Decentralized Storage

I. Executive Summary

There are many cases of data center fires that hamper the productivity of company performance. We believe that with a single entity, our data will not be accessible if the storage crashes. The key is to decentralize storage (dStorage) so that if one storage node crashes, other nodes can be accessed by users. For now, many decentralized storage services like IPFS, Filecoin, and Arweave. The platform is to store our data in a peer-to-peer or blockchain network with each feature and pricing. dStorage has a persistence and challenge mechanism. Every node in the persistence mechanism replicates the data, and new pieces of data are appended to the end, and contract-based persistence mechanisms guarantee persistence through contract agreements. For a challenge mechanism to make sure data is retained, a dStorage could issue a cryptographic challenge to the nodes to make sure they store the data as promised and users need to trust the pinning service provider that the provider has indeed pinned the file as per their terms and conditions.

Key takeaways that support our thesis:

- IPFS is a free distributed storage system for storing files, applications, and various data in massive peer-to-peer networks with a DHT mechanism to upload and make user's file available and user can start to run a node with install an IPFS Desktop Application and upload the file to start hosting the files.
- After a user uploads the file, IPFS splits the user's file into multiple pieces and generates a content identifier (CID) as a hash used as an address for the user to query the file through the IPFS gateway network. When the user retrieves the data, the node that retrieves it keeps it in its cache to be accessed by other nodes that request the same file. IPFS nodes frequently clear this cache out to make room for new content, so the user needs to pin the data on an IPFS to keep the data. IPFS needs ~44s on average to retrieve newly created content, but only ~12 ms to retrieve cached content.
- The user also can use pinning services like Pinata for users who do not run their node to keep users' data on the network because the pinning services have IPFS nodes that are always online with a nearly instant retrieval speed. The users can directly use the Pinata web app to pin content to IPFS or Pinata API to ensure the file's reliability and accessibility. But Pinning services are paid services depending on the features the user wants to use.
- The user can use Filecoin with Filecoin native token ($FIL) to store the data and transact in the Filecoin network, Filecoin is a monthly-paid peer-to-peer network for storing data in the distributed contract-based storage marketplace. Because Filecoin is used with IPFS, the data is always available and can be recovered instantly, while also safe and verifiably backed up on the Filecoin network over time.
- When using Filecoin the user must make a deal with storage providers (miners) to store the data. The miners need to run Proof of Replication (PoRep) and Proof of Space and Time (PoSt) schemes to upload and "seal" the data in the Filecoin network. Because the data retrieval process requires "unsealing" data, it takes 1-5 hours to retrieve the data on Filecoin, making it only suitable for backup. But unfortunately, using Filecoin is complex because the user must create a storage deal and find a storage provider through the registry manually through the command-line interface.
- The user can also use Arweave with Arweave native token ($AR) to store the data and transact in the Arweave network. Arweave is a blockchain that provides affordable storage where users have to pay a one-time, up-front fee to store their data permanently.
- Unlike conventional blockchain, Arweave uses the blockweave protocol, where each node must store the entire blockchain. For the consensus mechanism, Arweave uses Succinct Proof of Random Access (SPoRA) SPoRA is just Proof of Work (PoW) to mine a block + Proof of Access (PoA) to the miners to prove a piece of past data in a selected random block with consideration for

retrieval speed. Arweave can reach up to 5000 transactions per second (TPS) and retrieve data in roughly 400 ms.

- The uniqueness of Arweave is Arweave has a storage endowment mechanism, it is intended that only about 14% of fees paid go to miners, with the remainder going to the endowment pool. The mechanism is miners in profit when network fee revenue exceeds their cost of maintaining the blockweave data network. However, when nodes become unprofitable, the endowment is used to subsidize the miners' operations.
- The user can directly use the ArDrive web app to store the data, but the user needs the Arweave wallet with some AR tokens to pay for the service, and developers can upload data to Arweave by ArDrive with its Javascript SDK.

II. Preface

A non-fungible token (NFT) is a token that is used as a unique identifier of assets that are recorded on a blockchain. Without a doubt, the most common asset that NFT holds has to be images.

Unlike what most people think, digital assets such as images and media files are not stored on-chain because storage on conventional blockchains like Ethereum are prohibitively expensive. For example, the Bored Ape below is ~500kb. At the cost of data storage of 640k gas per kilobyte and the current gas price of about 15 Gwei, storing the Bored Ape below would cost ~4.8 ETH.

What most NFT projects do is that they store the image on off-chain storage and store the link to the image as part of the NFT's metadata. Only the token id and the metadata, which are just texts, are stored on the blockchain. The example below is only ~400 bytes, which costs only ~0.004 ETH to store.

```
my-nft.json

{
   "description": "Friendly OpenSea Creature",
   "image": "https://opensea-prod.appspot.com/puffs/3.png",
   "name": "Dave Starbelly",
   "attributes": [
       { "trait_type": "Base", "value": "Starfish" },
       { "trait_type": "Eyes", "value": "Big" },
       { "trait_type": "Mouth","value": "Surprised" },
   ]
}
```

By moving large files to off-chain storage, we pay 1/1200 of the cost! This goes to show the importance of storage to the web3 industry. This report serves as a primer to storage and a survey to compare centralized and decentralized solutions, with a focus on decentralized solutions as we think it is a superior technology.

III. Table of Contents

IV. Overview

Storage is the preservation and retrieval of data on a computer or device. Storage has moved from floppy disks (1.44MB) to hard disks on our computer, and hard disks in a remote location, i.e., cloud storage.

Cloud storage has a lot to desire: it is easy to use, efficient and programmable. It is no wonder that the cloud storage market is valued at 70.19 billion U.S. dollars in 2021, and is estimated to grow to 83.41 billion U.S. dollars in 2022 and 376.67 billion U.S. dollars by 2029, with a forecast compound annual growth rate of 24 percent.

However, cloud storage providers are centralized, i.e. they are operated and offered by a single entity. Amazon Web Services (AWS) is a centralized cloud storage provider that powers many apps that we use today: Airbnb, Spotify, and even Dropbox use AWS as their storage. They are popular for a reason:

- The cost to store 1 GB of data on S3 is about $0.025 per month. For context, 1GB allows us to store 2000 Bored Ape images mentioned in the Preface. It gets even cheaper at $0.0138 per GB per month for infrequently accessed files.
- Further, file retrieval on S3 can happen within tens to hundreds of milliseconds
- It is easy to use and composable. I can build a data pipeline that uses S3 as storage.

The danger of centralized cloud storage, however, is they provide convenient opportunities for censorship, and they are single points of failure. When either of these happens, data becomes inaccessible. To name a few instances, AWS had an outage in 2017, GitHub experienced an extended interruption in June 2020 and Microsoft suffered multiple weeks of cloud service problems in October 2020.

An answer to these issues is decentralized storage (dStorage). dStorage, as opposed to centralized servers run by a single business or organization, is made up of a peer-to-peer network of user-operators who each keep a share of the total amount of data, resulting in a robust system for sharing file storage. These can be found in any peer-to-peer (P2P) network such as IPFS (Interplanetary File System) or blockchain-based application such as Arweave.

P2P network dated as far back as 1999 where Napster took the world by storm by offering a platform for users to share media files freely with their peers. However, Napster has a central server that serves as a place where peers can meet and exchange directories of files, so they can be shut down easily. Improving upon the shortcomings of Napster, BitTorrent provides an application that you can use to connect to any of

the millions of tracker servers out there. It is as if there are millions of Napsters out there instead of just one.

