Ben Golub

@golubbe

Storj.io

@storjproject

Decentralized Cloud?

How the Intersection of Blockchain, Decentralization and Open Source is Impacting Cloud Storage

Ben Golub

Executive Chairman @ Storj

VERISIGN

GLUSTER

plaxo

docker

Open Source & Web 1.0

Decentralized Cloud

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?

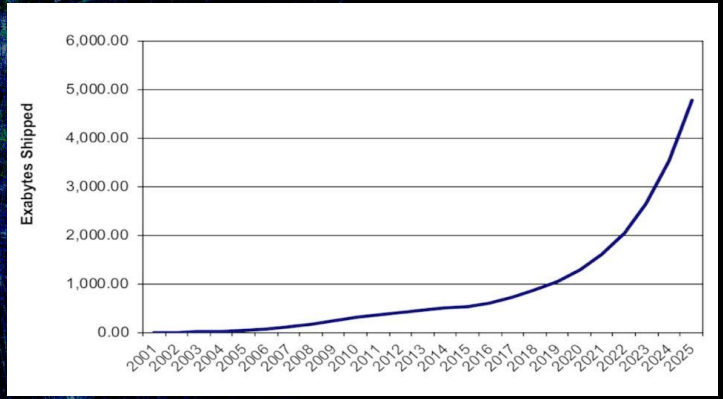
Do you care that your messages go across different routers each time you send them?

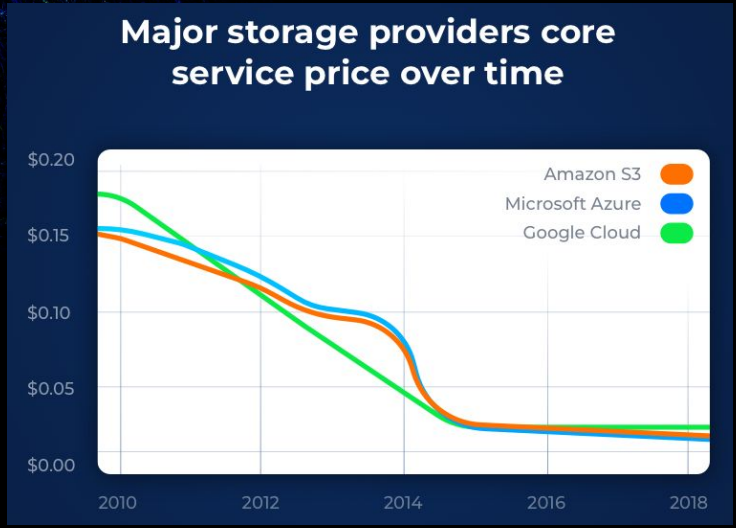
Do you want to go back to the days when communication networks were run by the large telcos?

