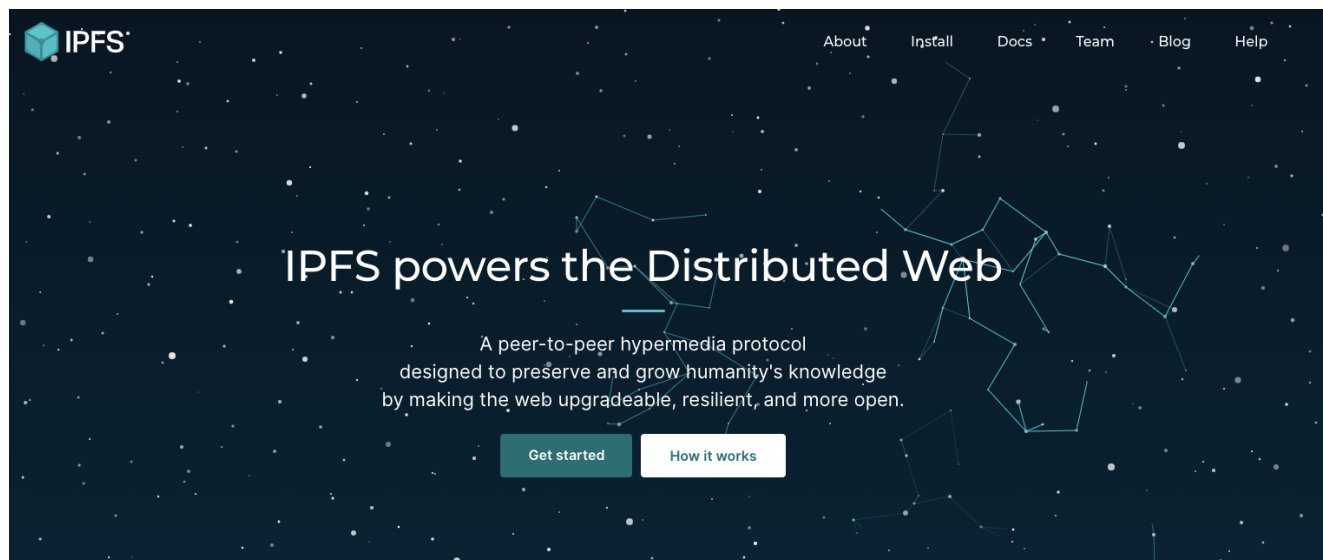


## Lesson 16 - Decentralised Storage

### IPFS



IPFS is a distributed system for storing and accessing files, websites, applications, and data.

### Concepts

IPFS is based on 3 concepts

1. Unique identification via content addressing
2. Content linking via directed acyclic graphs (DAGs)
3. Content discovery via distributed hash tables (DHTs)

### Content addressing

Instead of being location-based, IPFS addresses a file by *what's in it*, or by its *content*.

The content identifier above is a *cryptographic hash* of the content at that address. The hash is unique to the content that it came from.

Because the address of a file in IPFS is created from the content itself, links in IPFS can't be changed.

A *content identifier*, or CID, is a label used to point to material in IPFS. It doesn't indicate *where* the content is stored, but it forms a kind of address based on the content itself. CIDs are short, regardless of the size of their underlying content.

CIDs are based on the content's cryptographic hash.

That means:

- Any difference in the content will produce a different CID and

- The same content added to two different IPFS nodes using the same settings will produce *the same CID*.

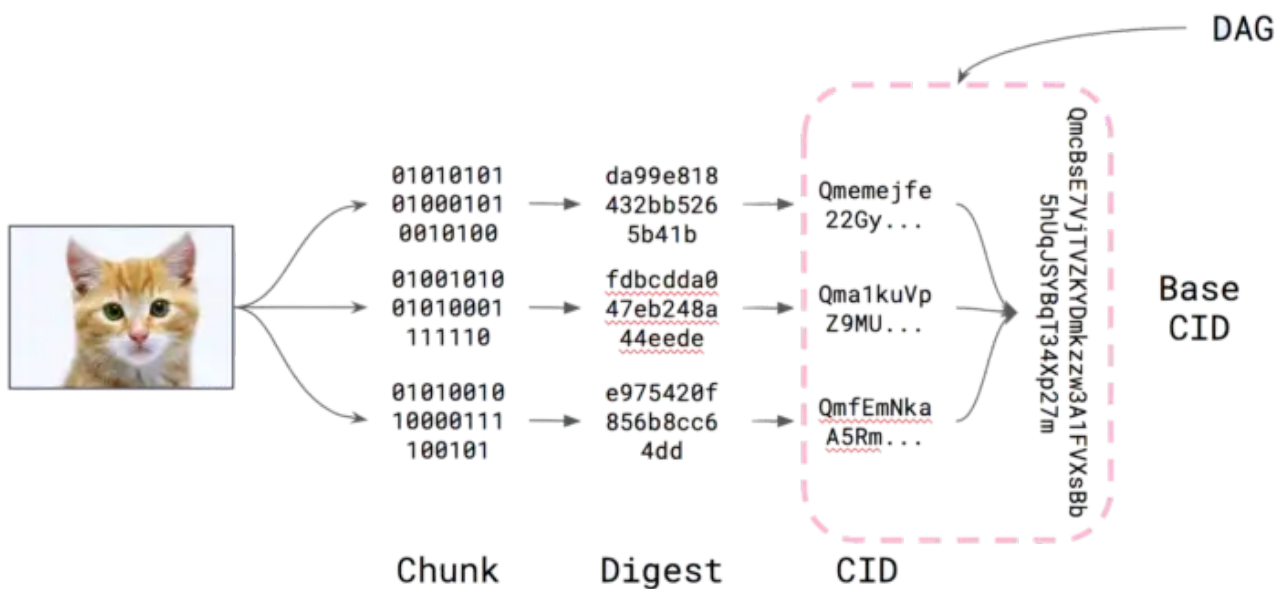
IPFS uses the `sha-256` hashing algorithm by default, but there is support for many other algorithms.

