DATA SCIENCE INSTITUTE

The World Data Science Institute is a primary source for Financial Data Science Education.

Blockchain (Simply Explained)

In this Data Science Report, we will describe in detail everything you need to know about Blockchain.

Meet Data Science Team

Anade Davis - Data Science Manager - LINKEDIN

William Munson - Project Lead/Quantitative Researcher - LINKEDIN

Shailee Desai - Quantitative Researcher - LINKEDIN

Harleen Kaur Bagga - Data Science Researcher - LINKEDIN

Martin Lardiwinata - Data Analyst - LINKEDIN

Table of Contents

What is Blockchain?

- Blockchain is a type of advanced database (also known and most commonly referred to as a digital ledger) of information.
- Think of it as a super secure database that is very difficult to hack (but not impossible).
- A digital ledger contains data that has been captured, recorded, and replicated into the Blockchain (aka database).

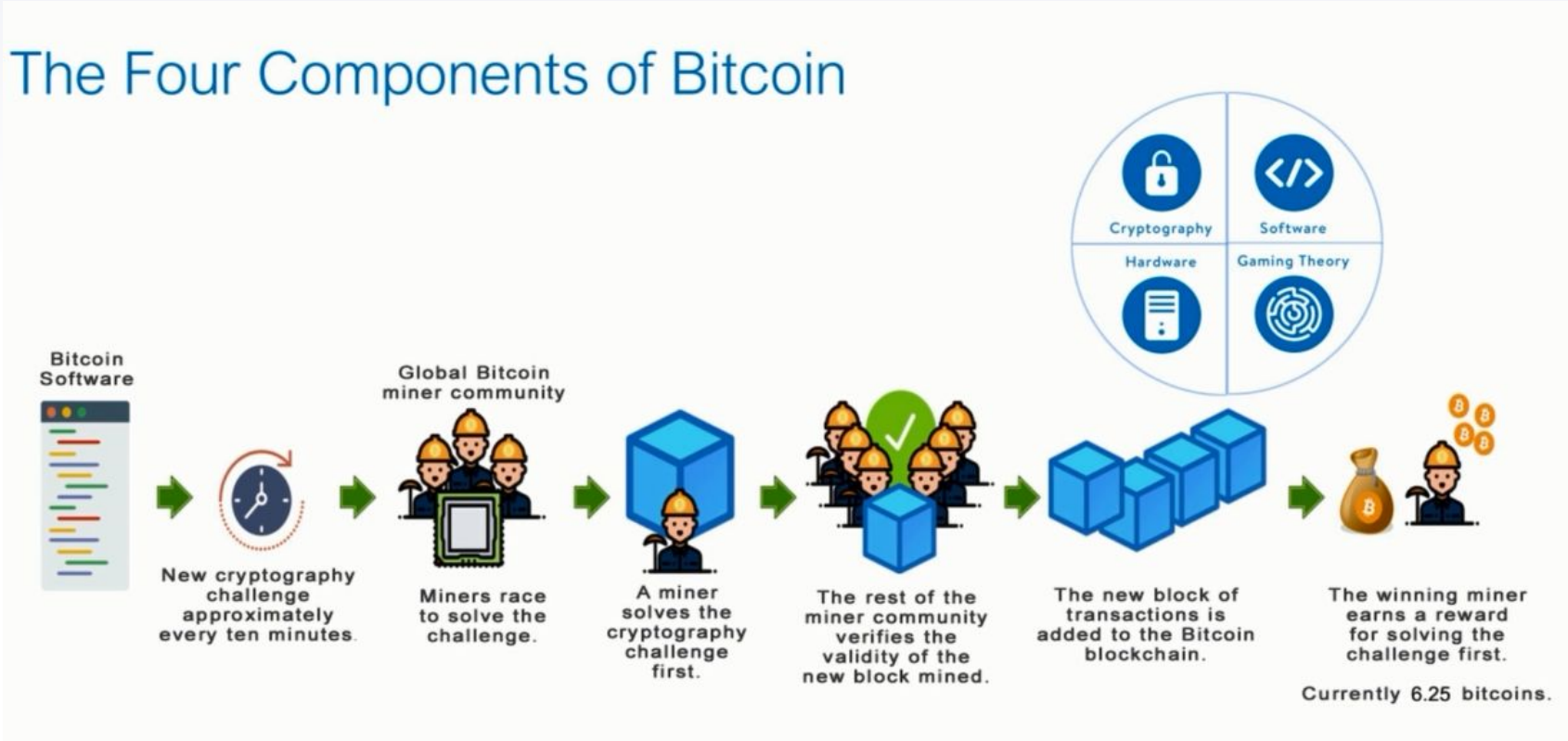
Definition of Blockchain

- " A Blockchain is a constantly growing ledger that keeps a permanent record of all the transactions that have taken place, in a secure, chronological and immutable way."
- In Simple Terms : In a world where we're dealing with malware, you're dealing with hackers, you are dealing with all sorts of counterfeit goods being put into supply chains all over the world. It's becoming more difficult to prove the authenticity, and guarantee the integrity of any item of value that you're buying.
- There is an answer and it's a technology called blockchain , because blockchain can provide you a single source of truth, that's permanent, verifiable and unchangeable.

Blockchain Use Cases in Finance?

