



Filecoin - A Blockchain Based Distributed Storage System

Presented by Rishabh Shenoy

Submitted to Prof. Sudhir Dhage

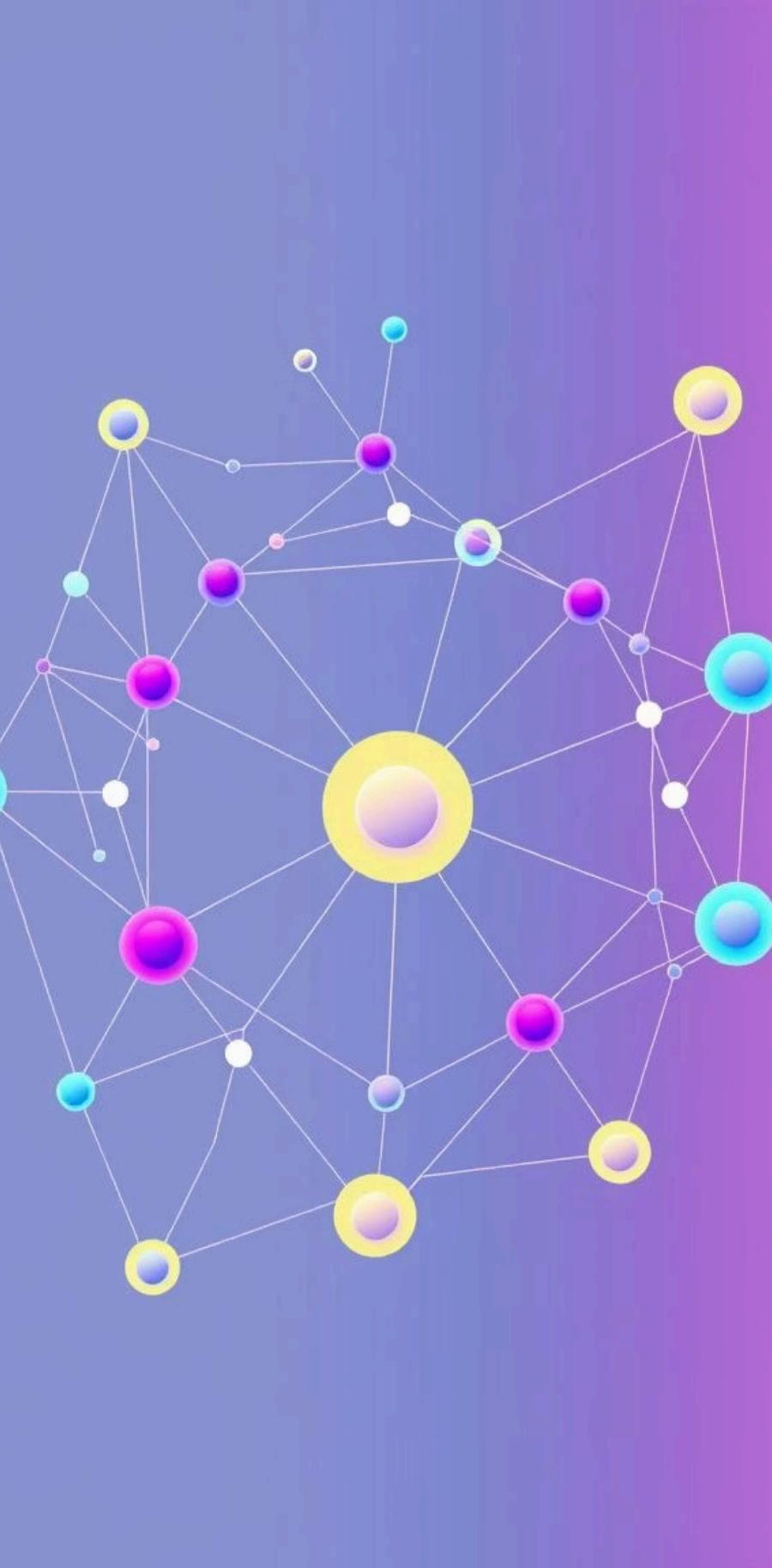
Date: September 15 , 2025

Abstract and Motivation

Filecoin, conceived by **Juan Benet** and developed by **Protocol Labs**, emerged in **2017** as a groundbreaking project aiming to revolutionize data storage.

Its importance lies in creating a **decentralized, peer-to-peer network for storing data**, offering a robust, secure, and efficient alternative to traditional centralized cloud storage. By leveraging blockchain technology, Filecoin establishes an open marketplace where anyone can contribute storage space or retrieve data, fostering a more resilient and democratic digital infrastructure.

