INTRODUCTION TO BLOCKCHAIN

Definition

- **Layman's definition**:

Blockchain is an ever-growing, secure, shared record keeping system in which each user of the data holds a copy of the records, which can only be updated if all parties involved in a transaction agree to update.

- **Technical definition**:

Blockchain is a peer-to-peer, distributed ledger that is cryptographically-secure, append-only, immutable (extremely hard to change), and updateable only via consensus or agreement among peers.

In short Blockchain → Distributed Ledger

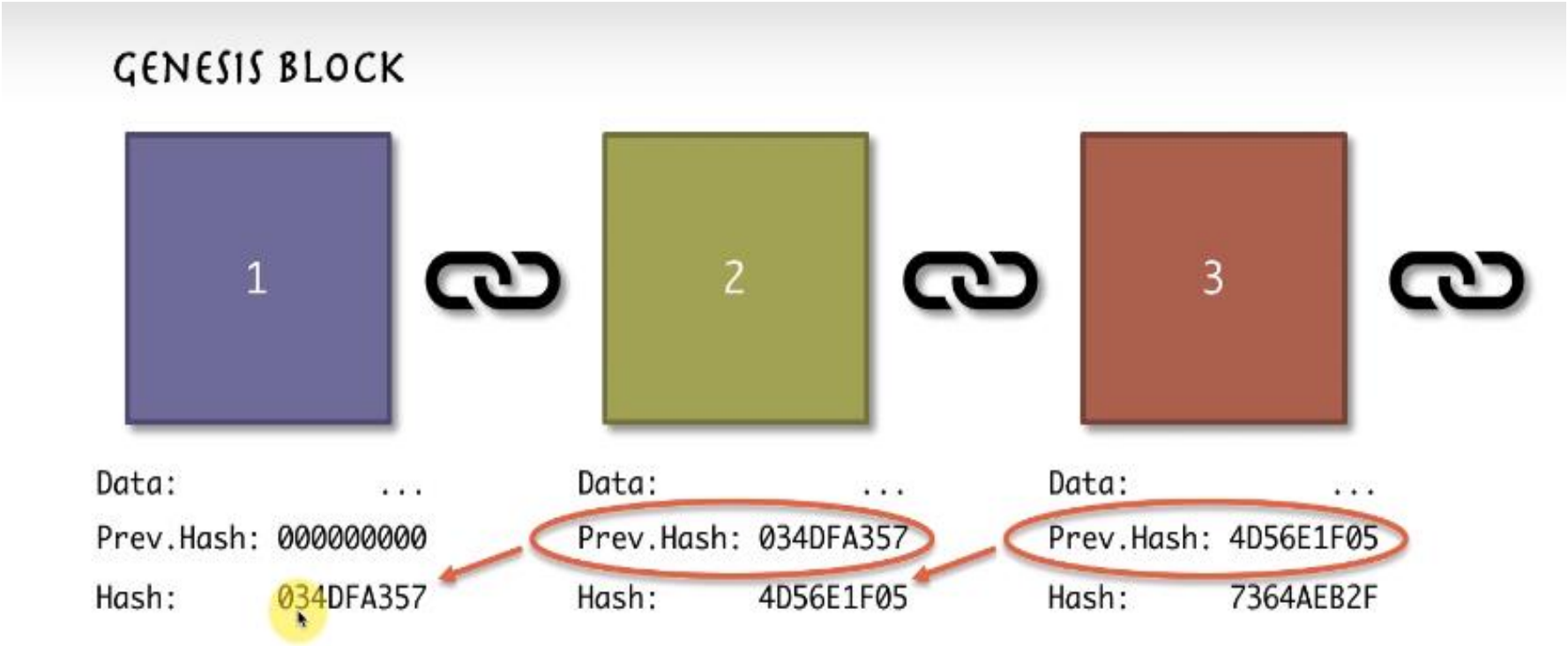
Key Terminology of Blockchain

- **Transaction**
- **Distributed Ledger**
- **Block**
- **Genesis Block**
- **Merkle root**
- **Hash Key**
- **Peer-to-peer network**
- **Node**
- **Consensus Mechanism/Protocol**
- **Smart contract**
- **Nonce**
- **Mining**
- **Wallet**

BLOCK IN BLOCKCHAIN

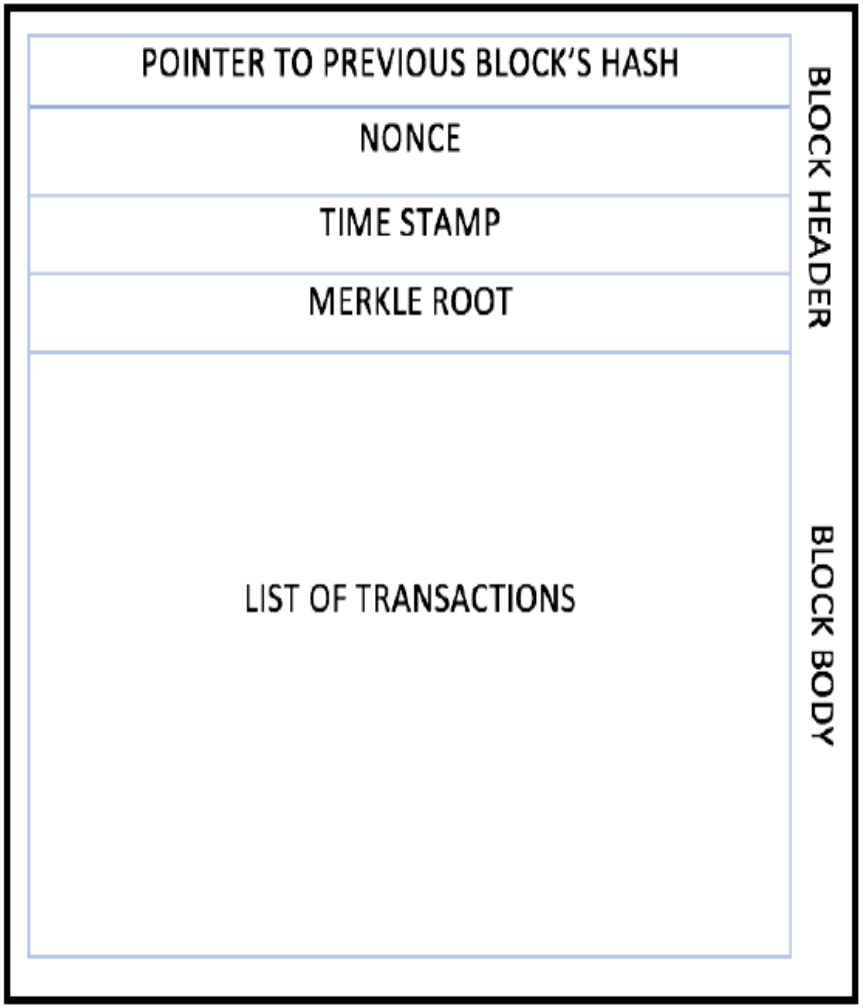
- **Transaction** - an asset transfer
- **Ledger** is the system of record for a business
- Business will have multiple ledgers for multiple business networks in which they participate.
- A shared ledger technology allowing any participant in the business network to see the system of record
- **A distributed ledger** technology allowing all participants in the business network to maintain a copy of that record/transactions
- In blockchain, a **Block** is a **container data structure** that contains a series of transactions
- **For ex. in Bitcoin:** A block may contain more than 500 transactions on average, the average size of a block is around 1 MB (an upper bound proposed by Satoshi Nakamoto in 2010)
  - May grow up to 8 MB or sometime higher (as of March 2018)
  - Larger blocks can help in processing large number of transactions in one go.

