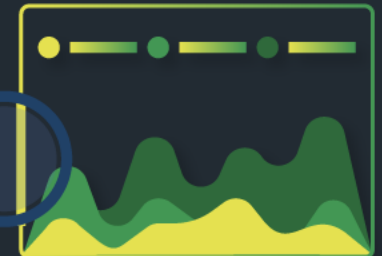
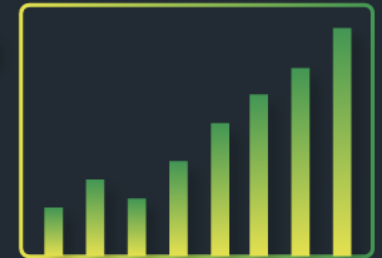
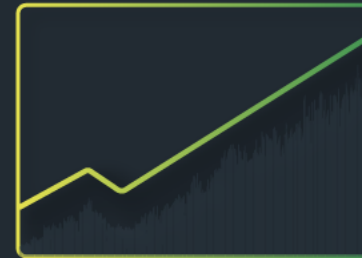
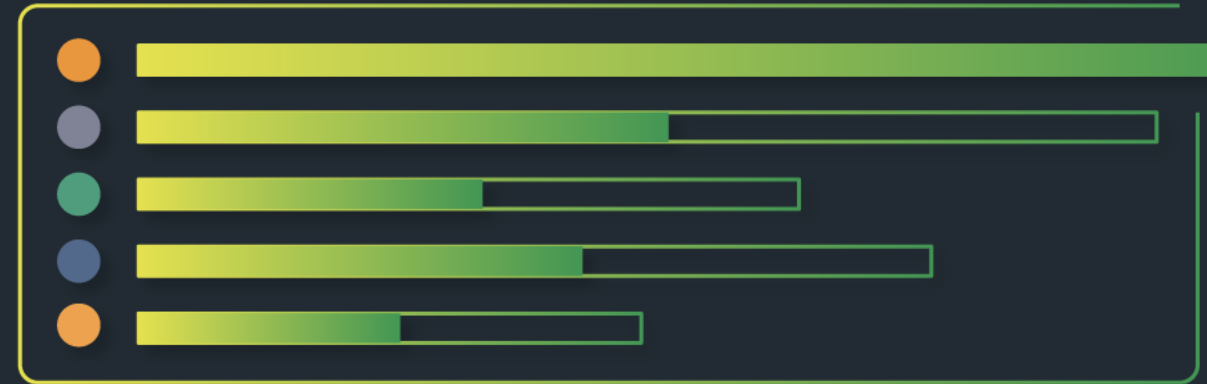