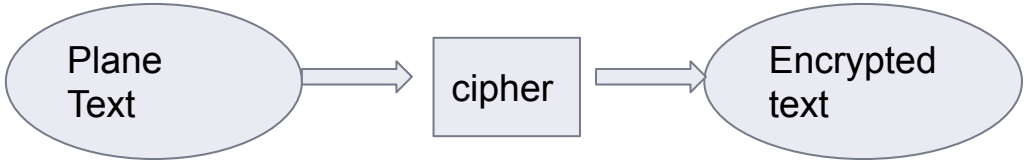
- Blockchain Technology Makes Transactions Easier for sending payments.
- Eliminates all the issues that occur when sending money internationally.
- Blockchain does not require any third party or intermediate for execution.
- Blockchain offers a higher level of security than traditional banks and financial institutions.

Blockchain Use Cases in Finance – Bitcoin



Cryptography

- Cryptography is a method of using advanced mathematical principles in storing and transmitting data in a particular form so that only those for whom it is intended can read and process it.
- Through encryption the message is encoded in a format so that it can't be understood by eavesdropper.



- Blockchain use Cryptography in two ways: **Hash function** and **Digital signatures**

Hashing

So, let's say we have a message to deliver to a client who agrees to pay us [$100]. We want to make sure that our client has received that exact message and agrees to pay the full amount.

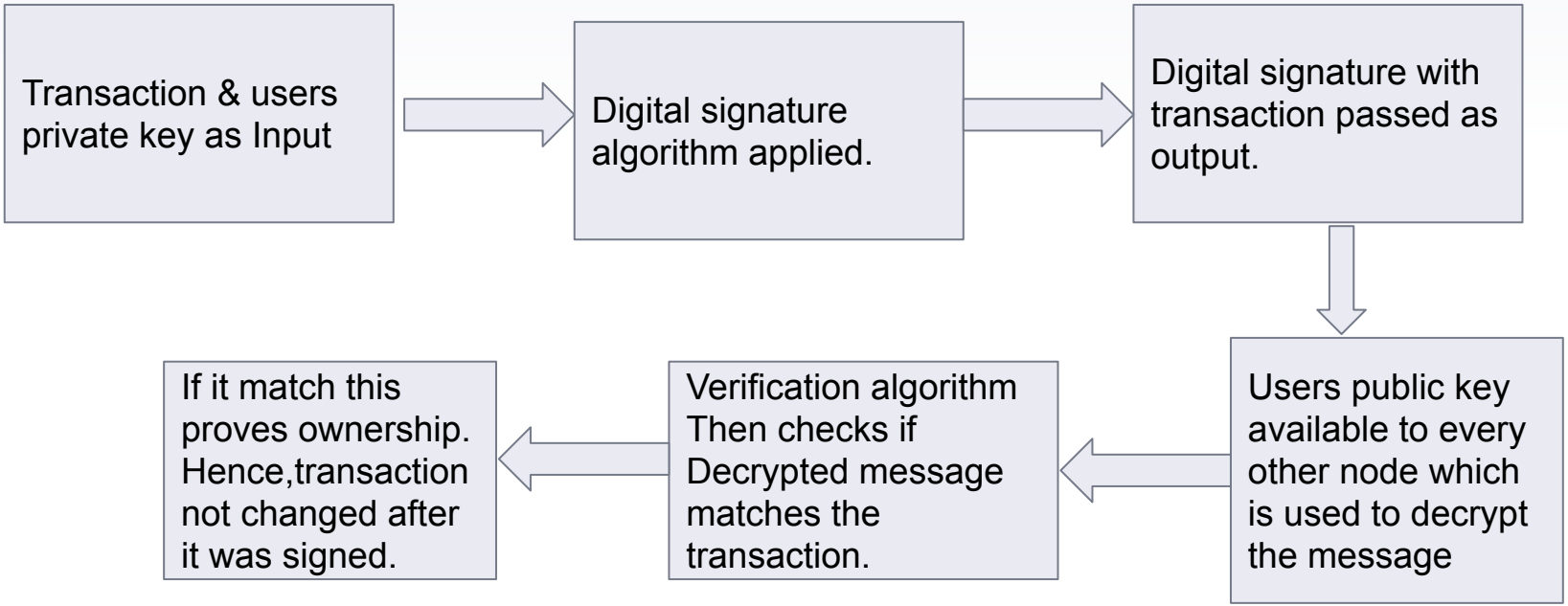
Hashing

- Now let's say the client tries to change something in the message before we deliver it to him? This is where hashing comes into play.
- To avoid letting this happen, we use hash code for the message. If anything is changed in the message, we give it a new hash code. That way, if the client tries to make any malicious changes, the hash code will prevent him from executing those changes.

Digital signature

- Digital signature is to establish a proof of ownership in a blockchain.
- Bitcoin uses it to prove that the sender of transaction actually owns the coins he wants to transfer to another person.
- **To prove ownership a pair of private key and public key is generated by the' user,**whenever the user wants to prove the ownership of some digital asset of transaction **he or she has to sign the transaction with their private key.**
- If any changes are made to transaction the digital signature gets changed and it is detected by the verification algorithm and clients are prevented from making fraudulent transaction.