Blockchain

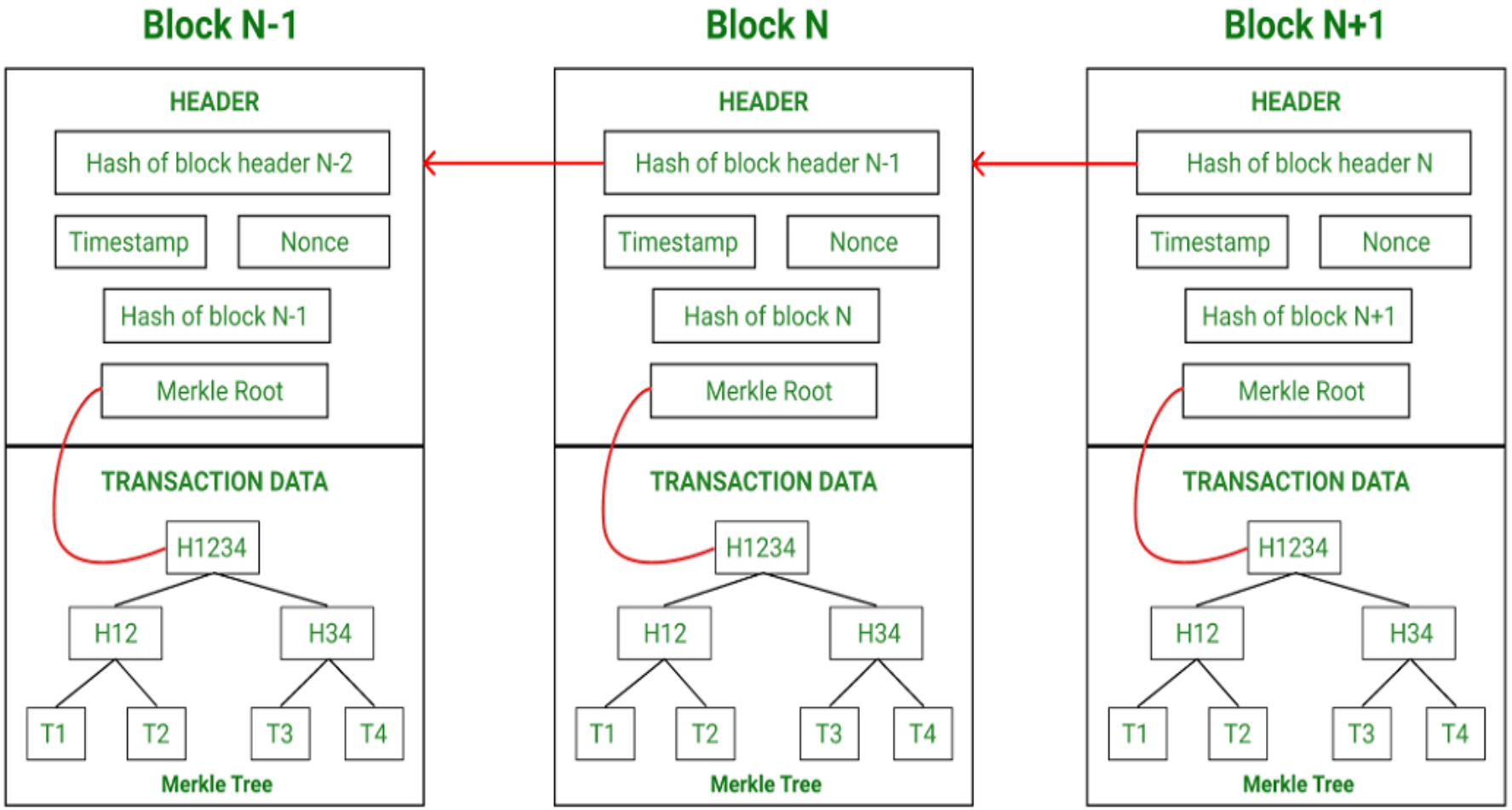

GENESIS BLOCK
1
2
3
Data: ...
Prev.Hash: 000000000
Hash: 034DFA357
Data: ...
Prev.Hash: 034DFA357
Hash: 4D56E1F05
Data: ...
Prev.Hash: 4D56E1F05
Hash: 7364AEB2F

Key Terminology & Block Structure

- The Block contains two parts
  - **the header** and
  - **the data (the transactions)**
- The header of a block connects the transactions – any change in any transaction will result in a change at the block header
- The headers of subsequent blocks are connected in a **chain – the entire blockchain needs to be updated if you want to make any change anywhere**

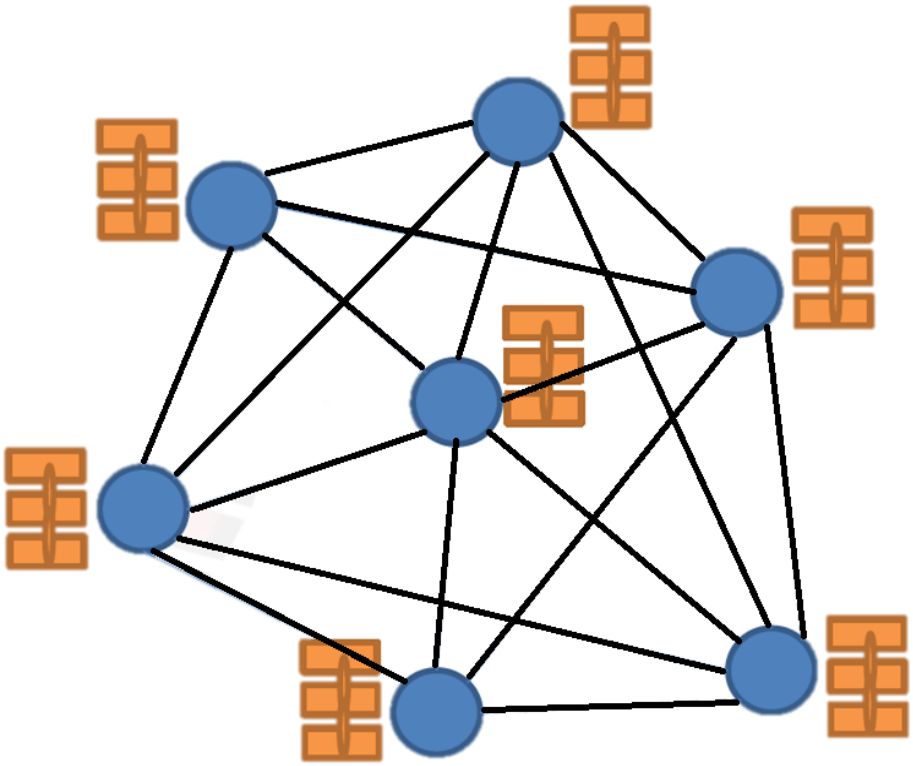


Blockchain Structure



Blockchain - The Notion of Distributed Consensus

- Every peer in a Blockchain network maintains a local copy of the Blockchain
- **Requirements**
  - All the replicas need to be **updated** with the last mined block
  - All the replicas need to be **consistent** – the copies of the Blockchain at different peers need to be **exactly similar**
- Ensure that different nodes in the network see the same data at nearly the same point of time.
- All nodes in the network need to agree or **consent** on a regular basis, that the data stored by them is the same.
- No single point of failure – the data is decentralized
- The system can provide service even in the presence of failures



51% accept

Blockchain Process

Transaction is proposed

Proposed transaction is broadcast to the network

Miners verify the transaction and bundle it into a block along with other transactions.

Miners compete to solve the complex mathematical puzzle (Mining)

The nodes verify the miner's work.