CoinGecko

The State of Decentralized Storage

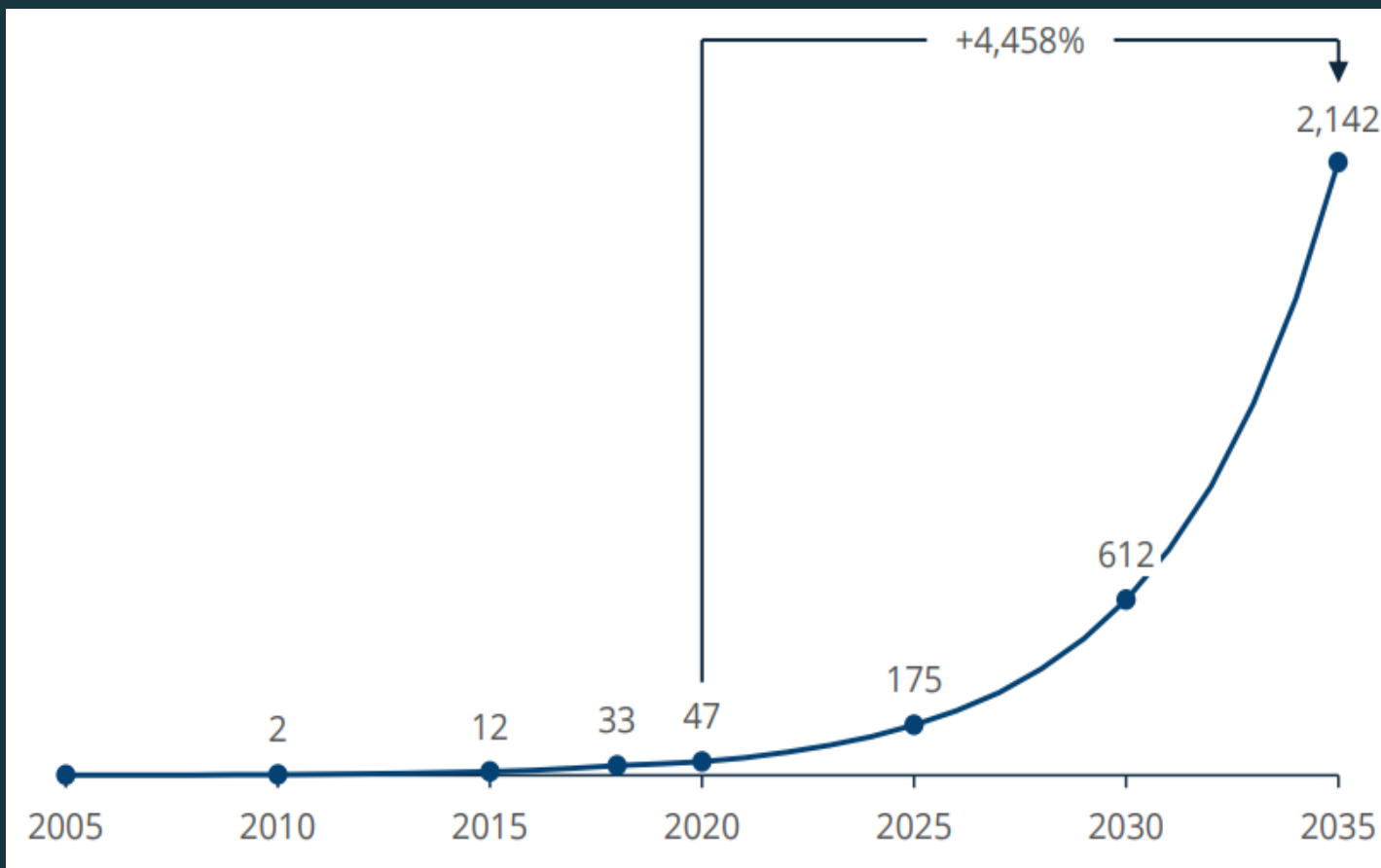
26 September 2022



The Global Data Landscape

Decentralized storage is still a small fraction of the pie

Amount of data generated globally (zettabytes)



In 2021, the world generated approximately 79 zettabytes of data¹

79B terabytes of data



BUT...

