Decentralized Platforms
Push Edge Networks
Closer to the Edge

The connection between edge computing and edge storage

storj.io
@storjproject
Ben Golub
Executive Chairman @ Storj

Open Source & Web 1.0

Decentralized Cloud
<table>
<thead>
<tr>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction:</strong></td>
</tr>
<tr>
<td><strong>State of storage:</strong></td>
</tr>
<tr>
<td><strong>What is Decentralization?</strong></td>
</tr>
<tr>
<td><strong>Example of Decentralized Storage:</strong></td>
</tr>
<tr>
<td><strong>What Does this Mean:</strong></td>
</tr>
<tr>
<td><strong>Security at the Edge</strong></td>
</tr>
<tr>
<td><strong>Getting Started</strong></td>
</tr>
</tbody>
</table>
Telecommunications
Then & Now
A long time ago, in a galaxy far, far away
(actually, earth circa 1993)

- Highly Centralized
- Higher Volume, Same Pricing
- Big Corporations were Trusted
- Providers Raked in Billions
Do You Ever Worry About...

No

Routers or Bridges

Models & Versions

Routers Going Down

Router Operators

Do you miss the days when communications were controlled by the large telcos?
Cloud Computing Today
Cloud Computing Today

Now

- Highly Centralized
- Higher Volume, Same Pricing
- Big Corporations are Trusted
- Providers Rake in Billions
Cloud Computing Today

**PROS**

- Fast & Reliable

**CONS**

- Security Concerns
- Data Mining
- Single Point of Failure
- 90% of Drives are 30% Utilized
- Most Drives Already at the Edge

Now
Our Goal

HDD Annual Capacity Shipments

Exabytes Shipped

Storage Pricing
“By 2022, more than 50% of enterprise-generated data will be created and processed outside the data center or cloud”

Gartner
What’s happening at the Edge

By 2022, more than 50% of enterprise-generated data will be created and processed outside the data center or cloud—Gartner
By 2022, more than 50% of enterprise-generated data will be created and processed outside the data center or cloud—Gartner

What’s happening at the Edge
Gartner Findings on Edge Computing

- The variety of use cases and requirements lead to sprawl of first-of-a-kind edge computing deployments, without any synergy and complicating efforts to secure and manage them.

- The scale of distributed computing and storage required by edge computing, as well as deployment locations that usually have no IT staff, combine to create new management challenges.

- With the processing and storage placed outside traditional information security visibility and control, edge computing creates new security challenges that need to be addressed in depth.

- Edge computing creates a sprawling data footprint across a distributed architecture that needs to be governed, integrated and processed.
Central Theses

Given, the increasing trend towards data collection, creation, analysis, and consumption at the edge…storage must increasingly happen at the edge.

The only way to do this securely, effectively, and economically, is to move towards a decentralized model of storage.
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...

- Crypto Currencies
- Blockchain
- Decentralized Systems

very rich
wow
such coin
What is a decentralized application?

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

Decentralized Systems
- 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
- Edge apps on Decentralized Storage Cloud
To create the world’s largest and most secure, resilient, performant, & economical cloud storage service - without owning or operating a data center.
IaaS Market Structure Disruption

Decentralized is the future

- Aggregate underutilized capacity across the globe. Similar business model to AirBNB, Lyft, Uber
- Market structure compared to centralized IaaS
  - Near zero CAPEX to build Exabyte network
  - Near zero cash OPEX to build EB network with variable cost structure
  - Hyperscale via incentivized marketplace of latent supply and developer demand to store more data
1 Year, 150,000 Node Operators, and 150PBs Later...
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
How it Works 2: What Happens to Files?

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.

