Decentralized Cloud?
How the Intersection of Blockchain, Decentralization and Open Source is Impacting Cloud Storage
A long time ago, in a galaxy far far away
(actually, earth circa 1993)

- Telecommunications was highly centralized
- Communication volume was increasing rapidly, but prices weren’t falling
- People thought that established providers were the only way to provide secure, reliable, robust, compliant communications
- Providers spent billions on capital, and raked in 10’s of billions in profits
- …oh, and open source was viewed as a cancer, and not for serious enterprises
Today

Do you worry about which routers/bridges are used when you send or receive information?

Do you worry about which version/model/router?

Do you worry about some of those routers going down?

Do you worry about who operates those routers?

Do you care that your messages go across different routers each time you send them?

Do you want to go back to the days when communication networks were run by the large telcos?
Today

- Open source is dominating
- Public cloud is dominating
- Open source is ⅔ of cloud workloads
- ...but cloud is highly centralized
- Storage volume (especially object storage) exploding
- Exabytes shipped each year is growing exponentially
- The cost of hard drives has decreased by about 50 percent (dollar-per-gb) over last 5 years
- BUT - the price of cloud storage has flatlined for the past five years
Today

- Cloud is generally fast, reliable. But
- Security, data mining, centralization concerns remain
- Providers spent billions on capital, and raked in 10’s of billions in profits
- 90% of drives are <33% utilized
Central Thesis

Open source and decentralization will fundamentally change cloud computing in the same way that OSS and Internet changed computing over past 2 decades.

Storage will lead the way.
Confidential - Storj Labs

Our Goal

Agenda

Central Thesis

What is Decentralized Cloud

Connection between Open Source and Decentralization

Example of Decentralized Storage: Storj. Can you make it enterprise grade?

What does this mean for Open Source

What does this mean for Storage Node Operators

What does this mean for developers? Use cases. Special Topics. Getting Started
Decentralized Cloud

Fundamentally different technical and economic model for delivering infrastructure

Built on open source and the principles that underlay the internet
Let’s clear one thing up...
What is a decentralized application?

**Centralized Apps**
- Central Authority
- Single Point of Failure
- Opaque
- Security by People
- Trust Me

**Decentralized Apps**
- No Central Authority
- No Single Point of Failure
- Transparent - Open Source
- Security by Math
- “Trustless” (really, trust open code and large community)
Let’s clear another thing up...

- Distributed Ledger
- Decentralized Storage Applications
- Regular apps on Decentralized Storage Cloud
<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payments</td>
<td>Bitcoin, Ethereum, ++</td>
</tr>
<tr>
<td>Compute</td>
<td>Dadi, Golem, Hypernet, SONM</td>
</tr>
<tr>
<td>Networking/CDN</td>
<td>Gladius, NKN, Orchid, Storj,</td>
</tr>
<tr>
<td>Storage</td>
<td>Storj, Sia, Maidsafe, Filecoin</td>
</tr>
<tr>
<td></td>
<td>Proprietary</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td><strong>Code Base</strong></td>
<td>Closed</td>
</tr>
<tr>
<td><strong>Decision Making</strong></td>
<td>Opaque</td>
</tr>
<tr>
<td><strong>Security Model</strong></td>
<td>Security through obscurity</td>
</tr>
<tr>
<td><strong>Community</strong></td>
<td>?</td>
</tr>
</tbody>
</table>
## Agenda

### Central Thesis

- What is Decentralized Cloud
- Connection between Open Source and Decentralization
- Example of Decentralized Storage: Storj. Can you make it enterprise grade?
- What does this mean for Open Source
- What does this mean for Storage Node Operators
- What does this mean for developers? Use cases. Special Topics. Getting Started
Example of Storj: Our Goal

To create the world’s largest and most secure, resilient, performant, & economical cloud storage service - without owning or operating a data center.
1 Year, 150,000 Node Operators, and 150PBs Later...
Our Goal

Current IaaS Infrastructure Is Highly Concentrated
V2: Petabyte scale storage for bleeding edge dApps

V3: Exabyte scale, enterprise grade storage for all Object Storage Use Case
Storj is a platform that delivers **Highly Distributed, Ridiculously Resilient** cloud storage.