Historically, the internet's increasing reliance on a few large corporations for data storage has raised concerns about data privacy, censorship, and single points of failure. Filecoin directly addresses these vulnerabilities, providing a critical piece of the puzzle for the vision of Web3—a decentralized, user-controlled internet.



What are Blockchain Based Distributed Storage Systems?

Blockchain-based Distributed Storage Systems leverage decentralized networks to store data, fundamentally changing how information is managed compared to traditional centralized models.

Centralized Storage Systems

- **Architecture:** Data resides on a few central servers managed by a single entity.
- **Vulnerabilities:** Susceptible to single points of failure, censorship, and data breaches.

Blockchain-Based Distributed Storage Systems

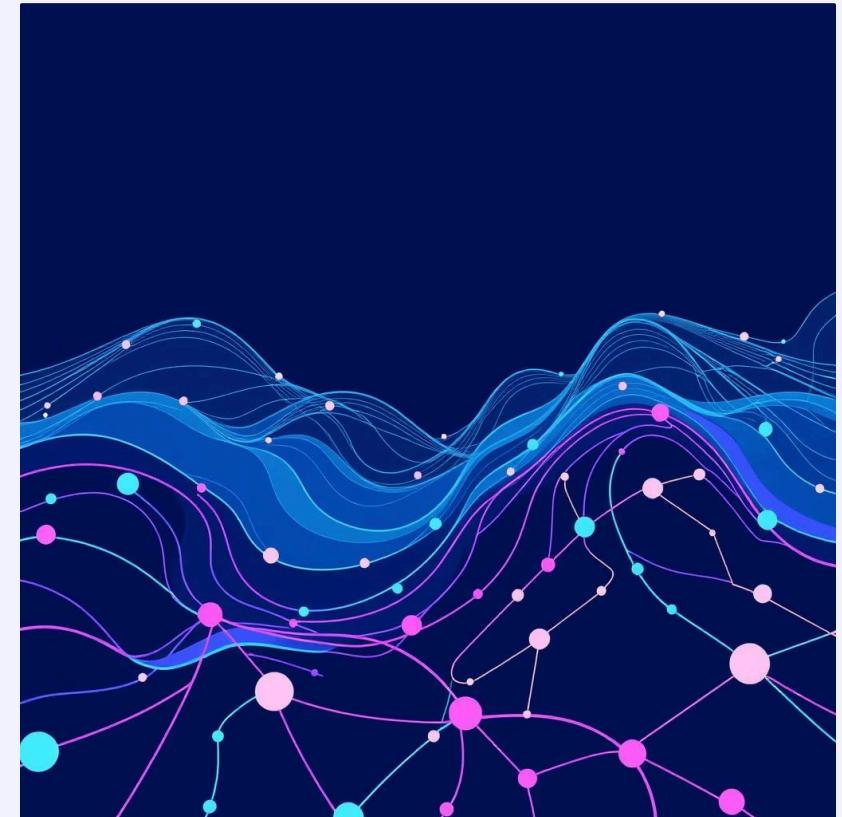
- **Architecture:** Data is fragmented, encrypted, and distributed across a network of independent nodes (peers).
- **Server Terms:** Operates on a peer-to-peer network; each node acts as a mini-server, contributing to a distributed ledger.
- **Advantages:** Enhanced security, censorship resistance, and data redundancy due to cryptographic proofs and network consensus.

Genesis of Decentralized Storage

From Data Deluge to Distributed Solutions

The exponential **growth of data** has strained traditional centralized storage models, leading to significant vulnerabilities and inefficiencies.

- Single points of failure
- High costs and limited scalability
- Censorship and control by large corporations
- Limited scalability under massive demand
- Censorship and corporate control over data
- Privacy risks & vulnerability to breaches
- Vendor lock-in reduces flexibility for users
- Lack of transparency in storage guarantees
- High infrastructure & maintenance costs



The emergence of **P2P networks** and **blockchain technology** provided a new paradigm, enabling secure and distributed systems. Filecoin was conceived to decentralize data storage, mirroring the success of Bitcoin's decentralized finance.

Filecoin Overview

The Decentralized Storage Revolution

Filecoin addresses the limitations of centralized data storage by creating a [decentralized, verifiable, and incentivized](#) network. It transforms cloud storage into an algorithmic market, powered by blockchain technology.

Problem Solved

Reliance on centralized providers, data breaches, vendor lock-in, and censorship risks.

Filecoin's Solution

A global network where anyone can store or retrieve data securely and verifiably.

