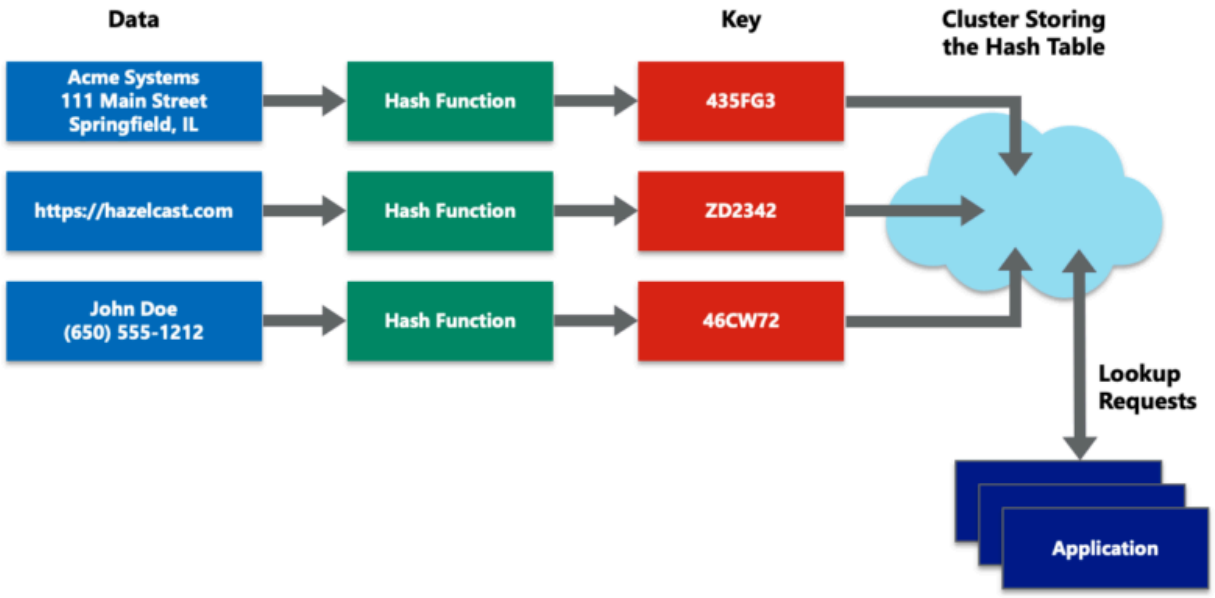
Distributed Hash Table

A distributed hash table is a decentralized data store that looks up data based on key-value pairs. Every node in a distributed hash table is responsible for a set of keys and their associated values. The key is a unique identifier for its associated data value, created by running the value through a hashing function. The data values can be any form of data.

Distributed hash tables are decentralized, so all nodes form the collective system without any centralized coordination. They are generally fault-tolerant because data is replicated across multiple nodes. Distributed hash tables can scale for large volumes of data across many nodes.



Advantages of Distributed Hash Table

- **Scalability** – DHTs are highly scalable, as they can store and retrieve large amounts of data without requiring a central authority or server to manage the system. This makes DHTs well-suited for large-scale distributed systems.

- **Efficiency −** DHTs provide an efficient way to store and retrieve data, as they use keys to determine the location of the data in the network. This allows DHTs to quickly locate and retrieve data without having to search the entire network.

- **Fault tolerance −** DHTs are highly fault-tolerant, as they can handle node failures without requiring a central authority to manage the system. If a node fails, the data it was responsible for can be redistributed among the remaining nodes in the network.

- **Decentralization −**DHTs are decentralized, as there is no central authority or server that controls the network. This makes DHTs more resilient and less vulnerable to downtime or attack.

- **Security −** DHTs can provide a secure way to store and retrieve data, as the data is distributed across multiple nodes in the network rather than being stored in a single location. This makes it more difficult for attackers to access or modify the data.

Disadvantage of Distributed Hash Table

- **Complexity −** DHTs can be complex to implement and maintain, as they require a large number of nodes to function properly. This can make DHTs more challenging to manage and maintain than other types of distributed systems.

- **Performance −** In some cases, DHTs may not perform as well as other types of distributed systems, particularly when the system is under heavy load or when the network is large and complex.

- **Security −** While DHTs can provide a secure way to store and retrieve data, they can also be vulnerable to certain types of attacks, such as distributed denial of service (DDoS) attacks or Sybil attacks.

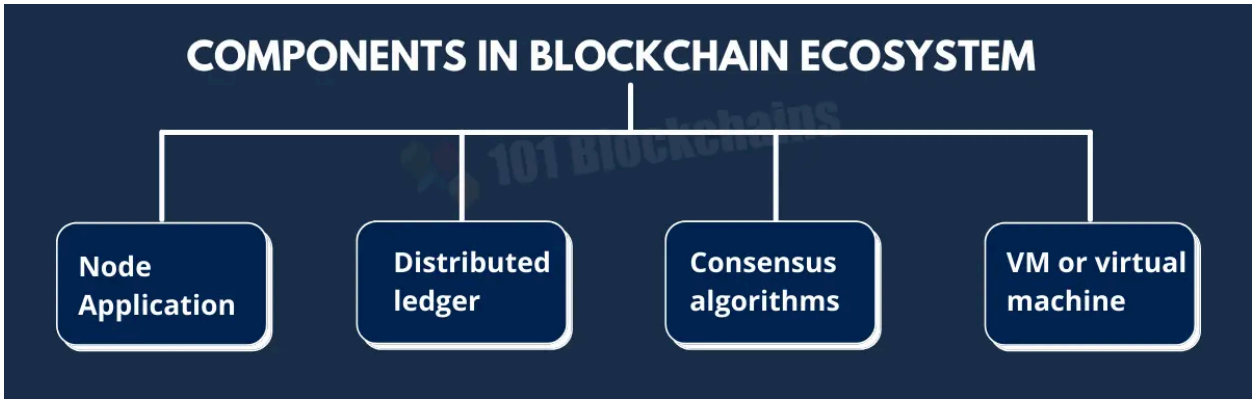
- **Compatibility −** DHTs may not be compatible with all types of data or applications, as they may require specific data structures or formats in order to function properly.