In 2019, Google, Amazon, Microsoft and Facebook stored at least 1.2M TB²

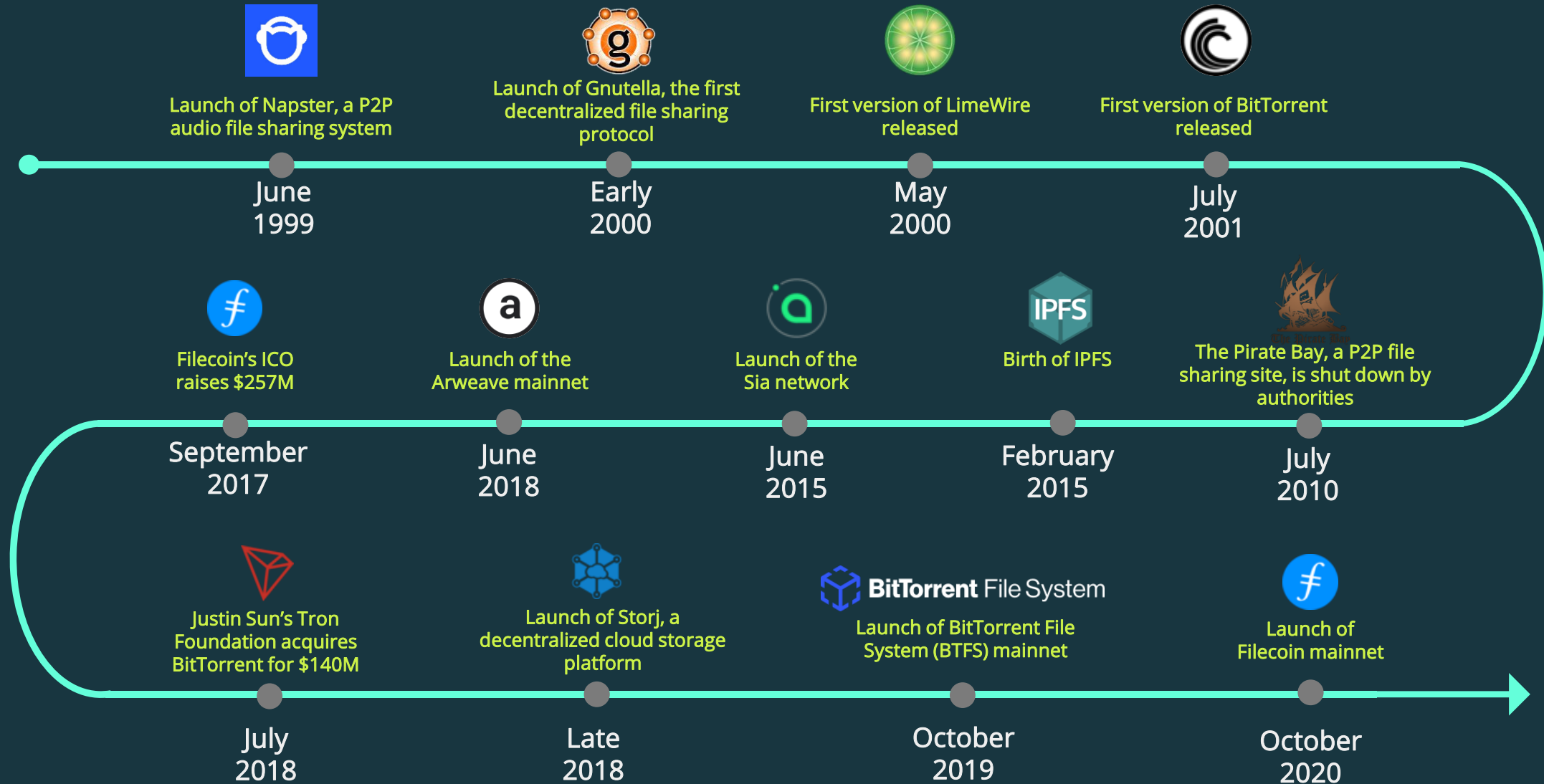


Even with a 5x increase in 2022, only **0.015%** of generated data is stored³









Sources:
¹ Statista
² Science Focus
³ An Honest Report on Web3 Data & Storage

History of Decentralized Storage



Why Do We Need Decentralized Storage?

The benefits of decentralized storage mimic those of blockchain systems

Key Aspects	Current State of Centralized Storage	How does Decentralized Storage Improve on This
 Data Availability & Resiliency	<ul style="list-style-type: none"> Data is typically stored on one main site, putting it at risk of data becoming unavailable if site goes offline Backup storage / site(s) can be provisioned but would usually cost more Typically works on HTTP, which points to a specific path. If data is no longer at the path it becomes a dead link and returns 404 error. 	<ul style="list-style-type: none"> Data is typically broken down into many parts and stored on multiple nodes, creating natural redundancy Uses advance techniques to reassemble data without needing all the parts, further improving availability On-chain storage is "always up" as long as there are miners / validators. In addition to supporting HTTP, mainly uses IPFS, which has a content-based approach – not dependent on path
 Security / Encryption	<ul style="list-style-type: none"> Data stored may (or may not) be encrypted Encryption keys are also stored on centralized databases, making them prime targets for hacks 	<ul style="list-style-type: none"> Most solutions offer auto-encryption and each individual part of the data can be encrypted separately
 Data Integrity	<ul style="list-style-type: none"> Tracing of any unauthorized changes to the data would require prior setup of a logging tool. If no such tool was implemented such changes may go unnoticed 	<ul style="list-style-type: none"> IPFS' approach uses hash to identify data Hashing helps ensure that data have not been corrupted / altered improperly, as doing so will change the hash
 Privacy	<ul style="list-style-type: none"> Certain storage providers would use the content of your data as a form of monetization, e.g. via serving ads 	<ul style="list-style-type: none"> As each node stores encrypted data, there is no way for the node to read the contents of the data
 Open / Censorship Resistance	<ul style="list-style-type: none"> Provider may undertake KYC on prospective clients Data stored in specific sites would be governed by local data laws / regulations 	<ul style="list-style-type: none"> Permissionless solutions where anyone can store data As nodes have no knowledge of actual content of data, there is no censorship possible
 Performance / Scalability	<ul style="list-style-type: none"> Typically would have multi-tier offerings based on performance Actual performance based on proximity to providers' data centers, which are typically built in locations with advance bandwidth infrastructure Sites are typically large datacenters run by providers, which requires significant investment from providers 	<ul style="list-style-type: none"> Beginning to see emergence of multiple tiers (though still not prevalent and still not as varied as centralized providers) Potentially could have greater coverage of nodes spread out across the globe, helping reach more far-flung areas Any miners / validator that meet much lower minimum hardware requirements can join in return for rewards

Who's Currently Using It?

Universally used by everyone, both in and out of the crypto space



NFTs

- Images, audio and video files are usually **too expensive** to store on the blockchain.
- But using decentralized storage, the content of the NFT (audio, images and its metadata) can be **stored off-chain using a unique hash**.
- The hash can then be **hosted on IPFS or other decentralized solutions** where it can be stored and accessed.



Developers

- Allows developers to **create tools and websites in an isolated environment**.
- Developers can create and host important files and **UIs that are censorship-resistant** and pivotal for **decentralized applications (dApps)**
- Networks can **store historic on-chain data** to **reduce the computational load** of validators.



Everyday Users

- Users can **permanently store important documents** and retrieve them even **if centralized services fail or cease to exist**.
- Businesses can **backup their data permanently** for future purposes.
- Ability to access websites or **information restricted by centralized entities**.

Decentralized Storage Projects by Data Persistence Mechanisms

Except Arweave, the decentralized storage projects today are contract-based

Blockchain-based Persistence

- Technically, every blockchain is a distributed database and can function as a decentralized storage network
- However, most blockchains are not built to store large amounts of data. They are designed more to store transactions, and are typically also append-only.
- They are also inefficient in the sense that every node on the network needs to keep a copy of the data.



Contract-based Persistence

- Instead of replicating data across every node on the network, a set of multiple nodes enter into a contract to store a piece data for a specific period of time
- The contract then can be renewed if the time period needs to be extended.
- Instead of the entire data set, the hash of where the data is located gets stored on-chain.