IPFS uses [multi hashes](#) (self describing hashes) to create CIDs

## Directed Acyclic Graph

IPFS and many other distributed systems take advantage of a data structure called directed acyclic graphs.

See this [article](#) for more details



## Distributed Hash Table

Nodes can store & share data without central coordination and use DHTs as a means to find each other.

A distributed hash table provides a lookup service similar to a hash table: key-value pairs are stored in a DHT, and any participating node can efficiently retrieve the value associated with a given key.

The main advantage of a DHT is that nodes can be added or removed with minimum work around re-distributing keys.

The [libp2p project](#) is the part of the IPFS ecosystem that provides the DHT and handles peers connecting and talking to each other.

## IPFS Network Properties

- Autonomy and decentralisation: the nodes collectively form the system without any central coordination.
- Fault tolerance: the system should be reliable (in some sense) even with nodes continuously joining, leaving, and failing.

- Scalability: the system should function efficiently even with thousands or millions of nodes.

The DHT is used in IPFS for routing, in other words:

1. to announce added data to the network
2. and help locate data that is requested by any node.

The [white paper](#) states:

Small values (equal to or less than 1KB) are stored directly on the DHT. For values larger, the DHT stores references, which are the NodeIDs of peers who can serve the block.

## Block Exchange - BitSwap Protocol

In IPFS, data distribution happens by exchanging blocks with peers using a BitTorrent inspired protocol: BitSwap.

Like BitTorrent, BitSwap peers are looking to acquire a set of blocks (want\_list), and have another set of blocks to offer in exchange (have\_list). Unlike BitTorrent, BitSwap is not limited to the blocks in one torrent.

BitSwap operates as a persistent marketplace where node can acquire the blocks they need, regardless of what files those blocks are part of. The blocks could come from completely unrelated files in the filesystem. Nodes come together to barter in the marketplace.

While the notion of a barter system implies a virtual currency could be created, this would require a global ledger to track ownership and transfer of the currency. This can be implemented as a BitSwap Strategy

## IPNS

IPNS gives us

- human readable links  
E.g. `/ipns/extropy` instead of

`QmPrPmbbUKA3ZodhzPWZnpFgcPMFWF4QsxXbkWfEptTBKl`

- A link to the latest version of content

## IPFS based websites

- websites that are completely distributed
- websites that have no origin server

- websites that can run entirely on client side browsers

See this [guide](#) on how to host a website on IPFS

## Structures used in IPFS

### 1. IPFS object

a data structure with two fields:

Data — a blob of unstructured binary data of size < 256 kB.

Links — an array of Link structures, these are links to other IPFS objects.

### 2. IPFS Links

A Link structure has three data fields:

- Name — the name of the Link.
- Hash — the hash of the linked IPFS object.
- Size — the cumulative size of the linked IPFS object, including following its links.

## IPFS Gateways

IPFS deployment seeks to include native support of IPFS in all popular browsers and tools. Gateways provide workarounds for applications that do not yet support IPFS natively.

The URL will be of the form

```
ipfs://{CID}/{optional path to resource}
```

Public gateway operators include:

- Protocol Labs, which deploys the public gateway <https://ipfs.io>.
- Third-party public gateways. E.g., <https://cf-ipfs.com>.

For example you can explore objects

```
https://explore.ipld.io/#/explore/QmSnuWmxptJZdLJpKRarxBMS2Ju2oANVrgbr2xWbie9b2D
```

## IPFS and NFTs

See this [guide](#) and these [best practices](#)

IPFS provide a command line [Minty](#) to help create NFTs.

---

# Alternatives to IPFS

## Swarm

See [white paper](#)







Swarm is a peer-to-peer network of nodes that collectively provide a decentralised storage and communication service.

This system is economically self-sustaining due to a built-in incentive system which is enforced through smart contracts on the Ethereum blockchain and powered by the BZZ token.