- **Limited functionality −** DHTs are primarily designed for storing and retrieving data, and may not provide additional functionality beyond these basic capabilities.

Blockchain and full Ecosystem Decentralization

Blockchain is a distributed ledger technology that allows the storage of data in blocks of information, with each block containing batches of transactions from one moment in time. Its open, decentralized and secure nature allows blockchain to disrupt traditional transactional systems.Blockchain has the potential to transform the ecosystem by becoming an indispensable part of our daily lives. Understanding the underlying technology and its impact is imperative for those looking to capitalize on this new era.

Blockchain Ecosystem



- **Node application**: It is a particular internet application that every internet-connected computer must download for participating in a blockchain ecosystem. After node application installation, a user becomes a participant in the blockchain network. Once one has a node application installed, it can participate in the ecosystem.

- **Distributed ledger**: this is the logical component and the data structure that is managed inside the node application. Once the node application is installed, one can view the respective ledger contents from that ecosystem. One can run as many node applications as likes and is permitted to use, and each will participate in their respective blockchain ecosystems.

- **Consensus algorithms**: The consensus algorithm is implemented as a part of the node application in the blockchain ecosystem. They provide the rules of the game for how the ecosystem will arrive at the single view of the ledger. Different ecosystems have different ways of attaining consensus. There are different consensus algorithms like PoW, PoS, etc, each method qualifies nodes as honest in their own way before participating in the consensus-building process.

- **Virtual machine**: It is the representation of the computer environment created by a computer program and operated with instructions programmed in a language. The virtual machine implementation happens alongside the node application. For example, in the Ethereum blockchain ecosystem, the EVM resides inside the node application.

Some other important aspects are:-

Storage

Data can be stored directly in a blockchain, and with this fact it achieves decentralization. However, a significant disadvantage of this approach is that a blockchain is not suitable for storing large amounts of data by design. It can store simple transactions and some arbitrary data, but it is certainly not suitable for storing images or large blobs of data, as is the case with traditional database systems.

A better alternative for storing data is to use distributed hash tables (DHTs). DHTs were used initially in peer-to-peer file sharing software, such as BitTorrent, Napster, Kazaa, and Gnutella. DHT research was made popular by the CAN, Chord, Pastry, and Tapestry projects. BitTorrent is the most scalable and fastest network, but the issue with BitTorrent and the others is that there is no incentive for users to keep the files indefinitely.

Users generally don't keep files permanently, and if nodes that have data still required by someone leave the network, there is no way to retrieve it except by having the required nodes rejoin the network so that the files once again become available.

Two primary requirements here are high availability and link stability, which means that data should be available when required and network links also should always be accessible. Interplanetary File System (IPFS) by Juan Benet possesses both of these properties, and its vision is to provide a decentralized World Wide Web by replacing the HTTP protocol. IPFS uses Kademlia DHT and Merkle Directed Acyclic Graphs (DAGs) to provide storage and searching functionality, respectively.

The incentive mechanism for storing data is based on a protocol known as Filecoin, which pays incentives to nodes that store data using the Bitswap mechanism. The Bitswap mechanism lets nodes keep a simple ledger of bytes sent or bytes received in a one-to-one relationship. Also, a Git-based version control mechanism is used in IPFS to provide structure and control over the versioning of data.

There are other alternatives for data storage, such as Ethereum Swarm, Storj, and MaidSafe. Ethereum has its own decentralized and distributed ecosystem that uses Swarm for storage and the Whisper protocol for communication. MaidSafe aims to provide a decentralized World Wide Web.

BigChainDB is another storage layer decentralization project aimed at providing a scalable, fast, and linearly scalable decentralized database as opposed to a traditional filesystem. BigChainDB complements decentralized processing platforms and filesystems such as Ethereum and IPFS.

Communication

The Internet (the communication layer in blockchain) is considered to be decentralized. This belief is correct to some extent, as the original vision of the Internet was to develop a decentralized communications system. Services such as email and online storage are now all based on a paradigm where the service provider is in control, and users trust such providers to grant them access to the service as requested. This model is based on the unconditional trust of a central authority (the service provider) where users are not in control of their data. Even user passwords are stored on trusted third-party systems.

Thus, there is a need to provide control to individual users in such a way that access to their data is guaranteed and is not dependent on a single third party. Access to the Internet (the communication layer) is based on Internet Service Providers (ISPs) who act as a central hub for Internet users. If the ISP is shut down for any reason, then no communication is possible with this model.

An alternative is to use mesh networks. Even though they are limited in functionality when compared to the Internet, they still provide a decentralized alternative where nodes can talk directly to each other without a central hub such as an ISP.

Blockchain has revived the vision of decentralization across the world, and now concerted efforts are being made to harness this technology and take advantage of the benefits that it can provide.