Permachive

Filecoin Discovery

Filecoin's Ecosystem & Arweave Comparisons

11.20.2022

GG Capital

A Digital Asset Research Firm

Combining ARchivers with Filecoin

When considering options for uploading data storage to a decentralized and permanent network, blockchains provide multiple options for companies & individuals to securely upload their content on a censorship-resistant network. Rahwa initially approached us for creating a set of archivers for her company Permachive, a decentralized library that archives immutable material around the world. We began this discovery by searching Arweave's ecosystem and decided to develop the Permachive repository based on separate programs from Bundlr's ARchivers. This set of resources allows for searching for keywords across Twitter, News Articles, and Youtube to then archive each dataset into Arweave's PermaWeb.

After a little more than a week of testing, we found that these ARchivers achieved our goal of creating a filtered stream of information, screenshotting discovered content using Puppeteer, a JavaScript library, and then uploading the data and attributes into the tags of an Arweave upload. Previously established programs, such as TwittAR, created a simplified method for uploading content to Arweave through Bundlr with minimal data usage and content sizes of <100KB per upload. However, the main drawback we found during testing was the increase in costs as the ARchivers parsed content and uploaded approximately 4.3 GB of data per day. With prices around $9.53, this averaged out to 0.8 Arweave (AR) tokens per day, leading to roughly 268.8 AR per year, or $2,561.66 in total for the year.

When asked to explore the Filecoin ecosystem following this initial discovery, what came to mind was finding the best solutions for decreasing costs. We concluded that Arweave's permanent storage would increase with costs over time, whereas Filecoin provides an open marketplace for using cheaper alternatives through chosen Storage Providers (SPs). The ideal option would be for Permachive to adapt the TwittAR Archivers for parsing data into a JSON file, then uploading the acquired data using third-party platforms that abstract away costs and infrastructure to provide easier access to users. We would suggest viewing simplified data storage options like Small Data Industries' Starling program, Textile and their "Buckets" for pinning InterPlanetary File System (IPFS) data, and Estuary for their two-fold storage system.

We envision that combining Arweave's tooling with Filecoin's costs allows for efficient data storage for public datasets. Updated findings can be viewed on our company's Notion page.

This report will cover the following sections:

Filecoin's Background & Resources

Official Links

Official website: https://filecoin.io/

Storage site: https://filecoin.io/store/#intro

Documentation: https://docs.filecoin.io/

Filecoin Block Explorer: https://filfox.info/en

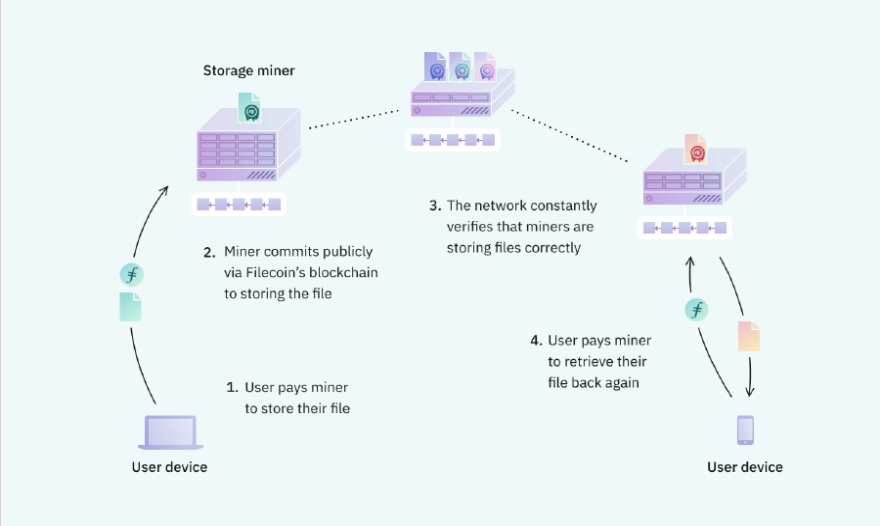
Filecoin Data Dashboard: https://www.filscout.com/en/

Filecoin Storage Provider Analytics: https://file.app/

Documentation

Filecoin aims to provide economic incentives through the $FIL token for users to store files over a set period. By providing an open marketplace for users to create deals with Storage miners or SPs, the network incentivizes SPs to create favorable offers within their Storage Deals, allowing for healthy competition between SPs across the world. The following quote simplifies this further:

> "Filecoin's blockchain records transactions to send and receive FIL, along with proofs from storage providers that they are storing their files correctly."



This diagram relays the underlying process for uploading and storing data through Filecoin's miners. As users pay a fee to upload data, this incentivizes miners to keep data protected and stored, while the network verifies this data is stored correctly.

The main benefit for users here would be the option to choose between various storage providers (listed at file.app) where costs/speed of access are negotiated between storage deals.

Comparing this to the miners, SPs can earn $FIL for providing empty storage to users as they pay for uploaded content. The infrastructure for this is provided through Filecoin and allows for a simplified method for selling data storage.

Filecoin Nodes & Providers

We can think of Filecoin Nodes/Clients as the validators/miners on this blockchain. They are also in charge of publishing various messages that can be used by the network to broadcast certain requirements, like proposing storage/retrieval deals to storage providers.

The Storage providers (SPs) execute the different storage/retrieval deals → collect FIL rewards once adding new blocks to the chain.

Storage Deals

Storage deals are the main component that Permachive would be using as they are agreements between the user and SPs for storing data on the network. The SP is in charge of proving that they are still storing data, without this proof, the SP will be slashed (similar to ETH miners being slashed for failing to maintain consensus).

Each SP<>client relationship depends on each storage deal, where the data must be stored completely during the lifetime of each deal.

Cryptographic Proofs

> "*Proof Of Replication (PoRep)*, storage providers demonstrate that they have received all the data and that they have encoded it in a way unique to that storage provider using their physical storage in a way that no other storage provider can replicate (so two deals for the same data cannot end up re-using the same disk). This proof is provided when the deal starts, and the *sealing* operation completes.
>
> Once a deal is active and during its full lifetime, the storage provider will use *Proof of Spacetime (PoSt)* to prove that it is *still* storing the data associated with a deal. For PoSt, random storage providers need to prove that random parts of the data they store are still there."