IPFS improves on BitTorrent by removing the need of a central tracker server by baking in the directory in the node. When you run an IPFS node, you participate by default in storing a part of the Distributed Hash Table (DHT), which maps from CIDs to people who have the content addressed by that CID. The hash table is distributed because no single node in the network holds the entire table. Instead, each node stores a subset of the hash table, as well as information about which nodes are storing other relevant sections.

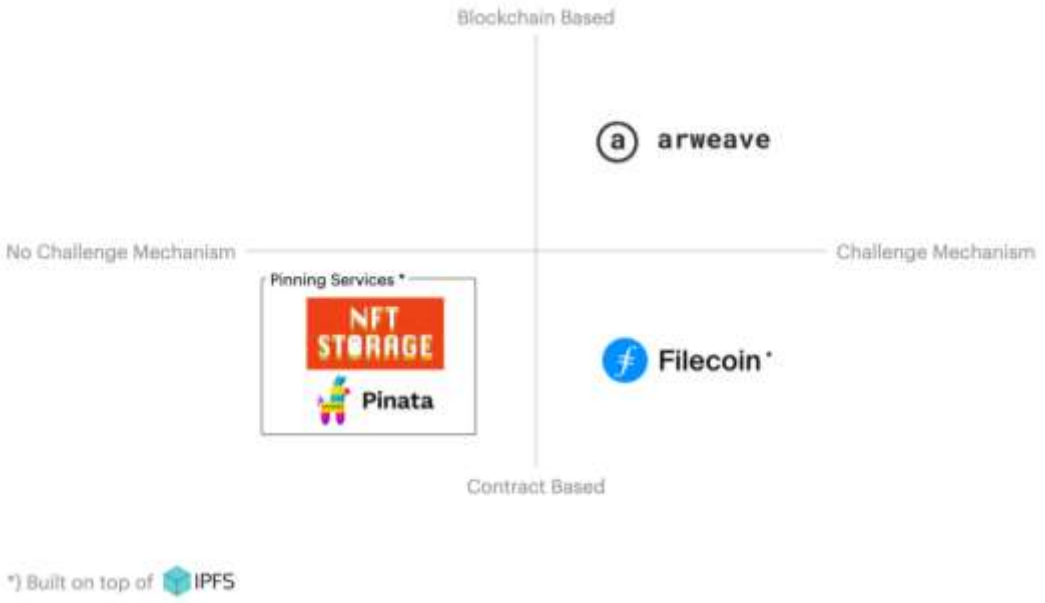
When we upload a file to IPFS, what actually happens is that IPFS adds an entry to the DHT that maps from CID to the IP address of our file. Anyone who wants to download our file would look up the CID in the DHT to get the IP address, and download the file directly from us.

This also implies that our file is persistently available if there is at least one IPFS node online that can serve the file. There are three ways for content to achieve such persistence:

- We upload the file to our own IPFS node and keep it connected to the Internet, just like an ordinary HTTP server
- When someone views our file using their node, their node will cache the file for an hour before the cache gets cleared. In theory, if our content gets at least one view every hour it could live on in the users' caches forever, but we cannot count on this method.
- Someone labels (“pins”) our file as important on their node. The node will then keep the file permanently and serve it whenever it is connected to the Internet, until the node owner unpins it.

The third method in particular gives rise to pinning services such as Pinata and NFT Storage. At its core, they are not very different from a web hosting business. They run IPFS nodes on the customers’ behalf so that the customers do not have to run the nodes themselves. While convenient, using such service is antithesis to the ethos of decentralization. The community has come up with alternative solutions to pinning services, most notably Filecoin and Arweave.

If we consider all solutions that attempt to address data impermanence on IPFS, they can largely be segmented by their persistence and challenge mechanism:



Persistence Mechanism

- Blockchain-based. In a blockchain-based persistence mechanism, every node replicates the data and new pieces of data are appended to the end. In addition, the blockchain must also have an incentive structure so that the validator is paid to add the data. Arweave is an example that uses a blockchain-based persistence mechanism.
- Contract-based. In this mechanism, persistence is guaranteed through contract agreements. Filecoin and pinning services for example only guarantees persistence of your files for the duration you pay them for. It is important to note that while Filecoin has a blockchain, it is for recording contracts between storage providers and users, and not storing the data.

Challenge Mechanism

- To make sure data is retained, a dStorage could issue a cryptographic challenge to the nodes to make sure they store the data as promised. Arweave's proof-of-access, for example, issues such a challenge to the node at both the most recent block and a random block in the past. If the node is unable to come up with a proof that they still store the data they promise to store, it gets penalized.
- In absence of a challenge mechanism, as with the case with pinning services, users need to trust that the provider has indeed pinned your file as per their terms and conditions.

Having provided an overview, we are ready to deep dive into how each solution mentioned above works in the next section.

V. Deep Dives

IPFS

As mentioned in the overview, IPFS is a distributed file storage system that allows computers all over the world to store and deliver files, webpages, applications, and data as part of a massive peer-to-peer network. IPFS is not a blockchain but rather works like BitTorrent with a decentralized directory called DHT and baked in incentives to make the DHT available to others. In order to upload and make your file available, you need to run a node. When you run the node, you store and share part of the DHT with others in the network.

How It Works

Uploading

- When the user puts a file onto the IPFS platform, IPFS splits it into multiple pieces and creates a hash from the file, with which a content identifier (CID) is constructed and returned to the user. Naturally, changing a file changes its hash, and consequently its CID which is used as an address.
- The CID can be thought of as a label used to point to its associated file in IPFS that the user can use to request the file through the IPFS network or even on a browser
- Upon successful file upload, IPFS will update the DHT to map the CID to the IP address of our IPFS node so that it can be discovered and accessed by others in the network

Retrieval

- When someone queries the file through an IPFS gateway, the query is passed to the DHT which returns the IP address of the node
- The node storing the file is then sent a request to retrieve the file

- When the file is retrieved, the node that retrieves it keeps it in its cache to be accessed by other nodes that request the same file

Garbage Collection

- IPFS uses garbage collection to free disk space on IPFS nodes by deleting cached data that it thinks is no longer needed
- To ensure that the data is not deleted, the owner needs to “pin” the data on an IPFS node, which is basically telling IPFS that the resource is important and should be saved
- Anyone can run their own node and pin the data to their own node, but this means the computer needs to always be connected to IPFS. This is why pinning services exist

Cost

IPFS is made free through subsidies from various IPFS storage providers

Retrieval Speed

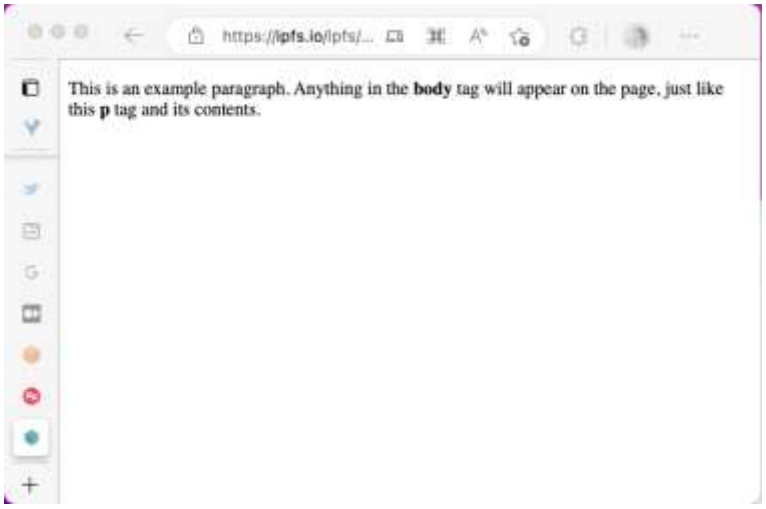
Tests reveal that it takes ~44s on average to retrieve newly created content, but only ~12 ms to retrieve cached content. Because of the nature of content addressing of IPFS, there is no indication of missing content, and waiting for a request timeout is the only way to tell whether the content exists.