PRIVACY FOR INDIVIDUALS ↔ TRUST FOR THE DATA ECONOMY

# Unstoppable data

Swarm continues where the blockchain ends, making the world computer real.

<b>Open &amp; Borderless</b> <p>Swarm is open source code, limited only by the people who use and maintain it - Join a community building the future of the web.</p> 	<b>Fault-tolerant</b> <p>Redundant storage with local replication ensures data availability even in the face of node dropouts or data loss.</p> 	<b>No downtime</b> <p>Swarm is decentralised and distributed, and so it's also always up, making it stable and reliable.</p> 
<b>Privacy first</b> <p>Swarm nodes provide cryptographic support enabling you to use network in an unrestricted way while building blocks like "Single-owner-chunks" enable extraordinary zero-leak privacy.</p> 	<b>DDoS resistant</b> <p>Swarm network is fully decentralised peer-to-peer network while it's resilience against DDoS grows with every additional node in the network.</p> 	<b>Credible neutrality</b> <p>Swarm is built to not discriminate - it's in the very nature of our design, and permeates throughout our mission.</p> 

For Swarm to properly function as a decentralised p2p storage and communication infrastructure, on very basic terms there must be network participants who:

- contribute bandwidth for incoming and outgoing requests
- provide storage for users to upload and retrieve data
- forward incoming requests to peers who can fulfil them if they can not serve the request themselves

Swarm introduces its own incentives system for ensuring correct network behaviour by rewarding nodes for serving these functions.

---

# Filecoin


See [Docs](#)

See [Primer](#)

## The missing incentive layer for IPFS

Filecoin adds incentivized, persistent storage to IPFS. IPFS users are able to reliably store their data on Filecoin right from the IPFS network — opening the network up to a world of applications and use-cases.

See IPFS apps: [awesome.ipfs.io](https://awesome.ipfs.io)

 [Learn more about IPFS](#)

Filecoin is a digital storage and data retrieval method, made by Protocol Labs and builds on top of InterPlanetary File System, allowing users to rent unused hard drive space.

The project was launched in August 2017 and raised over \$200 million within 30 minutes.

The Filecoin Network is made with storage providers and clients. They make deals and contribute to maintaining the Filecoin blockchain, obtaining storage services, and receiving rewards in the process.

## Filecoin Networks

Peers communicate over secure channels that they use to distribute information to the network (gossiping), to transfer data among themselves, and to discover other peers, maintaining a well-connected swarm in which information like blocks and messages flows swiftly even when many thousands of peers participate.

- [Mainnet](#), the only production Filecoin network.
- [Calibration](#), the primary testing network for Filecoin.
- [Wallaby](#), an early testing network for bleeding edge [Filecoin Virtual Machine](#) deployments.

## Filecoin Nodes

*Filecoin Nodes* or *Filecoin clients* are peers that sync the Filecoin blockchain and validate the messages in every block, which, once applied, provide a global state.

Filecoin Nodes can also publish different types of *messages* to the network by broadcasting them. For example, a client can publish a message to send FIL from one

address to a different one. Nodes can propose [storage and retrieval deals](#) to Filecoin storage providers and pay for them as they are executed.

## Filecoin Storage Providers

The storage providers provide services to the network by executing different types of deals and appending new blocks to the chain (every 30 seconds), for which they collect FIL rewards.

## Filecoin Deals

There are two main types of deals in Filecoin: *storage deals* and *retrieval deals*.

**Storage deals** are agreements between clients and *storage providers* to store some data in the network. Once a deal is initiated, and the storage provider has received the data to store, it will repeatedly prove to the chain that it is still storing the data per the agreement so that it can collect rewards.

If not, the storage provider will be slashed and they will lose FIL.

**Retrieval deals** are agreements between clients and *retrieval providers* (which may or not be also storage providers) to extract data that is stored in the network .

Unlike storage deals, these deals are fulfilled off-chain, using *payment channels* to incrementally pay for the data received.

## Proofs

See [article](#) about the proof system.

Storage providers must prove that:

- They store all the data submitted by the client
- They store it during the whole lifetime of the deal

## Proof of space

A proof-of-space is a piece of data that a prover sends to a verifier to prove that the prover has reserved a certain amount of space. For practicality, the verification process needs to be efficient, namely, consume a small amount of space and time. For soundness, it should be hard for the prover to pass the verification if it does not actually reserve the claimed amount of space.

One way of implementing PoSpace is by using [hard-to-pebble graphs](#).

The verifier asks the prover to build a labelling of a hard-to-pebble graph. The prover commits to the labelling.

The verifier then asks the prover to open several random locations in the commitment.

## Proof of spacetime

Proof-of-spacetime differs from proof-of-capacity in that PoST allows network participants to prove that they have spent a "spacetime" resource, meaning that they have

allocated storage capacity to the network over a period of time.

## Gas Fees

Executing messages, for example by including transactions or proofs in the chain, consumes both computation and storage resources on the network.

The gas consumed by a message directly affects the cost that the sender has to pay for it to be included in a new block by a storage provider.