PoRep allows for defending publicly verifiable claims of uploaded data. The research paper from Ben Fisch illustrates its purpose and importance to Filecoin, while PoSt defines

how verifiers benefit from a more efficient confirmation system compared to Proof-of-Work (PoW) here.

Gas Fees

This is a key section to understand for the underlying blockchain as a number of fees for each block are burned which is considered the Base Fee (defined as $FIL burned/unit of gas, measured in attoFIL/Gas).

Terms are listed as the following:

- GasUsage: cost for message execution measured in Gas.
- BaseFee: $FIL burned/unit of gas, measured in attoFIL/Gas.
- GasLimit: normal requirement for blockchains, not enough gas? no message will be sent and each transaction (in ETH layman's terms) will be reverted.
- GasFeeCap: important definition here is that the message sent must have a minimum balance of GFC*GasLimit. This places a barrier against a high BaseFee occurring, i.e. more FIL gets burned.

The total cost of a message for a sender will be:

- *GasUsage * BaseFee* FIL (burned) +
- *GasLimit * GasPremium* FIL (storage provider's reward) +
- *OverEstimationBurn * BaseFee* FIL

Storing Data on Filecoin

Using the Filecoin Store Overview, we can assess the defined software solutions for simplifying data storage for users

Their compiled list describes the following platforms:

- ChainSafe Files provides decentralized cloud storage with end-to-end encryption in a Dropbox-like interface. Check out the blog post by ChainSafe for a quick overview of the service.
- Estuary allows uploading and storing content on the Filecoin network directly from your browser, command line, and API.
- Space Storage by Fleek is an open-source, user-controlled, encrypted file storage and sharing platform using IPFS and Filecoin, tied to Ethereum accounts or common web OAuth options.
- Web3.Storage is an IPFS pinning service and Filecoin storage platform built-in-one. Upload files to Web3.Storage and access them instantly, safe in the knowledge that your data is securely backed up using the Filecoin network.

- Slate Slate is a search tool designed to help you remember and keep track of things you care about on the web.
- Lotus imports data and performs deals on the chain using its daemon and CLI. Lotus users get full control of the deals, the chosen providers, and the wallets used to pay. Make sure you are familiar with Lotus and have it installed and running.
- Starling uses Lotus to simplify decentralized storage for digital *preservationists* and archivists.

The following guide from Filecoin displays how to store data through Filecoin Plus. This would work for testing Filecoin's actual facilities and services they natively provide, however, this appears to be a secondary choice for fast uploads of large datasets. We would suggest that a developer understands this process to apply the acquired knowledge to the technical-focused elements of uploading data to Filecoin. From here, they can capitalize on current applications that are excellent examples of building on top of these mechanisms.

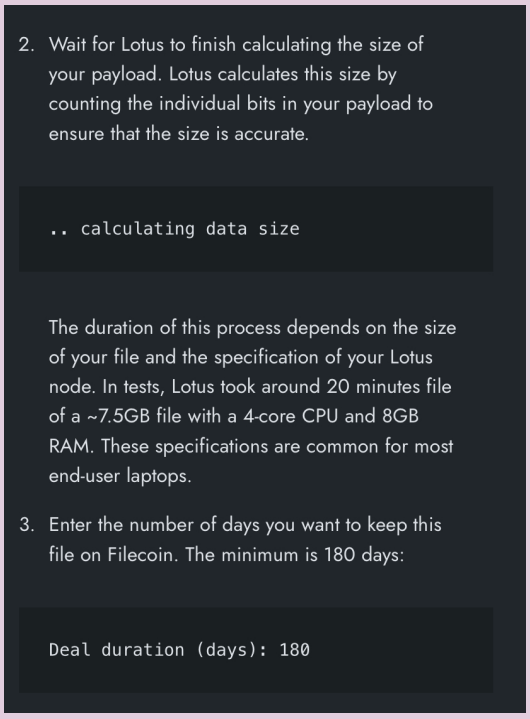
The important terms to consider on Filecoin for uploading data are the content IDs (Data CID), Miner IDs, and the accompanying Deal CID. These are displayed as the following:

Data CID: bafk2bzaceajz56zudni2hli7id6jvvpo5n4wj5eoxm5xwj2ipthwc2pkgowwu

Miner ID: f01000

Deal CID: bafyreict2zhkbwy2arri3jgthk2jyznck47umvpqis3hc5oclvskwpteau

One initial difference to consider for this is using Filecoin services vs other platforms. Take for example the following screenshot from this link:



This explains how a user that uploads a 7.5-8 GB file using the Lotus environment would take ~20 minutes for the upload to occur. The user would be able to specify the duration of the Storage Deal. The main takeaway here is that the upload time could hinder consistent streams of content from Twitter, News articles, or Youtube.

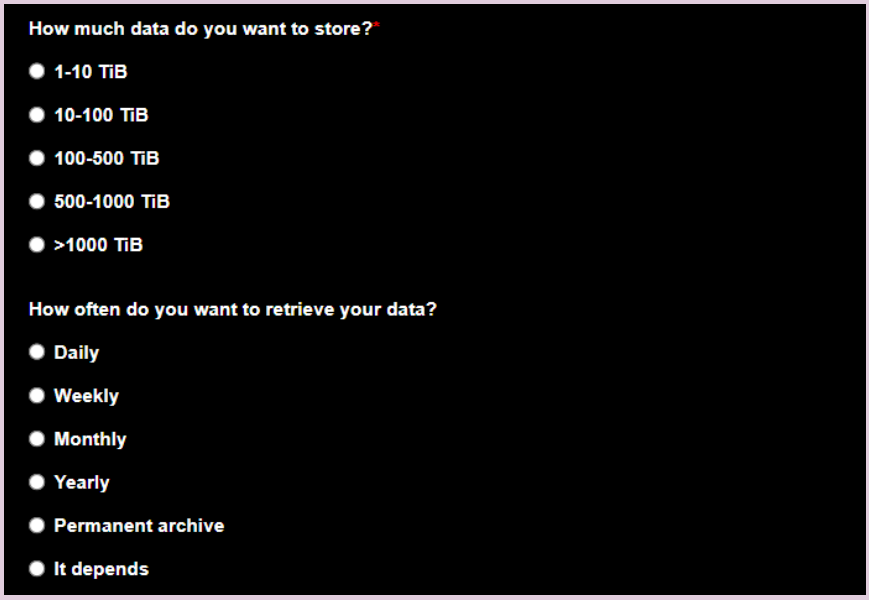
This testing process could become more efficient when using Estuary and Textile Buckets for the pinning of data to IPFS and verifying its status on Filecoin. Permachive would benefit from using the pinned content on IPFS as its access point instead of relying on data retrieval directly from Filecoin for displaying content.