In practice however it depends on how many nodes have the data that you need, how close they are in relation to you, and the bandwidth, among others. Check out the tux.art NFT minting platform and marketplace to see firsthand how much a dApp takes to load files stored on IPFS.

Using IPFS

To use IPFS, the user can install an IPFS Desktop application and upload the file to start hosting their files. To demonstrate how it works, we upload a simple webpage to the IPFS Desktop application and access the webpage using the CID returned by IPFS.

The webpage is now accessible via the browser
https://ipfs.io/ipfs/QmXwECqnS7eEeg8tvwL7w963MhqV5vvQwD5UZGkxZM2ZEE?filename=index.html
although when it may not be available to you since I am not running my IPFS node at the time you are reading this report.



For developers who are building a dapp, they can upload files to IPFS programmatically via the IPFS Gateway API to either their self-hosted IPFS node or a third-party IPFS node provided by companies such as Infura or Pinata

Pinning Services

The most common way to store and retrieve data on IPFS is the use of pinning services like Pinata for users who do not run their own node. These services run IPFS nodes and keep your data present on the network.

How It Works

To ensure a file is available, the data owner needs to pin it either on his own IPFS node or a pinning service. The former method is not difficult, but it can be inconvenient since users have to manage their own hardware. This problem gave rise to pinning services, paid services that allow users to upload the data to a remotely hosted IPFS node and retrieve it whenever they want.

Pinning services just run lots of IPFS nodes and allow users to pin data on those nodes for a fee to ensure the file's reliability and accessibility. There are several commercial pinning services such as Pinata, NFT.Storage, Temporal, Crust, and Infura but they work the same way.

On top of persistence, IPFS pinning services offer the following benefits:

- Speed. IPFS pinning services are often highly connected to other nodes in the IPFS network. This means that they have a better chance of quickly finding and retrieving data from the network than your nodes.
- Uptime. IPFS pinning services have IPFS nodes that are always online. Because these nodes are usually cloud-hosted, they act as a reliable way of keeping your data available even if your own IPFS node isn’t always online. This allows you and your users to access your content anywhere at any time, regardless of device.
- Redundancy. For those running their own IPFS nodes, IPFS pinning services provide a great way to keep data backed up

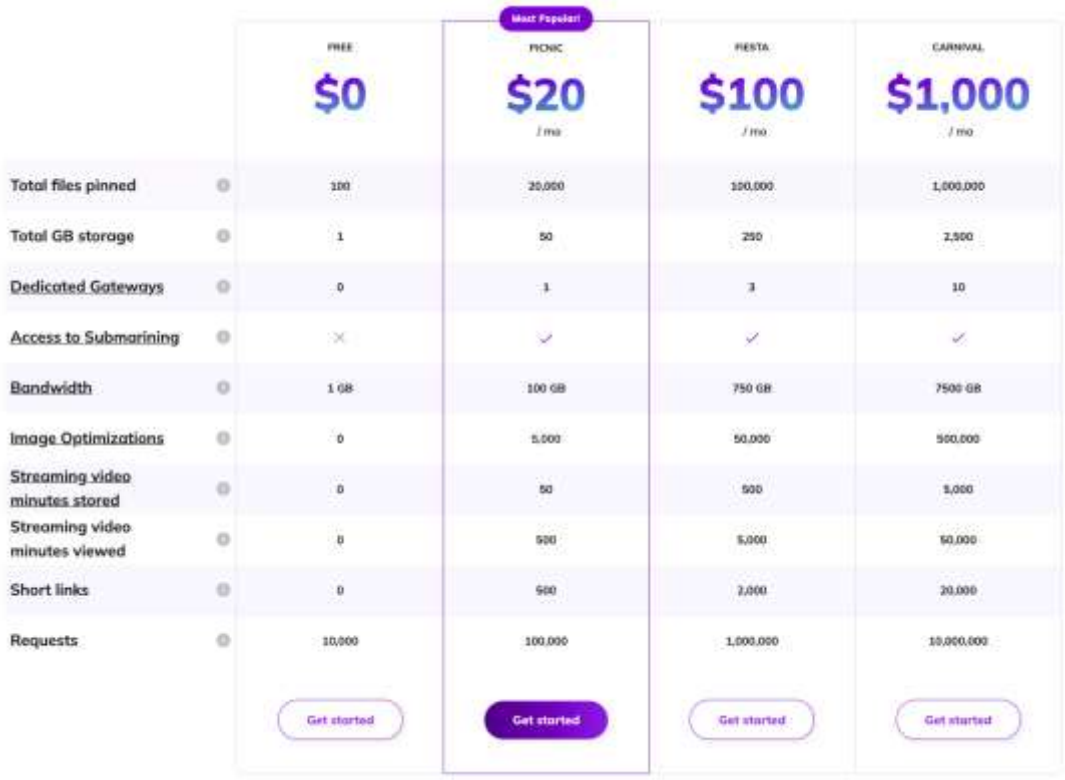
The major downside of using a pinning service is that it is centralized, which means it can go out of business or have any typical issue faced by a centralized entity. When that happens, your data gets unpinned and might get lost forever. For this reason, data owners should consider multiple backups by either using multiple service providers or managing their own pins.

Cost

The pinning service costs vary between providers, but their pricing model reminds us of web hosting services since their service is similar.

For reference, Pinata, one of the most popular pinning services, charges from $0 to $1000 per month, depending on the number of files pinned, storage size, bandwidth, and the number of requests, among others.

To store all 10,000 Bored Apes, Yuga Labs would only need the Picnic plan assuming that the average Bored Ape owner does not access their NFT more than 10 times per month.



Retrieval Speed

As mentioned, the speed of retrieval is nearly instant because IPFS pinning services are often highly connected to other nodes and can download from multiple nodes in parallel.

Using Pinning Services

The UX for pinning services is similar; users can pin content to IPFS via an interface or API.
For the purpose of demonstration, we will use Pinata's Pinmanager interface and the same file