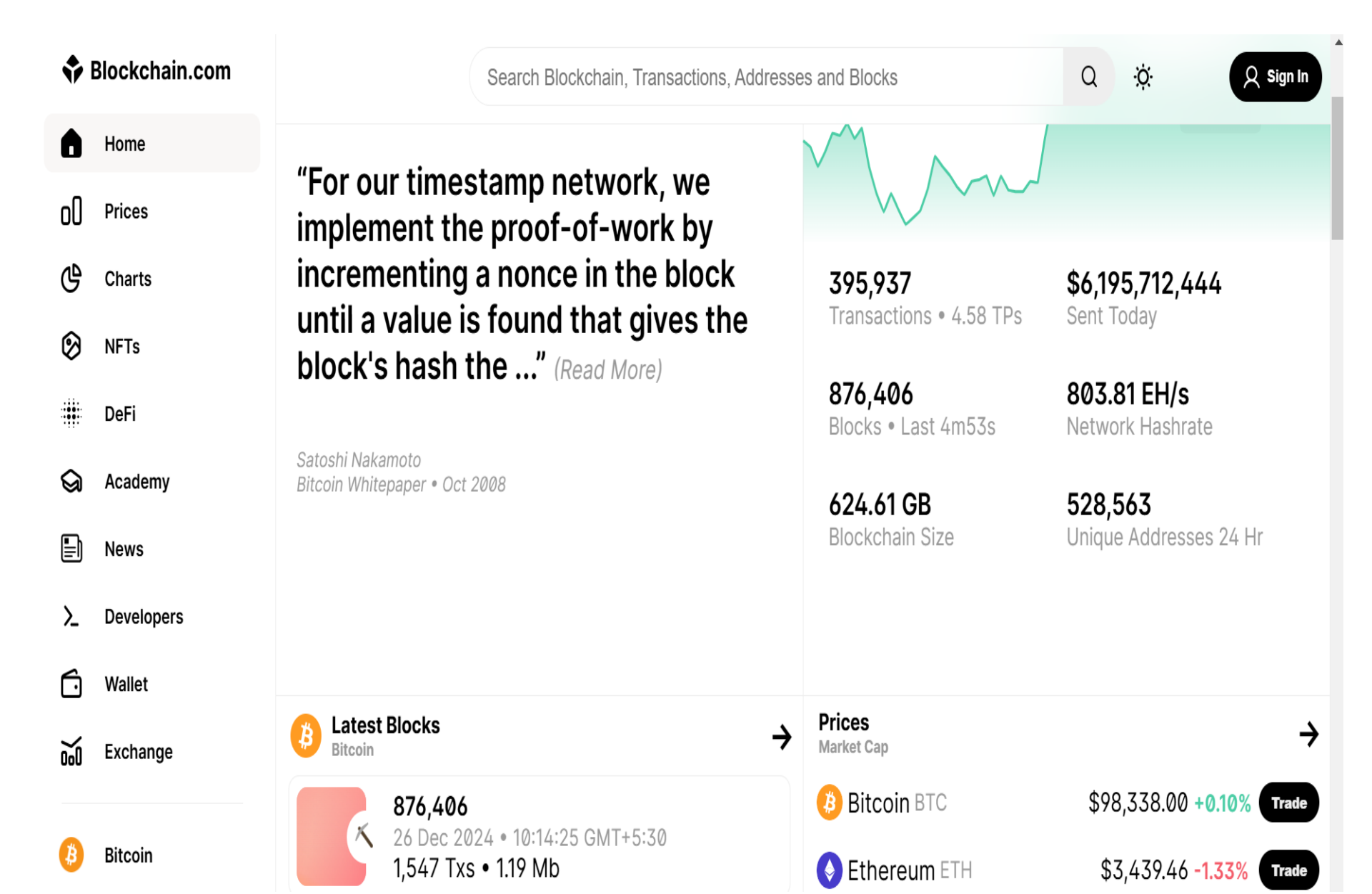
Block is added to the Blockchain.

The updated copy of the Blockchain is circulated throughout the network.

Transaction completion

How Blockchain works?

- **Step 1:** A node starts a transaction by first creating and then digitally signing it with its private key. A transaction can represent various actions in a blockchain.
- **Step 2:** A proposed transaction is propagated (flooded) by using a flooding protocol, called Gossip protocol, to peers that validate the transaction based on preset criteria. Usually, more than one node are required to verify the transaction.
- **Step 3:** Miners verify the transaction and bundle it into a block along with other transactions, and then propagated onto the network. At this point, the transaction is considered confirmed.
- **Step 4:** Miners compete to solve the complex mathematical puzzle. The puzzle requires much computational power to solve.
- **Step 5:** The nodes verify the miner's work. The miner who finds the correct hash broadcasts the block to the network. Majority of the nodes/miners need to approve/verify the block for it to be accepted into the blockchain. Once approved, the winning miner can collect his reward(Proof of Work).
- **Step 6:** Once the block is verified, the winning miner adds his block to the existing blockchain.
- **Step 7:** The updated copy of the blockchain is circulated throughout the network.
- **Step 8:** Transaction completion.

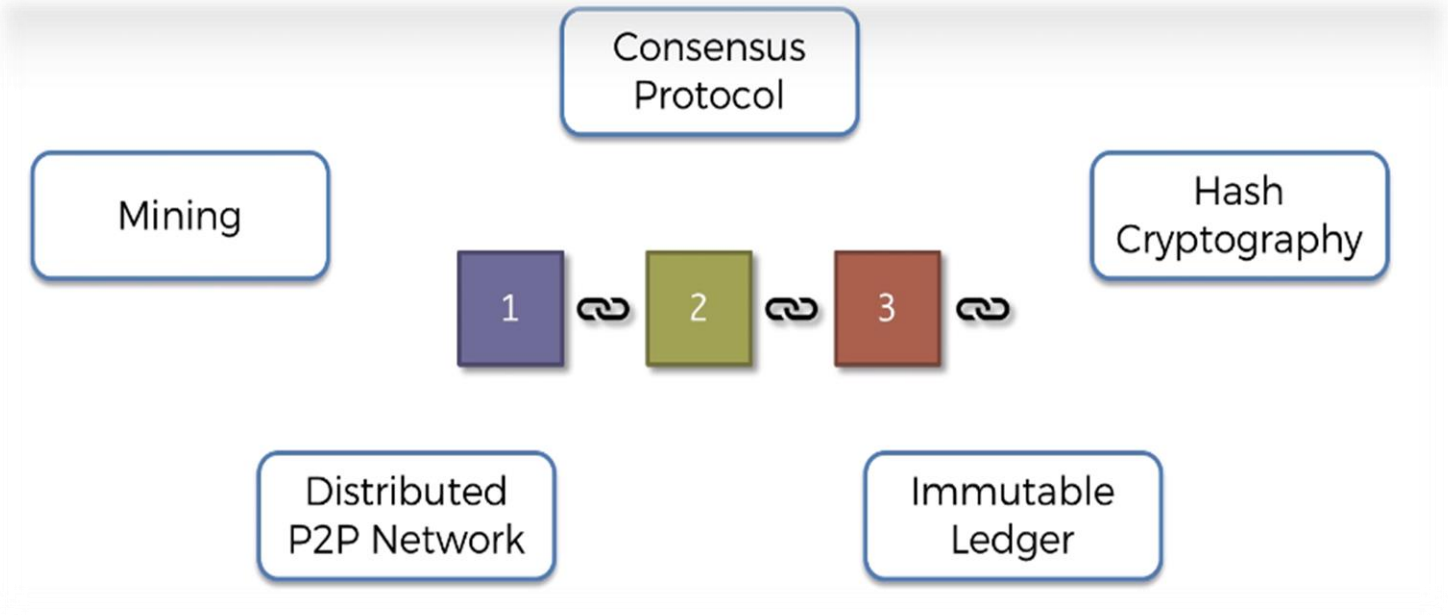


Bitcoin Blockchain statistics as on 26-12-2024 @10.30 AM

Key Elements of Blockchain

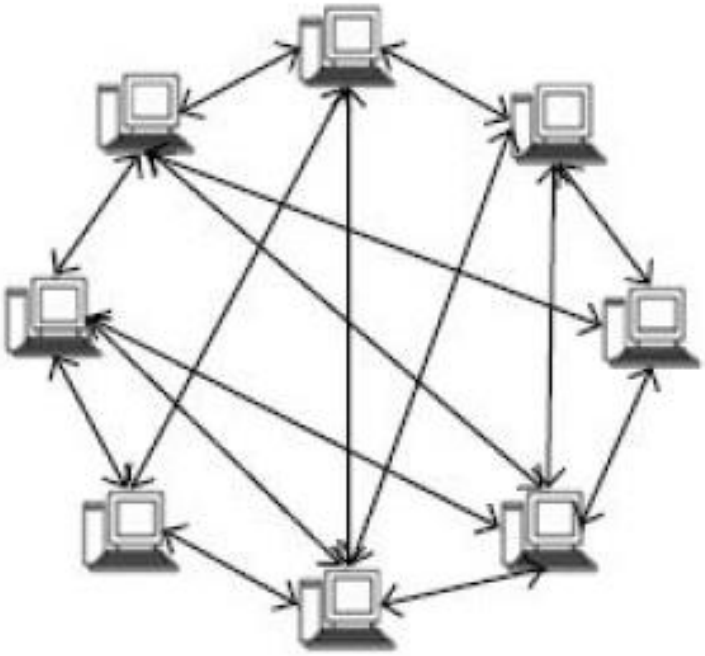
Definition:

Blockchain is a peer-to-peer, distributed ledger that is cryptographically-secure, append-only, immutable (extremely hard to change), and updateable only via consensus or agreement among peers.