Significance

Foundation for Web3, enabling censorship-resistant and resilient data storage.

Scalability

Exabyte-scale distributed storage network.

Global Participation

Dynamic market balances supply & demand. Open to individuals & enterprises.

Economic Security

Rewards honest storage, penalizes faults , Flexible contracts with transparency

Pillars of Filecoin

Core Principles



Openness

Permissionless network, accessible to all.



Incentive Engineering

FIL token rewards for storage, penalties for non-compliance.



Cryptographic Proofs

PoRep (Proof-of-Replication) and **PoSt** (Proof-of-Spacetime) ensure data integrity and availability.



Scalability & Security

Designed for petabytes of data with inherent blockchain security.



Privacy & Confidentiality

End-to-end encryption ensures miners cannot access plaintext data. Zero-knowledge proofs (zk-SNARKs) protect sensitive information.

Market & Coordination Layer

Transparent storage and retrieval markets enable fair price discovery.

Key Features of Filecoin



Decentralized & Open Network

Filecoin offers a peer-to-peer network for storing and retrieving data, fostering a censorship-resistant and accessible digital infrastructure.



Verifiable Storage

Through cryptographic proofs like Proof-of-Replication (PoRep) and Proof-of-Spacetime (PoSt), Filecoin guarantees data integrity and continuous availability.



Economic Incentives

An incentivized market allows users to pay for storage using FIL tokens, while providers earn FIL for reliable data storage, ensuring an efficient ecosystem.



Data Security & Privacy

Data is encrypted and distributed across multiple providers, protecting against loss, tampering, and unauthorized access with end-to-end encryption and zero-knowledge proofs.



Global Scalability

Designed to handle exabytes of data, Filecoin facilitates global participation, allowing individuals and enterprises to contribute storage or retrieve data seamlessly.

Impact and Legacy of Filecoin

Filecoin's introduction marked a significant shift in decentralized infrastructure, transforming how we approach cloud storage and inspiring a new generation of projects.



Decentralized Revolution

Pioneered a new model for cloud storage, aligning global economic resources with cryptographically verifiable storage commitments.



Catalyst for Innovation

Inspired countless projects in decentralized web, persistent storage, and open finance with its architectural innovations.



Addressing Core Issues

Solved fundamental challenges like long-term durability, auditability, and trustless provisioning in data storage.



Blueprint for Web3

Provides a foundational blueprint for future cloud and internet architecture, from research archiving to censorship-resistant publishing.



Influencing Next-Gen Platforms

Its blend of cryptography, market engineering, and decentralized consensus continues to inform new storage and application designs.

Architecture of Filecoin:

Filecoin's robust architecture is built upon a layered system, ensuring decentralized, verifiable, and efficient data storage.



Data Storage Layer

This foundational layer comprises **Storage Miners** who provide raw storage capacity. They seal client data into sectors and maintain its availability over time, constantly proving storage to the network through cryptographic proofs.



Blockchain & Consensus Layer

The heart of Filecoin, this layer is built on a custom blockchain that records all deals, validates proofs (Proof-of-Replication and Proof-of-Spacetime), and manages the network's state. It ensures transparency, immutability, and security across the system.

