51 percent attack can double their spending.

- **DDoS's attack**: A DDoS assault bombards the nodes with identical requests, clogging the network and knocking it down.
- **Cryptographic cracking**: Another reason blockchain technology is insecure is the encryption solution it employs. Quantum algorithms and computation are more than capable of breaching encryption systems. Blockchain systems, on the other hand, are increasingly including quantum-proof encryption methods..

Besides, There are some other disadvantages such as:

- **Time-Consuming**: To add the next block to the chain, miners must compute nonce values multiple times, which is a time-consuming operation that must be sped up before it can be utilized for industrial purposes.
- **Storage**: The fact that Blockchain databases are maintained on all network nodes causes a storage problem; as the number of transactions increases, so will the amount of storage required.

XII. Blockchain and Data Science

Data is at the heart of both of these technologies (Blockchain and Data Science). While Blockchain validates and records data, Data Science is concerned with obtaining valuable insights from data in order to solve problems. Both of these systems use algorithms to manage interactions with various data segments. As a result, Data Science is for forecasting and Blockchain is for data validation.

With their own set of advantages and disadvantages, Blockchain and Data Science can be a potent combo for efficiently managing data quantity and quality. Other Blockchain advancements and maturity will enable the study of more use cases, including Data Science.

The combination provides [23]:

- **Data traceability**: Blockchain facilitates peer-to-peer collaboration. For instance, if a published account fails to sufficiently describe a technique, any reviewer can examine the entire process and establish how the findings were obtained.
  Anyone can understand if data is correct to use, how to preserve it, how to update it, where it originates from, and how to use it appropriately thanks to the ledger's open channels. Finally, Blockchain technology will allow users to track data from start to finish.
- **Real-time analysis**: Real-time data analysis is extremely tough. The ability to monitor changes in real time is regarded as the most effective method of spotting fraudsters. However, real-time analysis was previously impossible. Companies can now discover any irregularities in the database from the outset, owing to Blockchain's distributed nature.
- **Data accuracy**: The data in Blockchain's digital log is stored on both private and public nodes. Before being added to subsequent blocks, the data is cross-checked and reviewed at the entry point. This process is a way of data verification in and of itself.
- **Data integrity**: Data scientists are using blockchain technology to assure the validity and traceability of data at every step along the chain. Its unchanging security is one of the reasons for its extensive use. The decentralized ledger of Blockchain protects data at every stage. Before anybody may access the data, the precise signatures must be supplied. Data hacking and leakage will be greatly minimized as a consequence.

Moreover, Blockchain has some the security features that is invaluable to Data Science:

- **Encoded transactions**: Blockchain encrypts every transaction that occurs in its ledger using complicated mathematical methods. These transactions are executed between parties as irreversible and unchangeable digital contracts.
- **Data lakes**: Data scientists typically utilize data lakes to maintain track of the information in their companies. Each block is allocated a unique cryptographic key when using Blockchain to monitor the provenance of data. This guarantees that everyone who uses the data gets the right key from the inventor, demonstrating that the data is valid, of high quality, and legitimate.

XIII. Introduction to Blockchain application in managing SME

In this later part of the report will be our proposal for a proof-of-concept project that implemented a system to help a small convenience store or convenience chain in managing the store, using some simplified version of the Blockchain's four aspects. We implemented a system that allows the managers or the shareholders to directly view the transactions that were made, with complete trust. This proof of concept system also helps SMEs to reduce the operation costs by giving them a simple tool that automatically does the accounting to eliminate the cost of hiring accountants. The system can provide data for the authorities to help them do taxes and other regulations without the need of humans.

The proof-of-concept system was built with the help from Flask and MongoDB (via Pymongo) with Python Programming Language.

XIV. Used theories

A. Python

Python is a high-level, interpreted programming language written by Guido van Rossum, a Dutch computer scientist, released to the public in 1991.

Python is a general-purpose language, so it has a lot of applications, namely:

- Web development, both backend with Flask, Django,.. and frontend with Py Script.