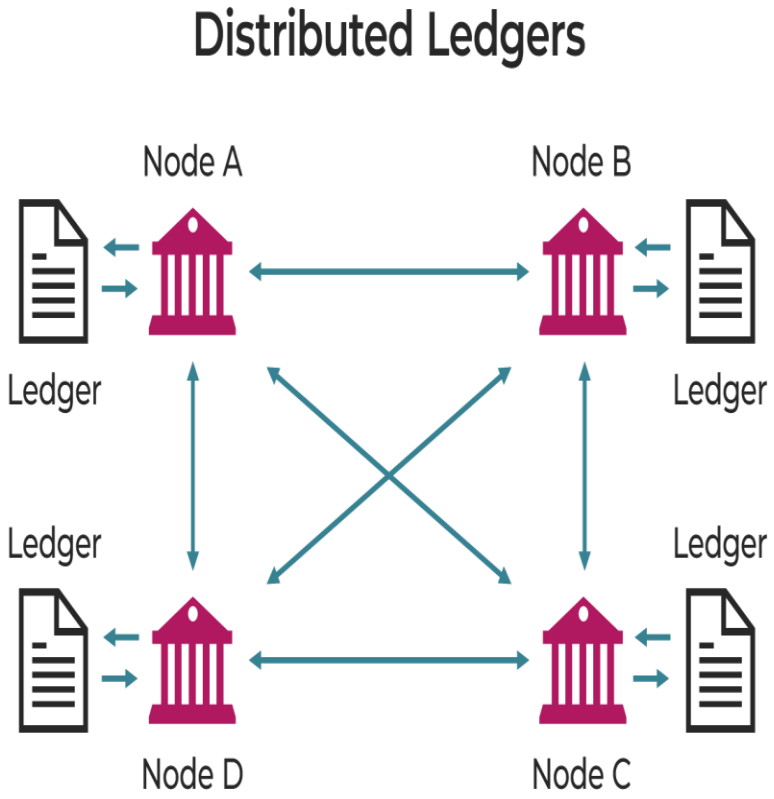
- Distributed Peer-to-Peer Network: There is no central controller in the network, and all participants talk to each other directly. This property allows for cash transactions to be exchanged directly among the peers without a third-party involvement, such as by a bank.

Types of Ledgers
Types of Nodes
Types of Blockchain
Layered Architecture
CAP Theorem



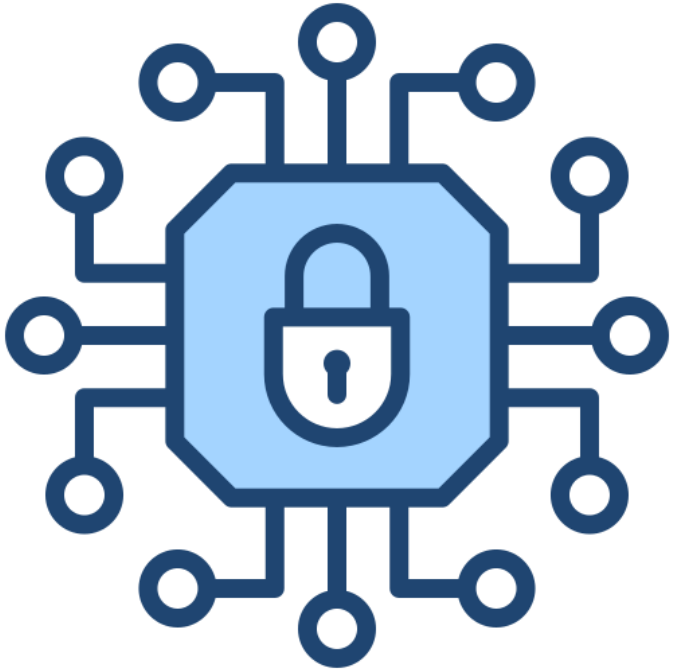
- Immutable Ledger: Which simply means that a ledger is spread across the network among all peers in the network, and each peer holds a copy of the complete ledger.

Decentralization
Distributed Ledger Technology(DLT)



- Cryptographically-Secure: Which means that cryptography has been used to provide security services which make this ledger secure against tampering and misuse. These services include non-repudiation, data integrity, and data origin authentication.

Cryptography
Hash Mechanisms
NONCE



- Append-only: Which means that data can only be added to the blockchain in *time-ordered sequential order*. This property implies that once data is added to the blockchain, it is almost impossible to change that data and can be considered practically immutable.
- Updateable Only Via Consensus: This is what gives it the power of decentralization. In this scenario, no central authority is in control of updating the ledger. Instead, any update made to the blockchain is validated against strict criteria defined by the blockchain protocol and added to the blockchain only after a consensus has been reached among all participating peers/nodes on the network.

Etherium

DApp

Smart Contracts

HyperLedger

Fundamental Components of Blockchain

- Node
  - Full Node
  - Partial or Lightweight or Light Node
- Ledger
  - Public v/s Distributed v/s Decentalized Ledger
  - Permissioned v/s Shared
- Wallet
- Nonce
- Hash
- Mining
- Consensus Protocol

Node

- A node is an electronic device (computers, mobile devices, servers, etc.) that is connected to the internet.
- In blockchain parlance, any computer or hardware device that is connected to the blockchain network is a node.
- As a principle, blockchain is open to all. Anyone can join the network, i.e., be a node and participate in the blockchain network to validate and create blocks.

A Full Node

- The node maintains a full copy of the transaction history of the blockchain.
- Computers run full nodes to help synch the blockchain.
- They also help the network by processing and accepting transactions/blocks, validating those transactions/blocks, and then broadcasting them to the network.
- **A full node may or may not be a miner node.**

A Partial or Lightweight or Light Node

- Nodes maintain only a partial copy of the ledger as they could be early users or those who do not have sufficient disk space for the full blockchain.
- Light nodes download only the block headers to validate the authenticity of transactions using a method called SPV or Simplified Payment Verification. They rely on the full nodes for the latest headers, account balance, and any transaction that affects their wallet.

Ledger

- A ledger, in blockchain technology, refers to a digital database of information that is immutable. Blockchain is commonly referred to as a public distributed decentralized ledger.

Ledger Is Public

- Anyone in the blockchain has access to the ledger and can read or verify the transactions therein.

Ledger Is Distributed

- All the nodes in the blockchain network have a copy of the blockchain ledger. The traditional database works in a client–server environment, while the blockchain works on the principle of replication in every node.

Ledger Is Decentralized

- Blockchain protocols are built such that no one node or group of nodes has excessive control over the ledger. There is no central control; hence it is decentralized with no single point of failure. So,while the traditional database works on central principle integrity (only the central body can validate the record), the blockchain works on the principle that anybody can validate the records.
- Also, the traditional database works on the CURD (Create, Read, Update, Delete) principle. In contrast, the blockchain works in the principle of Append-only, i.e., blocks are only added to the existing blockchain

Wallet

- A Wallet in the blockchain world is a digital wallet that allows users to manage cryptocurrency like bitcoin, litecoin, ether, etc. With a blockchain wallet, one can receive and send cryptocurrency.
- The term “wallet” is a misnomer, as real money is not stored.
- The wallet provides all the features that are needed for a safe, easy and secure transfer of funds between two parties.

Hot wallets

- A hot wallet offers online storage that you can access from a computer, phone, or tablet. A hot wallet has a security risk because it’s stored on the internet and is more susceptible to cyber-attacks.

Cold wallets

- A cold wallet doesn’t connect to the internet. You can store your cryptocurrency in an external drive, such as a USB device. You’ll receive a keycode to keep in a safe place. You lose the keycode, you may lose your cryptocurrency

Wallet (cont..)

Privacy Is Maintained

- Whenever a user creates a wallet, the public and private key associated with the wallet is also generated. It can be compared to how your email account works.
- The public key is like email id, and the private key is your email id password. Just like you share your email id to receive an email, you share your public key to receive funds.However, your identity and personal details are kept private.

Transactions Are Secure

- The private key is used to send funds as well as to open encrypted
- messages. This keeps the transactions secure.

Ease of Usage

- Wallets can be installed and accessed from the web, desktop, or any mobile device. Transfer of funds is relatively instantaneous, without any geographical constraints or intermediaries like banks.