Filecoin's Ecosystem

One of the best factors we found from our Filecoin research was the expansive amount of protocols that have built out helpful tooling for users and entities looking to store publicly accessible data. Another benefit to the ecosystem is the onboarding programs for both users and potential Storage Providers. For SPs, programs like Filecoin Discover create greater opportunities for miners. The majority of their interests are focused on preserving important datasets from current research so that future generations can use them to improve the world around us.

By sponsoring these crucial datasets the Filecoin team wants to decentralize these large amounts of data without risking loss or damage. This is primarily a method for incentivizing all miners to store data and earn additional FIL. A benefit here would be if Permachive was looking to use grants/subsidies to mine FIL, then use that rewarded FIL for storing additional data. The store page here displays current hardware reqs and what they will be shipping out from the program.

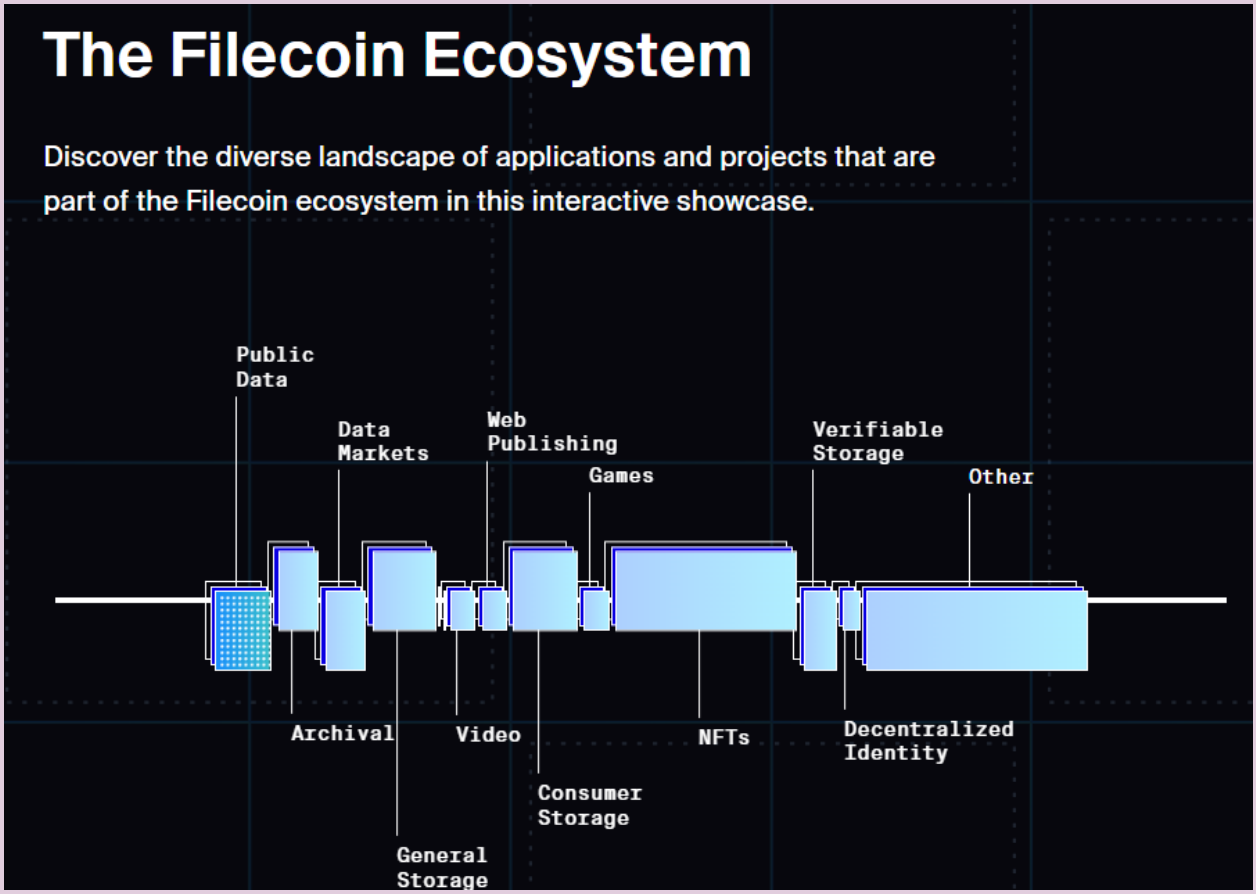
For simplified storing through Filecoin's own services, they offer the Data Onboarding program. This is where users can specify the amount of data being stored and the length of time desired for retrieval.



Thankfully, there are options available for retrieval speeds depending on third-party platforms and Permachive's outlook on pinned IPFS data vs retrieving data directly from Filecoin's SPs.

Collection of Ecosystem Projects

The Ecosystem database has a full list of information for users on everything from Public Data, Archival platforms, Data Markets, General Storage, and Privacy. Although the Filecoin ecosystem boasts over a hundred projects launched using their network, a select few have done an incredible job of assisting both users and miners with accessing the network's services.



Public Data

Filecoin Foundation - acts as the steward for helping users and programs get onboarded to Filecoin and assist with data storage. The ideal outlook here would be pursuing community grants to help sustain continued storage costs through Starling.

Filedrive Labs - provides storage services for larger, public datasets.

