

# Here's Why Tezos Storage and TezEdge Are So Important

 [medium.com/burritt-research/heres-why-tezos-storage-and-tezedge-are-so-important-f5f6836b750e](https://medium.com/burritt-research/heres-why-tezos-storage-and-tezedge-are-so-important-f5f6836b750e)

March 23, 2021

## Poorly Managed Blockchain Storage is a Big Problem



At its heart, blockchain succeeds by keeping accurate records across a wide network of computers. This distributed ledger technology makes sure that all blockchains across the network are accurate and verifiable. It also makes sure that no one is cheating by adding or changing those records.

But if blockchain storage is not managed correctly, it's a big, big problem.

Here's an example:

Think of blockchain like a sophisticated spreadsheet of your checkbook. In your spreadsheet, you record income (hopefully) for monies you earn and expenses for things you buy. Because your income and expenses have running balances — that is, balances that update with each new income or expense — the spreadsheet, over time, piles on the transactions. And after a while, the spreadsheet gets pretty big.

Now, assume that you want to make doubly sure that your spreadsheet is accurate, so you send it to a friend at the end of each month. You ask them to check and then verify all the transactions. (This is a very close friend!) If everything checks out, they copy the spreadsheet to their computer for safekeeping and let you know everything is ok.

Now imagine that you have a whole group of (really close!) friends that heard about your idea and now want to do the same thing with their checkbook spreadsheets. Everyone sends them to each other, goes through the verification process, keeps a copy, and lets the spreadsheet owner know that everything is ok.

But there's a problem. Your spreadsheet checkbook system is now so popular that your friends get flooded with more and more spreadsheets that are bigger and bigger. And those files aren't just clogging up your friends' computers: They're also so big that sending them around to everyone is taking way too long.

Pretty soon everyone's computers and internet connections are all filled up. Your friends get mad and quit. They stop checking the spreadsheets and they stop sending them back.

In a nutshell, the system grinds to a halt and fails.

## **Managing Those Fat and Slow Spreadsheets**

---

You ask your friends what happened and, to a person, they said that there were too many files that were getting too big and moving too slowly.

In other words, your system failed because it couldn't store and manage all those fat spreadsheets without bogging everything down.

Blockchain faces the same problem.

Think of your "friends" in the above example as nodes and think of the "spreadsheets" as all the transactions on the blockchain. In order for the system to work, each node (friend) must quickly and efficiently verify the blockchain data (spreadsheet). If they don't, the blockchain won't work.

## **Blockchain's Decentralization Means Security and Speed**

---

Before we get into how we solve this blockchain data problem in general and on TezEdge nodes, your friends in our failed spreadsheet checkbook system said there was one big positive to the system besides the verification of accuracy.

And that was decentralization: After the system failed, there were so many copies of everyone's spreadsheets distributed amongst so many computers that it was easy to get your information back. All you had to do was look at someone else's copy of the spreadsheet and all your info was there.

Plus, having so many copies of the same spreadsheet also meant that it was virtually impossible to change the transactions (one of your friends may like to cheat!) without someone noticing. All you would have to do is compare two copies.

But because the network is decentralized, the network of computers is so large that there are tons of storage available to the blockchain.

That means that this decentralized data storage system also makes the blockchain super-fast and — even more importantly — secure. After all, as we saw above, there are so many copies of the blockchain on so many computers, it would be virtually impossible to cheat or mess with the blockchain.

## Solving the Size Problem

---

Ok. So we have all these computers with the same copy of the blockchain running very efficiently.

But if you think back about our spreadsheet checkbook problem, it wasn't just lack of efficiency that was the problem.

It was size.

As those checkbook spreadsheets grew, they were not only cumbersome to move around the network, they were also way too big.

Here's how TezEdge solves the problem.

Instead of sending the entire spreadsheet around your network of friends for verification, what if you were able to send *just the changes to the files*? Without a doubt, that would take care of a good portion of the size problem.

With blockchain this is done by taking blocks already checked and verified in the system and translating them into a string of letters and numbers. This string of letters and numbers — called a “hash” (for chopping and mixing data) — is much easier to move around the network quickly and safely. And because the translation is only one way — a “fingerprint” that can't be reversed — it's very secure.

Now if you have this hash of the old transactions in your spreadsheet — and you've made sure it hasn't changed — all you have to do is check the new transactions in order to verify the entire spreadsheet. Your work would be easier and more efficient. And your files would be much smaller. As a result, the system would have solved its storage problem.