Delivered leveraging a global, **decentralized** network of **storage** nodes.

**Easy to use**, 25-100% **faster**, more **secure**, more **durable**, at a **fraction of price** of traditional cloud storage.
How It Works 1: Network Overview
Your files are encrypted and split into pieces client-side before being distributed across our network of high-performance storage nodes.
Erasure Coding: Mathematical means of splitting file into N pieces, of which any k can be used to reconstitute file.
Why is decentralized better?

**Durability**
- No single point of failure
- Each drive independently operated, located, powered, networked
- 51 independent drives would have to fail simultaneously, before repair, to lose file #1
- File #2 is on 80 different drives

**Security**
- Client-side encryption by default, on every file
- Decentralized access control/sharing
- Storj can’t see/mine data
- Hackers must find, locate, compromise 30 drives out of 100Ks
- Even then, blobs encrypted
- Start over again to compromise file #2

**Performance**
- Parallel uploads and downloads
- Erasure coding eliminates the long-tail of latency
- Streaming enabled out of the box
- Data served, stored at the edge
## Can Decentralized be Enterprise Grade?

<table>
<thead>
<tr>
<th>Phase</th>
<th>Current Status</th>
<th>Beta 1</th>
<th>Beta 2</th>
<th>Launch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing</td>
<td>Today</td>
<td>1-2 months</td>
<td>Q4</td>
<td></td>
</tr>
<tr>
<td>Durability</td>
<td>100%</td>
<td>99.999%</td>
<td>99.9999%</td>
<td>99.99999999%</td>
</tr>
<tr>
<td>Retrievability</td>
<td>99.93%</td>
<td>99.0%</td>
<td>99.9%</td>
<td>99.99%</td>
</tr>
<tr>
<td>Upload (Median time)</td>
<td>2.15s</td>
<td>1.25 AWS</td>
<td>ON PAR</td>
<td>.75 AWS</td>
</tr>
<tr>
<td>Upload (95th percentile)</td>
<td>2.24 s</td>
<td>1.25 AWS</td>
<td>ON PAR</td>
<td>.75 AWS</td>
</tr>
<tr>
<td>Download (Median time)</td>
<td>1.69 s</td>
<td>1.25 AWS</td>
<td>ON PAR</td>
<td>.75 AWS</td>
</tr>
<tr>
<td>Download (95th percentile)</td>
<td>1.82 s</td>
<td>1.25 AWS</td>
<td>ON PAR</td>
<td>.75 AWS</td>
</tr>
</tbody>
</table>

Durability = Segment health
(>30 pieces needed)

For more detail, visit https://storj.io/blog/2019/08/the-role-of-qualification-gates-in-getting-to-beta-and-beyond
Why is Decentralized Better? Economics

All the normal, user economic benefits of traditional cloud (scaling, low fixed costs, etc.)

Plus great supply-side economics:

- Doesn’t take billions to build out data centers
- SNOs: Idle capacity, no extra power, non-peak network

Result: Much lower prices for users, and prices decrease over time, and...

A new economic model for open source
## Central Thesis

### What is Decentralized Cloud

### Connection between Open Source and Decentralization

### Example of Decentralized Storage: Storj. Can you make it enterprise grade?

### What does this mean for Open Source

### What does this mean for Storage Node Operators

### What does this mean for developers? Use cases. Special Topics. Getting Started
When We Set up Marketplaces, We Create New Dynamics

If open source is the biggest driver of cloud usage, why not have decentralized networks programmatically pay open source projects to help drive growth?

The people who bring supply to the network (“SNO”) should be fairly incentivized & compensated, so they help build capacity.

The people who bring demand to the network should be fairly incentivized & compensated, so they help drive usage.

...If open source is the biggest driver of cloud usage, why not have decentralized networks programmatically pay open source projects to help drive growth?
Open Source Partner Program

Are you an OSS project that generates demand for object storage?  
Build a connector that gives users option to store on the network  
Network tracks usage and returns meaningful portion of revenue that your users generate to you

We can't see user data, and you can't either. But, we can track how much storage and egress is associated with your connector.