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

HDD Annual Capacity Shipments

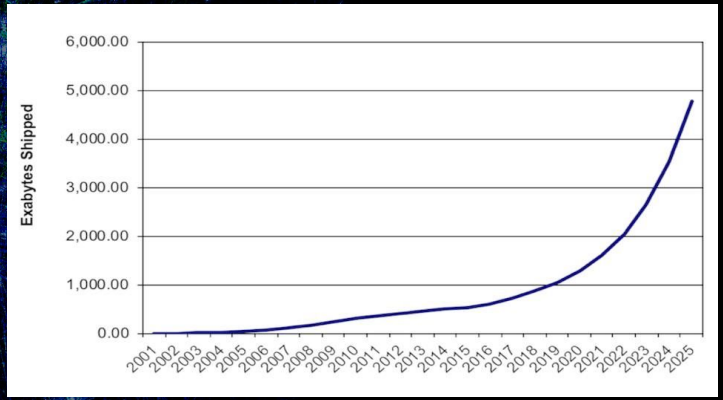


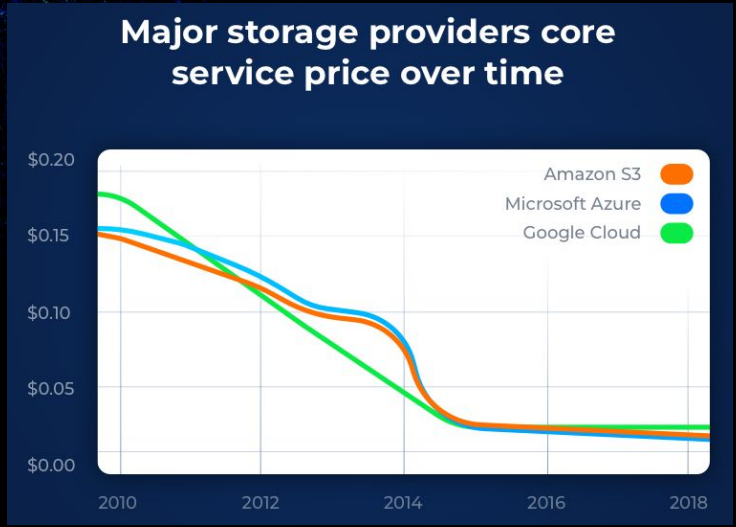


Today

- 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

HDD Annual Capacity Shipments





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

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...

very rich

wow

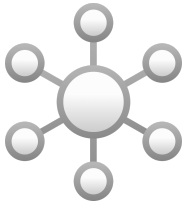
such coin

Crypto Currencies

Blockchain

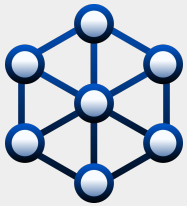
Decentralized Systems

What is a decentralized application?



Centralized Apps

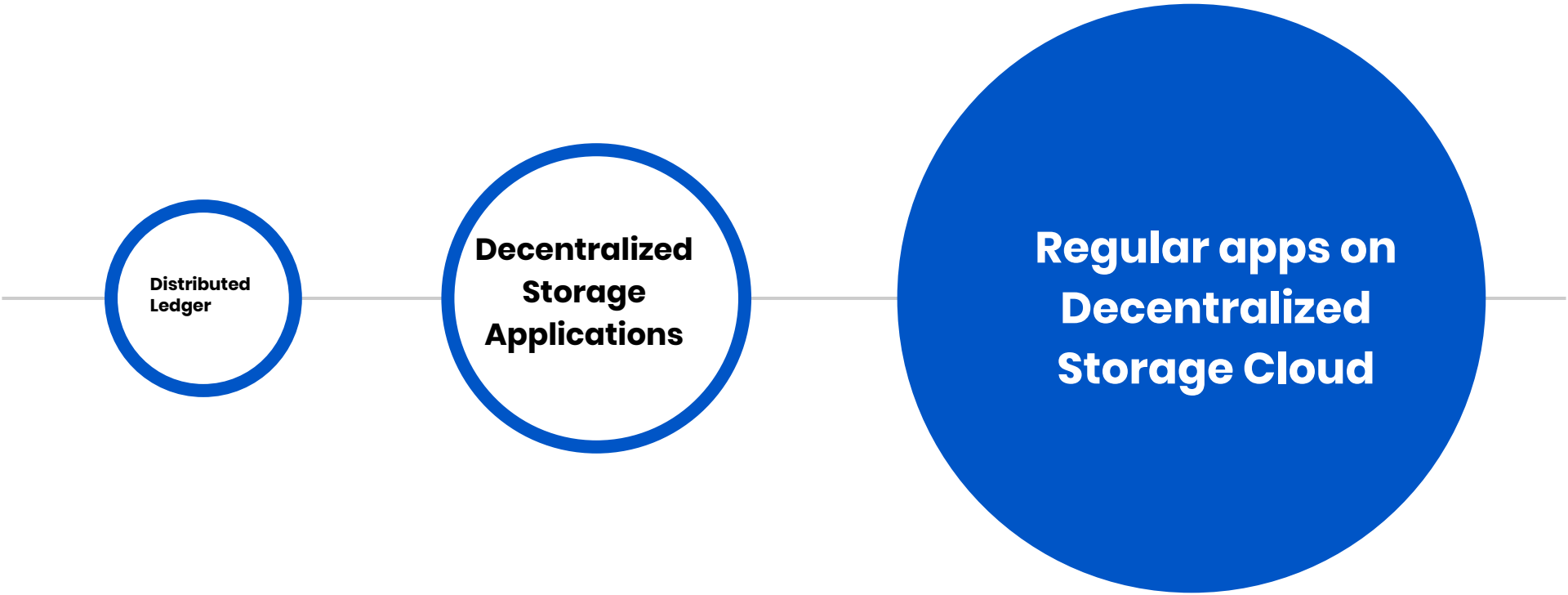
- Central Authority
- Single Point of Failure
- Opaque
- Security by People
- Trust Me