Archival

Starling Lab - academic research lab from Stanford & USC that uses open-source tools to store digital content. The primary area of focus is Journalism and discovering new purposes for blockchain networks with an academic background. The initial view of the lab is centered around "empowering journalists" through Fellowship programs and their established Framework.

Small Data Industries - works with art & media institutions for preserving digital content. This platform focuses on consulting and developing various collections. Although there is no direct

mention, they are tagged as creators of Starling Storage and would allow for direct testing with one of their established products.

Web3 Storage - a platform for storing data on Filecoin and making it accessible to users on IPFS. This project lets you sign up using Github and upload data through their platform by incorporating an API Token that allows users to directly input files.

The benefit here is that Permachive could potentially adapt its upload script for adding large numbers of files. One idea here would be to implement the same TwittAR Archivers as previously described and use the data parsed from the JSON file as the uploaded content through web3.storage.

Data Markets

CID Gravity - client management for pricing miners and displaying data points to potential customers. The benefit here is that CID acts as an integration for both Lotus and Filecoin miner environments (Lotus is used with Starling as their environment).

This is for miners that want to deploy business models for uploading data storage. By using their repository & interface, we can assess potential cost reductions and create a framework around miner statistics for SPs.

General Storage

Kotal - a platform for developing open-source tools and frameworks to deploy blockchain infrastructure. This acts as a simplified application for supporting node infrastructure and deploying access to your public networks. Their documentation & repository display a wide range of integrations from Ethereum mainnet to running your public node.

Protocol Labs - open-source R&D lab beyond the development of Filecoin, IPFS, and libp2p. Prominent background and expansion across different platforms, which has assisted in creating a resolute standard for testing P2P systems, designing IPFS protocols, and developing modular networking stacks.

Available Tooling

Using the following examples we can see how a small amount of technical effort would allow a developer (or anyone with a computer and Node.js) to set up a data storage system with low costs and efficient upload speeds. Although the protocols above are not a comprehensive list of accessible platforms built using Filecoin, those combined with the following three should provide a sustainable foundation for understanding how to store data, as well as implementing ARchivers into the Filecoin ecosystem.

Textile

Textile is a multifaceted organization that aims to build decentralized tooling for accessing IPFS endpoints and connecting private databases with public datasets. This program is still under heavy review as of Fall 2021, however, the documentation explains multiple features that are still available today.

One feature that Permachive can test is the Textile Bucket for rendering content on IPFS websites, creating private Buckets for storing data, and eventually storing content onto Filecoin for an extra layer of security. Sections here describe creating separate HTTP domains for the website itself or having the content pinned to a separate InterPlanetary Name System (IPNS) that could link to a Content ID (CID) for representing published content.

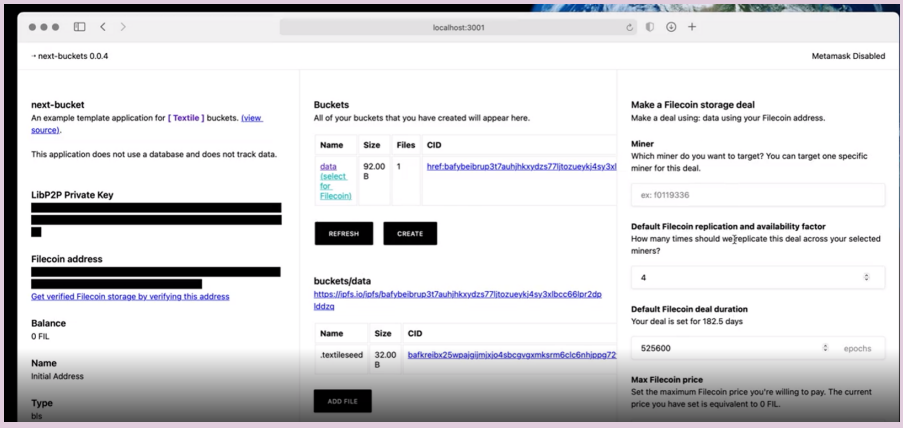
Buckets

The below explains how the Textile platform creates Buckets for uploading data to IPFS:

"Textile offers a simple hosted solution that developers can add to any website or application. Textile accomplishes this by using "Buckets" as a way to organize, index, and pin IPFS data, despite the IPFS protocol not being natively available in every mainstream browser.

Setting up the bucket is a simple matter of creating a local secrets file to store the keys needed to access a Textile Hub account. When you sign up for a Textile Hub account and start developing an app, you'll receive a public and private key. This key needs to be inserted into the text file in the format:

TEXTILE_HUB_KEY = XXXX
TEXTILE_HUB_SECRET = XXXX
IPFS_GATEWAY = https://ipfs.io"



The GitHub repository from Application Research explains how to upload using the Textile Bucket and how to archive this file onto Filecoin.

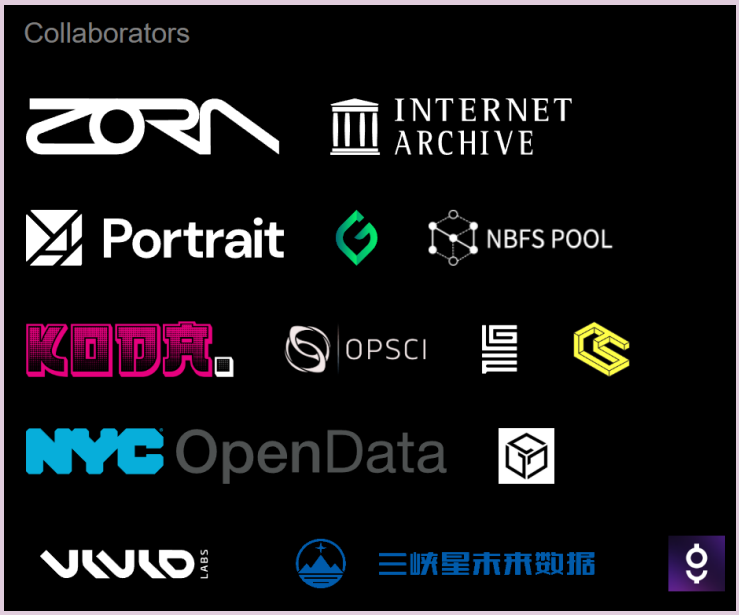
The most appealing feature here is the UI created for pinning each file onto IPFS, which can

be a simplified method for both IPFS uploading and verifying that this file is secured on Filecoin. While encryption is available, Textile is not a one-stop solution for data uploads and would require more time for development before full integration.