An amount of the fees is burned (sent to an irrecoverable address) to compensate for the network expenditure of resources, since all nodes need to validate the messages.

## How does Filecoin relate to IPFS ?

While interacting with IPFS does not require using Filecoin, all Filecoin nodes *are* IPFS nodes under the hood, and can connect to and fetch IPLD-formatted data from other IPFS nodes using libp2p.

However, Filecoin nodes don't join or participate in the public IPFS DHT.

## Filecoin tutorial

Protoschool offers a [tutorial](#) about Filecoin and storage.

---

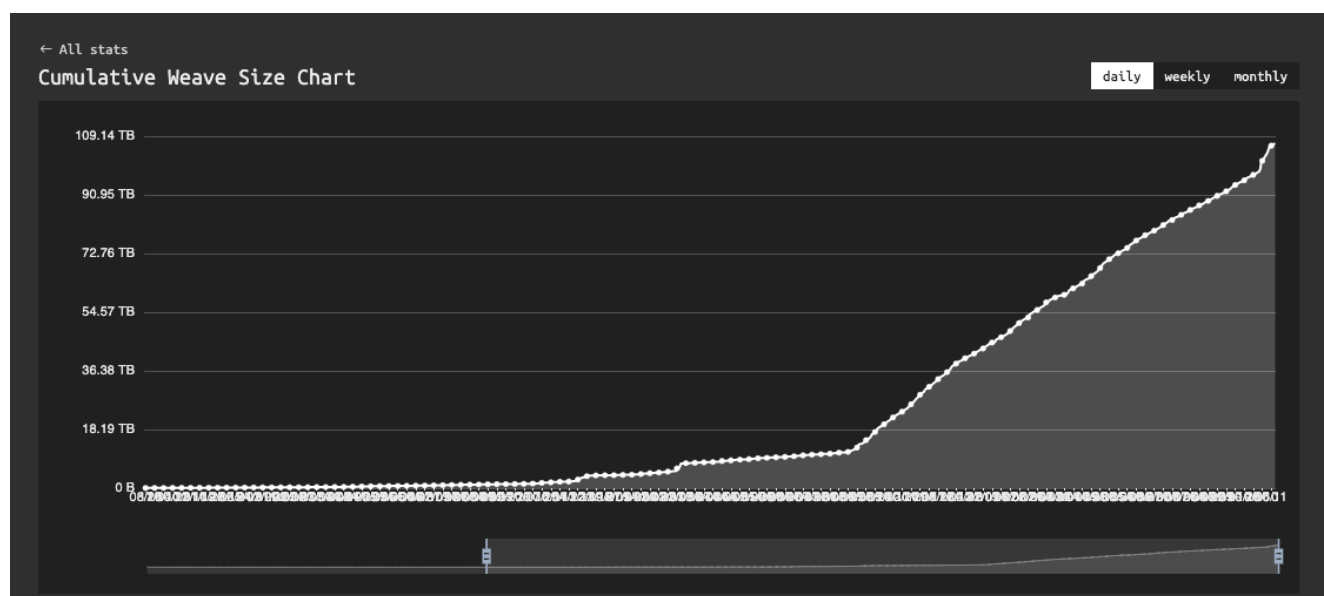


# Arweave

## Features

- Decentralised storage built on top of the Arweave protocol
- Permanent and resilient
- Proof of Access mechanism (similar to Filecoin's Proof of Space and Proof of Replication)
- Incentivises quick retrieval
- Data replicated and distributed across the network