- Game and software development with the help of Pygame or Tkinter framework.
- Artificial Intelligence, with numerous supported frameworks like Tensorflow, Keras, Pytorch,...
- And many more.

Python can run from general different operating system(OS), namely:

- Windows
- MacOS
- Linux-based
- ...

Python was originally built from C, a low-level, compiled programming language with blazing speed. To the current day, there a many implementation of Python, [16] namely:

- **CPython**: The Python's original and most-maintained implementation, developed in C. New language features are usually introduced here initially.
- **Jython**: The Python version that is written in Java. This implementation may be used to construct programs utilizing the Java class libraries or as a scripting language for Java applications. It's also frequently used to write tests for Java libraries.
- **Python for .NET**: This implementation utilizes the CPython implementation, but it is a managed.NET application that includes.NET libraries. Brian Lloyd is the father of this idea.
- **IronPython**: A Python alternative for.NET. Unlike Python.NET, this is a full Python implementation that produces IL and immediately compiles Python code to.NET assemblies. Jim Hugunin, the original creator of Jython, designed it.
- **PyPy**: Python implementation built entirely in Python. It has some sophisticated capabilities that other implementations do not, such as stackless support and a Just in Time compiler. One of the project's aims is to encourage linguistic experimentation by making it easy to tweak the interpreter (since it is written in Python).

Each of these implementations deviates from the language as specified in the Python documentation in some manner, or includes additional information not contained in the Python documentation itself.

B. Flask

Flask is a web framework, which means it offers users with tools, frameworks, and technologies to build online web applications. This online program might be as basic as a few web pages, a blog, or a wiki, or as complicated as a web-based calendar software or a commercial website.

Flask is a sort of micro-framework. Micro-frameworks are frequently frameworks that rely on few or no external libraries. This has both benefits and drawbacks. The framework is lightweight, with minimal dependencies to update and scan for security flaws, but you will have to do more work yourself at times or expand your list of dependencies by adding plugins.

Flask depend on two framework:

- Werkzeug: a WSGI utility library. [17]
- jinja2: A template engine

C. MongoDB

MongoDB, the most popular NoSQL database, is an open source document-oriented database. 'NoSQL' is an abbreviation for 'non-relational.' It denotes that MongoDB does not employ a table-like relational database structure like SQL or MySQL , but rather an entirely distinct system for data storage and retrieval. BSON is the name of this storage format (similar to JSON format).

Feature of MongoDB:

- **Document Oriented**: Unlike RDBMS, MongoDB preserves the primary subject in a limited number of documents rather than breaking it up into numerous relational structures. For example, it stores all of a computer's information in an one document called Computer rather than in distinct relational structures like CPU, RAM, Hard disk, and so on.
- **Indexing**: Without indexing, a database would have to scan every document in a collection to find those that fit the query, which would be time-consuming. As a result, indexing is critical for fast searching, and MongoDB uses it to examine enormous volumes of data at breakneck speed.
- **Scalability**: MongoDB grows horizontally via sharding (partitioning data across various servers). The shard key is used to split data into data pieces, which are then uniformly dispersed among shards spread across several physical servers. A functioning database can also be expanded with new computers.
- **Replication and High Availability**: MongoDB improves data availability by replicating data across numerous servers. It safeguards the database from hardware failures by providing redundancy. If one server fails, the data may be quickly accessed from other active servers that also have the data saved on them.
- **Aggregation**: MongoDB procedures process data records and return the calculated results. It's analogous to SQL's GROUPBY clause. Some aggregating expressions are sum, average, min, and max.etc

XV. Dataflow in the project

The full dataflow can be found at: This repository

(The following pipeline is the simplified version)

A. Customers

The system's main tasks for customers is to let them register and change their forgotten information. These task can be done as follow:

- **Register**: The customer will be ask their name, phone number and address. The input will be check in the database. The phone number is the unique identifier of a