Currency Conversion

- Wallets help you to transact across various types of cryptocurrencies like BTC, ETH, XMR, LTC, and others, without worrying about the currency conversion.

Nonce

- The blockchain header contains 32-bit whole number called a NONCE. The nonce is randomly generated when a block is created, which then generates a block header hash.
- The hash is a 256-bit number wedded to the nonce. It must start with a huge number of zeroes (i.e., be extremely small).

Hash

- A hash function can take data of any size, perform an operation on it, and return a “hash” that is a data of a fixed size.
- Inthe Bitcoin, blockchain hashes are 256 bits or 64 characters. The hashing algorithm used in the blockchain is called SHA-256.

Miners & Mining

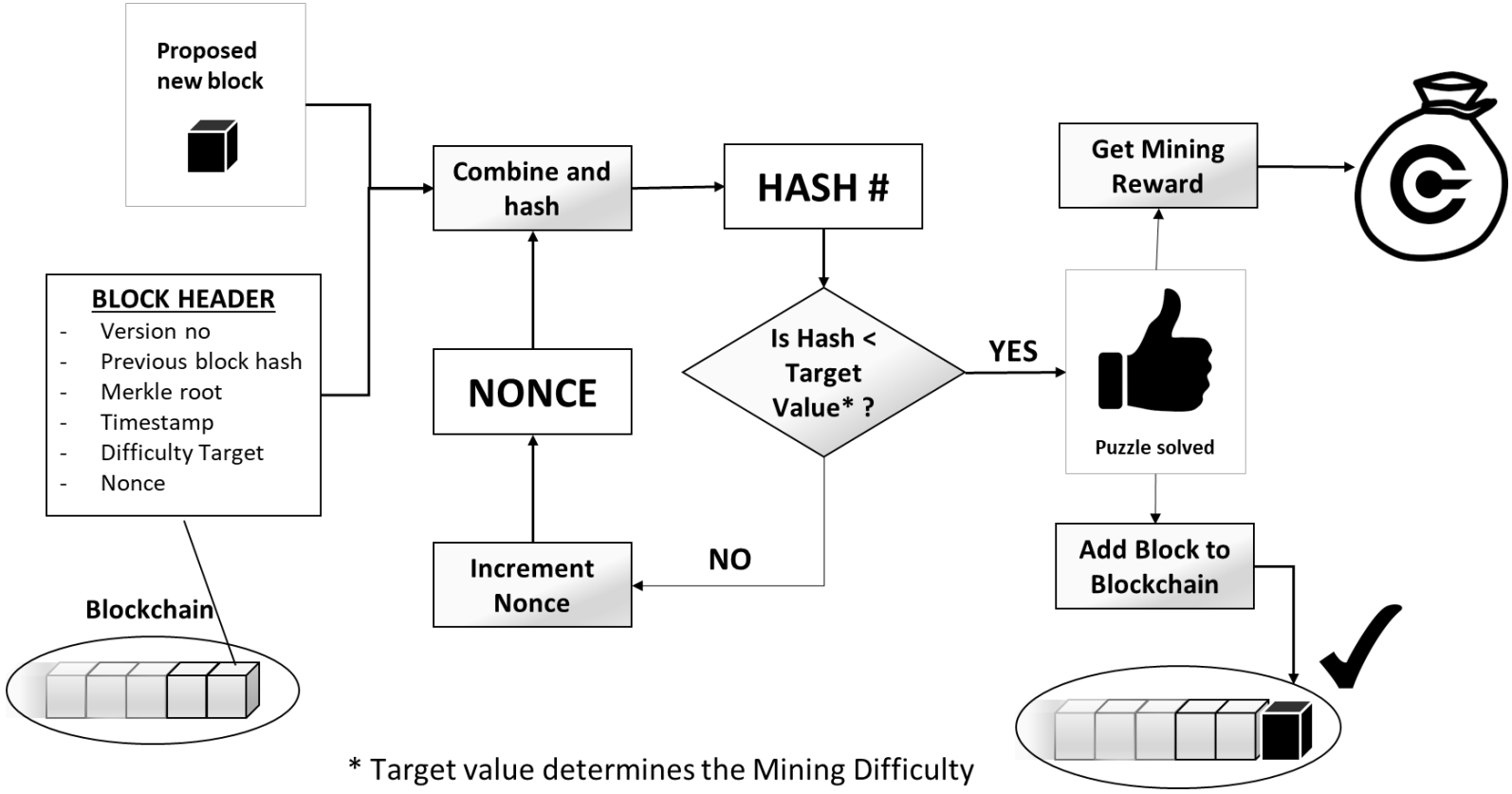
- Miners create new blocks on the chain through a process called mining or forging.
- In a blockchain, every block has its own unique nonce and hash, but also references the hash of the previous block in the chain.
- Miners use special software to solve the incredibly complex math problem of finding a nonce that generates an accepted hash. Because the nonce is only 32 bits and the hash is 256, there are roughly four billion possible nonce-hash combinations that must be mined before the right one is found.
- Making a change to any block earlier in the chain requires re-mining not just the block with the change, but all of the blocks that come after. This is why it's extremely difficult to manipulate blockchain technology.
- When a block is successfully mined, the change is accepted by all of the nodes on the network and the miner is rewarded financially

Mining Process

Mining – the mechanism to generate the hash

- The mechanism needs to be complicated enough, to make the blockchain **tamper proof**
- **Bitcoin Mining:** $H_k = Hash(H_{k-1} \,||\, T \,||\, Nonce)$
- Find the nonce such that H_k has certain predefined **complexity** (number of zeros at the prefix)

The header contains mining statistics – timestamp, nonce and difficulty



* Target value determines the Mining Difficulty

Consensus Protocol

- Consensus protocols are a set of rules whereby nodes in a network can achieve agreement on the data value or state of the network such that it benefits the network as a whole and does not focus on individual interests.
- In the decentralized world of Blockchain technology, all the participating nodes must agree on a single source of truth.
- Consensus protocols are used in blockchain to ensure that all transactions are validated before being added to the blockchain.
- **Smart contracts** are digital contracts(i.e Encoded in programming language) stored on a blockchain that are automatically executed when predetermined terms and conditions are met.
- Consensus is essential for the proper functioning of smart contracts on a blockchain

- The consensus algorithms available today, or that are being researched in the context of blockchain, are presented here.
- The following is not an exhaustive list, but it includes all notable algorithms.

Consensus Mechanism
Proof of Work
Proof of State
Proof of Elapsed Time
Proof of Activity
Proof of Burn
Proof of Capacity
Proof of Importance
And others....