Features and Technical Specifications

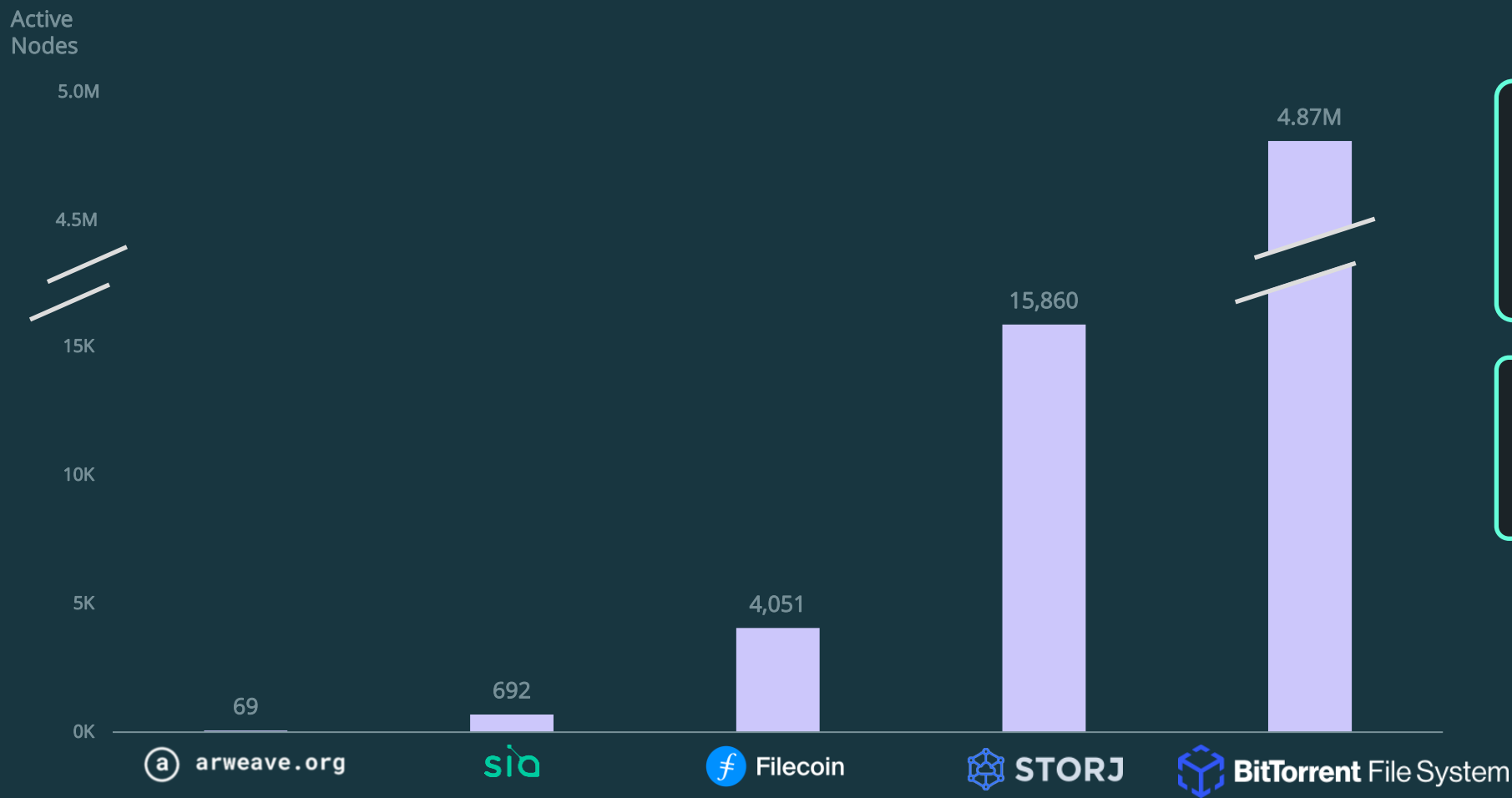
Each decentralized storage solution comes with its own unique features