Estuary

Estuary tech is a platform designed for uploading public datasets to Filecoin, which are then pinned to IPFS nodes for simple retrieval and display through a user's browser. Ideally, Permachive would use them to receive verified storage deals and upload directly to Filecoin through their command line (CLI) or by connecting the Estuary API. The only current drawback to this platform is that the program itself is still in its alpha testing phase and further testing would require an application. However, signups are currently available and Permachive's goal of uploading meaningful & publicly accessible datasets is a good fit for their mission.

One ideal factor about this group when considering key partners in data storage is their current list of collaborators. Described as the Estuary Ecosystem, the organizations and projects, such as Internet Archive and Mirror, have a well-aligned message and purpose with Permachive's outlook.



According to their documentation, we would have up to 30 GB/upload and would need to have a minimum of 3.57 GB before storing on Filecoin.

One drawback to this method would be the in-between moments where the stored data is kept on IPFS before sending to a "curated list of Filecoin storage providers". Although this would not pose a major issue during a filtered stream of content, there could potentially be a disconnect between uploaded content via IPFS and what is verifiably stored on Filecoin.

This differentiates from Starling which would allow for specific options to be created regarding storage costs and providers and have an upfront method for storing on both IPFS and Filecoin. Choosing between these platforms would depend on testing post-application, however, Starling appears as the easiest setup regarding go-to-market *IF* Permachive already has files ready in advance.

Starling

Starling has the most interesting dynamics for using CLI commands to upload specific folders and files to Filecoin.
- Starling capitalizes on the Lotus environment for data storage.
- Lotus allows users to accept storage deals on Filecoin through a simple CLI.
The benefit of this combination is that users receive full control of deals and chosen SPs while having a clear view of the status of their data.

Installation requirements are the following for easy uploading:

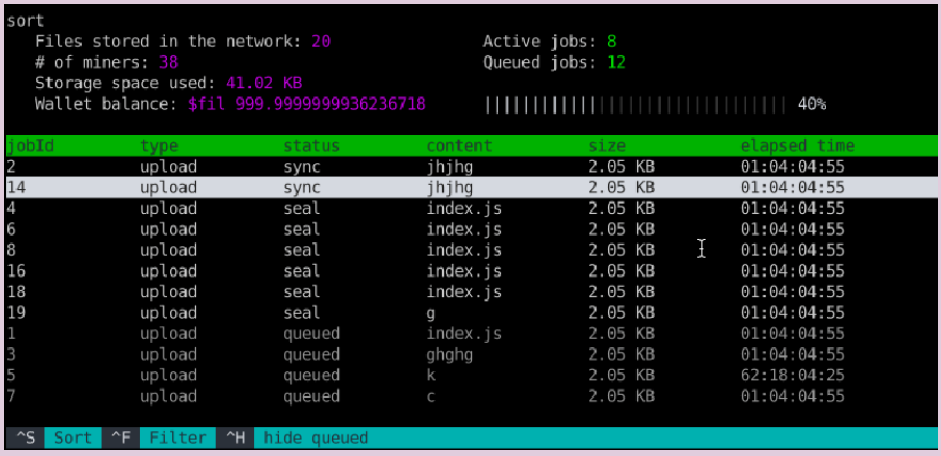
Filecoin download -> Node.js Download -> run *npm install -g starling*

Main component here is the splitting of data as they require files to be in chunks smaller than 256 mebibytes (MiB) which is equal to 268.43 megabytes (MB) or 0.268 Gigabytes (GB). For initial testing, this would comply with the <100 KiB storage plans that we had set up for Bundlr/Arweave.

Using this example, Permachive could set up a Filecoin wallet, connect it to Starling, and continuously queue jobs/storages. The following command for "Store" would define each process for making storage deals to Filecoin miners and represents the CLI display below.

starling store [path] - creates deals for the desired path based on the price/TB that the user specified at installation. This can be updated using the *config* command.

Adapting this with the frontend interface would be a helpful method for keeping track of Permachive's storage as well as overall costs. This is achieved through the *starling monitor* command.



On-chain Costs

Comparing Costs to Arweave

When comparing Arweave to Filecoin, the largest discrepancy is the current pricing for uploading data. While Arweave boasts permanent storage versus Filecoin's "pay-as-you-go" framework, there is a distinct sacrifice that users have to take when choosing permanent storage. We have based these costs on the original testing results in our previous report(s):

"The largest file size on the Twitter archivers for compressed images was ~120 KB, which would only cost a few cents. This was a rare occurrence in our testing with most files remaining under 100 KB. Our current estimate would suggest allocating for at minimum 1 AR token/month, and/or no more than $10/month.

The News Article archiver upload cost varies per upload. If there is a page with few pictures, the cost could be no more than 75 KB. At most, this would be more expensive than Twitter uploads, with an estimate of $30 a month, or around 2-2.5 AR.

The videos found with the Youtube archiver were generally shorter in length, no longer than a minute, ranging from 120 KB to 1.4 MB. Similar to the News Article Archiver, this would require further testing over a longer period of time to receive a more accurate estimate. However, with these file sizes, we can estimate that the user must hold at least 3-4 AR per month in the wallet file."

Archiving on Arweave

Given the current price of AR at $9.53 and a daily rate of 4.3 GB archived per day.

Archiving costs in USD: $7.63/day, $53.37/week, at most $236.35/month (31 days).

The highest daily cost to archive was a little under 1 AR and the least was around .6-.7 AR. Ideal budgeting would allow for at least 1 AR available per day. Recommended 1 AR/day cost in USD: $9.53/day, $66.71/week, at most $295.43/month (31 days).

Average archiving cost: 0.8 AR/day for 4.3 GB/day, 5.6 AR/week, minimum 22.4 AR, and at most 24.8 AR/month at 120.4 GB (31 days).