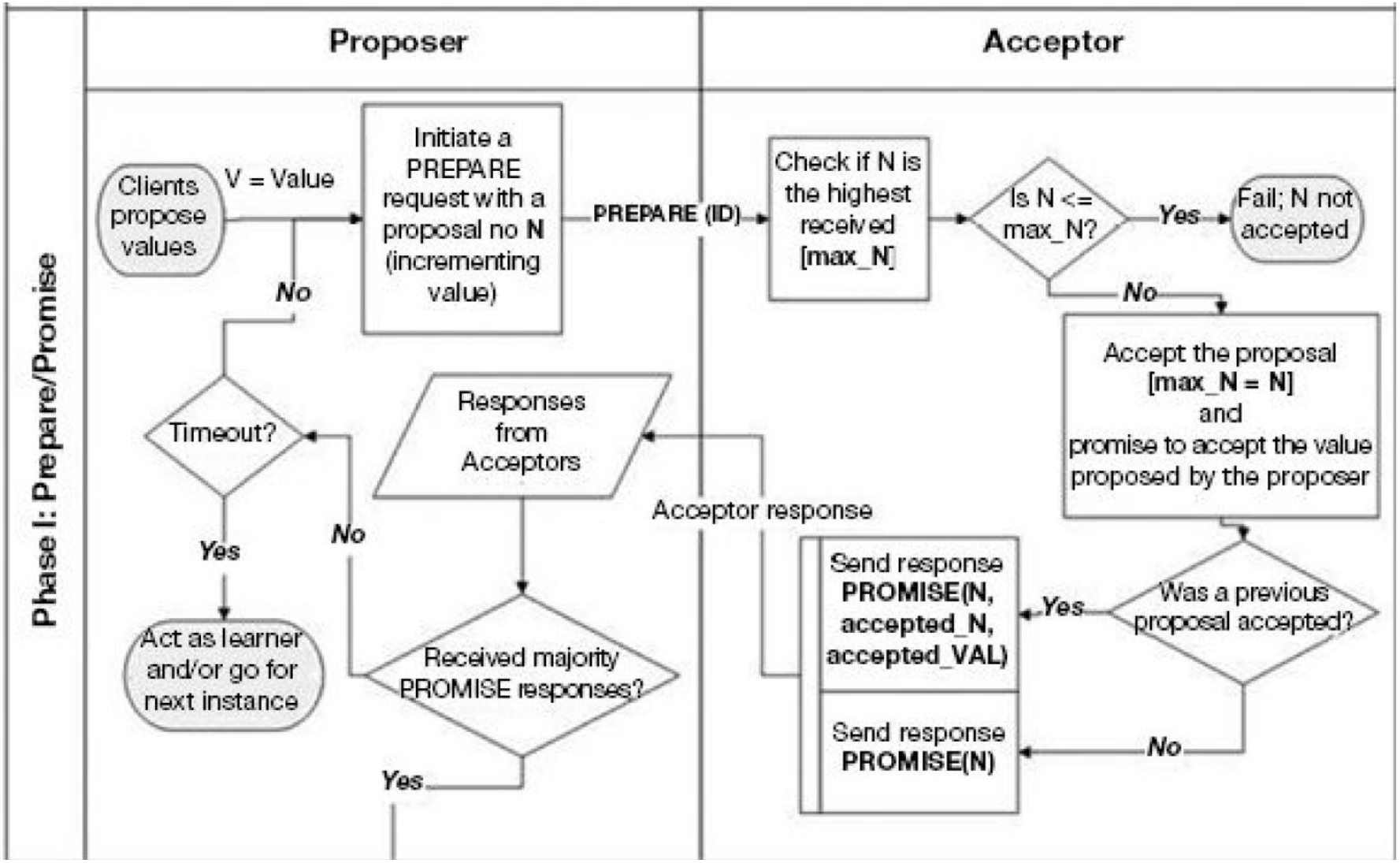
- **Proof-of-Work (PoW)** algorithm is used in Bitcoin protocol, while
- **Proof-of-Stake (PoS)** algorithm is in the Ethereum Casper Protocol.

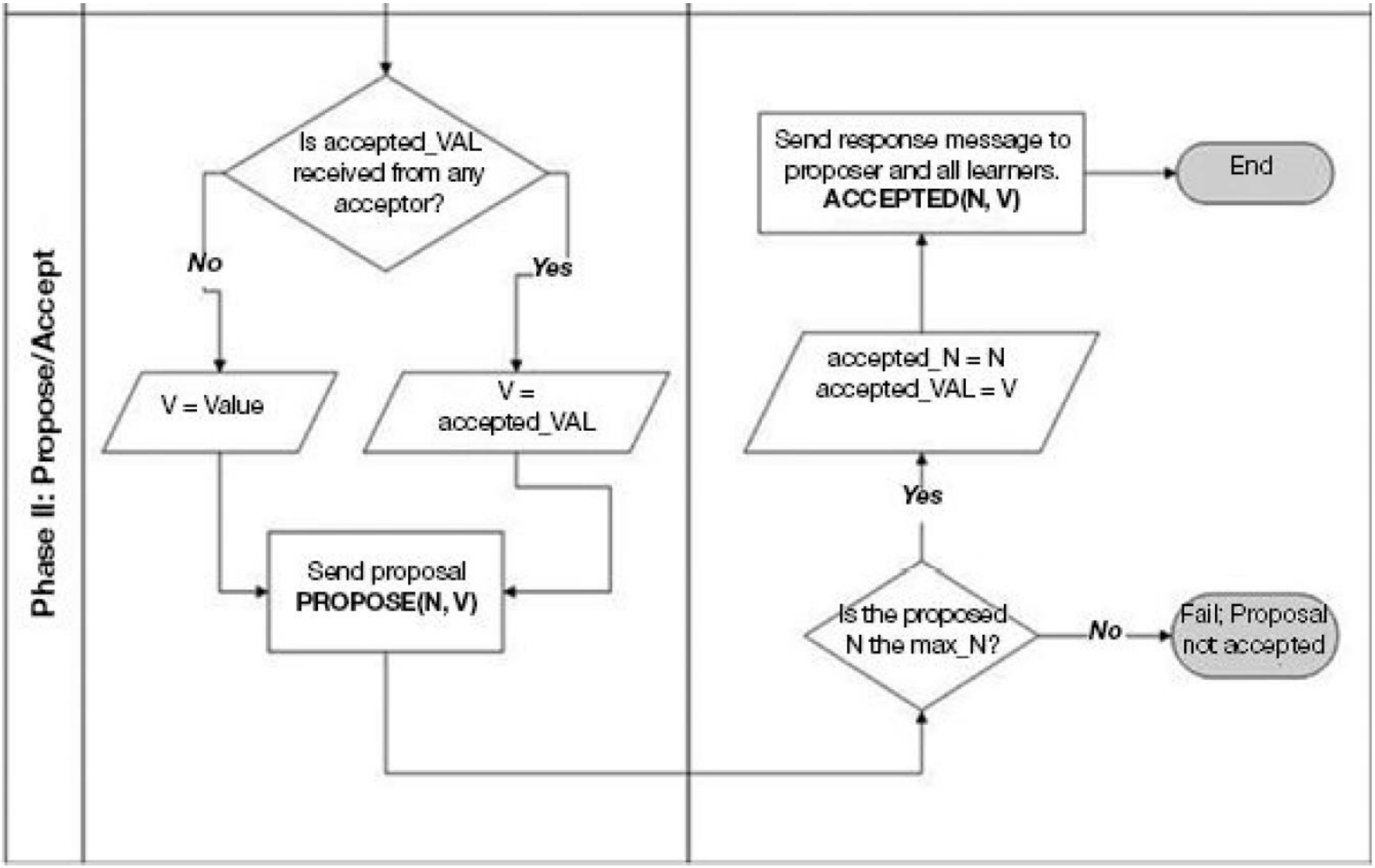
Types of consensus mechanisms

- All consensus mechanisms are developed to deal with faults in a distributed system and to allow distributed systems to reach a final state of agreement. There are two general categories of consensus mechanisms. These categories deal with all types of faults (fail stop type or arbitrary). These common types of consensus mechanisms are as follows:
  - **Traditional Byzantine Fault Tolerance (BFT)-based**: With no compute-intensive operations, such as partial hash inversion (as in Bitcoin PoW), this method relies on a simple scheme of nodes that are publisher-signed messages. Eventually, when a certain number of messages are received, then an agreement is reached.
  - **Leader election-based consensus mechanisms**: This arrangement requires nodes to compete in a leader-election lottery, and the node that wins proposes a final value. For example, the PoW used in Bitcoin falls into this category.
- Many practical implementations of consensus protocols have been proposed. **Paxos** is the most famous of these protocols. It was introduced by Leslie Lamport in 1989.
- With Paxos, nodes are assigned various roles such as Proposer, Acceptor, and Learner.

PAXOS Consensus in Distributed Systems

- Some of the challenges faced in distributed systems are Clock drift, Concurrency, Message Transmission, Component Failure.
- PAXOS was the first real-world fault-tolerant consensus algorithm introduced by Lynch and Liskov in the 1990s and later mathematically proven by Leslie Lamport and used by internet companies like Google and Amazon to build their distributed services.
- The primary PAXOS mechanism works under the principle that if the majority of the nodes agree on a value, then consensus is reached. It has three roles:
  - **1) Proposer**: A proposer receives client requests called ‘values’ and sends these proposed values to acceptors.
  - **2) Acceptor**: Receives messages from proposers and learners. They view the proposed values and inform the proposer whether they accept or reject the proposed value. They also inform the proposer if another value was already accepted.
  - **3) Learner**: Listens to all the acceptor’s decisions and delivers values in an ordered sequence. If any gap is found, the learner should contact the acceptors and repeat the decision. For example, say learners noted IDs 1 to 6, and the next instance delivered is value 8. The learner reverts to acceptors to repeat the procedure.