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...



Examples of Decentralized Cloud

Payments	Bitcoin, Ethereum, ++
Compute	Dadi, Golem, Hypernet, SONM
Networking/CDN	Gladius, NKN, Orchid, Storj,
Storage	Storj, Sia, Maidsafe, Filecoin

Similar to Open Source

	Proprietary	Open Source	Decentralization
Code Base	Closed	Open	Open
Decision Making	Opaque	Transparent, open	Transparent, open, algorithmic
Security Model	Security through obscurity	Many eyes	Many eyes, many actors, no single point of failure
Community	?	Critical to creating	Critical to creating, scaling, operating

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

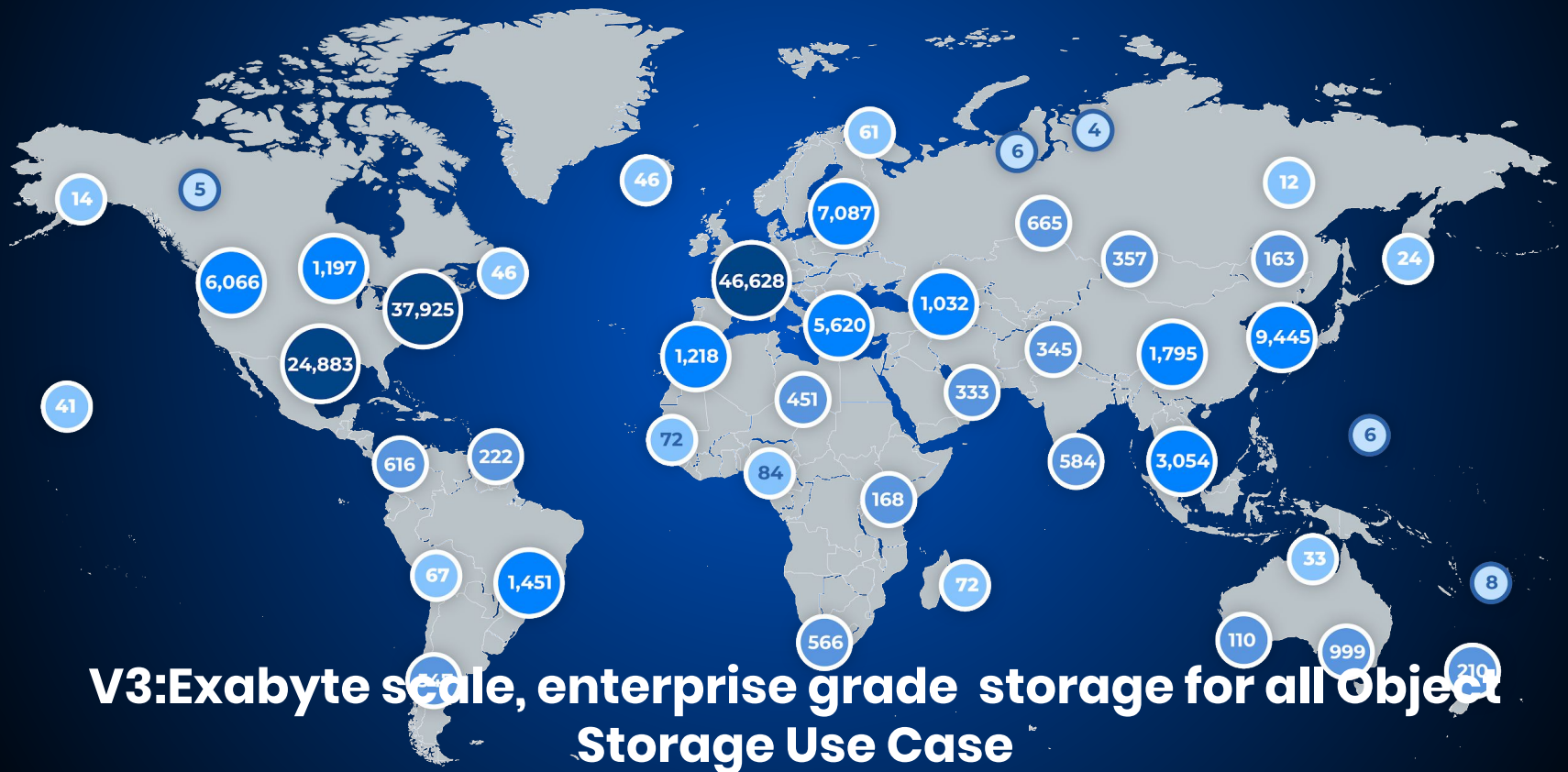
To create the world's largest and most secure, resilient, performant, & economical cloud storage service - without owning or operating a data center.