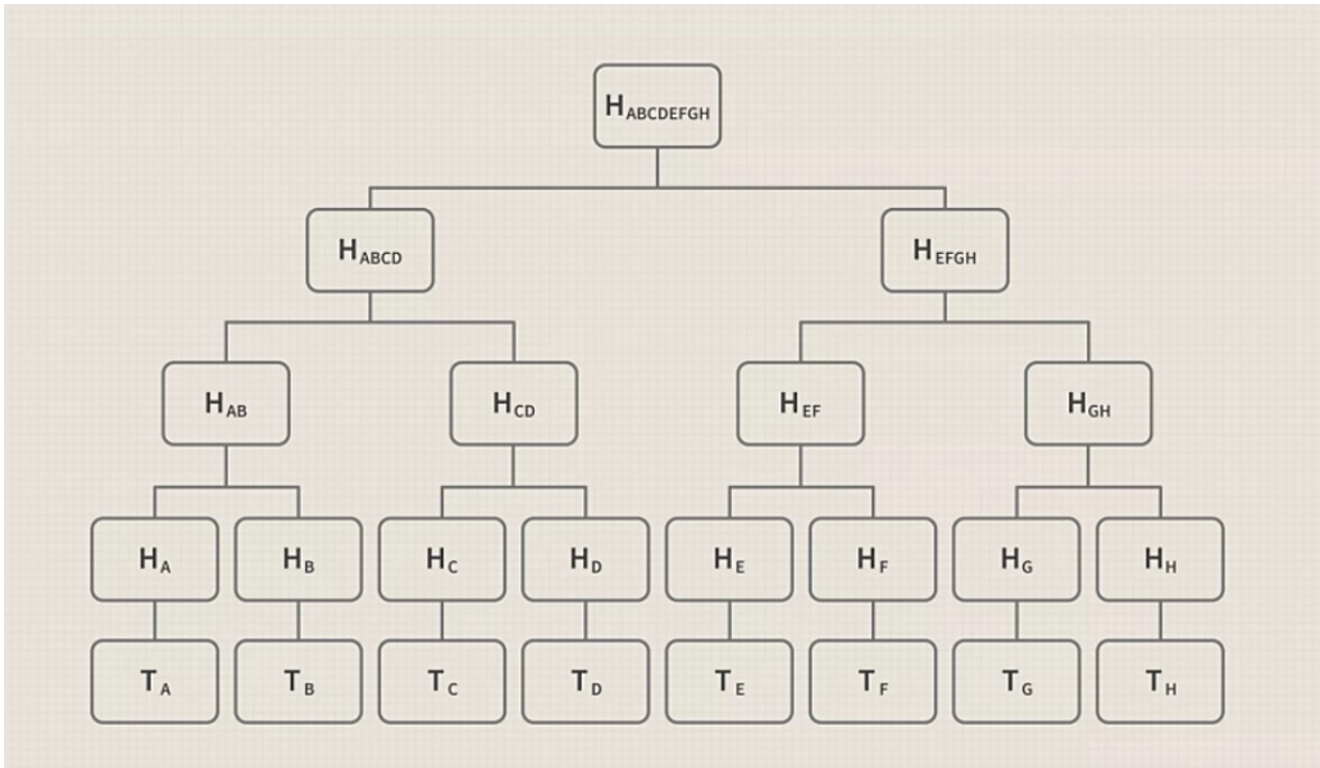
## Better Hash with Merkle Trees

---

What if we could take all those hashes that are storing all that already-verified information and combine them together? Would that make the blockchain even more efficient and reliable?

It would and it does. Through the use of Merkle Trees — named after the computer scientist Ralph Merkle — prior stored hashes are concatenated (combined) together to make a higher and more informed hash called the root.

In other words, Merkle Trees make better hash. Take a look:



Because Merkle trees can combine prior hashes (called branches and leaves) they can make retrieving information about transactions very quick and efficient. And because the amount of information moving around the network is small, the users can get the information they want without hogging a ton of bandwidth.

To get an even more vivid idea of how important Merkle trees are, imagine running a blockchain without hashes. Every time a new transaction wanted to be added to a block, the entire blockchain would have to be sent around to all the computers and get verified. This would require such a massive amount of computing power and bandwidth — bitcoin currently weighs a massive 334GB — that the blockchain would be essentially unusable.