How it Works 3: Erasure Coding

Encrypt

Split
(each segment into 80+ pieces, of which any 30 needed to reconstitute)

Distribute
(each piece on different, independent drive in global network)
How it Works | System View

1. Retrieve Optimal Nodes
2. Encrypt & Erasure Code Object
3. Upload Pieces to Nodes
4. Store Metadata on Satellite
5. Ongoing Audit & Repair
6. Retrieve Piece Location
7. Stream Data in Parallel
8. Decode & Decrypt
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
## V3 Readiness Gates

### All Exceeded: Now in Production

<table>
<thead>
<tr>
<th>Phase</th>
<th>Current Status</th>
<th>Pioneer 1 Beta 1</th>
<th>Pioneer 2 Beta 2</th>
<th>Early Access Production</th>
<th>General Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timing</strong></td>
<td></td>
<td>Aug 22</td>
<td>Nov 19</td>
<td>Jan 28</td>
<td>March 19</td>
</tr>
<tr>
<td><strong>Durability</strong></td>
<td>100%</td>
<td>99.999%</td>
<td>99.9999%</td>
<td>99.9999999%</td>
<td>99.9999999%</td>
</tr>
<tr>
<td><strong>Availability</strong></td>
<td>99.95%</td>
<td>99.0%</td>
<td>99.9%</td>
<td>99.95%</td>
<td>99.95%</td>
</tr>
<tr>
<td><strong>Upload 10 MB (95th percentile)</strong></td>
<td>2.14 s</td>
<td>1.25 AWS</td>
<td>ON PAR</td>
<td>ON PAR</td>
<td>ON PAR</td>
</tr>
<tr>
<td><strong>Download 10 MB (95th percentile)</strong></td>
<td>1.64 s</td>
<td>1.25 AWS</td>
<td>ON PAR</td>
<td>ON PAR</td>
<td>ON PAR</td>
</tr>
<tr>
<td><strong>Active Nodes</strong></td>
<td>8,200</td>
<td>1,500</td>
<td>3,000</td>
<td>4,000</td>
<td>5,000</td>
</tr>
<tr>
<td><strong>Vetted Node Churn</strong></td>
<td>1.13%</td>
<td>5%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Capacity</strong></td>
<td>25 PB</td>
<td>2 PB</td>
<td>4 PB</td>
<td>6 PB</td>
<td>6 PB</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
Insert edge Security Slides from our blog post
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
## 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>
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)
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)
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>
Macaroons: Decentralized Access Control
Rich, contextual, and decentralized delegation for access control

- Flexible, decentralized authorization credentials
- Bearer credentials - like cookies!
- Root can create, remove, extend, restrict privileges for files, folders, paths, etc.
- Down the chain, 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)
Central Theses

Given, the increasing trend towards data collection, creation, analysis, and consumption at the edge…storage must increasingly happen at the edge.

The only way to do this securely, effectively, and economically, is to move towards a decentralized model of storage.
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
An intuitive cloud experience for developers

Get started in just 3 steps

Create Account
Create Project
Create API Key
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
When We Set up Marketplaces, We Create New Dynamics

Supply

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

Demand

The people who bring demand to the network should be fairly incentivized and 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?
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
Related Use Cases for Decentralized Storage

The primary use case is for standard, enterprise grade, object storage use cases.

However, specialized 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
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/)
Macaroons: Decentralized Access Control
Rich, contextual, and decentralized delegation for access control

- Flexible, decentralized authorization credentials
- Bearer credentials - like cookies!
- Root can create, remove, extend, restrict privileges for files, folders, paths, etc.
- Down the chain, 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)
Central Theses

Given, the increasing trend towards data collection, creation, analysis, and consumption at the edge…storage must increasingly happen at the edge.

The only way to do this securely, effectively, and economically, is to move towards a decentralized model of storage.
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
An intuitive cloud experience for developers

Get started in just 3 steps

Create Account
Create Project
Create API Key
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
When We Set up Marketplaces, We Create New Dynamics

Supply

Those bringing supply (Node Operators) should be fairly incentivized & compensated, so they help build capacity

Demand

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?
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
Related Use Cases for Decentralized Storage

The primary use case is for standard, enterprise grade, object storage use cases.

However, specialized 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
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 uses 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/)
Growth of Edge Networks
Edge Networks Are:

- Semi-Decentralized
- Rapidly Growing
- The Future
Security & Edge Networks
Security Concerns

- More Endpoints
- Failover Management
- Access Control
What Decentralization Brings to the Table
Decentralization

- End-to-End Encryption
- Access Control
- Distributed Endpoints
- Regionless
- Inherent Privacy
Decentralized IAM
Decentralized IAM

- Macaroon-Based API Keys
- Privacy by Default
- Hierarchically Deterministic
Q&A