This 22.4 AR/month, turns into 268.8 AR/year or $2561.66/year on 4.3 GB/day being archived. On a scale of 200 years for permanence (according to Arweave's view on permanent storage), this reaches approximately $512,333.00 in total spending.

Archiving on Filecoin

Using the data from file.app, we can determine that:

Filecoin cost per byte * 1073741824 = 302244.3328391884 attoFIL per GiB

With $FIL at $4.33 this seems much cheaper for a calculation of per GB per month, with an average cost for storing at 0.000000026113910357305878 FIL ⇄ $0.0000001 USD

This would be converted to **26.1B attoFIL** for storage deals **or $0.0012/1 TB per year.**

The below table describes the conversions for FIL<>attoFIL (unit for calculating data uploads).

FIL (Filecoin)	attoFilecoin (aFIL)
1 FIL	1,000,000,000,000,000,000 attoFilecoin
10 FIL	10,000,000,000,000,000,000 attoFilecoin
100 FIL	100,000,000,000,000,000,000 attoFilecoin
1,000 FIL	1,000,000,000,000,000,000,000 attoFilecoin
10,000 FIL	10,000,000,000,000,000,000,000 attoFilecoin
100,000 FIL	99,999,999,999,999,991,611,392 attoFilecoin
1,000,000 FIL	999,999,999,999,999,983,222,784 attoFilecoin
10,000,000 FIL	10,000,000,000,000,000,905,969,664 attoFilecoin
100,000,000 FIL	100,000,000,000,000,004,764,729,344 attoFilecoin

Although this result is shocking, this is an ideal display of the benefit to Filecoin's open marketplace. One measure to keep in mind is that current prices are still completely speculative and should not represent accurate pricing on a 200 year timeline.

Arweave could reach $40/AR and Filecoin could reach $200/FIL in future cycles, which should be considered in all future models when deciding costs between platforms.

Table of Costs

Filecoin costs according to file.app check out to $0.000014/GB per year. For 1GB of storage over 200 years, this calculates to $0.00028/GB. With $FIL at $4.33, this puts *potential* archiving costs at $0.0012/1 TB per year.

Data used in this table is applying the price of Arweave at $9.53, with archiving costs at 22.4 AR/month. The previous archiving costs for the ARchivers were 0.8 AR/day or $7.624/day and there was an average of 4.3GB/day being archived across three programs. This is 268.8 AR/year or $2,561.66, and $512,332.00 after 200 years.

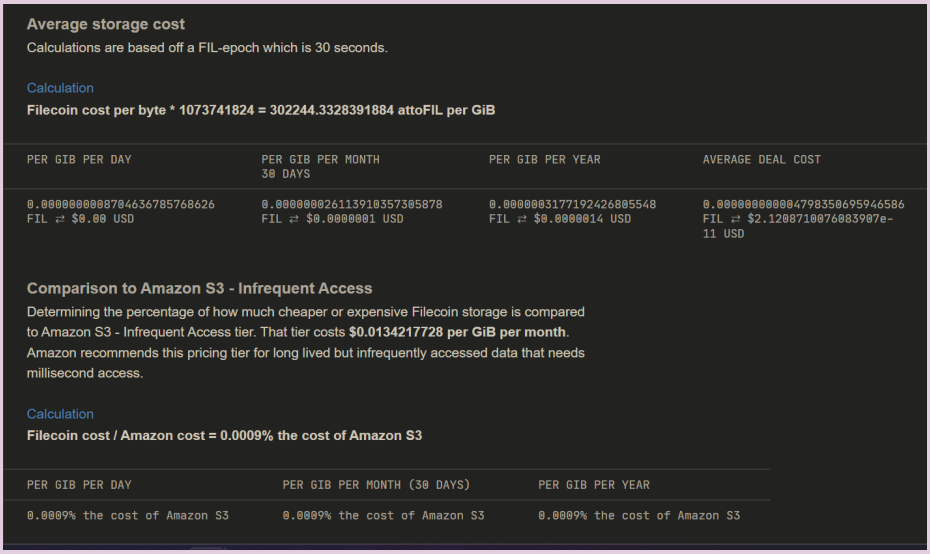
AWS costs are split between the different standard vs intelligent tiers and whether it's frequent or infrequent access. For example, S3 Standard is $0.023/GB per month and $55.20/GB after 200 years.

The following results applied various amounts of data storage across a 200-year timeline and clearly displayed the benefit of storing data via Filecoin compared to both traditional platforms like Amazon's AWS S3 Tiers and crypto-native platforms like Arweave.

Data Amount	Timeline	Arweave Cost	Filecoin Cost	AWS S3 Standard (Frequent)	AWS S3 Standard (Infrequent)	AWS S3 Intelligent (Frequent)	AWS S3 Intelligent (Infrequent)	AWS S3 Glacier (Instant Retrieval)
50KB	200yrs	$5.97	$0.000000014	$0.00276	$0.0015	$0.00276	$0.0015	$0.00048
100KB	200yrs	$11.94	$0.000000028	$0.00552	$0.003	$0,00552	$0.003	$0.00096
1MB	200yrs	$119.14	$0.00000028	$0.0552	$0.03	$0.0552	$0.03	$0.0096
10MB	200yrs	$1,191.47	$0.0000028	$0.552	$0.30	$0.552	$0.30	$0.096
100MB	200yrs	$11,914.70	$0.000028	$5.52	$3.00	5.52	$3.00	$0.96
1GB	200yrs	$119,147.16	$0.00028	$55.20	$30.00	$55.20	$30.00	$9.60
4.3GB (Daily Data)	200yrs	$512,332.80	$0.0012	$237.36	$129.00	$237.36	$129.00	$41.28
100GB	200yrs	$11,914,716	$0.028	$5,520.00	$3,000.00	$5,520.00	$3,000.00	$960.00
1TB	200yrs	$119,147,000	$0.28	$55,200.00	$30,000.00	$55,200.00	$30,000.00	$9,600.00
1.44TB (Yearly Data)	200yrs	$171,571,680	$0.40	$79,488	$43,200.00	$79,488.00	$43,200.00	$13,824.00

Filecoin Data

This app hosts a real-time updated display of current costs and comparisons between Filecoin storing and AWS. All current calculations are based on the FIL-epochs, which provides a better incentive to store large datasets prior to any bull runs occurring.