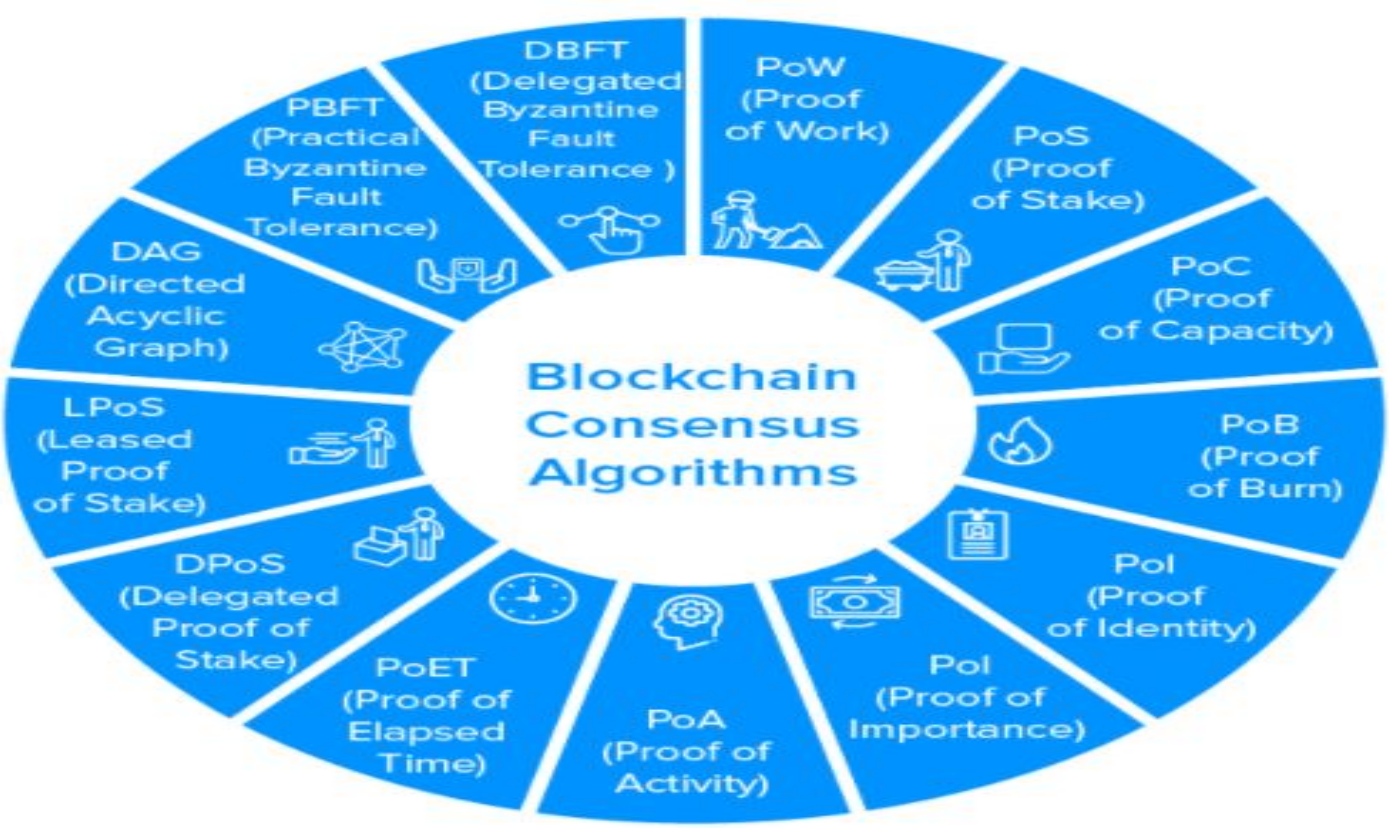
Digital Signature



Consensus Algorithms

- Consensus algorithms are used to ensure that a transaction is valid without the use of a central authority
- It's a form of resolution where individuals need to support the majority decision, whether they liked it or not.
- Let's take an example: Imagine a group of six people that want to make a decision about a project that benefits them all. Every one of them can suggest an idea, but the majority will be in favor of the one that helps them the most. Others have to deal with this decision whether they liked it or not.
- Consensus algorithms works in the similar way it do not merely agree with the majority votes, but it also agrees to one that benefits all of them. It's like a win in the network.

Types of Consensus algorithm



Proof of Work in detail

- Before a transaction is made, there needs to be a consensus as to who makes the transaction.
- Say we have 10 carriers. In order to reach a conclusion for who will deliver the message to a client, we give each carrier a sudoku puzzle. Whoever finishes first is the one who will deliver the message.