1 Year, 150,000 Node Operators, and 150PBs Later...

Current IaaS Infrastructure Is Highly Concentrated

STORJ

V2: Petabyte scale storage for bleeding edge dApps



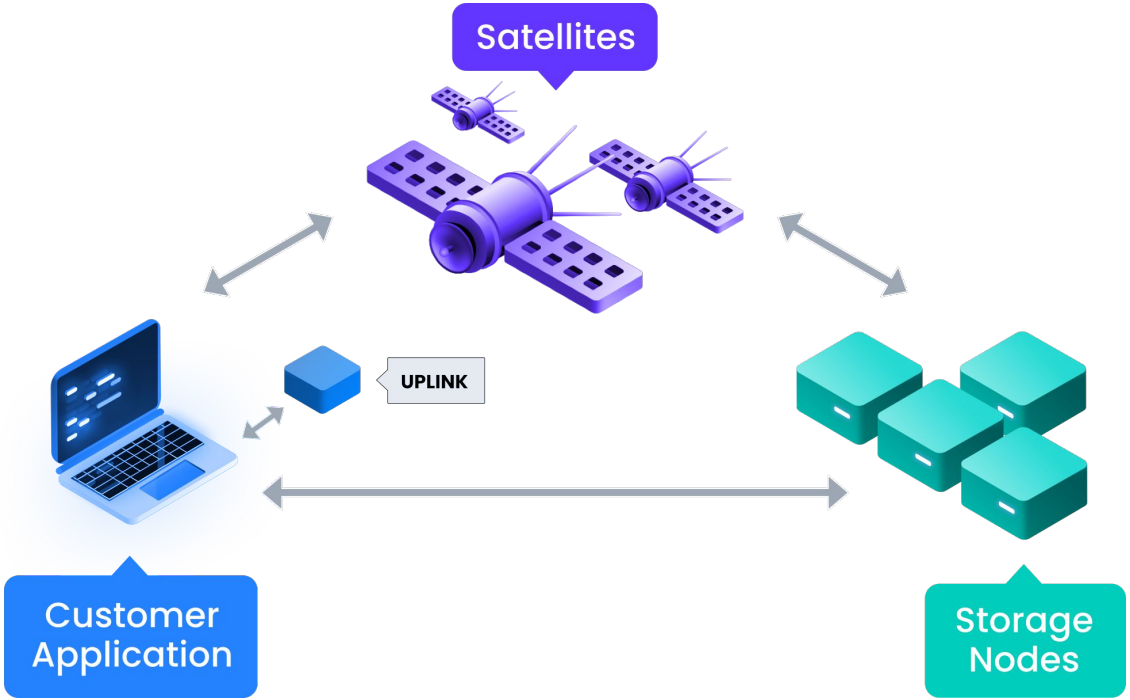
V3:Exabyte scale, enterprise grade storage for all Object Storage Use Case

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

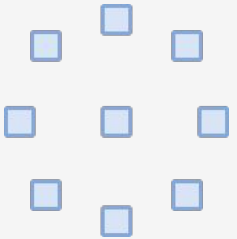


How it Works 2: What Happens to Files?