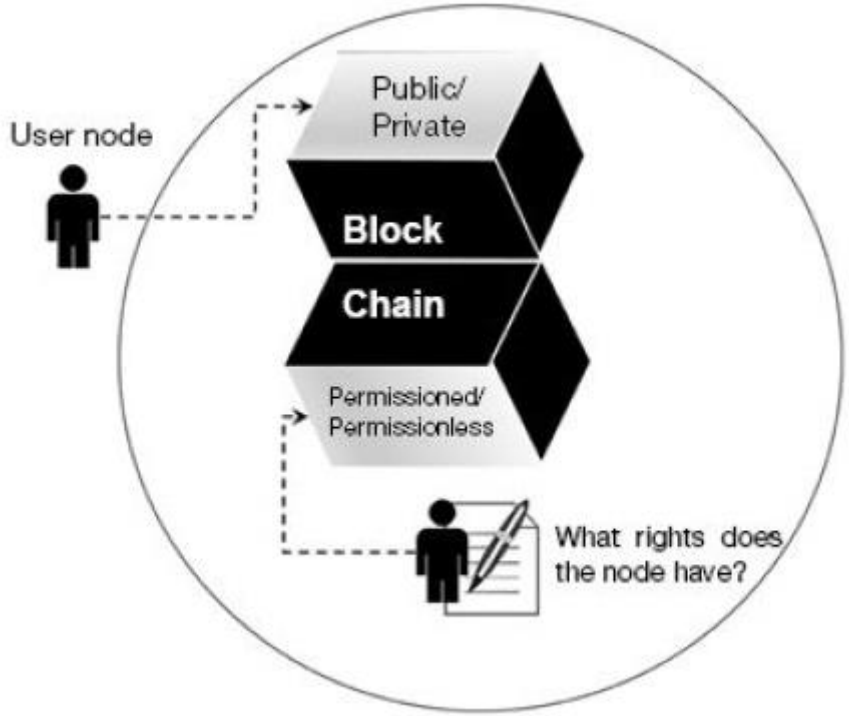

Phase I: Prepare/Promise
Proposer
Acceptor
Clients propose values
V = Value
Initiate a PREPARE request with a proposal no N (incrementing value)
PREPARE (ID)
Check if N is the highest received [max_N]
Is N <= max_N?
Yes
Fail; N not accepted
No
Accept the proposal [max_N = N] and promise to accept the value proposed by the proposer
Was a previous proposal accepted?
Yes
Send response PROMISE(N, accepted_N, accepted_VAL)
No
Send response PROMISE(N)
Acceptor response
Responses from Acceptors
Timeout?
No
Yes
Act as learner and/or go for next instance
Received majority PROMISE responses?
Yes


Phase II: Propose/Accept
Is accepted_VAL received from any acceptor?
No
Yes
V = Value
V = accepted_VAL
Send proposal PROPOSE(N, V)
Is the proposed N the max_N?
No
Fail; Proposal not accepted
Yes
accepted_N = N
accepted_VAL = V
Send response message to proposer and all learners. ACCEPTED(N, V)
End

Classification of Blockchain

Classification of Blockchain

- Public v/s Private v/s Consortium Blockchains v/s Hybrid Blockchains
- Permissionless v/s Permissioned



Public Blockchains

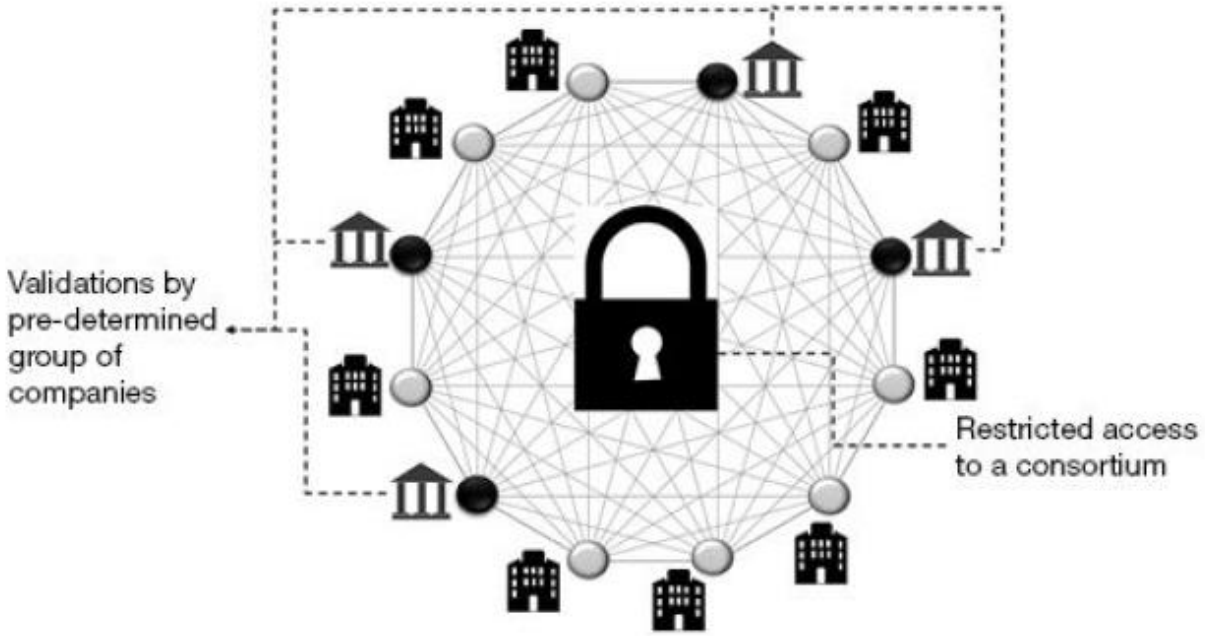
- Public blockchains are open, decentralized networks of computers accessible to anyone wanting to request or validate a transaction (check for accuracy).
- Those (miners) who validate transactions receive rewards.
- Public blockchains use proof-of-work or proof-of-stake consensus.
- Example : Bitcoin and Ethereum

Private Blockchains

- A Private Blockchain is just like a relational database i.e. fully centralized and owned by a single organization.
- Private blockchains are not open, they have access restrictions.
- People who want to join require permission from the system administrator.
- Example: Hyperledger is a private, permissioned blockchain.

Consortiums Blockchain

- Validation is conducted by known and identified members of the limited network of nodes
- Greater privacy since the information from verified blocks is not exposed to the public.
- There are no transaction fees.
- A consortium platform is more flexible.
- Example: Voting-based system, it ensures low latency and superb speed.