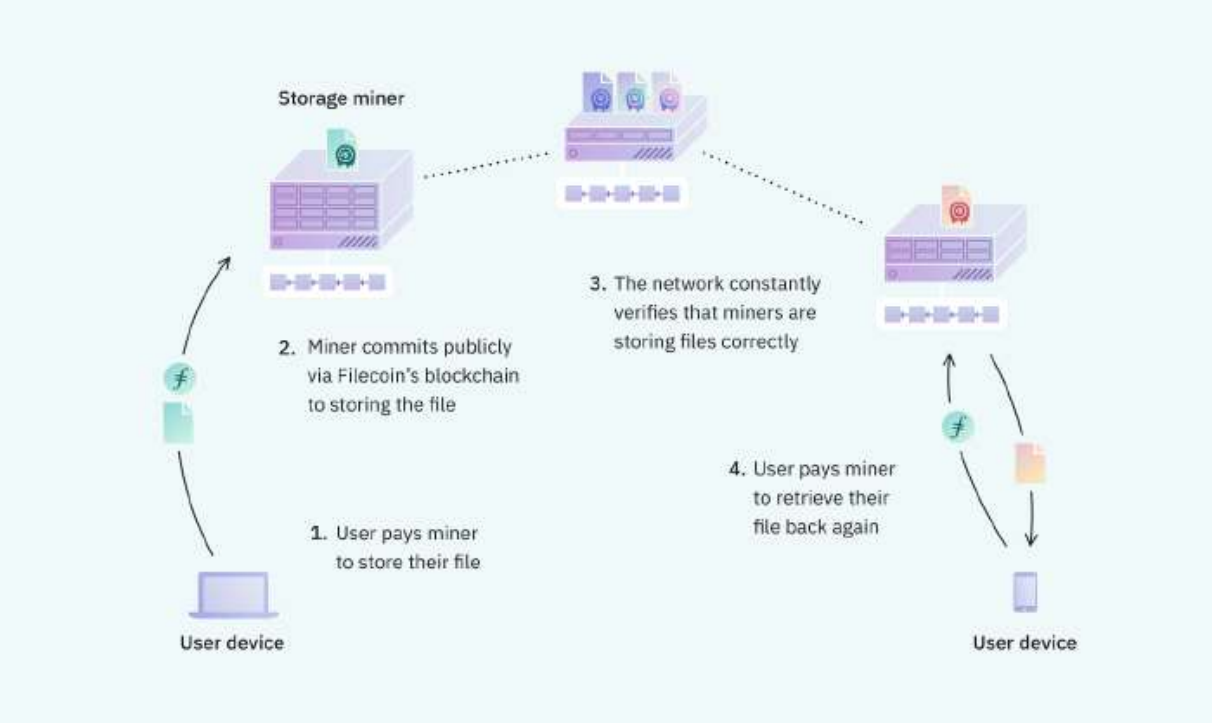
Filecoin

Developed by the same company that brought us IPFS, Filecoin offers an alternative to pinning services to provide data persistence that is missing in IPFS, by creating a distributed contract-based storage marketplace. More formally, Filecoin is a peer-to-peer network blockchain that stores files, with built-in economic incentives to ensure files are stored reliably over time.

Filecoin is not intended to replace IPFS. In fact, Filecoin is used together with IPFS to benefit from the best of both worlds. In such a scenario, the user can back up data on the Filecoin network while also ensuring content is discoverable in the IPFS network. This way, data is constantly available and can be retrieved quickly, while also making sure that it is safely and verifiably backed up on the Filecoin network over time.

How It Works

- The user cuts a storage deal with the storage providers (miners), including the storage duration and the price
- Once the deal is struck, the user can pay the fees and upload his files to Filecoin
- The network then records the transaction and miners need to demonstrate using Proof of Replication (PoRep) that they have encoded the data in a way unique to that miner. This encoding process is called “sealing.”
- Throughout the duration of the deal, the network regularly issues a challenge to make sure that miners are storing the data correctly in a scheme called Proof of Space and Time (PoSt). If it couldn’t provide the “sealed” file, the miner would be penalized
- In return, the storage providers earn units of $FIL



Cost

Filecoin cost per GB per month is about 0.000000026113910357305 13 FIL or USD 0.00000012. This is 0.00087% the cost of S3's infrequently accessed layer.

Retrieval Speed

Because the data retrieval process requires "unsealing" data, it takes 1-5 hours to retrieve the data on Filecoin, making it only suitable for backup. To get performance that is similar to other hot storage solutions, most users utilize Filecoin with a caching layer such as IPFS. These hybrid and multi-tiered storage solutions use IPFS for hot storage and Filecoin for affordable, frequent, and versioned backups.

Using Filecoin

Using Filecoin directly is technical and cumbersome. The user would have to set up his own Filecoin node called Lotus, get a FIL address, import your data into Lotus, find a storage provider through the registry manually from https://filrep.io and create a storage deal, all done through the command-line interface.

For instance, to create a storage deal, you need to provide the CID of the data imported, the number of days you want to keep the file on Filecoin and the miner IDs that you want to work with. The command-line will return an offer from the miner IDs, which you can accept or reject.

The community has built tools and services that make interacting with Filecoin easier. One popular choice is ChainSafe Files, which provides a Dropbox-like interface and abstracts out the process of finding a storage provider and cutting a deal.

For app builders, there are "middlemen" providers such as Estuary that handle the complex storage contract cutting process in the background and provide an easy-to-use API. When you store data using Estuary, that data will go to an Estuary node. Initially, all data is stored in IPFS for fast retrieval ("hot storage"). While this data is stored, each node will look through a curated list of Filecoin storage providers and initiate storage deals to move IPFS into long term storage that is persistent.

Token

The Filecoin token ($FIL) is the native crypto token of the Filecoin network, Token holders can use $FIL to participate and transact in the Filecoin network. In particular, users pay miners in $FIL to store or distribute data and to retrieve their information. Storage providers also post $FIL as collateral to provide a minimum level of guarantee of their service, which gets slashed should a deal with a customer fall through.

According to data from https://dashboard.starboard.ventures/circulating-supply, the current circulating supply of Filecoin (FIL) tokens is approximately 410 million. The maximum supply of FIL tokens is 2 billion, so this means that approximately 20% of the total token supply is currently in circulation. Over the past year, there have been various movements of tokens, including mining, burning, vesting, and locking. The net effect of these movements has been an increase in the circulating supply, from 220 million to 410 million, resulting in an inflation rate of approximately 90% over the past year.

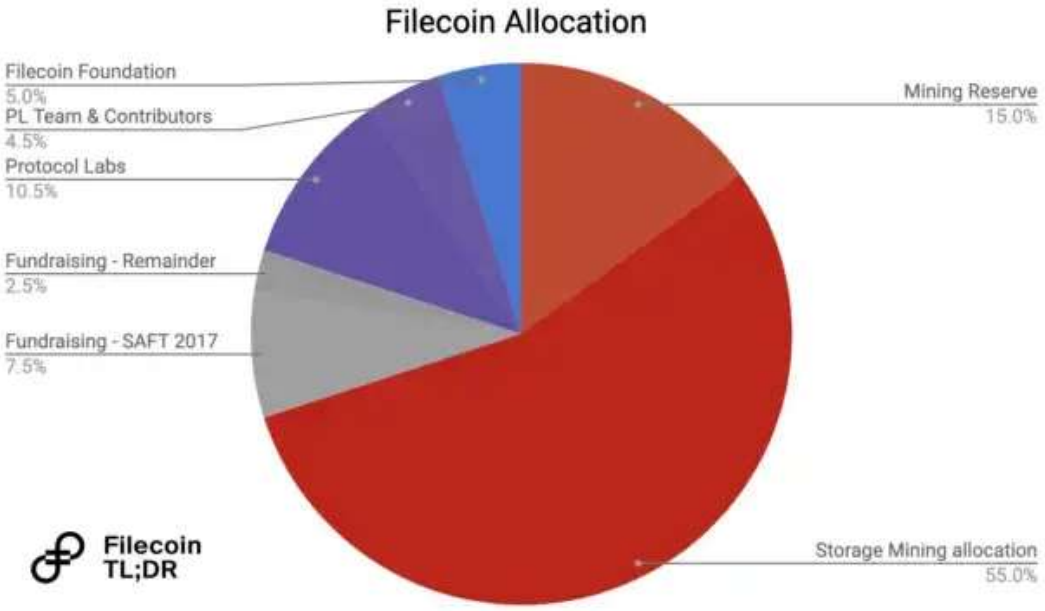
Distribution

FIL has a maximum supply of 2 billion tokens, of which 600 million were pre-allocated at the project's inception. The pre-allocated tokens were distributed according to the following percentages:

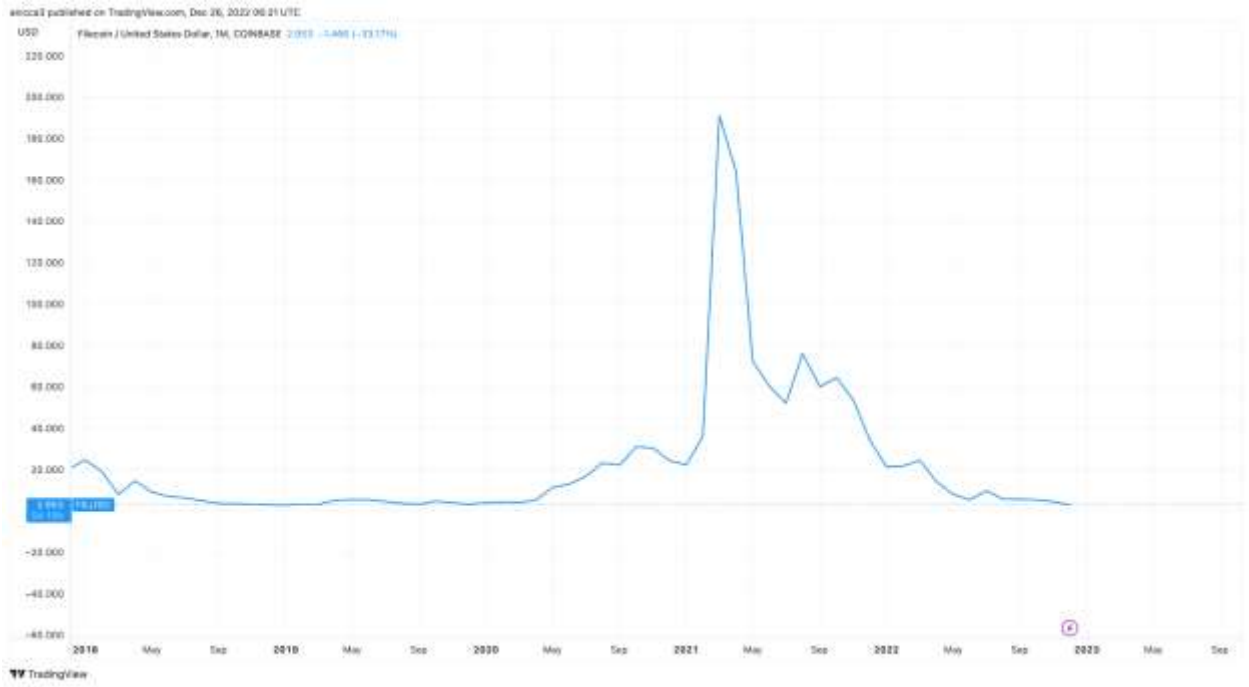
- Protocol Labs received 10.5% with a 6-year linear vesting period
- Protocol Labs team members and contributors received 4.5% with a similar vesting period
- 2017 SAFT investors received 7.5% with a vesting period ranging from 6 months to 3 years
- 2.5% was set aside for future fundraising or ecosystem development

- the Filecoin Foundation received 5% with a 6-year linear vesting period

The remaining 70% of the total token supply will be released to Filecoin miners over time as a reward for their contributions to the network, including providing data storage services, maintaining the blockchain, distributing data, and running applications.



FIL has experienced significant price fluctuations from just a few dollars to $164 in March 2021 and back to only $3 as of December 2022 demonstrate the high volatility of this cryptocurrency.



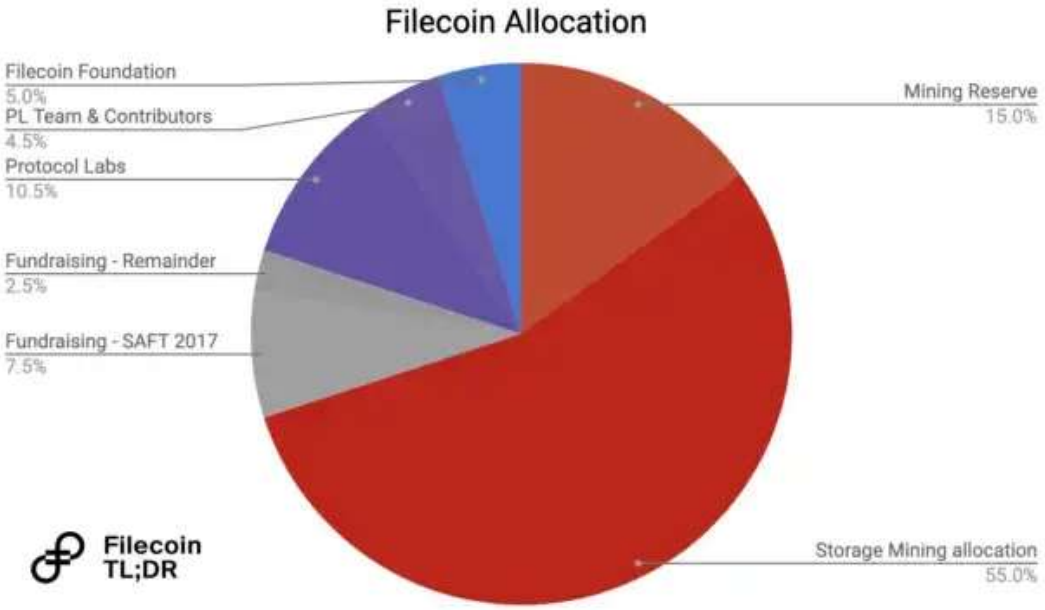
FIL tokens can be purchased on a variety of cryptocurrency exchanges, including Binance, Huobi, and Kraken. It is also possible to purchase FIL through decentralized exchanges such as Uniswap.

Token Supply

Starboard Blog GeckoSupply

Circulating Supply	350,970,724 FIL
Protocol Circulating Supply	408,910,000 FIL
Max Supply	1,965,995,138 FIL
Total Supply	1,965,995,138 FIL
% Supply Change From 1 Year Ago	+150%
Emission Type	Premined-Rewards: The term "pre-mined" means a portion of the coins has been mined and distributed.

Circulating Supply:The total amount of $FIL tokens in circulation according to API statecirculatingsupply definition (Minus burned token)
Protocol Circulating Supply: The total amount of $FIL tokens in circulation according to the Protocol's definition (Plus burned token)



Token Holder (Rich List)

The addresses that hold a large number of tokens or Whales of $FIL

Normal Account

Address	Balance	Balance Ratio $(\frac{Balance}{Current\ total\ Supply})$
f086971	45,410,783.5731 FIL	2.3%
f047684	33,551,052.437562958 FIL	1.7%
f01986715	20,685,479.121347427 FIL	1%
f01259647	7,999,999.857831205 FIL	0.4%
f0128006	7,222,289.448561907 FIL	0.3%