Your files are encrypted and split into pieces client-side before being distributed across our network of high-performance storage nodes.

Encrypt

Split



Distribute

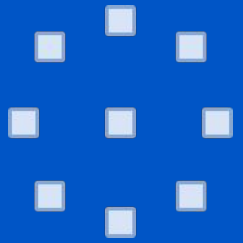
How it Works 3: Erasure Coding

Erasure Coding: Mathematical means of splitting file into N pieces, of which *any* k can be used to reconstitute file

Encrypt

Split

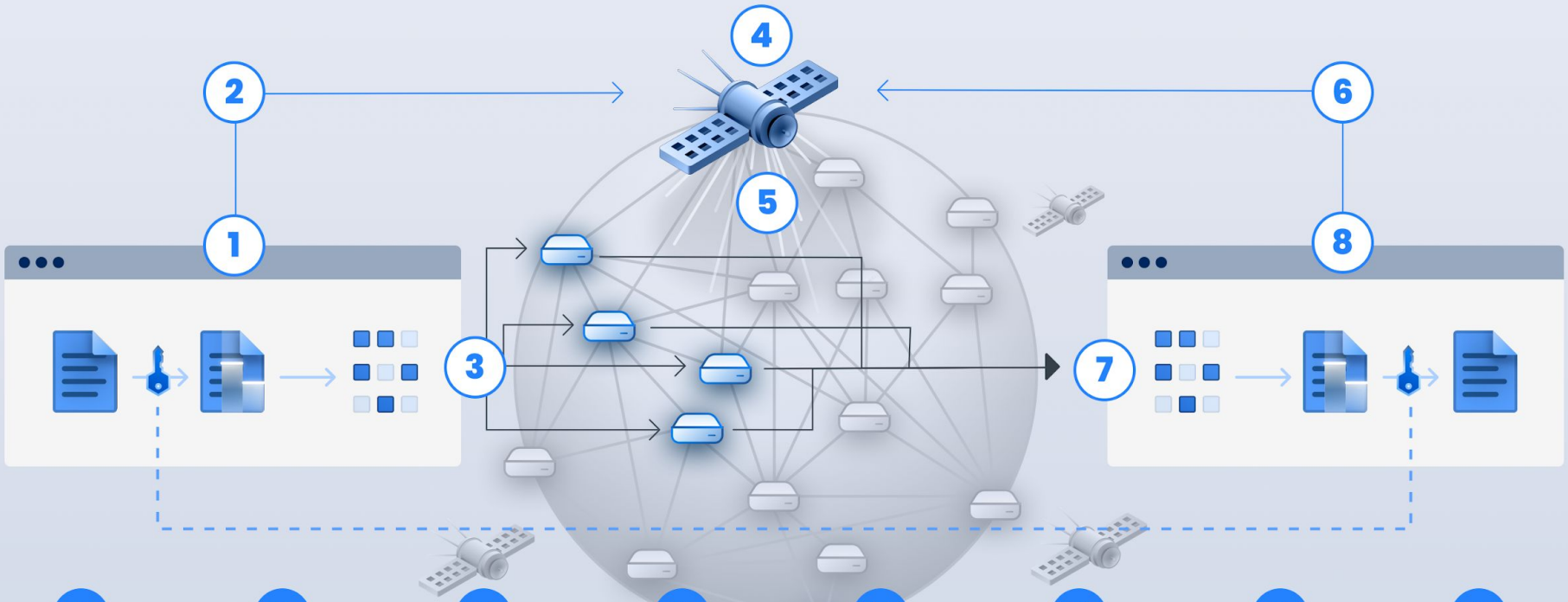
(each segment into 80+ pieces, of which any 30 needed to reconstitute)