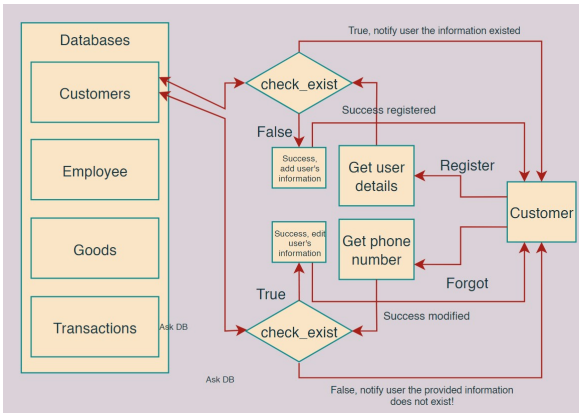


Fig. 12. Customers pipeline

customer. If the phone number exists in the database, that means the customer is currently a customer of the store, return the status to the customer. If the customer is not a customer, add the customer information to the database and prompt the successful status to them.

- **Forgot or change**: if a customer forgot or they changed their information, they can input their new or want to change information to the system. The system should look if that customer information exists. If they existed, change their information to new information, otherwise, tell them that they are not yet a customer.

B. Employee

The system's main tasks for employee is to let them login and logout from the system. Only technicians or owners can add or remove employee information. Only one employee can login to the system at one. These task can be done as follow:

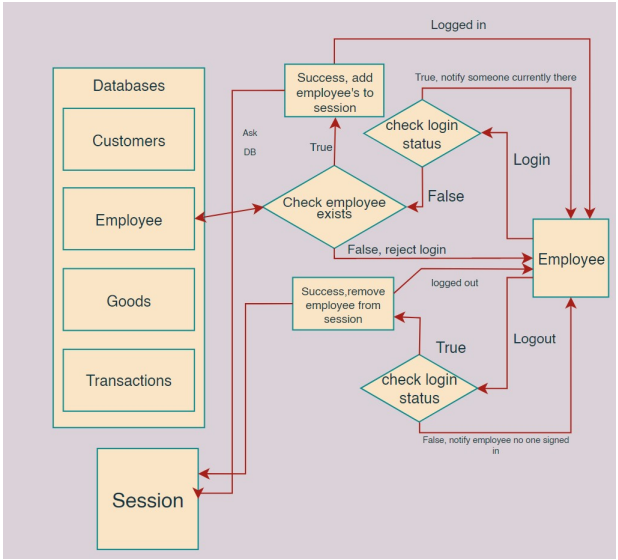


Fig. 13. Employee pipeline

- **Login**: Check login status, if yes, tell employee that someone is in. if false, check if employee information is in the database, if not, reject login, if yes, add them to Session info, prompt that they have successfully login.
- **Logout**: Check login status, if not, tell them that no one is logged in. If yes, remove current information from the Session, tell them that they have logged out.

C. Querying data

Querying data can only be done by owner or employee.These task can be done as follow:

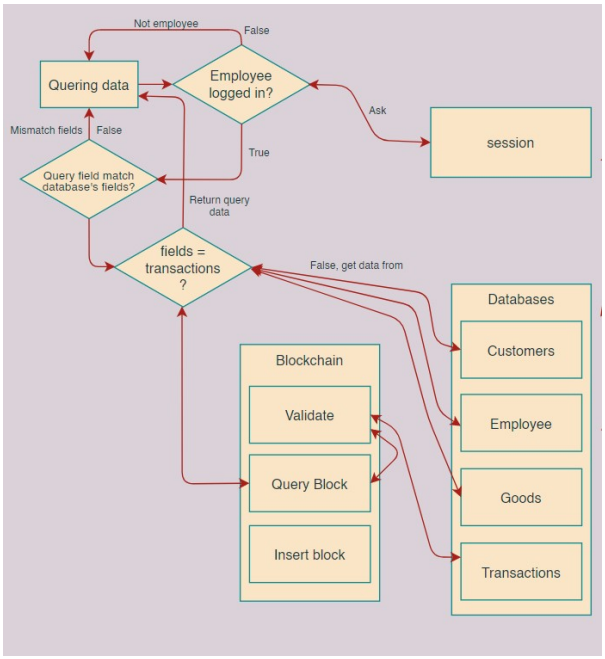


Fig. 14. Querying data pipeline

First, the system will check if there is anyone logged in, if no, the system will not allow querying data. Else, the system will check if the query field match the field in database, if not, the system will return "Mismatch fields" to notify user. if the field is correct, the system will check if the fields is transaction, if not, the system will take data directly from the database. If the fields is transaction, the system will contact the Blockchain manager, then pass the query to this module. The module will validate the data in transactions database, if the data is correctly verified, the system will return the data, otherwise, system should still provide data back, but with a notification that the data have been compromised.

D. Buying decision

Buying decision involves almost all participants. Customer, employee and goods information are required.

When making a transaction, system will check if any employee logged in, if not, it deny the transactions. If yes, it will ask for customer information, if customer is not a current customer, ask them to be customer first. if they are customer, check the good's stock, if not, prompt the customer

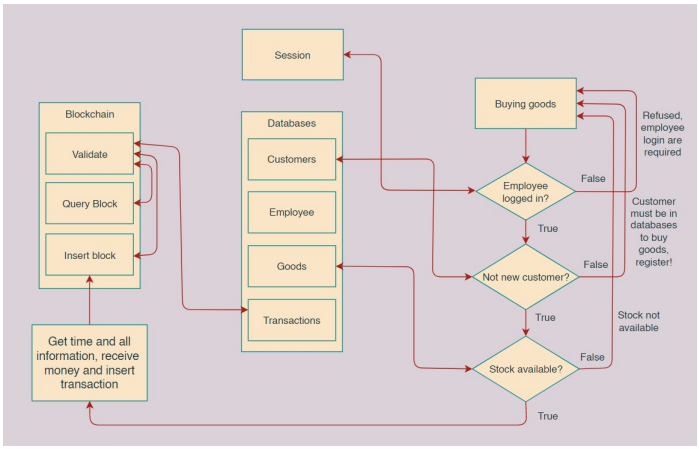


Fig. 15. Buying decision data pipeline

and employee. if good still have stock, proceed gathering all information of employee, customers, product information, then pass all information to Blockchain manager. The Blockchain manage module will connect the transaction to the previous transaction, but it will try to verify the current Blockchain first, if success, it will add the transaction to the Blockchain, otherwise, it will still add but prompt the employee that the Blockchain data have been modified.

XVI. Project/Report contribution

TABLE I
Project/Report contribution

Name	Project contribution	Task(s) envolved
To Duc Anh	40%	Writing report, system design, implemented, code base, convert report to LaTeX.
Le Tien Bang	10%	Writing report
Nguyen Minh Hue	25%	Writing report, design slide, making data, take part in inserting and querying data from MongoDB
Nguyen Thuy Linh	25%	Writing report, design slide, making data, take part in inserting and querying data from MongoDB

REFERENCES

[1] Tran, D.A. and Krishnamachari, B. (2022). Blockchain in a nutshell. arXiv:2205.01091 [cs]. [online] Available at: https://arxiv.org/abs/2205.01091 [Accessed 31 May 2022].

[2] Deloitte Malta. (n.d.). 2020 Global Blockchain Survey. [online] Available at: https://www.deloitte.com/mt/en/pages/technology/articles/2020-global-blockchain-survey.html [Accessed 31 May 2022].

[3] Budman, M., Bhat, R. and Bordoloi, S. (2020). Time For Trust: How blockchain will transform business and the economy. [online] PwC. Available at: https://www.pwc.com/timefortrust.

[4] Addison, S., Yang, S.-S. and Harry, C. (2008). Interest-Based Peer Selection in P2P Network. [online] https://www.researchgate.net. Available at: https://www.researchgate.net/publication/4345934_Interest-Based_Peer_Selection_in_P2P_Network [Accessed 1 Jun. 2022].