	Current Market Cap	Consensus Algorithm	Data Replication & Retrieval	Encryption	Smart Contract Execution	Minimum Hosting Requirements
 Filecoin	~\$1,520M	Proof of Spacetime (PoSt) & Proof of Replication (PoR)	Users choose the number of copies to be replicated	Users choose whether to encrypt their stored data or not	Utilizes Filecoin Virtual Machine (FVM)	CPU: 8 cores or more RAM: 137GB or more Hard Drive: 1.1TB or more
 arweave.org	~\$546M	Succinct Proof of Random Access (SPoRA)	Via recall data stored by miners. Data is replicated over 16 times across the blockweave	Users choose whether to encrypt their stored data or not	'Lazy' SmartWeave contracts that are executed and validated by users, not the network	CPU: 6 cores or more RAM: 8.6GB or more Hard Drive: 4TB or more
 STORJ	~\$81M	Proof of Availability (PoA)	Via Reed-Solomon erasure coding. Data is split into 80 pieces and only 29 is needed for retrieval	Automatically encrypted using the AES-256 algorithm by default	Does not have smart contracts	CPU: 1 core or more RAM: 2GB or more Hard Drive: 550GB or more
 sia	~\$214M	Proof of Work (PoW)	Via Reed-Solomon erasure coding. Data is split into 30 pieces and only 10 is needed for retrieval	Automatically encrypted using the Threefish algorithm by default	File contracts between renters and storage providers, automatically enforced by the network	CPU: 4 cores or more RAM: 8GB or more Hard Drive: 64GB or more
 BitTorrent File System	~\$851M	Proof of Stake (PoS)	Via Reed-Solomon erasure coding. Data is split into 30 pieces and only 10 is needed for retrieval	Users choose whether to encrypt their stored data or not	Utilizes BitTorrent-Chain Virtual Machine (BTTCVM)	CPU: 1 core or more RAM: 1GB or more Hard Drive: 32GB or more
 amazon S3	\$1.5T ¹	N/A	Users choose specific files to replicate within or across different regions	Users can enable server-side encryption using the AES-256 algorithm	N/A	N/A

¹ Source: Amazon market cap [Barron's](#)

Active Nodes

Is it really decentralized?



Not all nodes are created equal. **Technical or hardware requirements for some nodes are considerably higher**, such as **Filecoin and Arweave**. There will be different tradeoffs present for different systems.

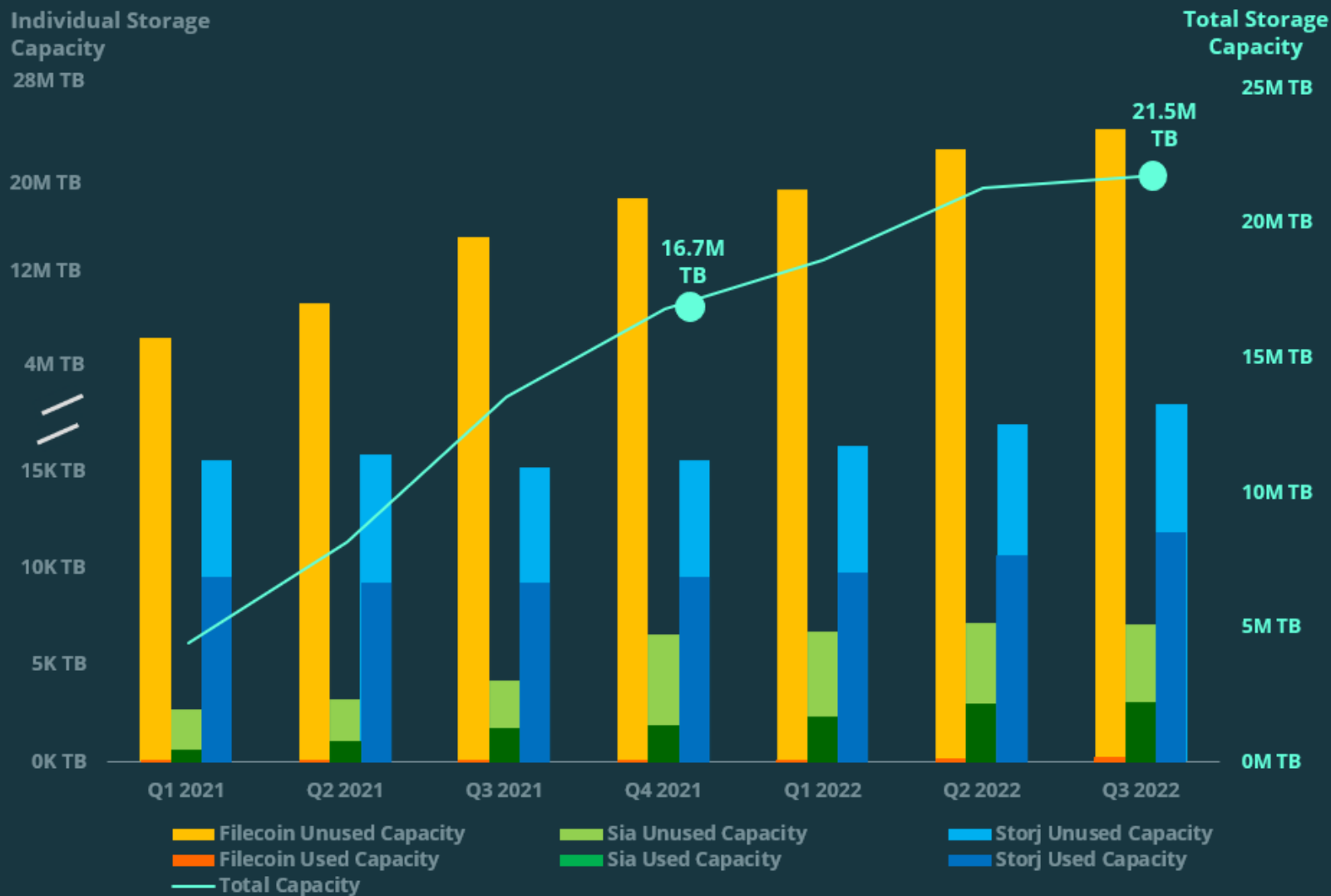
Ultimately, decentralized storage services provide censorship resistance if users' data are split and distributed widely enough.