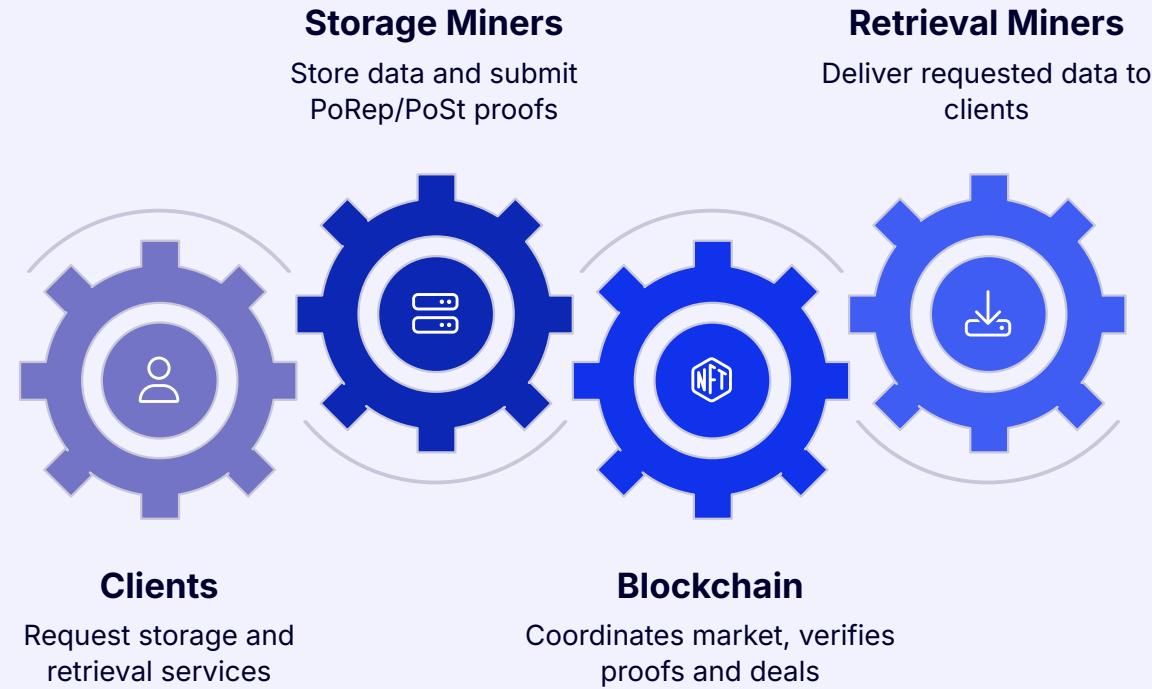


Storage & Retrieval Market Layer

This layer facilitates the interaction between **Clients** and miners. It's a dynamic marketplace where clients can find storage providers, negotiate deals, and retrieve data. **Retrieval Miners** specialize in delivering data quickly, fostering a competitive and efficient environment.

Network Structure

Filecoin Architecture



Metadata Management in Filecoin



Filecoin's decentralized nature relies heavily on efficient and verifiable metadata management. This involves several critical aspects:

- **Content Addressing & Data Identification** Filecoin uses Content Identifiers (CIDs) derived from the content's cryptographic hash, ensuring data integrity and uniqueness. All references to data within deals and proofs rely on these immutable CIDs, forming the foundation of metadata.
- **Deal & Contract Management** The network meticulously records all storage deals on the blockchain. This metadata includes details such as the client's address, the storage miner's ID, the data's CID, storage duration, agreed-upon price, and deal status.
- **Proof Metadata & Verification Processes** Crucial to Filecoin's integrity, metadata associated with Proof-of-Replication (PoRep) and Proof-of-Spacetime (PoSt) is managed on-chain. This includes timestamps of proofs, miner attestations, and verification results, allowing the network to continuously audit data availability and storage integrity.
- **Network State & Miner Reputation** Filecoin manages dynamic metadata about the overall network state, including active storage miners, their committed sectors, and historical performance. This contributes to a robust reputation system, influencing future deal selections and network health.

Data Storage Unit of Filecoin



Sectors

The fundamental unit of storage in Filecoin, representing a fixed-size chunk of data (e.g., 32 GiB or 64 GiB) that storage miners commit to store.



Client Data

Encrypted and often fragmented data provided by users, which is then sealed into sectors by storage miners for secure and verifiable storage.



Sealing

A cryptographic process where a storage miner transforms client data into a unique, verifiable sector, ensuring data integrity and commitment before it's stored on the network.



Proof of Replication (PoRep)

A zero-knowledge proof that cryptographically guarantees a storage miner has created a unique physical copy of the client's data and stored it on their dedicated storage space.



Proof of Spacetime (PoSt)

A continuous, verifiable proof that a storage miner is still storing the client's data over a period of time, ensuring ongoing data availability and preventing data deletion.



Storage Deals

On-chain agreements between clients and storage miners specifying the terms of data storage, including data size, duration, price, and the associated sectors.

Does Filecoin Employ a Consistency Model?

In the realm of distributed systems, a **consistency model** defines the rules for how data changes are propagated and perceived across multiple nodes. It dictates the guarantees offered to read and write operations, particularly in scenarios involving concurrent access and failures. Different consistency models exist, ranging from strong consistency (where all nodes see the same data at the same time) to eventual consistency (where data may diverge temporarily but eventually converges).