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

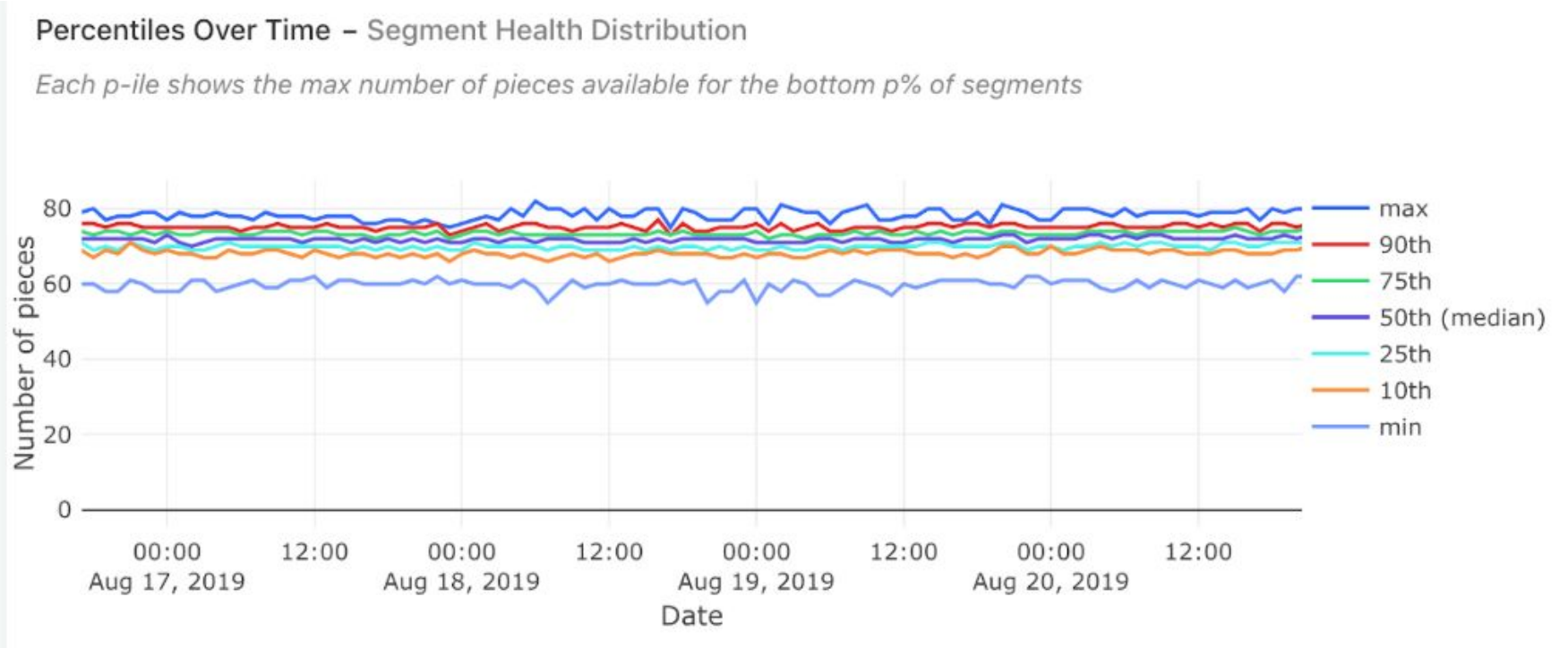
- Parallel uploads and downloads
- Erasure coding eliminates the long-tail of latency
- Streaming enabled out of the box
- Data served, stored at the edge

Can Decentralized be Enterprise Grade?

Phase	Current Status	Beta 1	Beta 2	Launch
Timing		Today	1-2 months	Q4
Durability	**100%**	99.999%	99.9999%	99.9999999%
Retrievability	**99.93%**	99.0%	99.9%	99.99%
Upload (Median time)	**2.15s**	1.25 AWS	ON PAR	.75 AWS
Upload (95th percentile)	**2.24 s**	1.25 AWS	ON PAR	.75 AWS
Download (Median time)	**1.69 s**	1.25 AWS	ON PAR	.75 AWS
Download (95th percentile)	**1.82 s**	1.25 AWS	ON PAR	.75 AWS

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

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

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

When We Set up Marketplaces, We Create New Dynamics

Supply

The people who bring supply to the network (“SNO”) 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?