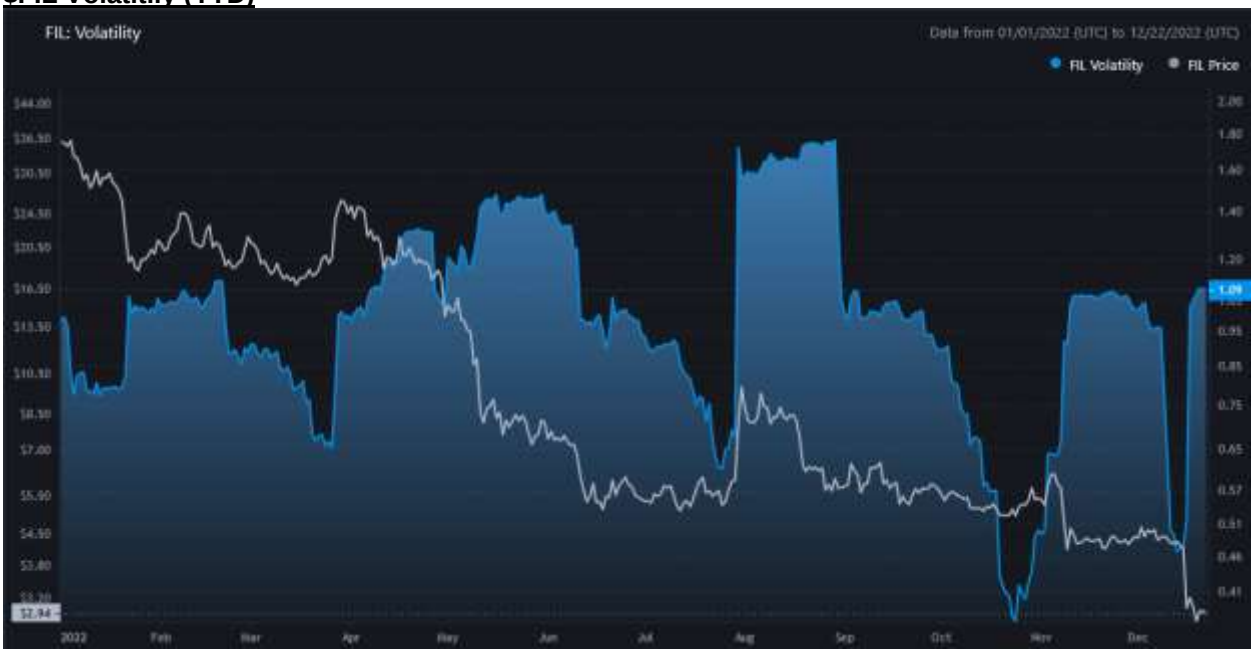
Miner Account

Address	Balance	Balance Ratio $(\frac{Balance}{Current\ total\ Supply})$
f0127595	1,735,896.3216 FIL	0.09%
f0688165	1,617,612.0622 FIL	0.08%
f0123261	1,523,162.6465 FIL	0.07%

f01173170	1,480,904.4945 FIL	0.07%
f0128559	1,340,889.5976 FIL	0.06%

Token Price Action

$FIL Volatitily (YTD)



Filecoin Market

Market

Pair	Volume	Exchange	Type
FIL/USDT	**$10.44M**	**Binance**	**CEX**
FIL/USDT	**$6.36M**	**OKEx**	**CEX**
FIL/USD	**$4.56M**	**Coinbase**	**CEX**
FIL/USDT	**$0.72M**	**Huobi**	**CEX**

Collabs

OpenSea?

Arweave

Arweave is a blockchain that provides affordable storage where users have to pay a one-time, up-front fee to store their data permanently. In contrast to Filecoin where users have to negotiate a storage deal—the storage duration and the cost—with the miners, Arweave has a fixed storage duration of 200 years and the price is determined by Arweave's model.

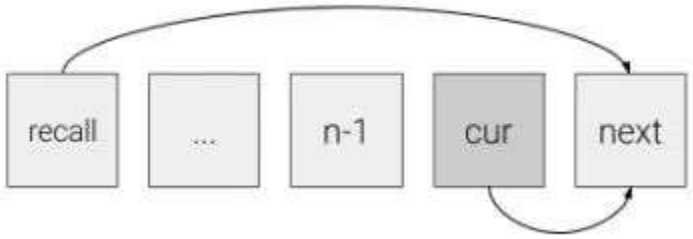
How It Works

Arweave is made possible with the following innovations:

Blockweave

Unlike conventional blockchain where each node must store the blockchain in its entirety, Arweave does not require miners to store every previous block. It does so by using its consensus mechanism, Proof of Access, which does not only require the miner to have access to the previous block but also a random block ("a recall block"). If the particular miner does not have access to the recall block then it is not eligible to mine the data.

As a result, Arweave's blockchain looks more like a weave than a chain and is aptly named the blockweave.



However, if the node does not store the entire blockchain then certainly there exists a probability that a block is not being stored by any of the miners thus resulting in data loss. Amber Group has done research to calculate the probability of a drop block:

		Replication Rate (%)								
		20%	30%	40%	50%	60%	70%	80%	90%	100%
Nodes	50	1.4 E-05	1.8 E-08	8.1 E-12	8.9 E-16	1.3 E-20	7.2 E-27	1.1 E-35	1.0 E-50	0.0 E+00
	100	2.0 E-10	3.2 E-16	6.5 E-23	7.9 E-31	1.6 E-40	5.2 E-53	1.3 E-70	1.0 E-100	0.0 E+00
	150	2.9 E-15	5.8 E-24	5.3 E-34	7.0 E-46	2.0 E-60	3.7 E-79	1.4 E-105	1.0 E-150	0.0 E+00
	200	4.1 E-20	1.0 E-31	4.3 E-45	6.2 E-61	2.6 E-80	2.7 E-105	1.6 E-140	1.0 E-200	0.0 E+00
	250	5.9 E-25	1.9 E-39	3.4 E-56	5.5 E-76	3.3 E-100	1.9 E-131	1.8 E-175	1.0 E-250	0.0 E+00
	300	8.5 E-30	3.4 E-47	2.8 E-67	4.9 E-91	4.1 E-120	1.4 E-157	2.0 E-210	1.0 E-300	0.0 E+00
	350	1.2 E-34	6.1 E-55	2.3 E-78	4.4 E-106	5.3 E-140	9.8 E-184	2.3 E-245	0.0 E+00	0.0 E+00
	400	1.7 E-39	1.1 E-62	1.8 E-89	3.9 E-121	6.7 E-160	7.1 E-210	2.6 E-280	0.0 E+00	0.0 E+00

The replication rate is the ratio of the average number of blocks stored by the miner so for example if the network has like 10,000 blocks and each of these miners store 9,000 blocks then the replication rate is 90%. According to ViewBlock, Arweave has 70 online nodes. However, most nodes store 100% of the data in the network leading to virtually zero chance of a drop block.

Consensus Mechanism

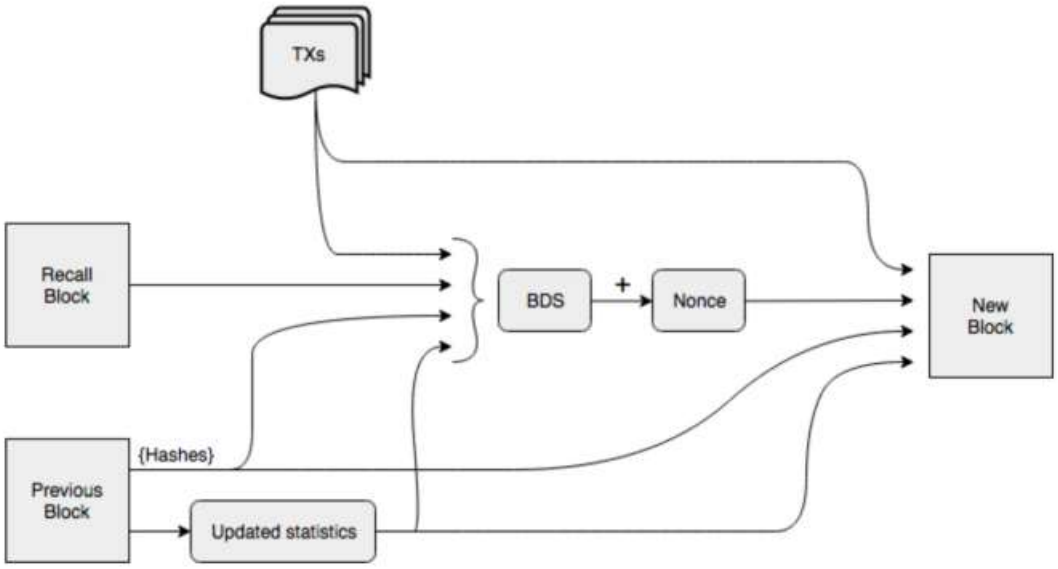
Arweave uses the classical Proof of Work (PoW) to mine a block, but with a unique twist. In a mechanism called Proof of Access (PoA), the miner can only begin mining a block after they have proven that they own a piece of past data in a selected random block, known as a "recall block." This means that a miner's block probability is determined by the probability of having a randomly recalled block and the probability of being the first to find the hash.