Filecoin does not adhere to a traditional strong consistency model in the same vein as a centralized database. Instead, it offers a verifiable consistency model centered on cryptographic proofs and blockchain immutability. This model ensures:

- **Data Integrity and Immutability via Content Addressing:** Filecoin uses Content Identifiers (CIDs) derived from a cryptographic hash of the data itself. This means the content's address is intrinsically linked to its value. Any alteration to the data results in a different CID, making data tampering immediately detectable and ensuring that clients always retrieve the exact data they stored. This forms the foundational layer of Filecoin's consistency.
- **Verifiable Storage with Proof-of-Replication (PoRep):** PoRep guarantees that a storage miner has uniquely replicated a client's data and committed it to their dedicated storage space. This cryptographic proof is submitted to the Filecoin blockchain, establishing a verifiable record of data placement at a specific point in time. It's a crucial component for ensuring the initial consistent state of data storage.
- **Ongoing Data Availability with Proof-of-Spacetime (PoSt):** PoSt extends PoRep by continuously proving that a storage miner is still storing the client's data over time. Miners periodically submit these proofs to the blockchain. If a miner fails to provide a valid PoSt, it indicates a potential breach of the storage agreement, which can lead to penalties and a re-evaluation of the data's consistent availability. This mechanism ensures ongoing consistency in data presence.

Data Structures in Filecoin

Pieces

Atomic units of client data, can be chopped and distributed across miners.

Sectors

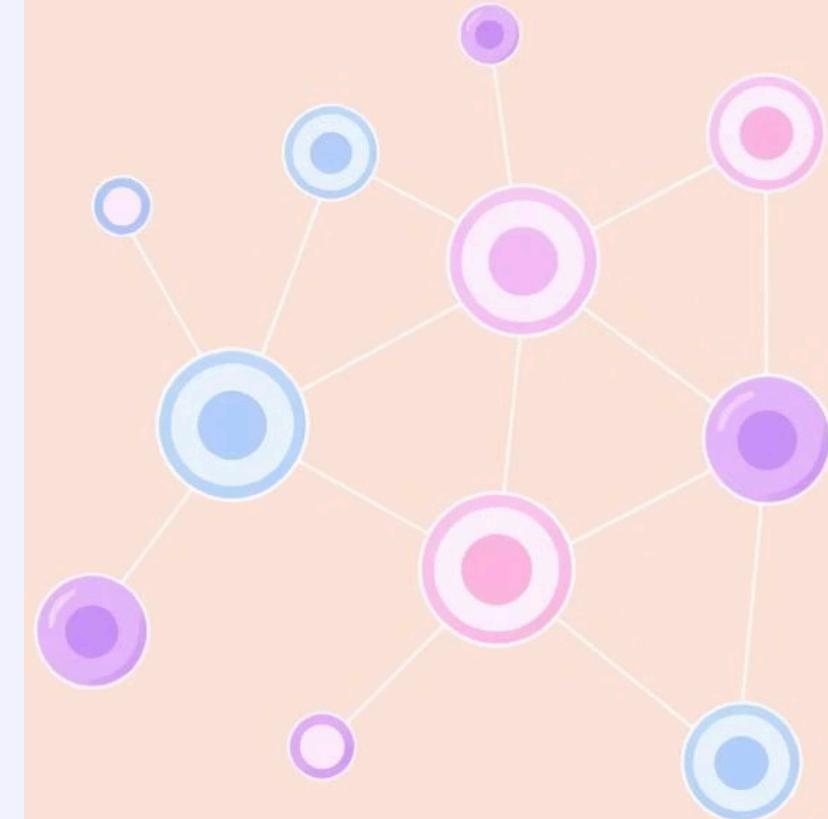
Physical slices of miner storage, pledged and cryptographically sealed.

Allocation Tables

On-chain data structures mapping pieces to sectors/miners, recording proof status, pledge status, and proof/fault counters for auditability.

Orders and Deal Records

Data structures reflecting every step of storage/retrieval: bid/ask/deal orders, proof receipts, micropayment channels



Interaction between Components in Filecoin:

Filecoin's decentralized storage ecosystem operates through intricate interactions, ensuring data integrity, availability, and incentivized participation.

1

Onboarding & Role Commitment

Participants (Clients, Storage Miners, Retrieval Miners) join the network. Storage Miners pledge collateral (FIL) for storage space.

2

Client Storage Request & Matching

Clients submit bid orders for storage, and Storage Miners publish ask orders. The network autonomously matches compatible bids and asks.

3

Deal Confirmation & Data Movement

Client encrypts and sends data to the Storage Miner. Both cryptographically sign a deal, committing it to the blockchain and updating the AllocTable.

4

Sealing & Continuous Proofs

Storage Miners seal data using Proof-of-Replication (PoRep) and continuously submit Proof-of-Spacetime (PoSt) to demonstrate ongoing storage.

5

Network Verification & Incentives

The network validates all proofs. Missing or invalid proofs lead to penalties on miner collateral and trigger fault-repair mechanisms.