The main takeaway here is that Filecoin offers nearly 0.0006% of the current costs for Amazon S3. An initial assumption is that this is much cheaper month to month than Arweave but lacks permanent storage.

Miner Free Space Example

The app does a great job of displaying analytics of Filecoin SPs including, Address Locations and the amount of Free Space each miner has available. For example, this miner in GB/Europe has 562.25 TiB available and average costs would be based on deals.

f01606849	Europe [GB]	Power	1.71 PiB
		Free space	562.25 TiB
		Used space	230.81 TiB
		Cost	0 FIL ⇄ $0.00 USD per GiB per day
		Verified cost	0 FIL ⇄ $0.00 USD per GiB per day
		Total deal count	7448
		Average deal cost	0.000000000000002685 FIL ⇄ $0.00 USD per deal
		Total storage cost	0.000000000001999788 FIL ⇄ $0.00 USD

There are a few discrepancies here on whether the above data is accurate but by confirming sources through previous discussions, we can specify how each of these results were reached.

Sources

https://filrep.io/api - provides us ~2k miners and one of the sources of data via API.

https://bafzbeidtcvgkyhof2jqy35xyevtxefqkfubkrh7lvrenn53uz3srea3rcu.textile.space/ - provides us ~2k miners and is another source of data via API.

https://storage.filecoin.io - provides us with how much real data is being used by apps on the Filecoin network.

Fee & Revenue Comparisons

Beyond decentralization and censorship-resistant networks, blockchains allow for value to be derived from current services and platforms and directed to users and validators on each of these networks. Even with Amazon's cheap costs, there is no revenue generation for users which provides zero incentivization to users.

Judging from current trading fees, you could argue that there's a benefit to holding FIL vs AR just by the large amount of daily fee intake they possess. Over time, that could become a net-negative effect for Permachive as Filecoin implements the pay-as-you-go business model vs Arweave's permanent storage.

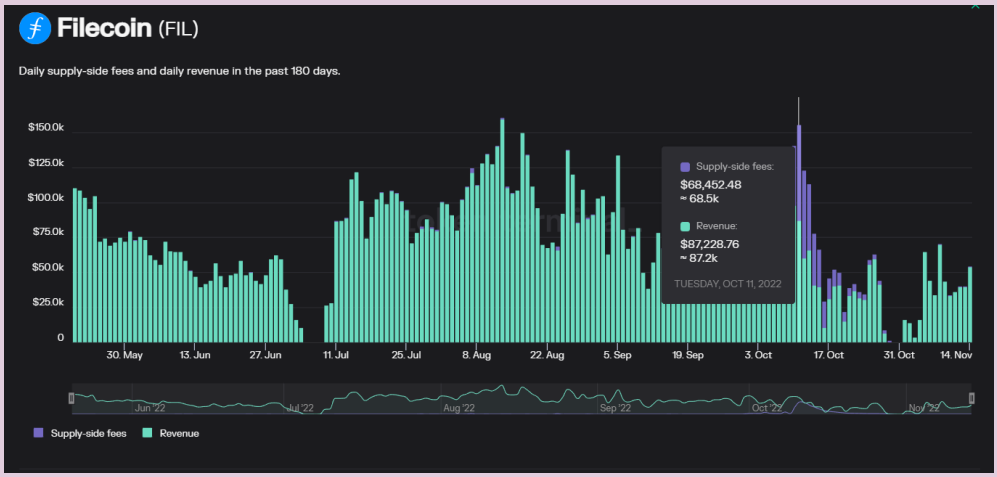
To further demonstrate this relationship, we have used Token Terminal for its real-time comparisons of supply-side fees and revenue across the most widely-used networks.

Competitive Landscape for Filecoin vs Arweave:

Current fees and revenue display that both validators and FIL holders are earning relatively large amounts from the overall network.

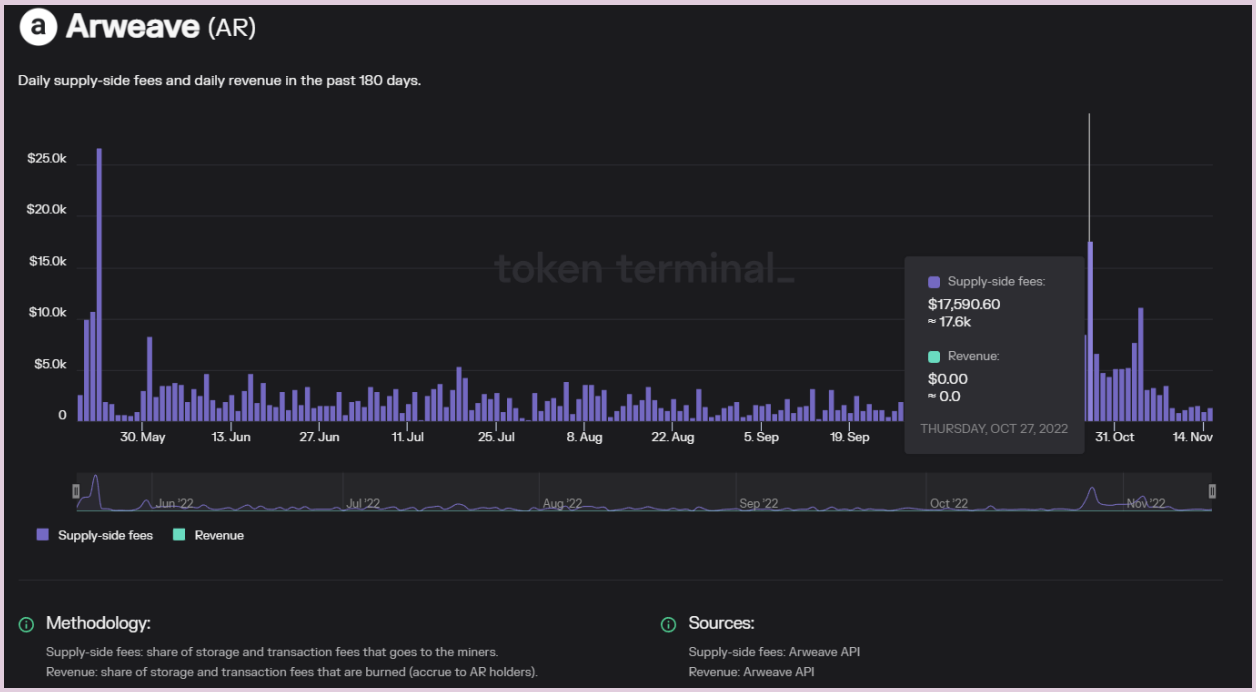
> "Supply-side fees: share of transaction fees that goes to the validators.
>
> Revenue: share of transaction fees that are burned (accrue to FIL holders)."

