MACHINA

Immutability of Data and Storing Data On-Chain

The fascination behind immutable data storage lies in the intrinsic aspiration of human beings to express themselves without being personally beholden to any particular party; retroactively observe the changes and trickle of time without losing priceless information in the stream of entropy; collectively share the value of knowledge and preserve pristineness thereof.

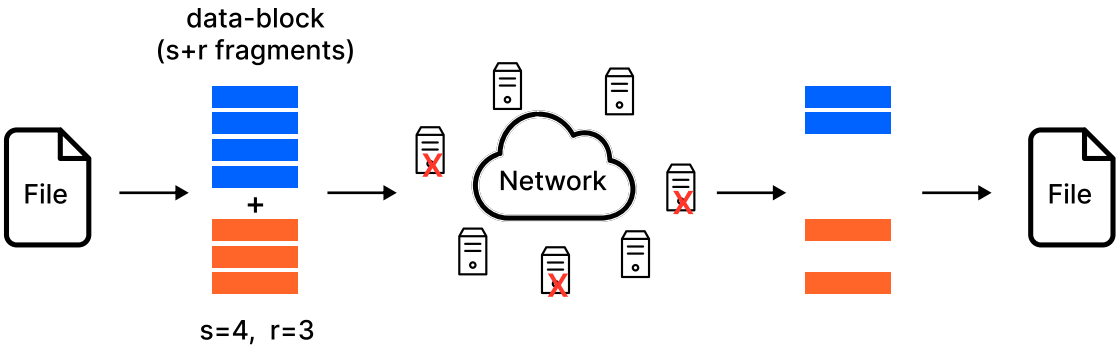
We are quite frequently met with the opposing force of censure and silencing that halt free speech and impair any nonconforming exploits. Blockchain technology in its primordial concept aims to upheave any oppressing instantiations and enable people to freely exchange, communicate, and retain contractual agreements to the predetermined immutable conditions.

Therefore data, its integrity, and availability are crucial aspects that ought to be preserved to the deserving degree. Machina strives to provide people with the on-chain, secure, and reliable way to store, transmit, and exchange data in an environment that is immutable and freely accessible, and/or readable.

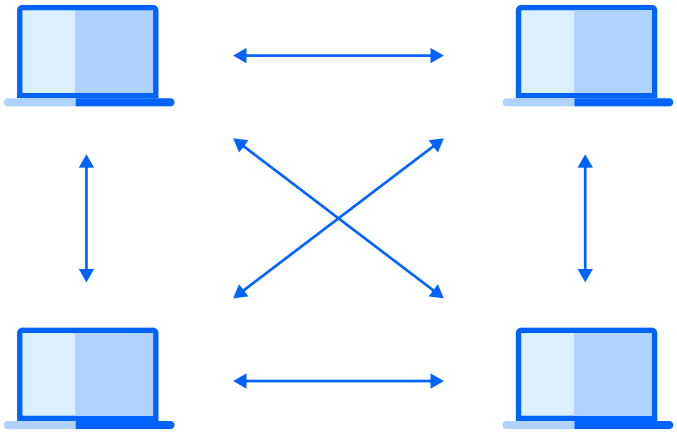
Storing data on the blockchain is a two-edged sword that requires a valid and sagacious approach toward securing and distributing it in a decentralized modus.
This dichotomy of blockchain data storage, namely the secured and decentralized nature of it, is achieved through the distribution of data across peers. Peer to peer storage approach is nothing new outside of the blockchain milieu.

Normally, it refers to the process of sharing stored data amongst several peers, that is several computers hold parts of data on their machines.
Data can be retrieved, restored, and integrity retained in the same way as traditional single server storage solutions.

Essentially peer to peer storage represents a viable alternative to data centers and to some capacity decreases the likelihood of the whole system succumbing to a slew of potentially precipitous attacks such as DoS attacks. Most of them rely on the use of erasure codes to introduce redundancy to the data and reduce the computational overhead.



Machina - How to Translate Peer-to-peer Storage to Blockchain



https://protonmail.com/blog/centalized-vs-p2p-protondrive/

Adding “blockchain factor” to the peer-to-peer data storing complicates things quite a bit. Firstly, it introduces a need to make data stored immutable. The immutability factor refers to the idea of data retaining its untampered state. Secondly, in order to make data storing decentralized, and privacy-conformant there is a need to involve more players in the game.

The more players join the network the more decentralized, and less prone to single-entity (or 67%) attacks the system will be.