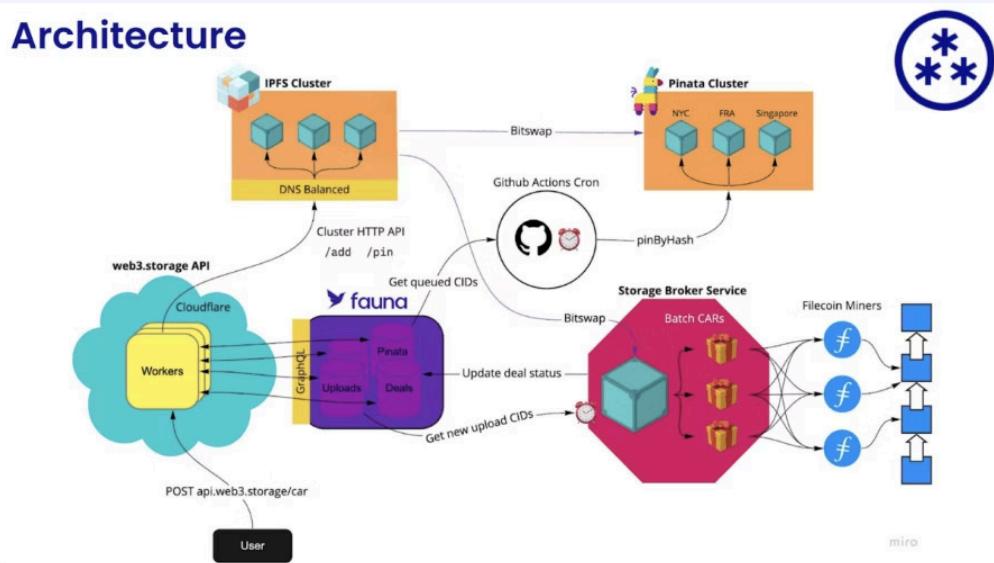
6

Client-Side Retrieval Process

Clients submit retrieval bids. Retrieval Miners respond, and a micropayment channel is established for efficient, chunk-by-chunk data delivery.

Some Other Components in Filecoin:

Architecture



Architecture and Components in Filecoin distributed system

Market Settlement & Autonomic Repair

At every protocol epoch, the Network's Manage Protocol: Monitors all active storage and retrieval deals in the orderbooks.

Shared Data Structures & Transparency

On-chain and off-chain sets tracking all open, matched, or completed orders.

Client and Miner Autonomy, Privacy, and Flexibility:

All data is encrypted end-to-end by Clients; Miners cannot decrypt or misuse content.

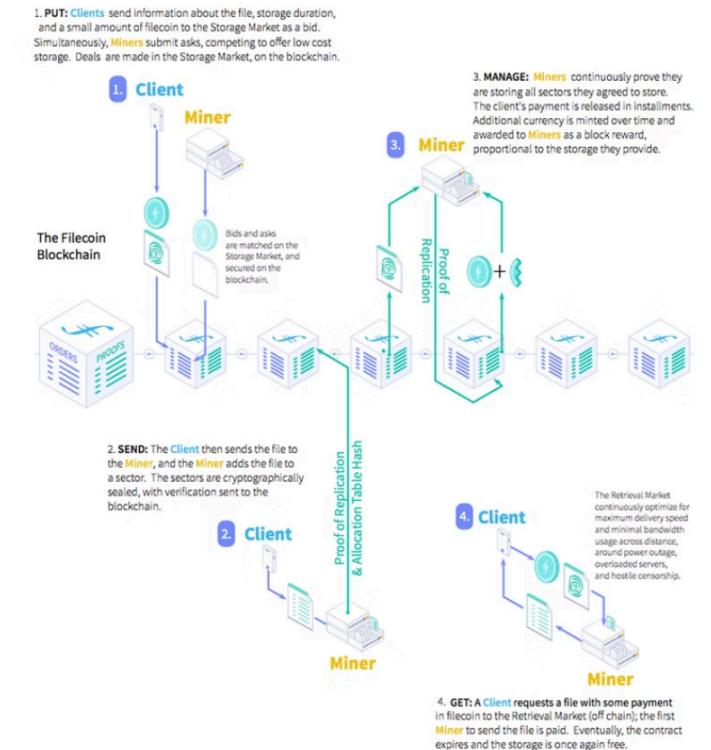
How Data is Organized

File System Design

Filecoin employs a robust file system design to ensure data integrity and availability:

- **Data Organization:** Files are broken into **pieces**, which are then sealed into **sectors** for storage by miners. **Allocation tables** track data distribution.
- **Access Methods:** Simple **Put/Get** operations, similar to traditional file systems, but decentralized.
- **Encryption:** Data is encrypted end-to-end for privacy and security.
- **Fault Tolerance & Scalability:** Achieved through data replication across multiple miners and distributed architecture.
- **Hot/Cold Storage:** Supports both high-speed retrieval (hot) and archival storage (cold) needs.