Sign up and start building today
Overview

Central Thesis

What is Decentralized Cloud

Connection between Open Source and Decentralization

Example of Decentralized Storage: Storj. Can you make it enterprise grade?

What does this mean for Open Source

What does this mean for Storage Node Operators

What does this mean for developers? Use cases. Special Topics. Getting Started
About the SNOs (Storage Node Operators)

Most SNOs are Good
- Must be vetted first
- Continual uptime monitoring
- Content audits
- Incent good behavior

Assume Some SNOs Are Bad
- Dis-incent bad behavior
- Encryption throughout
- Kick out bad actors
- Highly resilient to bad/incompetent SNOs

...but even Jon SNO knows nothing
(everything encrypted)
How do Storage Node Operators get paid?

Reputation matters.

What’s most important to your node’s reputation:

- **Uptime** - don’t turn your node off without a graceful exit
- **Response Time** - Faster hardware is more likely to serve CDN use cases, and thus get paid more!
- **Audits** - Never lose data and never fail an audit

For a complete list of statistical factors and their weight in the node reputation system, see: [https://storj.io/blog/2019/01/reputation-matters-when-it-comes-to-storage-nodes/](https://storj.io/blog/2019/01/reputation-matters-when-it-comes-to-storage-nodes/)
Our Goal

• A minimum of one (1) processor core dedicated to each storage node service
• A minimum of 500 GB with no maximum of available space per node
• 2 TB of bandwidth available per month; unlimited preferred
• 5 Mbps bandwidth upstream
• 25 Mbps bandwidth downstream
• Online and operational 99.3 % of the time per month (MAX total downtime of 5 hours monthly)

Recommended minimum hardware requirements

- Node Operators | minimum requirements
Agenda

Central Thesis

What is Decentralized Cloud

Connection between Open Source and Decentralization

Example of Decentralized Storage: Storj. Can you make it enterprise grade?

What does this mean for Open Source

What does this mean for Storage Node Operators

What does this mean for developers? Use cases. Special Topics. Getting Started
Where does distributed storage win?

Distributed storage is best for the following types of data:

- Large files over 1MB up to TB size files
- Static data, infrequently changed
- Write once, read many files (WORM)
- Private data
- High volume egress
## Decentralized Storage Use Cases

<table>
<thead>
<tr>
<th>Platform/Service</th>
<th>Description</th>
<th>Decentralized Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archival Storage</td>
<td>Long term storage of large files required for business continuity or based on regulatory compliance.</td>
<td>Low cost and always available high-throughput bandwidth means storage is economical and recovery is rapid.</td>
</tr>
<tr>
<td>Database Backup</td>
<td>Regular snapshot backups of databases for backup or testing are an entrenched part of infrastructure management.</td>
<td>Streaming backup eliminates the need to write large database snapshots to local disk before backup or for recovery.</td>
</tr>
<tr>
<td>Private Data</td>
<td>Data that is highly sensitive and an attractive target for ransomware attacks or other attempts to compromise or censor the data.</td>
<td>Client side encryption and industry-leading access management controls and highly distributed network of storage nodes reduce attack surface and risk.</td>
</tr>
<tr>
<td>Multimedia Storage</td>
<td>Storage of large numbers of large multimedia files, especially data produced at the edge from sources like security cameras that must be stored for long periods of time with low access.</td>
<td>Rapid transit leveraging parallelism makes distributed storage effective for integrating with video compression systems to reduce volume of data stored.</td>
</tr>
<tr>
<td>Multimedia Streaming</td>
<td>Fluid delivery of multimedia files with the ability to seek to specific file ranges and support for large number of concurrent downloads.</td>
<td>Native file streaming support and distributed bandwidth load across highly distributed nodes reduce bottlenecks.</td>
</tr>
<tr>
<td>Large File Transfer</td>
<td>Transiting large amounts of data point to point over the internet.</td>
<td>High-throughput bandwidth takes advantage of parallelism for rapid transit; Client-side encryption ensures privacy during transit.</td>
</tr>
</tbody>
</table>
# Decentralized Storage Use Cases