[5] Stoica, I., Morris, R., Karger, D., Kaashoek, M.F. and Balakrishnan, H. (2001). Chord. ACM SIGCOMM Computer Communication Review, 31(4), pp.149–160. doi:10.1145/964723.383071.

[6] Romano, D. and Schmid, G. (2017). Beyond Bitcoin: A Critical Look at Blockchain-Based Systems. Cryptography, 1(2), p.15. doi:10.3390/cryptography1020015.

[7] Damgård, I. (1999). Commitment Schemes and Zero-Knowledge Protocols. Lectures on Data Security, pp.63–86. doi:10.1007/3-540-48969-x_3.

[8] Reis, G.A., Chang, J., Vachharajani, N., Rangan, R. and August, D.I. (2005). SWIFT: software implemented fault tolerance. [online] IEEE Xplore. doi:10.1109/CGO.2005.34.

[9] Lamport, L., Shostak, R. and Pease, M. (1982). The Byzantine Generals Problem. ACM Transactions on Programming Languages and Systems, 4(3), pp.382–401. doi:10.1145/357172.357176.

[10] Nakamoto, S. (2008). Bitcoin: a Peer-to-Peer Electronic Cash System. [online] bitcoin.org. Available at: https://bitcoin.org/bitcoin.pdf.

[11] Castro, M. and Liskov, B. (2002). Practical byzantine fault tolerance and proactive recovery. ACM Transactions on Computer Systems, 20(4), pp.398–461. doi:10.1145/571637.571640.

[12] Gilad, Y., Hemo, R., Micali, S., Vlachos, G. and Zeldovich, N. (2017). Algorand. Proceedings of the 26th Symposium on Operating Systems Principles. doi:10.1145/3132747.3132757.

[13] IBM (2022). What are smart contracts on blockchain? [online] IBM. Available at: https://www.ibm.com/topics/smart-contracts.

[14] Majaski, C. (2021). Understanding Distributed Ledgers. [online] Investopedia. Available at: https://www.investopedia.com/terms/d/distributed-ledgers.asp.

[15] Asif, R., Ghanem, K. and Irvine, J. (2020). Proof-of-PUF Enabled Blockchain: Concurrent Data and Device Security for Internet-of-Energy. Sensors, 21(1), p.28. doi:10.3390/s21010028.

[16] docs.python.org. (2022). 1. Introduction — Python 3.10.4 documentation. [online] Available at: https://docs.python.org/3/reference/introduction.html [Accessed 5 Jun. 2022].

[17] P.J, E. (2010). PEP 3333 – Python Web Server Gateway Interface v1.0.1 — peps.python.org. [online] peps.python.org. Available at: https://peps.python.org/pep-3333/.

[18] Malhotra, D., Awadhesh, Singh, K. and Saini, P. (2022). Deloitte Survey: Blockchain Reaches Beyond Financial Services With Some Industries Moving Faster. [online] Deloitte Croatia. Available at: https://www2.deloitte.com/hr/en/pages/press/articles/blockchain-2017.html.

[19] Casey, M.J. and Wong, P. (2017). Global Supply Chains Are About to Get Better, Thanks to Blockchain. [online] Harvard Business Review. Available at: https://hbr.org/2017/03/global-supply-chains-are-about-to-get-better-thanks-to-blockchain.

[20] Sutherland, M. and Wöcke, A. (2011). The symptoms of and consequences to selection errors in recruitment decisions. South African Journal of Business Management, 42(4), pp.23–32. doi:10.4102/sajbm.v42i4.502.

[21] Beck, R. and Müller-Bloch, C. (2017). Blockchain as Radical Innovation: A Framework for Engaging with Distributed Ledgers. [online] Available at: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/86612670-480d-4e1d-8b39-b00e3cfe4234/content.

[22] Golosova, J. and Romanovs, A. (2018). The Advantages and Disadvantages of the Blockchain Technology. 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). doi:10.1109/aieee.2018.8592253.

[23] (Liu, Jiameng, et al. "Blockchain for data science."Proceedings of the 2020 The 2nd International Conference on Blockchain Technology. 2020.)