Proof of stake in detail

- Unlike proof of work there is no puzzle to compete for and no reward for it.
- **The creator of a new block is chosen from a pool of users that have staked a certain amount of cryptocurrency.**Miners take feat from every transaction .
- Here **one would need to own 51% of the cryptocurrency of the chain** which is very expensive for any major blockchain network.
- Since nobody is competing to solve every block ,**no massive energy is required**.
- The penalty of harming the network is possible at the cost of losing the money staked.
- **Developers who want a cheaper and greener form of consensus are switching to proof of stake.**

Some cryptocurrencies with their consensus protocols

Cryptocurrency	Currency	Consensus Protocol
Bitcoin	Bitcoin(BTC)	Proof of Work
Ethereum	Ether(ETH)	Proof of Work Planning to switch to proof of stake
Litecoin	Litecoin(LTC)	Proof of work
Cardano	Cardano(ADA)	Proof of stake
Polkadot	Polkadot(DOT)	Proof of stake
Bitcoin Cash	Bitcoin Cash(BCH)	Proof of work
Stellar	Lumens(XLM)	Federated Byzantine algorithm

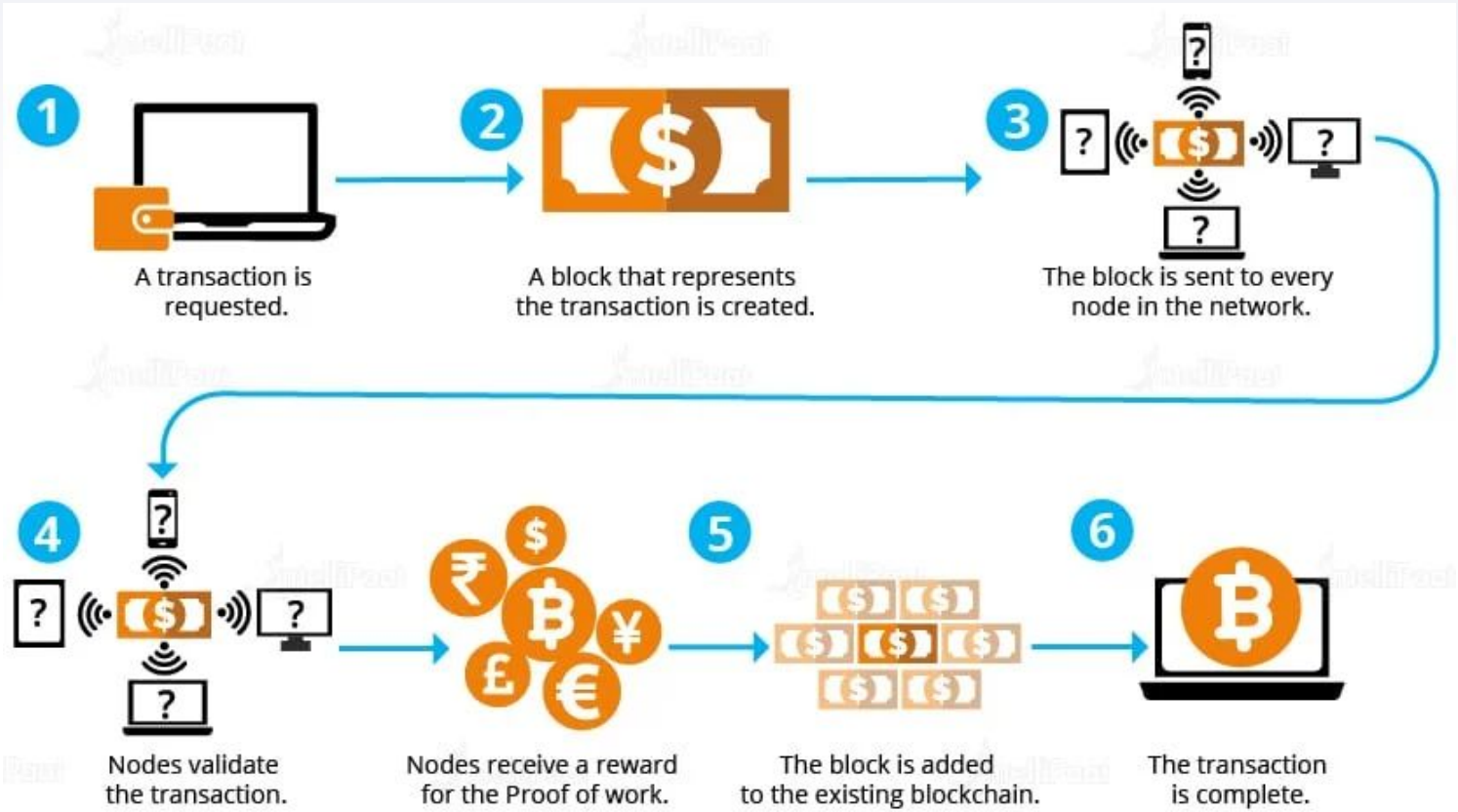
Double-Spending

Double-spending occurs when a party essentially “copy-and-pastes” or re-uses an electronic transaction.