Currently adding approx 6 TB per month



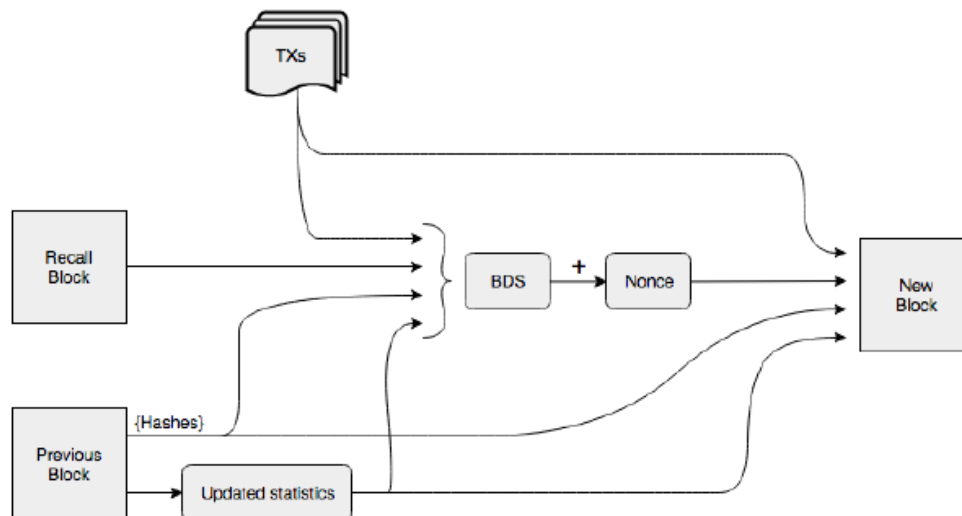
## Storage details

- Mechanism to distribute tokens to sustainably incentivise perpetual storage or arbitrary quantities of data.
- Based on a logarithmically decreasing \$/GB-h.
- Storage media assumed to take 434 yrs to reach max theoretical data density limit of  $1.53 \times 10^{67}$  bits/cm<sup>3</sup> (currently  $1.66 \times 10^{12}$ )
- Storage medium reliability increases.
- Incentive for cheap storage grows as humanity's demand for data is growing.
- Predictions for future safekeeping of data based on ultra-low cost of storage medium.
- Arweave expected to be 'nested' inside future storage systems.

## Chain Architecture

- Block constructed from previous block, recall block and transactions
- Recall bloc
  - A deterministic but unpredictable choice of block from the "weave's" history
  - Prevents data from being lost if it is rarely accessed
  - Lesser known blocks increase miner fees

Figure 6: Block construction from previous block, recall block, and transactions



## Block Construction

- Waiting pool for all new transactions.
- Mining pool for tx's that are validated and likely known by other nodes.
- Transactions removed from mining pool once block is mined.
- Tx's can be reintroduced if there is a fork.
- Block Data Segment (BDS) serves as the 'puzzle' in the PoW mechanism is a hash of:
  - Independent hash of previous block
  - Entire contents of recall block
  - Transactions and other metadata for the candidate block

## Creation process:

1. Assemble relevant metadata.
2. Rank and validate transactions.
3. Generate new BDS.
4. Find nonce (PoW).
5. Propagate new block in memoised form.

## Network Mechanism Design

- Made up of miners and users.
- Users pay tokens (AR) to add data to the network.
- Miners receive tokens for mining new blocks.
- Each tx has optional recipient and data fields meaning they are very flexible.
- Token used for encoding data into system and rewarding miners (thus non-zero financial value).
- Maximum 66 million AR (55 million in genesis block, 11 million through block mining rewards).

## Wildfire Mechanic Incentive Mechanism

- Type of Adaptive Interacting Incentive Agents (AIIA) game.
- Dominant Strategy Incentive Compatible (DSIC) nature of network ensures truth telling is a weakly dominant strategy.
- Peers are ranked on:
  - Generosity - sending new tx's and blocks.
  - Responsiveness - responding promptly to requests for information.
- Similar to Bittorrent's optimistic tit-for-tat algorithm.
- Scores are gossiped to higher-ranked peers allowing a node to rationalise its bandwidth allocation.
- Data/blocks/tx's are preferentially gossiped to more responsive peers in parallel, incentivising a stronger network.
  - Less strong peers are provided the data sequentially.
- Slow nodes are ultimately excluded from the network as peers remove them from their peer list.
  - Enforces pro-social behaviour.
  - Reduces wasted bandwidth.
  - Not assumed each node running same AIIA agent.

## Fork Resistance and recovery

- Caused by:
  - Propagation delay
  - 'Edge' validation performed on data (to protect the network) before being propagated quickly.
  - Propagating blocks quickly will raise chance of receiving mining rewards.
- Forks are recovered to the majority fork immediately in a similar fashion to Nakamoto consensus-based blockchain protocols.
- A 'cumulative\_difficulty' field in each block is used to resolve forks.

