APPLICATION OF MERKLE TREE IN BLOCKCHAIN

Hierarchical data representation in computer science uses trees where nodes containing values are linked to a subsequent list of reference nodes. Edges connect nodes and may or may not have child nodes. The top node is the root/ parent, nodes with the same parent are siblings, and nodes without children are leaf nodes. Bitcoin and cryptocurrencies use Merkle trees, also called hash trees, to encode and encrypt data contained in a Blockchain. Merkle tree was a concept presented by Ralph Merkle in 1979. These trees are created by calculating hashing pairs of nodes until only one pair is left. It is a complete binary tree, and each node is to hash the value from its child node. The tree is constructed using a bottom-up approach where every transaction is hashed then they are subsequently paired and hashed together until there is only one in the entire block. This now makes it easy to prove that a set of data was used to generate a root hash without necessarily having to store all the original data.

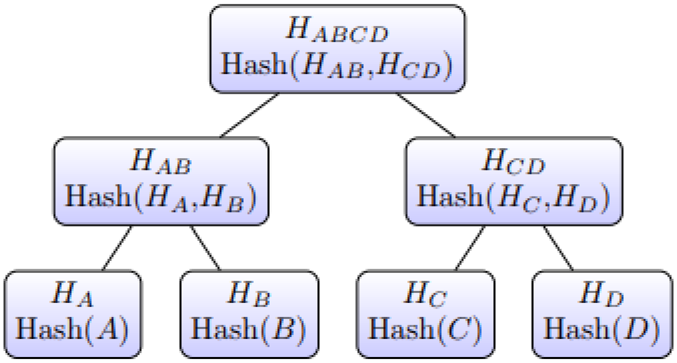

Figure 4: Simple binary hash tree, where H_{ABCD} is the Merkle root.

The above figure entails that if we need to clarify if data A was used to generate the root hash, we just check Hash$_B$ (data B) and Hash$_{CD}$ (data C and D). If any of the data changes, the root hash will automatically change. It provides a similar view of the Merkle tree indicating how each node hashes its value from its child node.

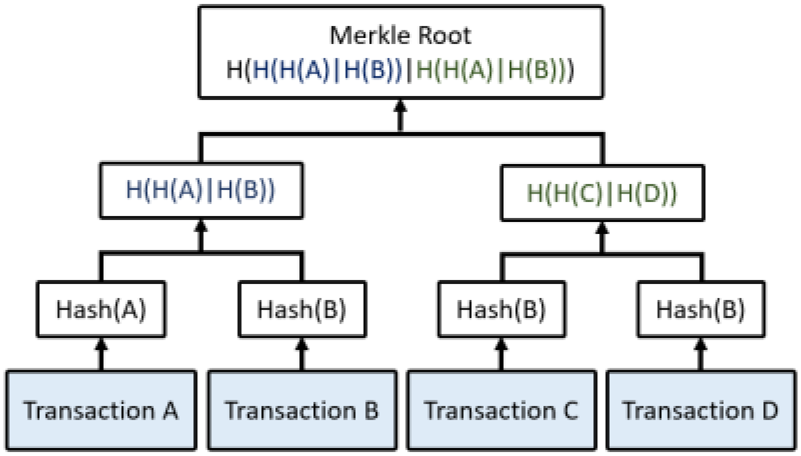

Figure 3. Merkle tree structure.

This concept can be applied to very large sets of data since it will not compromise on the scalability of the systems and will not require centralization of services. Merkle trees have been an essential key to data verification throughout the history of computers. Their structure helps to verify the consistency of data content. Its architecture helps to speed up security authentication in big data applications.

In Bitcoin and cryptocurrencies, Merkle trees serve as a summary of all the transactions that have taken place in that block. This is done by producing a digital fingerprint of the entire set of transactions. The Merkle root of a given block is stored in the block header and it is combined with other information. All this information is then hashed again to produce the block's hash which is not included in the relevant block, but in the next block (as the previous block's hash). To get the leaves of the Merkle tree, Bitcoin uses the transaction hash, that is, the transaction ID (TXID) of every transaction included in the block. As a reminder, the TXID is a unique string of characters given to every transaction that is verified and added to the Blockchain.