<table>
<thead>
<tr>
<th>Platform/Service</th>
<th>Description</th>
<th>Decentralized Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software Distribution</strong></td>
<td>Storage and transfer of binary files to be downloaded for software applications, updates, or add-ons</td>
<td>Highly performant bandwidth enables rapid transit of files; Access management and encryption reduce unauthorized access and use</td>
</tr>
<tr>
<td><strong>Content Delivery Network (CDN)</strong></td>
<td>High volume delivery of content, especially large files and multimedia direct to end user or as origin to feed a CDN</td>
<td>Decentralized architecture provides better response times for the consumer experience, as well as efficiency in transport and peering costs with hyper-local</td>
</tr>
<tr>
<td><strong>Compliant Storage</strong></td>
<td>Data subject to regulatory compliance requiring restrictions to storage related to privacy, governance or data residency</td>
<td>Programatically targeting a subset of storage nodes that meet regulatory compliance requirements or that are geofenced within a physical boundary area</td>
</tr>
<tr>
<td><strong>Hybrid Cloud</strong></td>
<td>Flexible ability to provide elastic capacity to on-premise data storage</td>
<td>Enables enterprises to monetize excess storage capacity when not needed and provides secure, private cloud storage on demand</td>
</tr>
<tr>
<td><strong>Machine Learning</strong></td>
<td>Storage transit for processing of large data sets from disparate data sources and types</td>
<td>Decentralized architecture provides better response times for data processing, which can translate into the ability to process more data within time limits, as well as efficiency in transport and peering costs</td>
</tr>
<tr>
<td><strong>VR/AR</strong></td>
<td>Virtual reality and augmented reality are both latency sensitive and bandwidth demanding with large file sets.</td>
<td>Distributed storage provides better response times toward end users, as well as efficiency in transport and decreased peering costs</td>
</tr>
<tr>
<td><strong>IoT Data</strong></td>
<td>Connected devices generate massive amounts of data</td>
<td>Small IoT files can be packed into large blocks for efficient storage while individual message files can be accessed via streaming to specific data ranges</td>
</tr>
</tbody>
</table>
An intuitive cloud experience for developers

Get started in just 3 steps

Create Account
Create Project
Create API Key
Macaroons: Decentralized Access Control
Rich, contextual, and decentralized delegation for access control

- Flexible, decentralized authorization credentials
- Bearer credentials - like cookies!
- Caveats, restrict capabilities and can only be appended, and not removed. Similar to how a blockchain is constructed, HMACs are chained (whereby each caveat contains a hash referring to previous caveats)
Drop-in S3 Compatibility

Upgrade from Amazon S3 without rewriting code....
Or Use Advanced Capabilities with native library

The S3 Gateway allows you point your application towards the Storj Network, without changing any code!

You can even reconfigure the AWS CLI tool to talk with the Storj Network
Create better applications

- Each encrypted and sharded file has a unique hash which serves as a private key.
- Key based architecture of data enables sovereign data ownerships.
- Native end-to-end encryption unlocks new opportunities for user ownership of data.
- Storj improves security and privacy for companies and customers.
Related Use Cases for Decentralized Storage

Other blockchain-based use cases have emerged that extend the value of distributed storage:

- Blockchain-based proof of file integrity
- Non-Fungible Tokens (NFTs) for digital object provenance
- Digital Chain of Custody
- Supply chain management
- Contract-based programmatic file transfer
Example: Proof of File Integrity

**Challenge:** Ensuring that a document is authentic and has not been altered

**Solution:** A hash-output of a file is stored on the blockchain, creating a cryptographic proof that a file has not been altered since a specific point in time.

The blockchain is an immutable, public datastore well-suited for cryptographic timestamping proofs for any file / digital content.
We’re in the Midst of a Major Transformation

“The network is the computer”

Scott McNealy, 1983
We’re in the Midst of a Major Transformation

The network is the marketplace
Central Thesis

Open source and decentralization will fundamentally change cloud computing in same way that OSS and Internet changed computing over past 2 decades.

Storage will lead way
Thank you!

For more info:
OSPP: storj.io/partners
Contribute: github.com/storj/storj
Stats: bit.ly/2ZgB1QJ
White paper: storj.io/whitepaper/
Get Started as Developer: tardigrade.io