This system incentivizes miners to store as much data as possible in order to increase their chances of being selected to mine a block and earn rewards. It also means that rarer blocks will become more popular among miners, as they can compete with a smaller number of miners by storing rarer blocks instead of well-replicated ones.

However, this consensus mechanism does not provide incentives for miners to retrieve data quickly, so miners often outsource storage to remote storage pools that are cheaper but slower to access and centralized.

To address this issue, Arweave switches to a new consensus mechanism called Succinct Proof of Random Access (SPoRA), which combines PoA and PoW with consideration for retrieval speed. This incentivizes miners to store data locally and in multiple locations, increasing decentralization and improving retrieval speed.

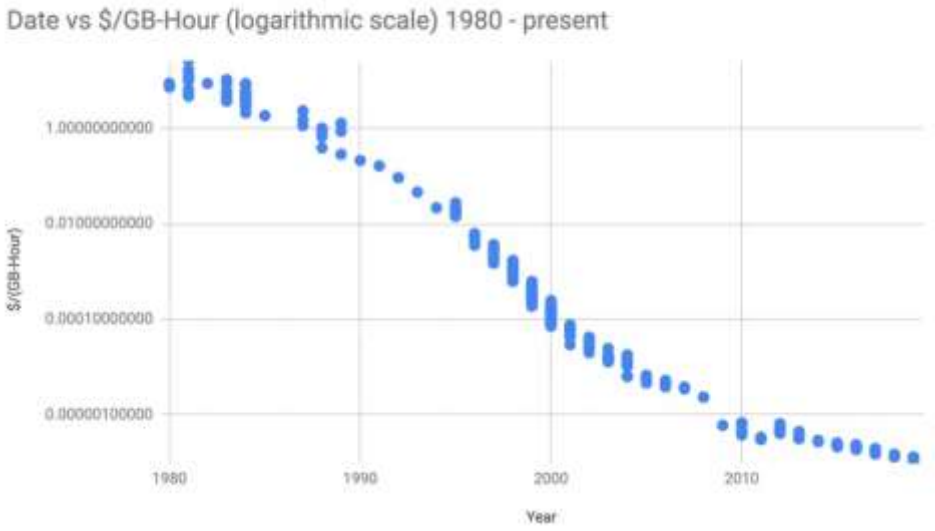
Due to problems in data retrieval speed, Arweave has a SPoRA mechanism solution to force miners to store data for fast retrieval data. Arweave switches its consensus mechanism to Succinct Proof of Random Access (SPoRA). SPoRA is just PoW + PoA with consideration for retrieval speed, so the mechanism triggers the miners to store data locally and in different locations so that they are closer to the user, thereby increasing decentralization.



Predictive Pricing Model

Arweave users pay a one-time, up-front fee to store their data permanently. Over the past 50 years, the average annual rate of decline in storage costs has been 30.6%. Arweave’s transaction pricing model conservatively estimates that these costs will decline by just 0.5% per annum. If the actual cost reduction turns out to be higher than 0.5%, the user’s storage duration gets extended proportionally.

Figure 1:



Storage Endowment

The Arweave's economic mechanism is that when a piece of data is added to the Arweave network, the user pays a "principle" upfront, on which "interest" in the form of storage purchasing power is accrued. It is intended that only about 14% of fees paid go to miners, with the remainder going to the endowment pool. The mechanism is miners in profit when network fee revenue exceeds their cost of maintaining the blockweave data network. However, when nodes become unprofitable, the endowment is used to subsidize the miners' operations.

Cost

We can use the Arweave fee calculator at http://arweavefees.com/ to get a price estimate for Arweave. To store 1GB for 200 years, it costs 0.15204 AR which is USD 1.43 at time of writing. For context, for the same cost, you can store the same amount of data for only about 5 years on AWS S3 even assuming that they will reduce their cost by 0.5% every year.

Retrieval Speed

Arweave created scalable infrastructure to be the backbone for the data economy because it can reach up to 5000 transactions per second (TPS) and retrieve data in roughly 400 ms.

Using Arweave

To use Arweave as a Dropbox alternative, you can use ArDrive, which is a file storage tool built on top of Arweave. You will need the Arweave wallet with some AR tokens to pay for the service.
To use ArDrive, you will need to connect your wallet, create an account and a new drive as follows.
After the initial setup above, you can upload and manage your files as you would on Dropbox.
For developers, they can upload data to Arweave with their Javascript SDK. In order to do so, the developers will need the Arweave wallet with some AR tokens. The process is straightforward in contrast to Filecoin because Arweave does not require the user to cut a deal with any particular miner and the term of storage is fixed at 200 years.

Token

Arweave has a native Arweave token ($AR), Users of the network can pay a one-off endowment to use and provide permanent and immutable data storage. The AR token rewards network miners and pays for data upload fees within the network. Users that wish to include a transaction in a block have to pay transaction fees denominated in AR tokens.

Token Supply

Arweave Supply

Circulating Supply	50,108,502 AR
Max Supply	66,000,000 AR
Total Supply	64,598,643 AR
% Supply Change From 1 Year Ago	-
Emission Type	Programmatic Burn: The developer can use this smart contract to burn $AR in order to add $bAR minting while the users upload data to the permaweb e.g

Token Holder (Rich List)

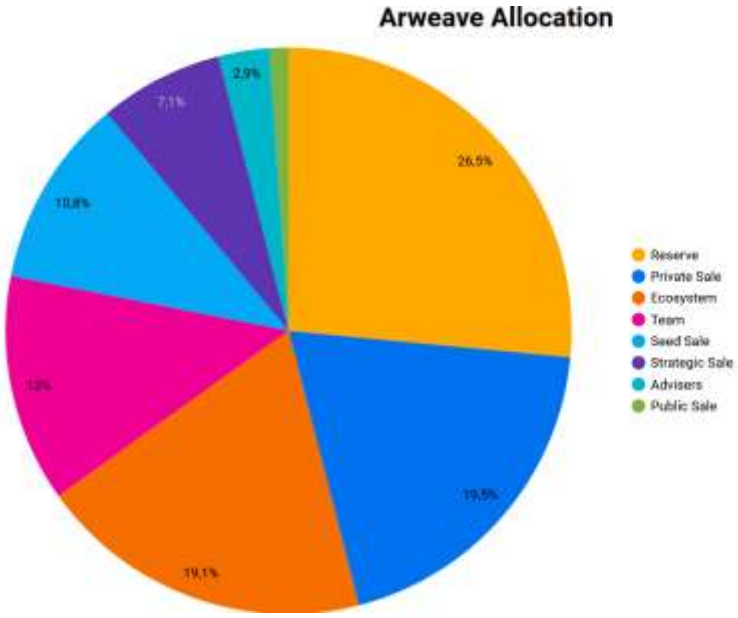
The addresses that hold a large number of tokens or Whales of $AR.