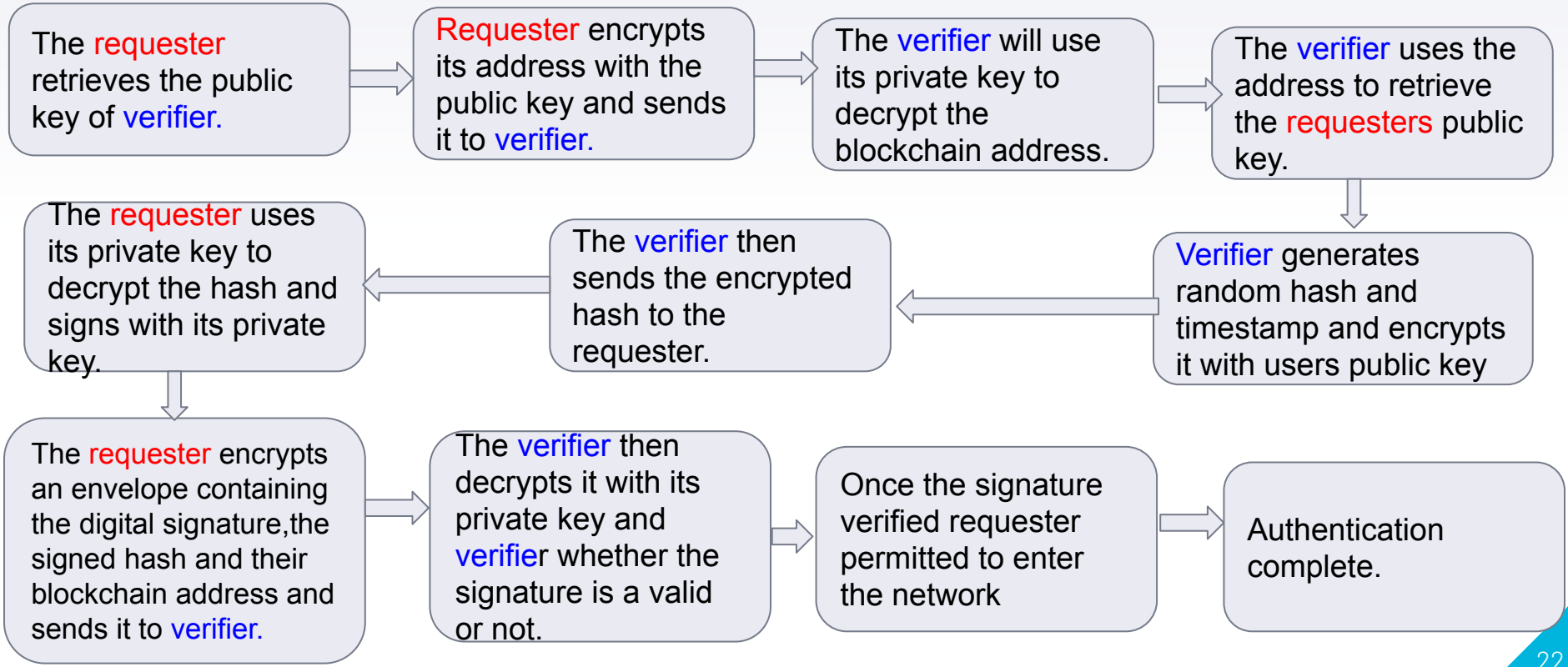
Distributed Ledger

- A *distributed ledger* is a type of database that is shared, replicated, and synchronized among the members of a decentralized network.
- It is a system of record that allows different entities to trust each other and track goods and services even as they change hands.
- Distributed ledger can be any kind of database.**Blockchain is a type of distributed ledger.**
- Distributed ledgers **relies on consensus protocols for it's functioning.**Participants in the network govern and agree by consensus on the updates to the records in the ledger.
- All records in the distributed ledger has a timestamp and unique cryptographic signature which helps to make the ledger an **auditable, immutable history of all transactions in the network.**

How Blockchain works



How Blockchain Authentication works



Tokens

- Tokens represent a digital “unit of possession” that can be exchanged between two people.
- This could represent anything, such as Bitcoin

Miners/Validators

- Cryptocurrency mining is the process by which recent transactions are checked and new blocks are added to the blockchain.
- Miners and validators are participants in the blockchain network who are involved in validating transactions and avoiding double-spending.
- Miners compete against each other to keep the history of blockchain network valid.
- The blockchain network is designed so that everyone on it is rewarded with cryptocurrency for doing the right thing.
- The only way miners can receive a reward is by doing the honest thing and submitting valid blocks.If a miner submit fake or altered block the other miners will discover that and then earn the reward themselves by submitting the honest block.

Smart Contracts

- Smart contracts are self-executing contracts containing the terms and conditions of an agreement among peers.
- It is an Ethereum blockchain application that can facilitate the exchange of money, content, property shares or anything of value without the need of any third party (like government in case of traditional contracts) for monitoring.
- Users set ethers to interact with self operating computer programs and they will run exactly as coded without any possibility of censorship, downtime fraud or interference.

Smart Contracts

- The conditions of such contracts are written in software code(if then condition based), so this allows for an exact consensus on how things should be executed. It's a self executing code which runs until a certain goal is reached.It removes the need for any intermediary.

- If all the validators share the exact conditions that the client has provided, they reach the consensus on the contract

Blockchains VS Databases

Blockchains and databases function differently.

Blockchain (Advanced Database)	Centralized Database
No central authority	A small group of individuals have authority
Modifying data is almost impossible	Data is easily accessible and modifiable
Data is open to the public	Data is hidden
Low cost	High cost
Low speed (Bitcoin blockchain has 4.6 transactions per second)	High speed
Users don’t trust each other	Trust between users

Cryptocurrency

- **The monetary incentive built in a blockchain network is cryptocurrency.**
- Whenever a valid block of transaction is submitted and accepted to the blockchain ,the miner who submitted it is rewarded with newly minted cryptocurrency.
- This way blockchain maintains the integrity of the chains transaction history and submit the honest history of transaction in every block.