Filecoin Revenue Share compared to Arweave:



Compare this to Arweave that's earning less than half of the amount of fees to AR miners and zero revenue accruing back to AR holders.

Arweave Fees & Revenue Share over past 180 days:



Filecoin x IPFS x Privacy

IPFS Collaboration

> "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" - Filecoin Store

IPFS combined with Filecoin allows for people already accustomed to IPFS to benefit from Filecoin's ecosystem. There are various methods for uploading data to Filecoin and these can range from deciding on efficient costs to complex processes. However, the main focus for Permachive should be the actual data retrieval to an IPFS site. There are a few platforms already created that provide this service, including Textile for serving content to IPFS websites and Estuary for uploading large public datasets as described above. For pinning data to IPFS files, this can occur in an extended step-by-step process through Filecoin Plus but we have viewed other options to find current advantages through available tooling that has been adapted to fit our current uploading requirements.

This blog post provided an effective summary to explain this connection. One method for relating the two programs is that IPFS stores files on a content ID (commonly known as a CID) and that is what becomes stored on IPFS. To replicate the regular content we see day-to-day, tools like Textile have to be implemented for organizing and pinning the data to IPFS. Textile has a method for organizing data into "Buckets", which are then indexed onto IPFS. This integration would merge together with other frontend languages like React or Javascript to actually view such content via internet browsers.

The next step after pinning data to IPFS would be to confirm that the content is stored on Filecoin. What Filecoin explains as a simple method for uploading this data is:
- filing data into Textile with their Buckets mechanism,
- using Estuary for pinning the Buckets from Textile,
- completing a storage deal with any Storage Provider that the user decides on.

The ideal note about Estuary is that this uploading method could be integrated with current applications using the Estuary API.

Although IPFS technology allows for simplified data uploading across a decentralized network, it does not confirm privacy for both users uploading and SPs, which could be a concern if uploading private data. This next section of the documentation explains the levels of privacy and encryption that IPFS includes and excludes from its program.

Privacy Concerns

The following is a conversation between me and Rahwa on potential identifiers when uploading to both IPFS and Filecoin:

Rahwa: "Another thing to add that's important is discoverability/anonymity and accessibility (eg IPFS can make the IP address of the original uploader visible, Filecoin) (Filecoin is primarily built for public access but you can use a mixer/encryption key to seal the data but in the event that encryption key has a vulnerability it may later be discoverable)."

Me: "IIRC the storage provider used would be what an IP address would be affected by but a connection would be made from the light nodes like Estuary and Lotus for the IPFS access. I think SPs would be affected so most likely best to use a VPN but I'll need to see where points of failure would be beyond IP address tracking.

A leading example for encrypting data prior to storage would be Chainsafe, which would be fully up to the user at the end of the day.

They would allow for a greater level of anonymity based on using an ETH wallet or GitHub for the initial signup, additional measures would depend on the uploader so would say using a Vpn is a requirement prior to using any of the platforms fully."

After this initial prompt, we found that Filecoin should not be viewed as a completely private storage option unless the user decides to encrypt all data prior to uploading. With the possibility of IP addresses being tracked through an IPFS node itself, it would be advised that uploaders take full precautions to maintain their anonymity if looking to upload private data. Our initial assumption is that IP address tracking would be abstracted away if using Estuary nodes or that deciphering would only concern SPs, however, we have not confirmed this and would need to actively test out separate nodes to determine any points of failure.

The most common example provided by both IPFS and Filecoin is by using a connected platform that hosts encryption like Textile. If we view IPFS as the main risk factor, encrypting data prior to upload, accessing a public IPFS gateway, or running our own private IPFS network could be a different option. Then, storing data could be pinned to an IPFS address on this network prior to being sealed by SPs on Filecoin.

Conclusion

After reviewing the extensive ecosystem and background of Filecoin, we view this blockchain as an ideal data storage solution for both personal users and entities that need to upload secured content at affordable costs. Even though our initial research into Arweave resulted in a Proof of Concept for filtered streams of data directly uploaded to the PermaWeb, we concluded that Arweave's permanent storage would have potentially negative effects on company growth as costs increase over time. In comparison to Filecoin, we found that an open marketplace provides both healthy competition to occur between Storage Providers and cheap alternatives for users depending on their location and Storage Deals.

Our initial summary is that the ideal option for Permachive is to adapt the TwittAR Archivers for parsing filtered streams of data into a JSON file, then uploading the acquired data using third-party platforms that abstract away costs and infrastructure to provide easier access to users. With multiple programs available for grants and onboarding, Filecoin's ecosystem can be an incentive for capitalizing on cheap costs and efficient storage solutions. The tooling provided to first-time users is accessible and allows for proper integration with current applications.

This is nowhere near a comprehensive report on the ecosystem nor does this reflect the complete work in development by both the Filecoin team and its collaborative protocols. However, we do believe that any user or company would find value in beginning with the compiled resources within this report and deciding whether Filecoin is the right fit for their data storage needs. Any further updates we provide can be found on the Notion page for this report.

For any questions, further discussions, or contacting us for research, feel free to reach out to Yohan GG on Twitter or email us at yohangg@ggcapital.io and kidp@protonmail.com. Our company's information can be found at GG Capital and on our Twitter.