Sources: [ViewBlock](#), [SiaStats](#), [Filfox](#), [StorjStats](#), [BTFS Scan](#) (Data as of 8 Sept 2022)

*BTFS documentation indicate that 1) node count may include both renters and hosts, and 2) there is no limit to the number of nodes that can be run from a single public IP

Capacity vs Usage

Decentralized storage capacity has increased exponentially in the past 2 years



As the NFT season of 2021 took off, there was a surge of demand for decentralized storage, resulting in a massive increase of available storage. By the end of 2021, the total storage capacity breached 16.7M TB, increasing by more than 4x from 2020.


Filecoin currently has the largest capacity compared to other decentralized storage solutions, with network storage power of over 21M TB. That's more than 40x the capacity of BitTorrent's BTFS network, the 2nd largest decentralized storage provider.

However, most of this storage currently remains unused. As of Q3 2022, only 1% of Filecoin's total capacity is actively being used. On the other hand, usage on other smaller solutions is much higher. For example, ~64% of Storj's total capacity is currently utilized.

*Total capacity data includes Arweave, Filecoin, Storj, and Sia. BTFS excluded due to incomplete data
 **Arweave usage is always equal to its capacity

Cost of Decentralized Storage

Decentralized Trumps Centralized

Decentralized Providers	Monthly Price per TB	Centralized Providers	Monthly Price per TB
 Filecoin	\$0.0002	 iCloud	\$6.00*
 arweave.org	\$1.09	 Google Drive	\$5.00
 STORJ	\$4.00	 OneDrive	\$7.00
 sia	\$0.94	 Dropbox	\$5.00
 BitTorrent File System	\$3.01	 amazon drive	\$7.00*

Decentralized storage is much cheaper

- In terms of pricing, **demand for decentralized storage trumps centralized storage**. Even in the decentralized sector, there are **outliers especially Filecoin**.
- Filecoin Plus, an incentive program that boosts rewards for legit, verified deals, has seen storage providers **offering near-zero or zero fees** to compete for block rewards.
- These rewards are often **subsidized by Filecoin** as they aim to grow the network.

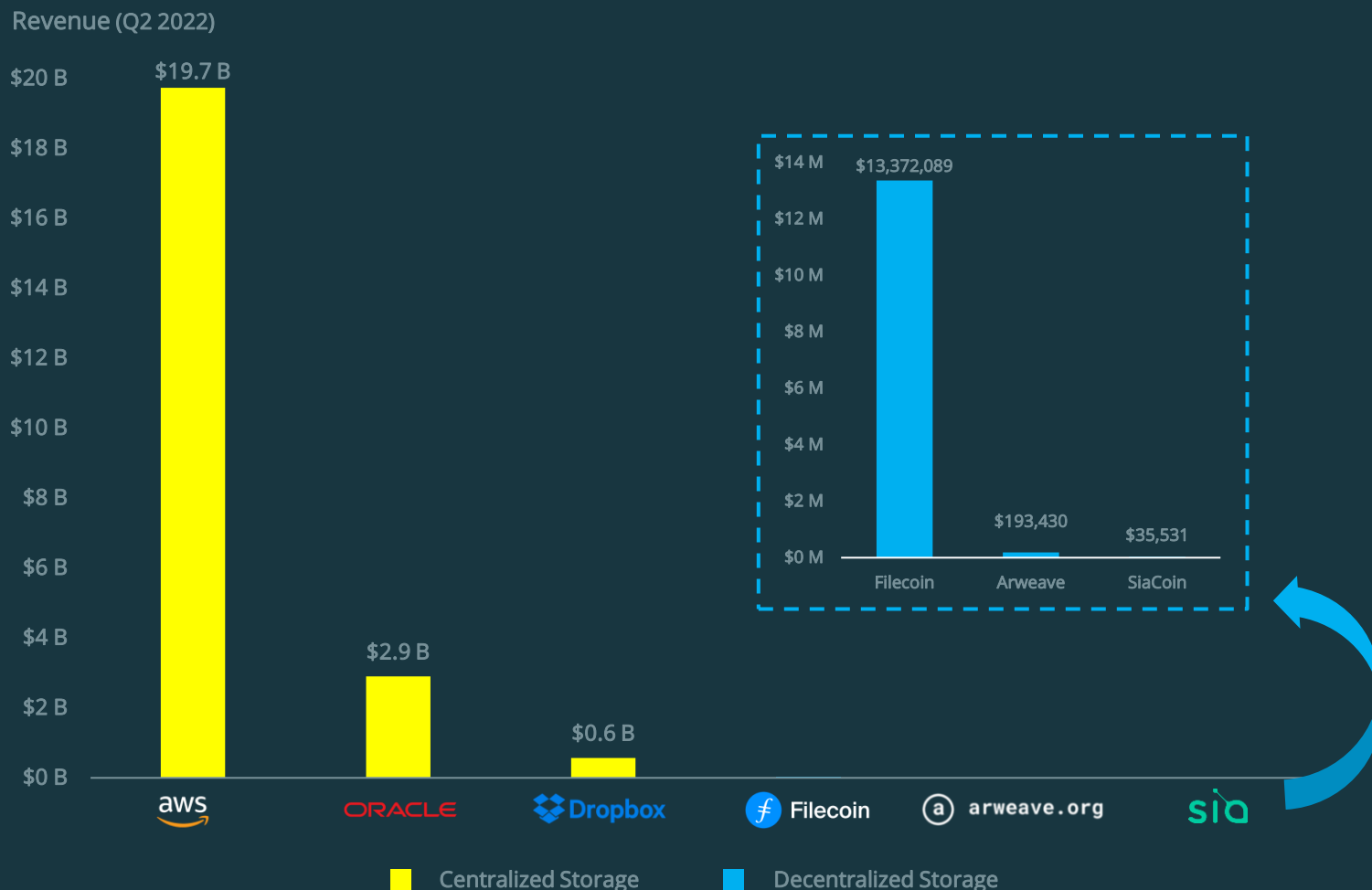
The catch:

- Bandwidth - **Upload (ingress)** and **retrieval (egress)** fees are involved as well.
- Storj charges **\$7 / TB to upload / download** while Sia costs **\$0.41 / TB to upload, and \$2 / TB to download**.
- Filecoin charges a **market price** that's quoted by the storage or retrieval miners.

* Only 2TB packages available. Figure is derived by dividing the cost of package by half.

Protocol Revenue

Filecoin is far ahead of its peers, but is still a long way behind centralized services



The bulk of decentralized storage protocols' **revenue comes from their network fees** and are closely intertwined with the price of their coins (in USD terms).

Filecoin's revenue has taken a hit due to the **slumping FIL price**. For context, FIL closed at \$5.32 by the end of Q2 2022, **-85%** compared to the start of the year. Yet, Filecoin is still **by far the largest decentralized storage solution** in revenue.

Decentralized storage however still **pales in comparison against its centralized counterparts**. AWS for instance raked in close to \$20 billion in Q2, **more than 1,000 times** the revenue of Filecoin, Arweave and Siacoin combined.


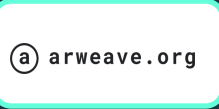


Decentralized storage networks are **still growing**, and we can expect **more services and applications being built** on top to provide **additional revenue streams** in the future.

Revenue data for other decencentralized solutions such as Storj and Bittorrent unavailable.

Sources: [TokenTerminal](#), [SiaStats](#), [CNBC](#)

Decentralized Apps on Decentralized Storage

What's being built on storage L1s?

	Web3 storage	Dev Tooling	Data Market	Consumer Storage	Marketplaces	Socials	Others
	<ul style="list-style-type: none"> Web3.Storage NFT.storage Filedrive 	<ul style="list-style-type: none"> Fleek Estuary Lighthouse 	<ul style="list-style-type: none"> Ocean Filehive 	<ul style="list-style-type: none"> ChainSafe Files Slate 		<ul style="list-style-type: none"> Chingari 	<ul style="list-style-type: none"> Filfox Filscan Filecoin Green
	<ul style="list-style-type: none"> Kyve Permafrost Via 	<ul style="list-style-type: none"> Arweave. Design Gitopia TestWeave 	<ul style="list-style-type: none"> Amplify Koii Meson. network 	<ul style="list-style-type: none"> Ardrive Akord Evermore 	<ul style="list-style-type: none"> Verto Pianity 	<ul style="list-style-type: none"> Metaweave Decent.land Glass Koii 	<ul style="list-style-type: none"> Weve Sarcophagus ArVerify Traxa
		<ul style="list-style-type: none"> Uplink CLI 		<ul style="list-style-type: none"> Drivex Arq FileZilla 			<ul style="list-style-type: none"> Fastly
	<ul style="list-style-type: none"> Skynet 			<ul style="list-style-type: none"> Filebase VUP Arzen 		<ul style="list-style-type: none"> SkyFeed 	<ul style="list-style-type: none"> SkyID SkySend

Most ubiquitous use of de-storage is for **storing NFTs and Web3 data**.

However, apps span **many different** current and emerging **use cases**, some which may overlap. Others include content distribution networks, decentralized IDs, oracles, payment systems, e-mail and more.

Arweave leads in terms of native applications and smart contract capabilities, while **Filecoin has many prominent users** (e.g. MagicEden, OpenSea, Audius, etc.) **and stores important data** (e.g. Shoah Foundation, Internet Archive).