Three pillars of Blockchain

1. Decentralization
2. Transparency
3. Immutability

Decentralization

- In a centralized system of money transfer you trust the bank and they look after your money and charge you fees in exchange. This industry is quiet complex and enormous.
- In a decentralized system we take the bank out of middle and use computer code to connect directly among people.
- Blockchain is based on a peer to peer network where two or more computer are connected and share resources without the need of centralized server.
- We take the power and the authority from one source and share it among everyone in the network.
- Blockchain is the ingenious structure which helps every individual of the system in a decentralized network to keep their history of transactions organized and identical to everyone else.

Decentralization

- **The power of decentralization comes from the magnitude of network.**
- Instead of a bank telling you that you don’t have money the code in decentralized system can check itself and everyone else on the network, so if you were going to lie you have get more than half the network lie for you.
- **Blockchain and cryptocurrencies are technical tools and incentives on which we create decentralized system.**

Transparency

- All the information about the transactions between users is stored in public ledger.
- This makes the blockchain completely transparent.
- The most interesting thing is that a person's identity is still kept hidden via complex cryptography, while making transaction information public.
- This means system remain secure while maintaining accountability at the same time.
- Transparency makes it quite difficult to fake transactions,or to engage in any kind of shady business.

Immutability

- The information stored on the blockchain is tamperproof.
- **A unique hash is added to every block, and each block is connected to the previous one as it also contains the hash from the block before it.**
- If a malicious user tries to falsify or change information the block's hash will no longer match the one reflected in following block.
- This makes tampering with information almost impossible.
- Thus in blockchain you cannot fiddle with information without being caught.

Wallets

- Wallet is a place where you store cryptocurrency.Bitcoin uses Elliptic curve Digital signature (ECC)algorithm to generate wallets.
- This curve gives you two keys public key and private key.
- Your private key should not be shared with anyone as it is essentially your signature which gets attached to transaction you make in such a way that proves you made it.
- Wallet uses your public key to generate address that you can use for purchasing your bitcoins but it doesn’t allow people to search your transaction history.
- A miner in a blockchain verifies the transaction .Then the rest of the network update their ledgers to reflect the change.Once that is done balance is updated.

Bitcoin transaction

- A Bitcoin transaction is simply a transfer of value between **two wallets**, which is recorded on the blockchain.
- It doesn't matter if I send the bitcoins to my neighbour or someone in the other side of the world ,it happens at the same speed.
- Due to consensus rules every computer on the network agree unanimously on every transaction.
- Actions on bitcoin transaction are irreversible one ,so they should be done with extra care.

Types of blockchain and their relevance

1. Public blockchain (permissionless)

2. Private blockchain (permissioned)

3. Consortium blockchain (semi-decentralized)

4. Hybrid blockchain (combination of public and private blockchain)

Public (Permissionless) Blockchain

The name says it all - public blockchains allow anyone to use their network.

Examples of public blockchains include:

1. Bitcoin
2. Ethereum

Public (Permissionless) Blockchain

Advantages:

- Information is open and transparent
- Trust is not needed
- Almost impossible to hack into

Disadvantages:

- Slow due to its accessibility
- Energy consumption
- Scalability concerns

Private Blockchain

- Blockchain network that works in restrictive environment like close network or under control of single entity
- Much smaller scale than public blockchain
- Instead of everyone being able to provide computing power, private blockchains usually operated inside company or organization
  - Example : Multichain and Hyperledger project, Corda

Private Blockchain

Advantages

- Usually very fast because of their limited size
- Controlling organization sets permission level, security and accessibility

Disadvantages

- Many people consider private blockchain aren't true blockchain because it is more centralized
- The source code is often closed. Users can't independently audit or confirm it.

Consortium blockchain

- Also known as federated blockchain
- It has private and public blockchain feature
- Is a private blockchain with limited access to a particular group, eliminating the risk of just one organization
- Controlled by preset nodes
  - Example : Energy Wen Foundation, R3

Consortium blockchain

Advantages

- Tend to be more secure than private blockchain
- More scalable and efficient than public blockchain network

Disadvantages

- Less transparent than public blockchain since only selected organizations have access
- If a member node is breached, can impact the blockchain regulation

Hybrid blockchain

- Combine elements of private and public blockchain
- It lets to use permission-based system alongside public permissionless system by allowing public to access specific data stored in blockchain
- Typically transaction and records are not public, but can be verified through smart contract
  - Example : Dragonchain

Hybrid blockchain

Advantages

- It protect privacy but allow communication to third party
- Transaction tend to be cheap and fast

Disadvantages

- Isn’t completely transparent
- There is no incentive for user to participate or contribute to the network

Hyperledger Fabric

- A widely used primary Blockchain,primarily used in enterprise.
- Helps make transaction between multiple businesses efficiently and seamlessly.
- **Has modular design**:Business can plugin in different functionalities to suit their particular needs.
- Fabric records history of transaction in chronological ledger.
- **Deals with asset. An asset can be anything with monetary value.**
- **Assets are represented as** a collection of **Key-value pairs with state changes recorded as transaction in ledger.**
- Fabric modifies assets using Chain code and permissions.