Filecoin: Lifecycle of a File



Comprehensive Services



Storage & Retrieval

Primary service for decentralized data storage and access.



Client-Side Caching

Enhances retrieval speeds for frequently accessed data.



Replication & Self-Healing

Automatic data duplication and repair for resilience.



Distributed Namespace

Global, unique identifiers for stored data.



Authentication

Cryptographic methods ensure secure user and data verification.



Economic Security

Incentive mechanisms protect against malicious behavior.

Performance Optimisation and Security

Performance Optimisation

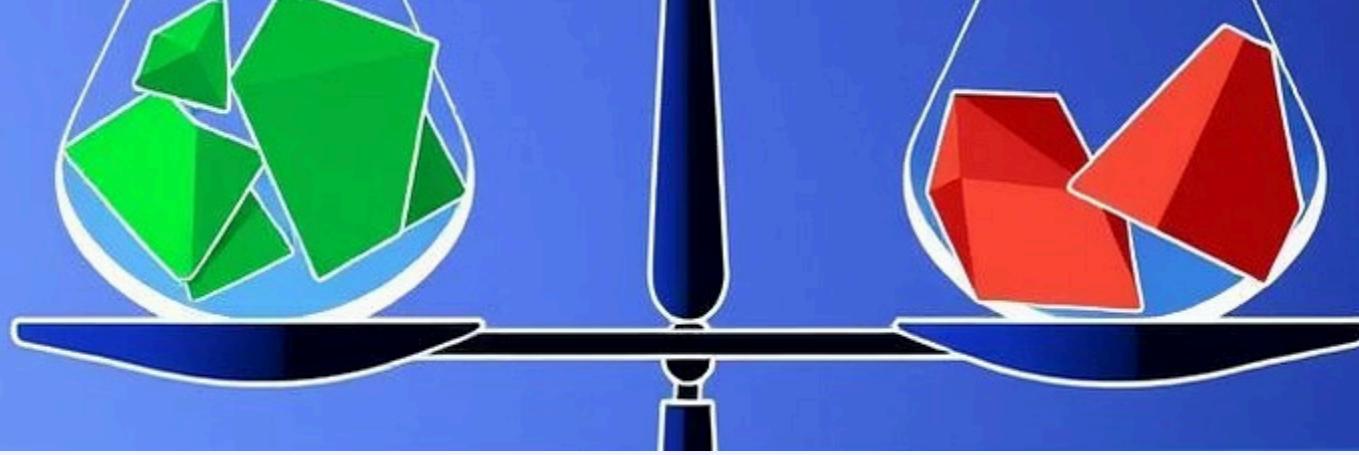
- **Sealing Optimisation:** Use GPU/parallel processing to speed up PoRep & PoSt generation.
- **Batching Proofs:** Combine multiple proofs to reduce overhead & chain load.
- **Client-side caching:** Frequently accessed data cached locally via IPFS for low latency.
- **Efficient retrieval:** Retrieval Market with micropayment channels ensures fast streaming.
- **Hybrid storage layers:** Hot storage (fast access) + cold storage (archival, cheaper).
- **Content routing:** DHTs + locality-aware miner selection reduce retrieval delays.
- **Enterprise tuning:** Use NVMe/SSD for hot sectors, HDD for archival data.

Security

- **End-to-end encryption:** Data encrypted at client-side, miners store only ciphertext.
- **Zero-knowledge proofs (zk-SNARKs):** Ensure storage validity without revealing data.
- **Proof-of-Replication (PoRep):** Guarantees each replica is unique & miner-specific.
- **Proof-of-Spacetime (PoSt):** Continuous proofs ensure long-term storage integrity.
- **Collateral & slashing:** Dishonest miners lose FIL tokens for invalid/missing proofs.
- **Public verifiability:** All proofs stored on blockchain → open auditing possible.
- **Access control:** Smart contracts (File Contracts) enable programmable permissions.