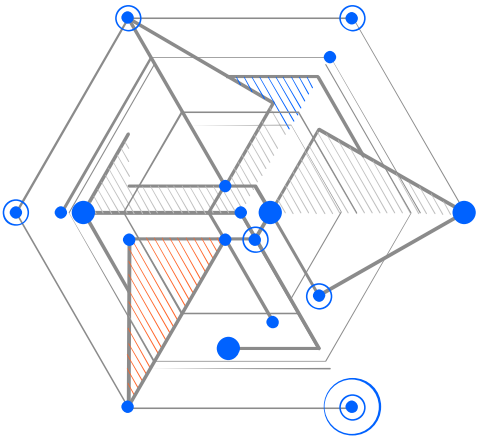
Generally, the immutability factor is inherently present on-chain and is unalterable by definition. What does pose an issue is the check of data integrity. In order to make sure that the data hasn't been tampered with in any way in the process, there is a need for multiple nodes or validators to "approve" that the data has remained intact.

Validation of files as such is a very cumbersome task that would present a very hard brunt to bear for the network. Therefore to effectively minimize the time, energy, and computational overhead required - Machina will employ a method called File Fingerprinting.

This method is used to map data accordingly, efficiently track and categorize it. Fingerprinting allows to ensign each dataset with a short text string or, in basic terms, endow it with a unique identifier. So instead of validating a complete, full-blown, full-scale file; each validator will only be entitled with a task to approve or disprove the integrity of data by a given short text string. Consensus validation is then made based on the reached agreement on the fingerprint.

File Fingerprinting allows to efficiently scale up the process of validation and check of data integrity without muddling and retarding the network.
The process of validation and reaching consensus can be twofold and normally differs in the way agreement is reached and the network is kept at bay from succumbing to the common intrusions. This calls for two of the most popular choices of validation or consensus algorithms: Proof of Work and Proof of Stake. Both present a viable option to validate whether given bits were reliably preserved.

All of the currently proposed solutions of decentralized data storage utilize PoW consensus to validate or inversely invalidate the integrity of data. Solutions such as Arweave, Storj and Filecoin specifically employ PoW consensus and the traditional mining schema with respective rewards in the native digital currency. Despite PoW being inherently more decentralized, although with some asterisks, consensus algorithm than traditional PoS, it also comes with a plethora of problems of its own. To elaborate a few:

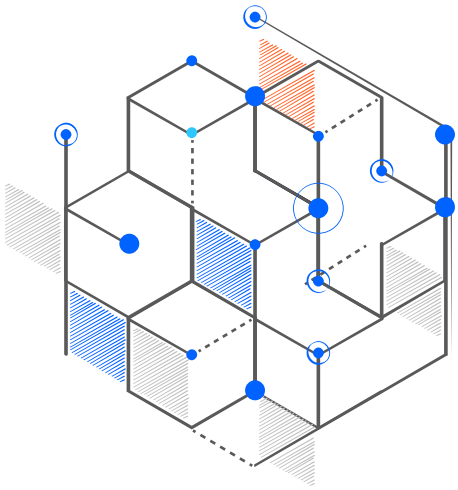


- Intrinsically the highest energy spent.
- Typical mining exploits and unfairness that stems from it.
- Corked scalability and walled garden because of the requirements for the node.

Machina - Novel Way to Store Data Immutably, Securely, and Cheaply

Machina – is a sharded data storage structure that is designed to provide scalable, cost-efficient, and immediately retrievable information on-chain.

Machina leverages the unique properties of Nightshade sharding of NEAR and provides a seamless economic model utilizing stablecoins.



Since Machina employs sharded data availability heuristics, it can scale up the data stored proportionally to the number of validators joining the network. Consequently, Machina utilizes PoS consensus as the means of validating blocks. This is crucial in two ways: energy-efficient and computationally less burdening consensus allow to nullify the energy output needed for the network to run.

Secondly, stemming from the first point, PoS, unlike PoW, doesn't require a potential validator to possess a powerful machine in order to produce blocks, in essence allowing anyone to join the network and scale the storage size. Supplementarily, it also plays a deciding role in decentralization allowing more nodes to join the network making it increasingly more dispersed and less predisposed to the common intrusions.

Sharding - How It Helps to Store Data Efficiently

Machina will utilize the latest development of NEAR blockchain and directly leverage NEAR mainchain as the conduit. With the help of Machina, NEAR expands its storage capacity which will allow an even more sharded and decentralized structure.

The implementation of sharding on NEAR is crucial to set out a new standard for on-chain immutable storage. Sharding institutes a novel way to minimize and homogenize storage reserves. Before delving into the intricacies of sharding implementations for the purpose of storage economization, we shall first examine sharding in general. In broad terms, sharding represents separate instances of separate blockchains called sharded chains that operate in conjunction with a beacon chain.

Each sharded chain theoretically can be viewed as a blockchain instance that is operable with its own set of validators that validate this particular shard. The beacon chain essentially embodies the core of the structure that relays, bookkeeps, and stores all of the data. Sharded chains also do not possess any governance of their own; instead, their governance is underpinned by the beacon chain's decisions.