Chaincode and permissions

- Hyperledger fabric provides ability to modify assets using chaincode.
- Chaincode defines an asset or assets ,and the transaction instructions for modifying them.
- **Smart contracts deployed to the fabric Ledger executes chaincode where the business share business logic .**
- Members of each permissionless network interact with the ledger using chaincode either by adding new contract or invoking transaction.
- **State created by a chaincode is scoped exclusively to that chaincode** and can't be accessed directly by another chaincode.
- Given the appropriate permissions a chaincode may invoke other chaincode to access its state.

Chaincode and permissions

- To enable permission service Hyperledger provides Membership identity service that manages IDs and authenticates participants.
- Access control list provides additional layers of permissions through authorization of specific network operations.
- Assets in a chiancode are added, updated, and transferred using chaincode.

Nodes in hyperledger

Peer nodes

-Peer nodes are responsible for Executing and verifying transaction

Ordering nodes

-Ordering nodes are responsible for ordering and executing Correct history of transaction.

Nodes in Hyperledger

- These nodes help in increasing scalability and efficiency.
- Peer nodes are allowed to batch and process multiple transitions simultaneously.
- The networks consensus protocol which businesses in a network can customize is then implemented by the ordering nodes to create a single true record of transactions.

Two components of Hyperledger Fabric

- **Blockchain Log :** Stores the immutable sequenced record of transactions of a block.the purpose of log is to track the assets place of origin as it is exchanged among multiple parties.

  To track an assets provenance is to track when and where it is created,it is extremely important in the world of businesses as it ensures that a business selling a chain of items possess the chain of title verifying their ownership of it.

- **State Database**:Maintains the blockchain's current state. For speed and efficiency sake Hyperledger Fabric stores the current state as well and allows all the members of the network to query it.

Private Channels in Hyperledger fabric

- Having a distributed ledger meant that every party in business network will have the access to all transitions even if they weren't involved in it.

- **This was a huge demerit for businesses who were in the same business network as their competitors and didn't want to reveal their data and transactions to them.**

Private channels in Hyperledger

- Hyperledger introduced Private channels which allows **Restricted messaging** paths that provide privacy for specific subsets.
- All **data invisible to members** is not granted access.
- **It allows competing businesses interest and any groups that requires private confidential transaction to coexist on the same network .**
- **An example given in next slide will make this concept clearer.**

Example of Hyperledger Fabric

- Suppose there's a manufacturer that wants to ship chocolates to a specific retailer or market of retailers (i.e., all US retailers) at a specific price but does not want to reveal that price in other markets (i.e., Chinese retailers).
- Since the movement of the product may involve other parties, like customs, a shipping company, and a financing bank, the private price may be revealed to all involved parties if a basic version of blockchain technology is used to support this transaction.