Fault Tolerance and Impact of Filecoin

Aspect	Details
Fault Tolerance	<ul style="list-style-type: none">Decentralized Storage Network: Data is split and stored across multiple independent storage providers.Cryptographic Proofs (PoRep & PoSt): Storage providers must continuously prove they still hold the data, reducing risk of data loss.Redundancy & Incentives: Clients can contract multiple providers for the same data; providers are financially penalized for losing or corrupting data.No Single Point of Failure: Outages of one node do not compromise availability.
Impact	<ul style="list-style-type: none">Enhanced Data Integrity: Users gain confidence that their data is stored as promised.Lower-Cost Decentralized Cloud: Competition among storage providers drives prices down.Supports Web3 Ecosystem: Provides censorship-resistant storage for dApps, NFTs, and research data.Token Economy Influence: FIL token aligns incentives but also introduces market dynamics (price volatility affects storage deals).



Balancing Strengths and Challenges

Advantages & Disadvantages

Advantages

- **Scalability:** Elastic and global storage capacity.
- **Verifiability:** Cryptographic proofs ensure data integrity.
- **Incentive Compatibility:** Economic model aligns participant interests.
- **Decentralization:** No single point of failure or control.

Disadvantages

- **Complexity:** Technical challenges for new users/miners.
- **Latency:** Initial retrieval can be slower than centralized services.
- **Resource Demands:** Significant hardware and bandwidth for miners.
- **Adoption Barriers:** Integration with existing systems.

Use Case Application and Limitation of Filecoin

Aspect	Details
Use Cases	<ul style="list-style-type: none">• Decentralized Cloud Storage: Users store data on a distributed network instead of centralized providers.• Data Archival & Backup: Cost-effective long-term storage for enterprises and researchers.• NFT & Web3 Data Hosting: Permanent storage for NFT metadata, dApps, and blockchain data.• Censorship-Resistant Storage: Suitable for sensitive or public data where central control is risky.
Limitations	<ul style="list-style-type: none">• Complex Onboarding: Setting up nodes or deals requires technical knowledge.• Retrieval Latency: Data access may be slower than centralized clouds (no instant CDN).• Token Price Volatility: FIL price fluctuations can affect storage cost predictability.• Limited Ecosystem Maturity: Fewer mainstream integrations compared to AWS/Google Cloud.

Evolving Landscape

Future Directions & Ecosystem Role

Filecoin is constantly evolving, with significant developments underway:

- **Filecoin Virtual Machine (FVM):** Enables smart contracts and DApps on Filecoin.
- **Zero-Knowledge Proofs (ZK proofs):** Enhancing privacy and efficiency.
- **Challenges:** Improving user experience, optimizing retrieval, and fostering wider adoption.

It plays a unique role alongside alternatives:

Alternatives: Arweave (permanent storage), Storj (decentralized cloud), Sia (p2p storage), **IPFS** (content-addressed network), BTFS (BitTorrent File System).

Ecosystem Role: Filecoin complements **IPFS** by providing a persistent, incentivized storage layer, critical for the long-term vision of Web3.

The Path Forward: Decentralized Storage Landscape

Understanding the diverse approaches to decentralized storage is crucial for appreciating Filecoin's unique position and future potential. This table compares key distributed file systems:

System	Primary Focus	Incentive Mechanism	Data Durability/Redundancy	Key Differentiator
Filecoin	Decentralized Data Storage	FIL (Storage/Retrieval payments, miner collateral)	Proof-of-Replication, Proof-of-Spacetime, Sector sealing	Verifiable storage with active economic incentives
Arweave	Permanent & Perpetual Storage	AR (One-time payment for permanent data)	"Blockweave" (data replicated across all miners)	Truly permanent, "pay once, store forever" model
Storj	Decentralized Cloud Storage	STORJ (Payments for storage/bandwidth)	Erasure coding, encrypted shards across nodes	Enterprise-grade, highly distributed, developer-friendly
Sia	Peer-to-Peer Storage Marketplace	Siacoin (Contracts between users & hosts)	Erasure coding, direct host-client contracts	Direct peer-to-peer relationships, private by design

Each system offers distinct advantages, catering to different needs within the broader decentralized web ecosystem. Filecoin stands out for its robust verifiability and market-driven incentives for storage providers.

The Path Forward

Conclusion & Outlook

Filecoin is more than just a storage solution; it's a critical infrastructure for the decentralized future.

01

Relevance to Web3

Enables data ownership, censorship resistance, and foundational layer for decentralized applications.

02

Long-Term Influence

Reshaping how data is stored, accessed, and secured globally, fostering data freedom.

03

Integration with Emerging Technologies

Synergies with AI, IoT, and metaverse for robust, scalable data backbones.

Filecoin continues to pave the way for a more open, transparent, and resilient internet.