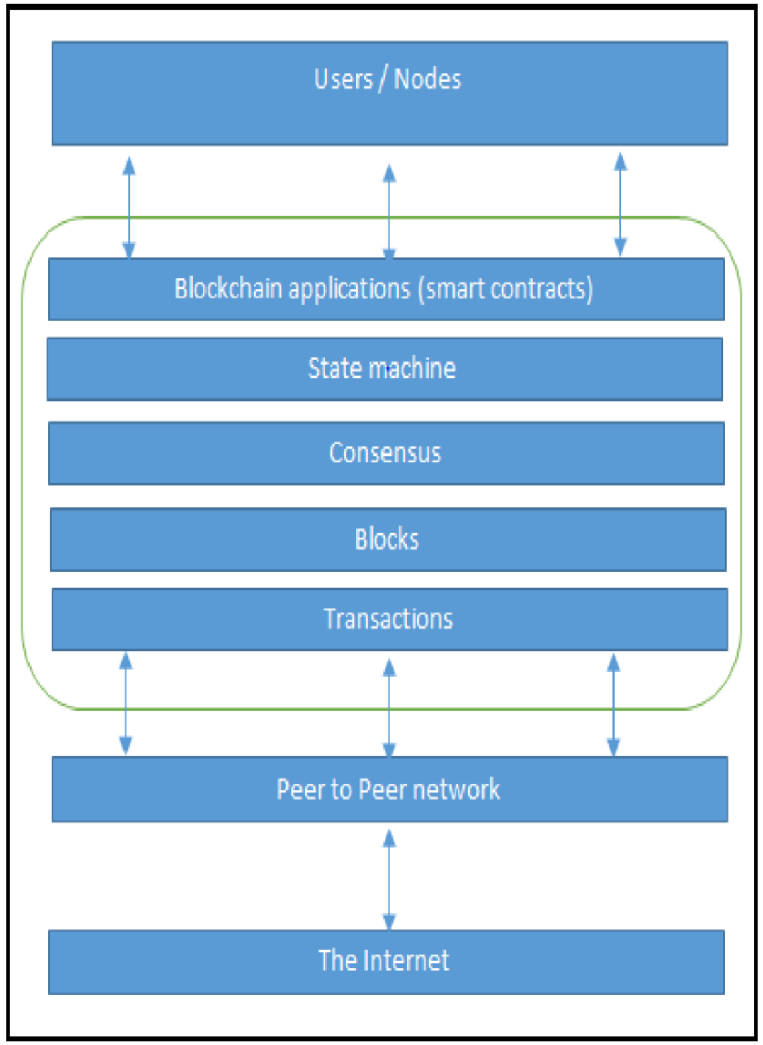
Hybrid Blockchain

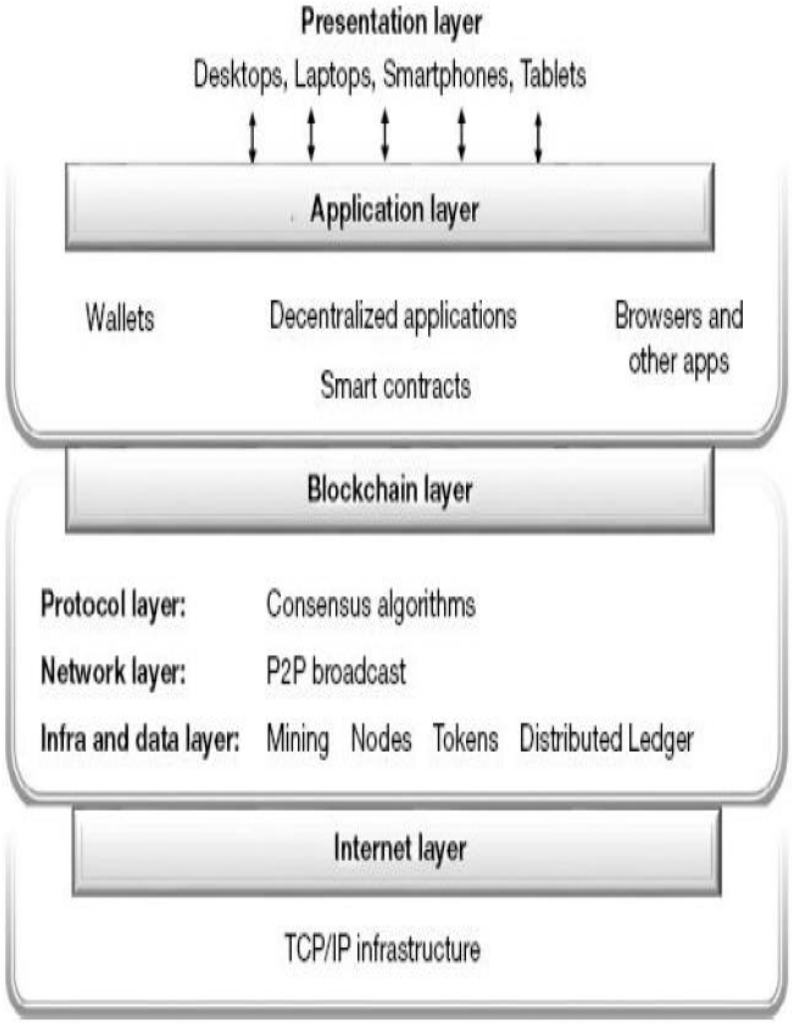
- Hybrid blockchain is best defined as a combination of a private and public blockchain.
- It has use-cases in an organization that neither wants to deploy a private blockchain nor public blockchain and simply wants to deploy both worlds' best.
- Example : Dragonchain, XinFin's Hybrid blockchain

Permissionless v/s Permissioned

- This category is based on the type of rights the user or node has within the blockchain network. The rights, if any, are defined by a central entity or group of entities.
- A Blockchain is considered **permissionless** if no such control entity exists, and all the nodes have equal rights to the network, i.e., they can all read, receive and send transactions and participate in the consensus mechanism.
- In a **permissioned** blockchain, the central entity or group restricts the roles that the nodes can play. It can vary from nodes having rights to only initiate transactions to those who validate transactions and to still others that deploy or execute smart contracts. In other words, only selected nodes will participate in the consensus mechanism for permissioned blockchain.
- In contrast, in a permissionless blockchain, all or majority nodes in the network need to agree on the validity of a record collectively.

Blockchain Layered Architecture





The Application Layer

- This is the layer than combines the business logic with user interactions.
- It consists of the Decentralized Application, better known as DApps running on a P2P network. It is the DApps that sets the communication between the Presentation and Blockchain layers. DApps connects to the Blockchain via Smart Contracts.

The Blockchain Layer

- This is where the core of the blockchain subsists. It consists of the consensus algorithms (PoW, PoS, pBFT, etc.), the medium, and interface for the P2P network that decide how data is packetized and transmitted between peers.
- It controls the mining layer, protocols that decide the consensus methods and participation, nodes that execute protocols, and the distributed ledger.

The Internet Layer

- Blockchain works over the internet. This layer ties all the networks together, i.e., the computers, IoT devices, smartphones, etc.

CAP Theorem

- **CAP theorem**, also known as Brewer's theorem, was introduced by Eric Brewer in 1998 as conjecture.
- In 2002, it was proven as a theorem by Seth Gilbert and Nancy Lynch.
- The theory states that any distributed system cannot have consistency, availability, and partition tolerance simultaneously.
  - **Consistency** is a property which ensures that all nodes in a distributed system have a single, current, and identical copy of the data.
  - **Availability** means that the nodes in the system are up, accessible for use, and are accepting incoming requests and responding with data without any failures as and when required. In other words, data is available at each node and the nodes are responding to requests.
  - **Partition tolerance** ensures that if a group of nodes is unable to communicate with other nodes due to network failures, the distributed system continues to operate correctly. This can occur due to network and node failures.

- To achieve fault tolerance, replication is used. This is a standard and widely-used method to achieve fault tolerance.
- Consistency is achieved using consensus algorithms in order to ensure that all nodes have the same copy of the data. This is also called **state machine replication**. The blockchain is a means for achieving state machine replication.
- In general, there are two types of faults that a node can experience. Both of these types fall under the broader category of faults that can occur in a distributed system:
  - **Fail-stop fault**: This type of fault occurs when a node merely has crashed. Failstop faults are the easier ones to deal with of the two fault types. Paxos protocol, introduced earlier in this chapter, is normally used to deal with this type of fault. These faults are simple to deal with.
  - **Byzantine faults**: The second type of fault is one where the faulty node exhibits malicious or inconsistent behavior arbitrarily. This type is difficult to handle since it can create confusion due to misleading information. This can be a result of an attack by adversaries, a software bug, or data corruption. State machine replication protocols such as PBFT was developed to address this second type of faults.
- In blockchains, consistency is sacrificed in favor of availability and partition tolerance. In this scenario, **Consistency** (**C**) on the blockchain is not achieved simultaneously with **Partition tolerance** (**P**) and **Availability** (**A**), but it is achieved over time. This is called eventual consistency, where consistency is achieved as a result of validation from multiple nodes over time. The concept of mining was introduced in Bitcoin for this purpose.

Benefits of Blockchain

- Decentralization → copy is maintained by all
- Transparency and trust → better communication between nodes
- Immutability → all transactions are made auditable
- High availability → removes Single Point of failure
- Highly secure → consensus mechanism & complex security algorithms
- Cost saving → no intermediaries

Limitations of Blockchain

- Scalability - Significant computing power is expended by miners leading to substantial energy consumption and wastage. Hence, it is not suitable for organizations that require instant transaction results within milliseconds.
- Adaptability – If a time-tested and fully functional database and the operational network are already in place, the benefits of replacing or introducing blockchain may not produce the required return on investment.
- Not every node has the capacity to maintain and run a full copy of the blockchain. This can potentially affect consensus and immutability.
- Privacy– Stronger players (nodes with higher computing power or with pooling) can take control of the network, impacting decentralization. In smaller blockchains, there is a risk of a 51% attack.
- Regulation standard

Applications of Blockchain

- Currency – Bitcoin
- IoT
- Health
- Finanace
- Media
- Aviation
- Voting
- Identity Management
- Stock trading
- Agriculture