To shortly summarize the main distinguishing points of state sharded blockchain compared to the traditional ones:

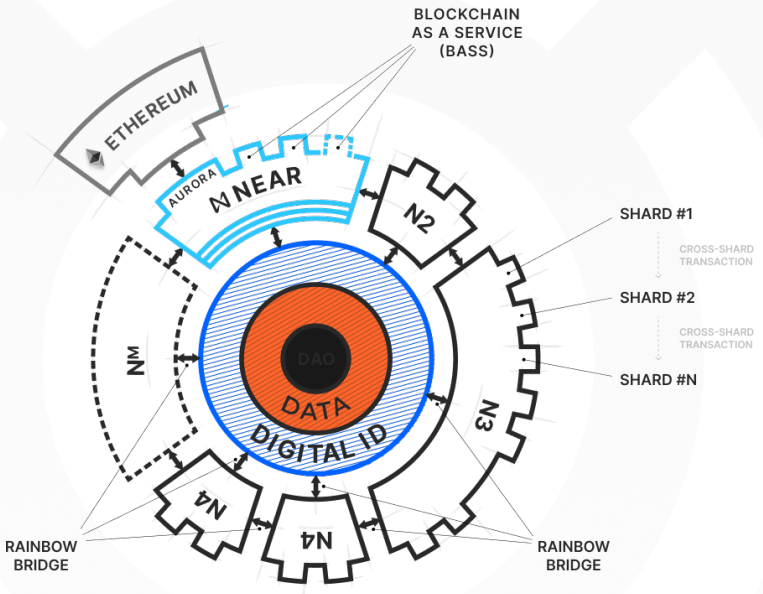
- Increased scalability due to parallelized processing/validation of transactions.
- More decentralization since more players participate in the validation.
- Less energy spent and fewer heavy computations needed - making state sharded blockchain more available, and all-around more adoption-friendly. With PoS consensus at the core of consensus reaching, Machina is exceedingly less environmentally damaging than PoW equivalents.

Sharded approach to the storage of data also allows the transmission of data across shards more efficiently. That's the reason why data structured is super cheap and scalable, and each shard can store a substantial chunk of data.

Machina - Shaping the New Mould of Storage on NEAR

The predictable price model inherent to NEAR also proves to be harmoniously intertwined with the idea of sharded data storage. Entities utilizing storage of Machina are not tied to the payment model of any specific currency or asset. Instead, payment for storage is deducted in stablecoins which ensures that developers and users alike have an easier time dealing with fees.

Machina will use file-based storage, storing every transaction directly into it, removing the runtime needed to actuate the retrieval or uploading period. The common issue with data-heavy smart contracts engulfs the current space and constricts the creative outpouring that can potentially stem from increasing the storage available.



Machina enables on-chain data to be accessible on-chain coupled with all of the security measures innate to the blockchain such as consensus and decentralization.

This will aid in ushering a new way of smart contract development that will allow a far wider scope of possibilities to innovate in.

By instituting a reliable, immediately retrievable, and always-online on-chain storage solution that can be utilized, with the help of EVMs such as Aurora, to also be applied to Solidity-based Smart Contracts.

To shortly summarize the main distinguishing points of state sharded blockchain compared to the traditional ones:

- Increased scalability due to parallelized processing/validation of transactions.
- More decentralization since more players participate in the validation.
- Less energy spent and fewer heavy computations needed - making state sharded blockchain more available, and all-around more adoption-friendly. With PoS consensus at the core of consensus reaching, Machina is exceedingly less environmentally damaging than PoW equivalents.

Sharded approach to the storage of data also allows the transmission of data across shards more efficiently. That's the reason why data structured is super cheap and scalable, and each shard can store a substantial chunk of data.

Machina - Initial Design and NEAR Repurposing

The plan is to make this data chain usable for large data storage almost immediately upon launch but then iterate on it subsequently, until it is able to supersede features of Filecoin and IPFS.

This design would rely heavily on validator selection and account creation to be done on the original NEAR Mainnet by locking tokens in a special contract. In this way, all of the stake and all of the balances on Machina would correspond to the locked tokens on the original NEAR Mainnet.

The initial design of Machina hinges on the idea of taking NEARcore, and modifying it in a way that would allow Machina to operate as intended.

Private Sharding Implementation

The very backbone of cryptographically secured solutions is to ensure that privacy of individuals or entities can remain in the identity-private state. This calls for a data-storing technology that can provide private, unbreachable, and unilaterally accessible information on request. Institutions may choose to decouple from the public shards and instead opt for private ones with the help of NEAR's Private Sharding and easily leverage the innateness of blockchain privacy to their merit.