## Content Policies

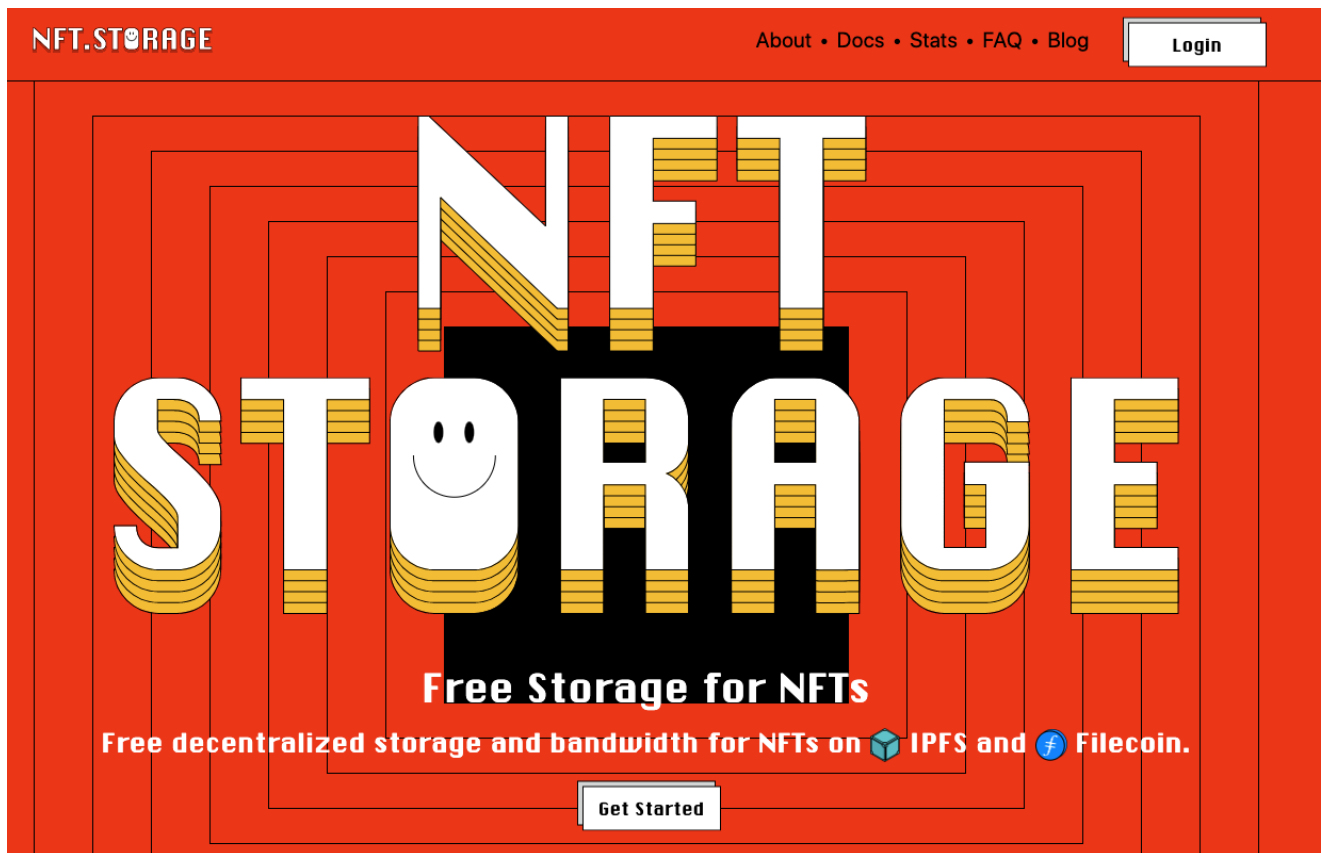
- Collective maintenance of the Arweave network necessitates a mechanism for miners to express their opinions on what content should and should not be hosted.
- Mechanisms used:
  - A democratic process of voting on content entry into the blockweave.
  - The individual ability of each node to choose what content to store on their machines.
  - The ability of gateway nodes to filter blockweave data that users are exposed to.
- Content matching protocols such as PhotoDNA (Microsoft) can be implemented.
- Differing content policies can cause forks.
- Gateways offer opportunity to access content with agreeable policy.



## Use cases

- Providing a reliable archive of record
    - Once data is added to the blockweave, it cannot be removed or altered, either intentionally or unintentionally.
  - Authenticity.
    - The blockweave offers 'proof of existence' of a specific piece of data at a certain point in time, based on the associated transaction's verifiable, reliable timestamp.
  - Provenance.
    - Each transaction is linked permanently to the previous transaction from the same wallet, meaning that end-users can consistently verify the true origin of the data inside any transaction by wallet address, including that of decentralised applications hosted on the Arweave network.
  - Decentralised application hosting.
    - The blockweave ensures more reliable access to applications than any centralised application-hosting platform, which commonly negatively impact application integrity.
  - Incentivised data storing and serving.
    - Arweave's unique mechanism design robustly encourages the rapid serving of data to all participants in the network, including end-users.
-

## NFT Storage



From [Docs](#)

NFT.Storage is a storage service that lets you upload off-chain NFT data for free, with the goal to store all NFT data as a public good.

The data is stored perpetually in the [Filecoin decentralized storage network](#) and made available over [IPFS](#) via its unique content ID.