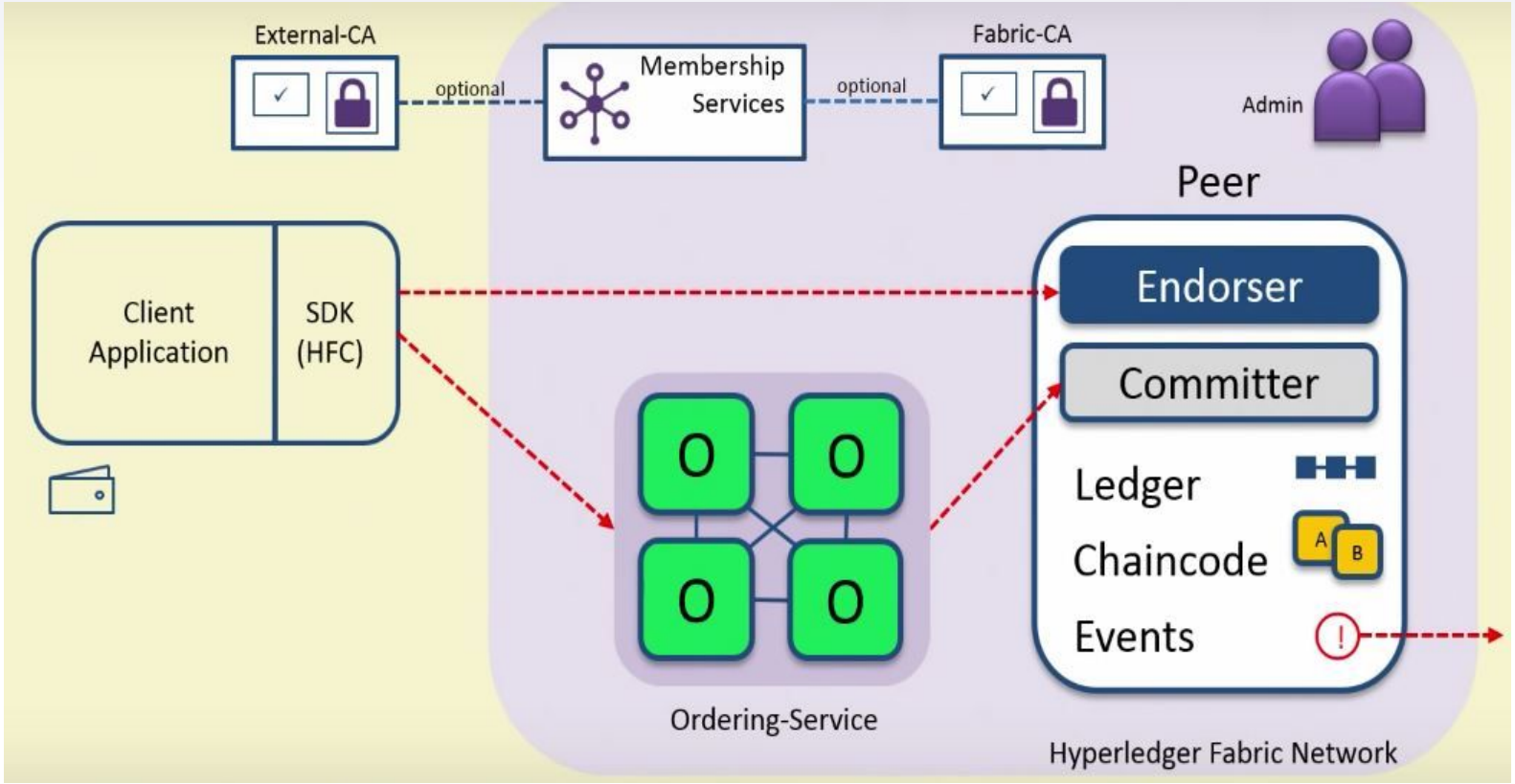
Example of Hyperledger Fabric

- Hyperledger Fabric addresses this issue by keeping private transactions private on the network; **only participants who need to know are aware of the necessary details.**

- Data partitioning(private channels/subnet) on the blockchain allows specific data points to be accessible only to the parties who need to know.

- This example illustrates that Hyperledger fabric can enable private confidential transaction to coexist on the same network something which was not possible in transaction through Bitcoin and ethereum.

Hyperledger Fabric architecture



Hyperledger Fabric architecture

Committer -It is responsible for final transaction added to blockchain.

Endorser-It is responsible for signing client's transaction.This it does so on the chaincode.

Chaincode - Smart contract

Ledger - Blockchain copy

Events - A kind of notification

Membership service- It provides certificate to the user that it can send transaction to blockchain.

Ordering service-Order transaction and then create block which it sends to committer.

Flow of Hyperledger Architecture

Transaction proposal: A hyperledger fabric client calls an application through an application SDK.The proposal is a request to invoke transaction with the intent to reading or updating the ledger.Client informs the endorser that she is sending transaction.

Execute transaction: Endorser will then decide whether to sign or reject the proposal on the basis of chaincode (smart contract).

Proposal response: The result of a function called read/write is set and sent to the client so that each of the result from each node can be compared as we need to come to the consensus that all the answers are **at least 51% in agreement**.

Flow of Hyperledger Architecture

Deliver transaction: Client then sends request to the ordering service to arrange the transactions chronologically by channel and create block of transactions per channel.

Validates transaction: The ordering nodes simply batches all the transactions and distribute them to committing peers so that we can begin consensus.Every node in the blockchain validates against the custom Endorsement policy and the transactions are written to the distributed ledger in the form of new block and is thus added to the blockchain.

Notification: Client is then notified by each peer with which it is connected about the success or failure of transaction

Flow of Hyperledger Architecture

The committing peer reaches consensus on the basis of two algorithms:

1.Kafka (Crash Fault tolerance):

- Kafka is a very popular event management service in apache
- Internally Kafka achieves consensus on the notion CFT(Crash Fault Tolerance) consensus that uses a “leader and follower” node configuration.
- In CFT if one node dies, due to a software or a hardware fault, data is preserved. In context of leader follower system ,the leader owns a partition, and the follower has a replication of the same. When the leader dies, the follower becomes the new leader.

Flow of Hyperledger Architecture

- Kafka can thus tolerate certain set of nodes to fail at any point of time and even with certain failures among these nodes you can achieve consistent order across all the nodes.

- In Kafka if not more than 50% of the nodes fail we can achieve consensus.

Flow of Hyperledger Architecture

2. SOLO(Single node development):

- Solo involves single ordering node.
- It is most typically used by developers experimenting with Hyperledger Fabric networks.
- Solo is not meant for production.It is not and never be fault tolerant.
- It is the simplest mechanism, which only broadcasts the transaction without establishing any real consensus.

Key benefits of Hyperledger fabric over other blockchains

1. Permissioned Blockchain Deployment
2. Trustless, Scalable and Performant
3. Need-to-know basis Access
4. Supported by Business leaders
5. Modular Architecture Plug-in Components Support
6. Protection of digital keys and sensitive data

Hyperledger vs Ethereum

Hyperledger	Ethereum
Ideal for B2B transactions	Ideal for B2C transactions
No consensus algorithms (users have to make their own) .	Proof of work consensus algorithms
No built-in cryptocurrency/tokens	Comes with its own token (Ether, or ETH)
Private ledger	Public ledger
Written in Go, Java, and Node.js	Written in Solidity

References and citations

Satoshi Nakamoto White Paper

https://academy.101blockchains.com/courses/enterprise-blockchains-fundamentals

https://courses.blockgeeks.com/course/introduction-to-blockchain-and-cryptocurrency/

https://courses.blockgeeks.com/course/ce101-intro-to-cryptoeconomics/

https://courses.blockgeeks.com/course/hyperledger-101/

https://courses.blockgeeks.com/course/hl102-the-fabric-network/

https://blockgeeks.com/guides/what-is-blockchain-technology/

Hyperledger White Paper