## Even Better Hash with Merkle Storage

---

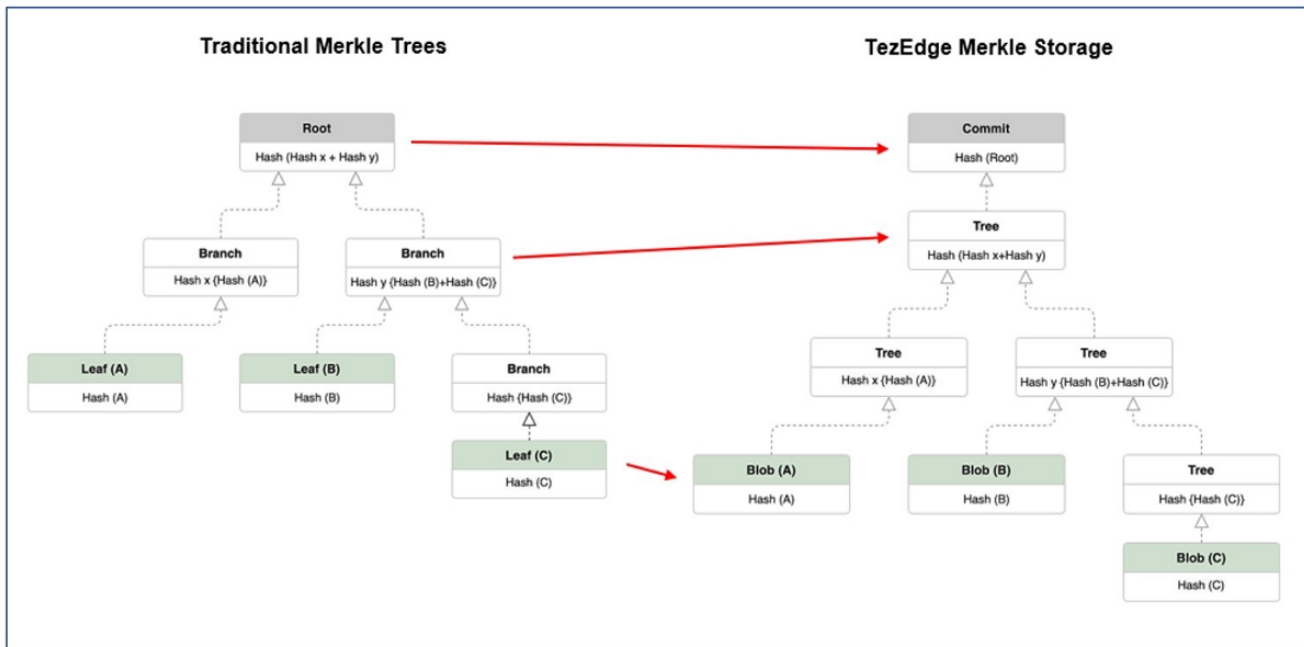
Using hashes to store data in blockchain is nothing new. But the way TezEdge does it is. By using a special form of hashing with Merkle trees, TezEdge nodes can process blockchain transactions very quickly.

Here's how.

First, TezEdge hashes transactions the same way that Irmin – Tezos' blockchain storage system – stores the data (a simple back-end database called RocksDB). And while Irmin is currently run with an OCaml implementation, the new TezEdge node uses Rust, which will increase both speed and security.

Second, TezEdge nodes use a key-value store called Merkle Storage. It combines Git-like commands – such as SET, GET, and COMMIT – with Merkle trees. The result is efficient data verification that is also easier to implement and debug.

Here's how traditional Merkle Trees and Merkle Storage look side-by-side:



As the diagram shows, with TezEdge Merkle Storage:

- Leaves become Blob Hash
- Branches become Tree Hash
- Root becomes Commit Hash (also known as context hash)

In order to add a block to the existing blockchain, all you have to do now is to repeat the list of Git-like actions in the same order to create an exact state of the blockchain. And this makes the process of adding a new block very fast.

To make things even better:

1. Merkle Storage uses heaptrack to improve allocations and optimize memory usage. As a result, data isn't getting space when it doesn't need it.
2. Instead of copying the entire tree all the way up to the Merkle Tree root, Merkle Storage simply finds the trees that need modification and does so using entries of blobs, trees and commits. While that doesn't preserve the entire tree state, the tree is easily rebuilt by repeating the prior Git-like commands. That saves memory and speeds things up.
3. Traditionally, actions against the blockchain were done continuously, requiring multiple database queries. With Merkle Storage actions are efficiently bundled together and applied in a single batch in memory and/or CPU cache. That makes action application even faster.

## Bottom Line

---

The key to a successful blockchain is the fast and efficient management of verifiable data. Without it, the decentralized block management process would grind to a halt very quickly. Capitalizing on the traditional hashing and Merkle Trees, TezEdge's Merkle Storage speeds up blockchain transactions by optimizing memory allocation, using Git-like commands, and batching application actions.

## More Info

---

For more information on this topic, take a look at this great article by

**[A deep dive into the Tezos storage — How the blockchain state is stored in the TezEdge node...](#)**

---

**[In a distributed network like blockchain, we want to be able to quickly verify that the data distributed from node to...](#)**

---

[medium.com](https://medium.com)

## **Burritt Research**

---

Custom Data Science Solutions



Written by



## **Wayne Burritt**

---

CEO of Burritt Research. Writer, developer, analyst. Passionate about blockchain and data science. burrittresearch.com



## **Burritt Research**

---

Since 2005, our team of developers, writers, and marketers has been helping people solve problems using data and blockchain. <https://burrittresearch.com/>

## **Learn more.**

---

Medium is an open platform where 170 million readers come to find insightful and dynamic thinking. Here, expert and undiscovered voices alike dive into the heart of any topic and bring new ideas to the surface. [Learn more](#)

## **Make Medium yours.**

---

Follow the writers, publications, and topics that matter to you, and you'll see them on your homepage and in your inbox. [Explore](#)

## **Share your thinking.**

---

If you have a story to tell, knowledge to share, or a perspective to offer — welcome home. It's easy and free to post your thinking on any topic. [Write on Medium](#)

[About](#)

[Help](#)

[Legal](#)

**●● Medium**