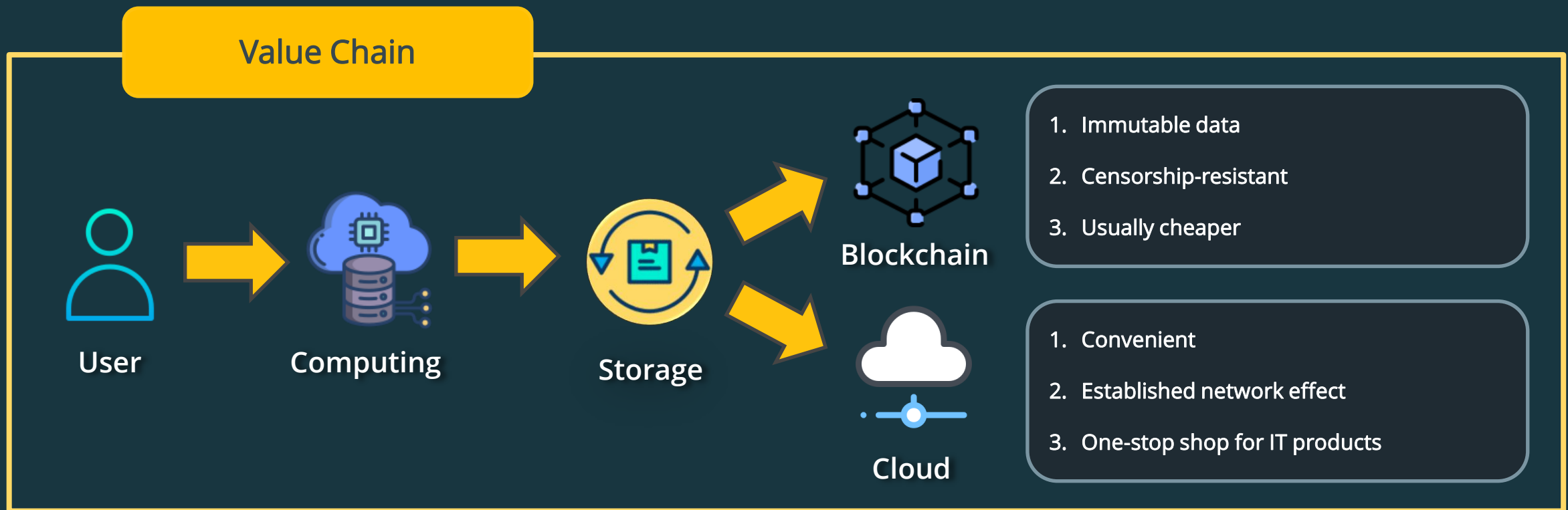
Further buildout of the Filecoin Virtual Machine in late 2022 could see more native applications built on Filecoin.

*BTFS apps still in development

Looking to the future

Decentralized storage is part of the value chain of online computing

To understand where decentralized storage is headed, we must look at the bigger picture. Data storage is merely a subset of the larger **value chain of online computing**.



Understanding the competition

Incumbents already have a head start after capturing the value chain

1

While decentralized storage providers focus on one aspect of the value chain, incumbents like Amazon already have a suite of cloud products designed for online computing, and not just for storage. Under storage alone, AWS has 9 types of different products, catering to different solutions. However, they have 227 products (as of 15 Sept 2022) across different lines like AI, IoT etc. This allows them to **cross-sell different products while offering a holistic solution under the umbrella of Cloud services.**



2

Incumbents have existing products that funnel users into their storage space. Platforms like Google and Microsoft offer email messenger services and word processing programs (e.g., Microsoft Word and Google Docs). Users **will naturally save their files on the most convenient platform which is usually the incumbent's native product.**



3

The regulatory space surrounding data management is strict. Compliance to the GDPR has become the main hurdle that every business faces when it comes to data compliance. As it stands, there are some arguments to suggest that **blockchain technology is not compliant due to the permanence of blockchain data**¹. At the very least, the legal state of decentralized storage is uncertain. As a business/consumer, would you be willing to take that risk? Or would you rather stick with existing cloud providers that are more familiar?



¹ Source: [Arweave News](#)

Conclusion and Key Takeaways

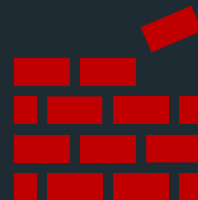
There is a lot of ground to cover before decentralized storage can become mainstream

Decentralized storage has its advantages, but it will take time before it even comes remotely close to the big boys. Outside of censorship concerns and costs, there is little incentive to migrate outside of centralized storage providers, especially when you consider the compliance concerns. The first mover advantage has allowed incumbents to secure a strong network effect. Not only that, incumbents like Amazon and Microsoft have a suite of complementary IT-related products to enforce user stickiness.

Bridging the gap will require innovative methods to capture other parts of the value chain and funneling users into decentralized storage platforms. Some opportunities include:



Integration with traditional software companies



Building products which can attract genuine users



Tackling niche areas that are dominated by cloud technology, such as edge computing



THANK YOU!

[@bobbyong](#)

[@coingecko](#)