Address	**Balance**	**Balance Ratio** $(\frac{Balance}{Current\ total\ Supply})$
dRFuVE-s6-TgmykU4Zqn246AR2PIsf3HhBhZ0t5-WXE (Reserves)	10,200,042.505032 AR	15.8%
4JOmaT9fFe2ojFJEls3Zow5UKO2CBOk7lOirbPTtX1o	9,459,190.333097 AR	14.6%
nQKiFZE11MiXjY18qib_M4vz_AHyO3cf6gxxbemtIJY	7,550,181.488874 AR	11.6%
4u5gMvlfVhkn_atzuagjO92H_xJLtVNjucSfEYBrL0E	2,542,589.003794 AR	4%
MBB9dcPWUG_t75ezcBwt7u3C0vCyu4tuwxjstlCpvIE	2,506,482.477938 AR	3.9%
Endowment Pool	44,707.64 AR AR	0.07

All of the investor lockup periods have expired

The 14M AR token reserved for future development is funds reserved for the exchange development to help to run the project for a long period of time

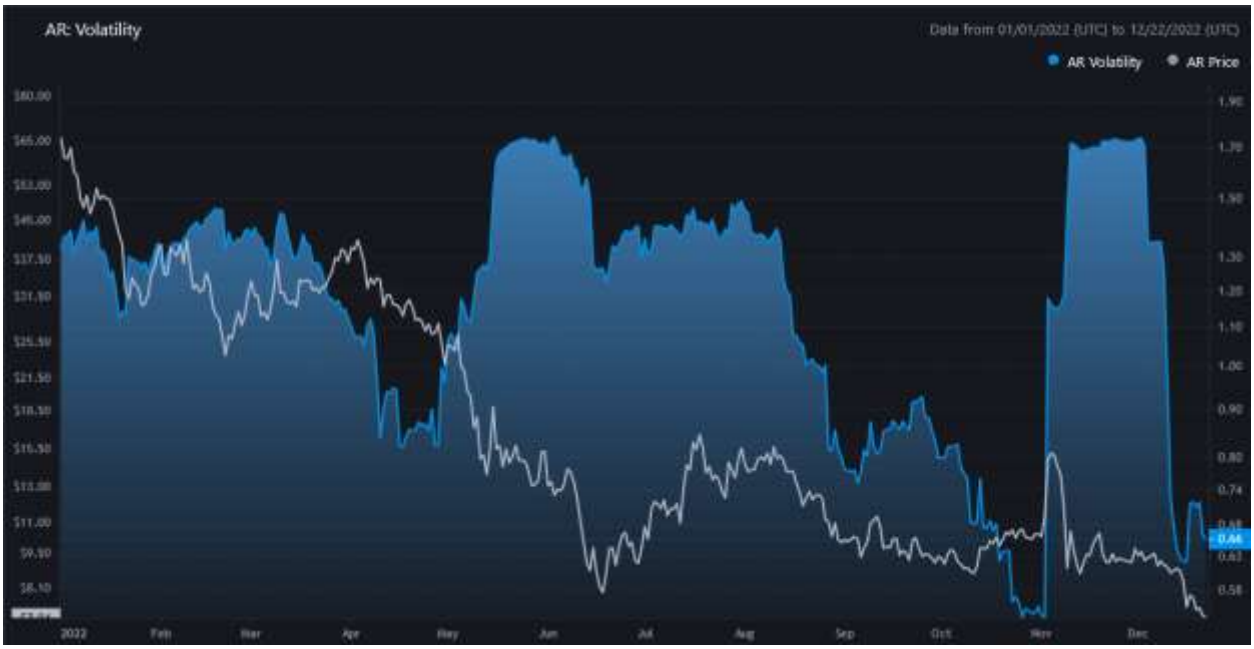
No fund has been paid to the miners from the endowment fund.



Arweave Token Allocation Table		
Allocation	**Quantity**	**Genesis block %**
Team	7,150,000	13.0
Techstars	330,000	0.6
Advisers	1,595,000	2.9
Adoption incentives	10,505,000	19.1
Project	14,575,000	26.5
Seed	5,940,000	10.8
Strategic	3,575,000	6.5
Private Sale	8,855,000	16.1
Community Sale Part 1	1,540,000	2.8
Community Sale Part 2	935,000	1.7
Genesis block total	**55,000,000**	**100.0**
Mining incentives	11,000,000	
Total supply	**66,000,000**	

Token Price Action

$AR Volatility (YTD)



Arweave Market

Market

Pair	Volume	Exchange	Type
AR/USDT	**$3.49M**	**Binance**	**CEX**
AR/USDT	**$0.68M**	**OKEx**	**CEX**
AR/USDT	**$53.67K**	**Huobi**	**CEX**
AR/USD	**$2.4K**	**Bitrex**	**CEX**

Collabs

Arweave has been integrated into Meta for Instagram collectibles storage. This integration is part of the latest blockchain-related developments on Meta. Meta uses Arweave as a platform for delivering a permanent data storage mechanism. The integration means Instagram (Meta) users can now issue digital collectibles for their posts, stored on Arweave

VI. Project Comparison

The following table summarizes the differences between the solutions we discussed in the report and provides metrics for further comparison:

	IPFS	Pinning Service	Filecoin	Arweave
Economic Model	Free	Monthly subscription	The user and storage providers strike a custom storage deal	One-time fee to store for 200 years

Data Availability	Data is available as long as the nodes are kept online	Data availability is guaranteed by the terms of service from the provider	Data availability is ensured by the following mechanisms: **Proof of Replication** Nodes prove that data has been replicated on dedicated storage space **Proof of Space and Time** Nodes continuously prove that they are storing data for the duration of the contract	Data availability is ensured by **SPoRA** Miners need to prove access to prior blocks to mine new ones
Contract Method	N/A	Various monthly plans depending on number of files, storage size and bandwidth required, among others	Many different storage contracts with varying terms on price, duration, replication factor, among others	One contract between user and protocol to stored permanently
Number of Nodes	200K nodes	200 IPFS nodes (Pinata)	3909 storage providers	70 online nodes, 321 offline nodes
Decentrality	Decentralized	Centralized	Decentralized if storing on multiple miners	Decentralized
Retrieval Speed	~44s on average to retrieve a newly created content, but only ~12ms to retrieve cached contents.	Nearly instant	1-5 hours to retrieve the data on Filecoin	Reach up to 5000 transactions per second (TPS), and can retrieve data in roughly 400ms
Cost	Free	$0-$1000 depends on the features	$0.00000012/GiB as of 05 December	$1.42/GB as of 05 December
Last Downtime	10 Nov 2022 on web app	23 Nov 2022, due to API and Web App	30 Nov 2022, due to Filecoin network upgrade	January 2022, on Arweave gateway

VII. Conclusion

dStorage offers a number of advantages over traditional centralized storage, such as increased resiliency and anti-censorship. While decentralized storage networks are still in their early stages, they have the potential to disrupt the current landscape of data storage and distribution. dStorage is not without its own challenges; it needs to solve for scalability, user adoption, and the need for efficient incentivization mechanisms to ensure long-term sustainability.

Based on our discussions, Arweave is the more promising and comprehensive decentralized technology than alternatives, especially considering that Filecoin cannot be used without IPFS, and yet IPFS requires a centralized pinning service to keep a file accessible. That said, Arweave is still nascent with only 70 online nodes at the moment, casting shadow to its security and continued feasibility. An argument can be made that, especially for app builders, using multiple pinning services to avoid a single point of failure and Filecoin for a backup is a less riskier approach.

Overall, there is no silver bullet for dStorage yet; the choice of technology depends on the problem we are solving for. One thing for sure though is that dStorage is set to play a significant role in shaping the future of data storage, and as such, should be considered by individuals and organizations alike as more robust alternatives to their centralized counterparts.