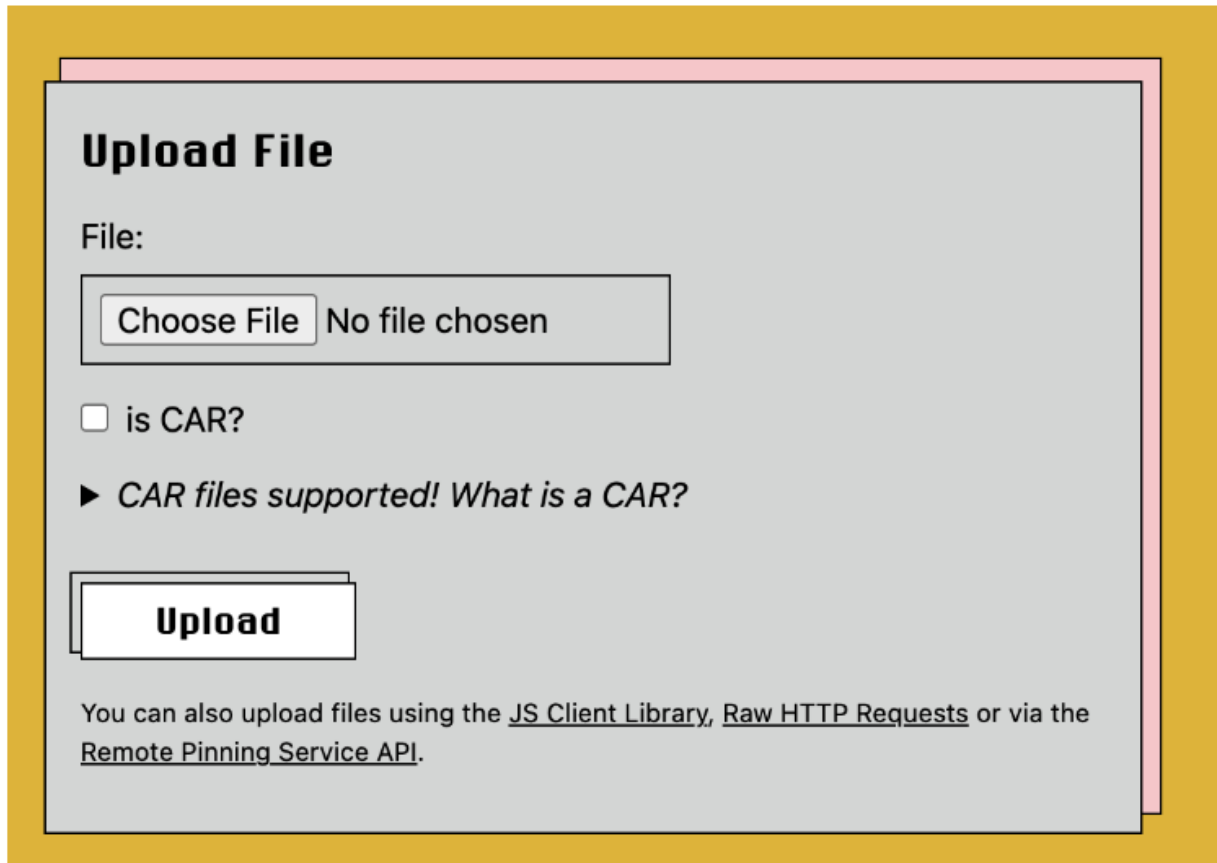
You can upload as much data as you want as long as it's part of an NFT (e.g., metadata, images and other assets referenced in a token or its metadata), although there is currently a limit of 31GiB per individual upload.

### Creating an account

Go to [Login](#)

### Uploading files via the website

1. Click **Files** to go to your [NFT.Storage file listing page](#).
2. Click the **Upload** button to go to the [File upload page](#).
3. Click the **Choose File** button to select a file from your device:

A screenshot of the 'Upload File' interface. The title 'Upload File' is at the top. Below it, the label 'File:' is followed by a button labeled 'Choose File' and the text 'No file chosen'. Below this is a checkbox labeled 'is CAR?'. Underneath the checkbox is a link that says '► CAR files supported! What is a CAR?'. At the bottom of the form is a large 'Upload' button. Below the button, there is a paragraph of text: 'You can also upload files using the [JS Client Library](#), [Raw HTTP Requests](#) or via the [Remote Pinning Service API](#).' The entire interface is set against a light gray background with a yellow border.

You can alternatively use the NFT [App](#)

You can interact using the node package

Install with

```
npm install nft.storage
```

there is some example client code [here](#)

## NFT Storage Gateway

The project created a new HTTP gateway that uses existing public IPFS infrastructure and cloud-native caching strategies to provide a high-performance, CID-based HTTP retrieval solution that is NFT-focused.

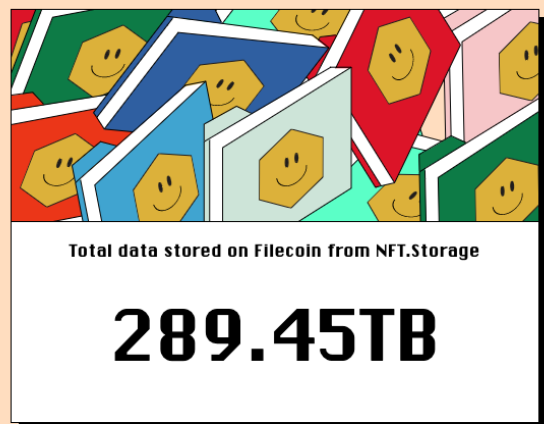
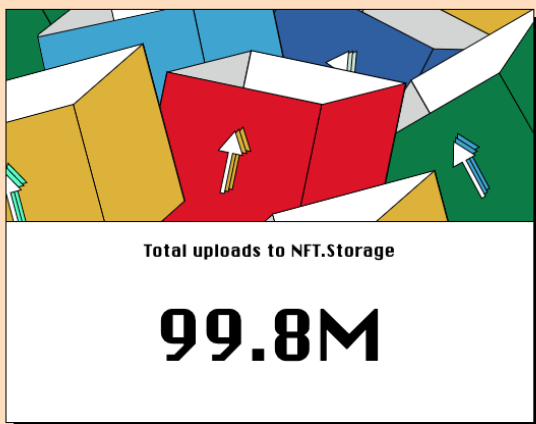
The NFT.Storage gateway is accessible at the URL `https://nftstorage.link`, and you can use it just by creating URLs with `nftstorage.link` as the host.