Similarly, the same approach can be taken to resolve one of the impasses that businesses face when trying to accommodate blockchain technology into their workflow - data storage. NEAR Private Sharding in essence allows the creation of separate blockchain instances without a gruesome development cycle needed to actually design a blockchain.

Alternatively, we can utilize the Private Sharding structure to implement separate storage instances for entities willing to keep their data out of public reach or confine data to a restricted circle of users. In private shards, blockchains can share information without compromising on security.

Each private shard gets its own name, similar to domains on the web making it easy to establish a storage space with an immediately identifiable nomenclature.

The idea of private sharding and storing data on them is to coalesce public and private accesses to the extent that would benefit both parties by securing pieces of data or revealing those pieces only to the limited selection of individuals, and with Machina increasing NEAR storage capacity significantly, will allow more data to be stored, segregate it more efficiently, and enable even more information, including sensitive one, to be preserved securely, and in privacy-conformant way.

Data Integrity and Consistency (can be redacted, deleted, revised)

In order to ensure that each node in the network reliably and consistently carries out its duty and there are no ill-intended nodes that impair the network, Machina incorporates a type of ranking system that aims to eliminate, sift through, and reward nodes according to the “behavior”. Essentially this system rewards validators that provide timely, consistent, and improved data access across a given epoch and penalizes those who slack, are being negligent, or try to swindle the network.

In essence, this allows us to conveniently keep the network in the best possible conditions by engaging and rewarding honest players. Each shard replication can be decided depending on the reliability of the nodes. Reliability also can be tuned by the penalties nodes receive for not approving/producing chunks in time. This replication factor will define the cost structure above the basic hardware costs.

What kind of data can Machina store:

- Databases that require sizable storage space and low enough fees to be sustainable. For example ZKP-protected KYC, archives of DLT data, blockchain governance decision details, infrastructure contracts (related to large-scale infrastructure solutions need to store a sizable amount of data).
- Content pieces such as media or news-related issuances.
- Non Fungible Tokens such as digital art pieces, unique authenticators, and GameFirelated files.

The idea of Machina is to allow parties to retrieve and upload data trustlessly without approval from third parties. With the help of Machina, users can upload data securely and worry not about their data being tampered with or accessed on unauthorized terms.

This has become especially salient in the light of countless data breaches that expose personal data, allow for identity theft, and endow more authority to single-governed entities to manipulate datasets.

Permanence and Impermanence of Data with Machina

Storage state defines the length of time a particular data is stored for. In essence we can divide storage states in two foundational categories: permanent and impermanent state. Impermanent data storage refers to data that has a certain “expiry date”. This type of data state can be utilized in order to automatically delete a certain sensitive data from the storage when or if it’s not needed.

There is also a specific set of information and data that requires to be and is deriving value from being short-lived. This also helps to efficiently save up storage space for data that doesn’t need to be stored for indefinite time.

Permanent storage delineates the ability to reproduce, retrieve, and interact with a given data irrespective of the time passed. Permanence of data refers not necessarily to the fact of data being stored in complete state for the time interminable. Instead, it pertains to the ability of data to be restored to the original state because storing complete, uncompressed files would mean an extravagant storage waste.

Permanent data storage is far more intensive than short term storage as it requires files to be in the immediately retrievable state all the time.

Therefore most of the time permanent data storage is used to store files such as legal contracts, literature, scientific studies, and other similar pieces of documents that are expected to serve a specific need in the foreseeable future.
Machina is being designed as a data storage structure that is able to store information reliably for a time designated by a user.
Machina incorporates the equilibrium of storage by introducing both permanent and impermanent types of storing data. It can be both - pay fixed USD value for impermanent storage and burn storage token for permanent. It also facilitates low-fee and seamless data uploading that enables a myriad of implementations that is currently unachievable due to high costs and limited storage.

Future Features

These features can be launched on the Machina as protocol upgrades in any order.

Remove gas entirely

Since the Machina will not have smart contracts it will not have arbitrary Turing-complete computations, which means the fees can be determined statically, which means we don’t need to attach gas. We will be able to completely remove gas as a concept and not require attaching anything to the transactions.

Cold vs hot data

To allow storing huge amounts of data we can charge different fees depending on when the last time the given data was accessed. Implementation-wise hot data can live in memory.

Glacier data

It is reasonable to assume that some data can be extremely rarely accessible, but should be still available. Most likely the underlying storage layer will not allow retrieval of such data quickly enough and will take 3-5 hours, because it will be running on something like AWS Glacier. In that case we should introduce a new action called Warmup that will turn glacier data into warm data in 3-5 hours. All other transactions targeted towards glacier data that was not warmed up would be expected to fail.