## NFT Market By the Numbers

### The Price of Missing NFTS ([reference](#))

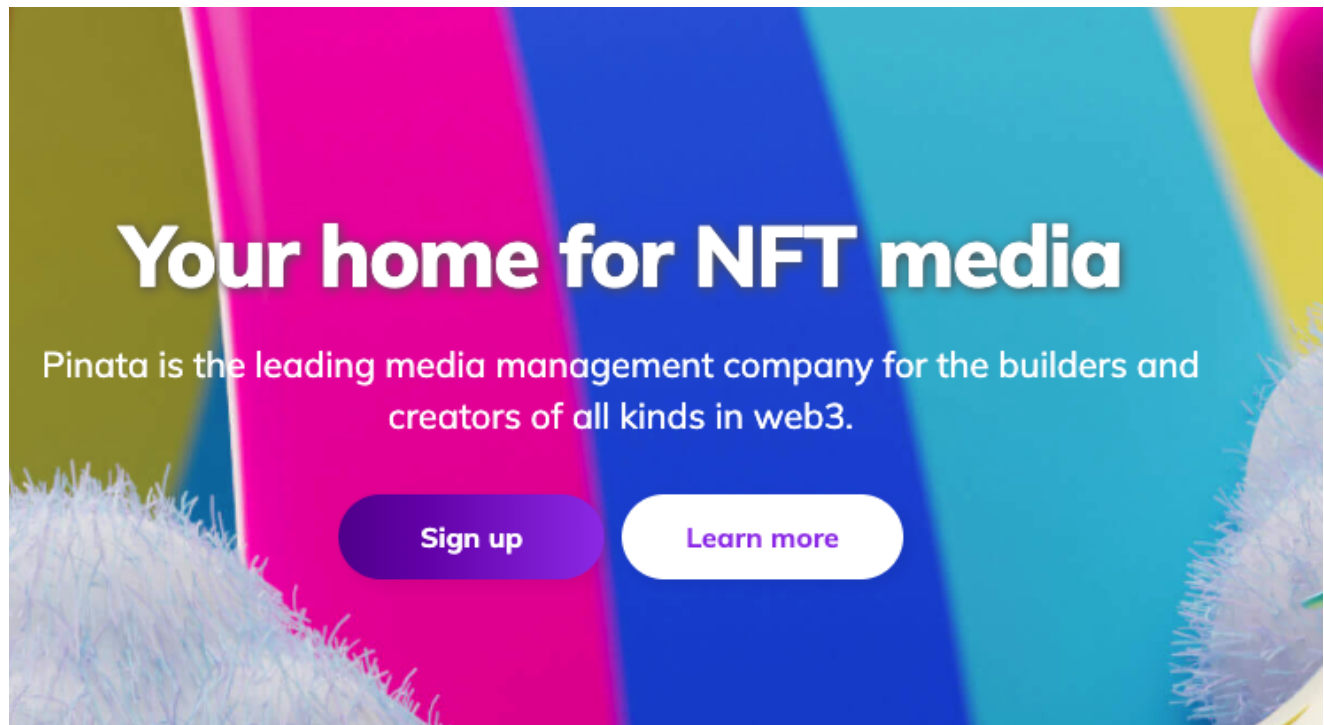


## NFT.Storage by the numbers



## Pinata

See [Docs](#)



### Flexible plans

For all builders & creators of web3

<p>FREE (FOREVER)</p> <p><b>\$0</b></p> <p>For creators who are just starting their web3 journey and want to experience what it's like storing their content on IPFS through Pinata.</p> <p>Select plan</p>	<p>Most Popular!</p> <p>PICNIC</p> <p><b>\$20</b> / mo</p> <p>For the rising artist or builder with a new app or PFP collection who wants to get their work in front of a wide audience and offer a fast, personalized experience.</p> <p>Select plan</p>	<p>FIESTA</p> <p><b>\$100</b> / mo</p> <p>For the early-stage marketplaces, more advanced web3 apps and games, and developers who work with a decent amount of data to scale without worrying about infrastructure.</p> <p>Select plan</p>	<p>CARNIVAL</p> <p><b>\$1,000</b> / mo</p> <p>For the brands, media companies, and web3 builders who need all the bells and whistles and can't afford to be slowed down by building things in-house and only getting half the benefits.</p> <p>Select plan</p>
---	---	--	--

### Features

- Dedicated gateway
- Built in URL shortener
- Built in CDN to optimise for video and images
- Allows sharing of unlockable content. (NFT verification / geo location)