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 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/

STORJ

Node Operators | minimum requirements

Recommended minimum hardware requirements

- 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)

STORJ

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

STORJ

Decentralized Storage Use Cases

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

STORJ

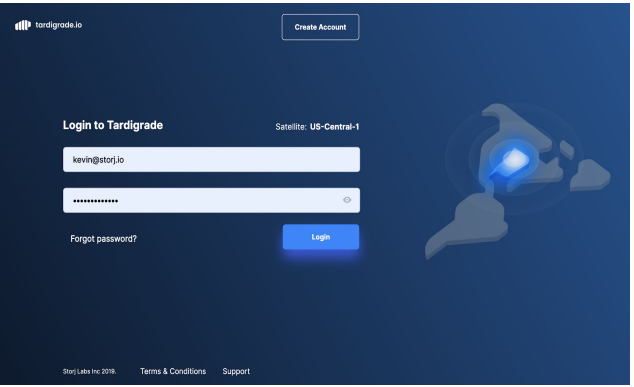
Decentralized Storage Use Cases

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

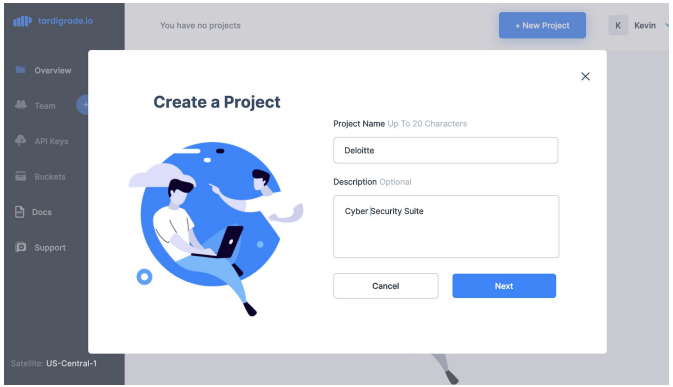
STORJ

An intuitive cloud experience for developers

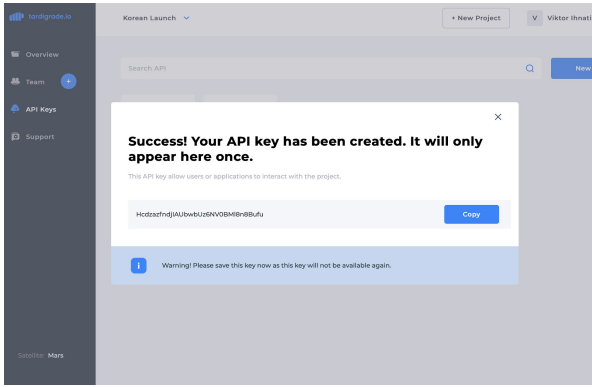
Get started in just 3 steps



Create Account



Create Project



Create API Key

STORJ

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)

STORJ.IO

Drop-in S3 Compatibility

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

```
gateway/0                12:32:54.918 | Starting Storj S3-compatible gateway!
                                      |
                                      | Endpoint: 127.0.0.1:9000
gateway/0                12:32:54.918 | Access key: 3yiSMdGCckS7tw2VGHcdp6g5W71B
                                      | Secret key: mwBfMSXdCPmYmzhnYkyrYfdFHz1
```

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

```
[Kevins-MacBook-Pro-2:bin kevinleffew$ aws configure
 AWS Access Key ID [****************W71B]: 3yiSMdGCckS7tw2VGHcdp6g5W71B
 AWS Secret Access Key [****************FHz1]: mwBfMSXdCPmYmzhnYkyrYfdFHz1
 Default region name [\]:
 Default output format [\\\\]:
[Kevins-MacBook-Pro-2:bin kevinleffew$ aws s3 --endpoint=http://127.0.0.1:9000/ mb s3://demobucket
```

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

STORJ

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

STORJ

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

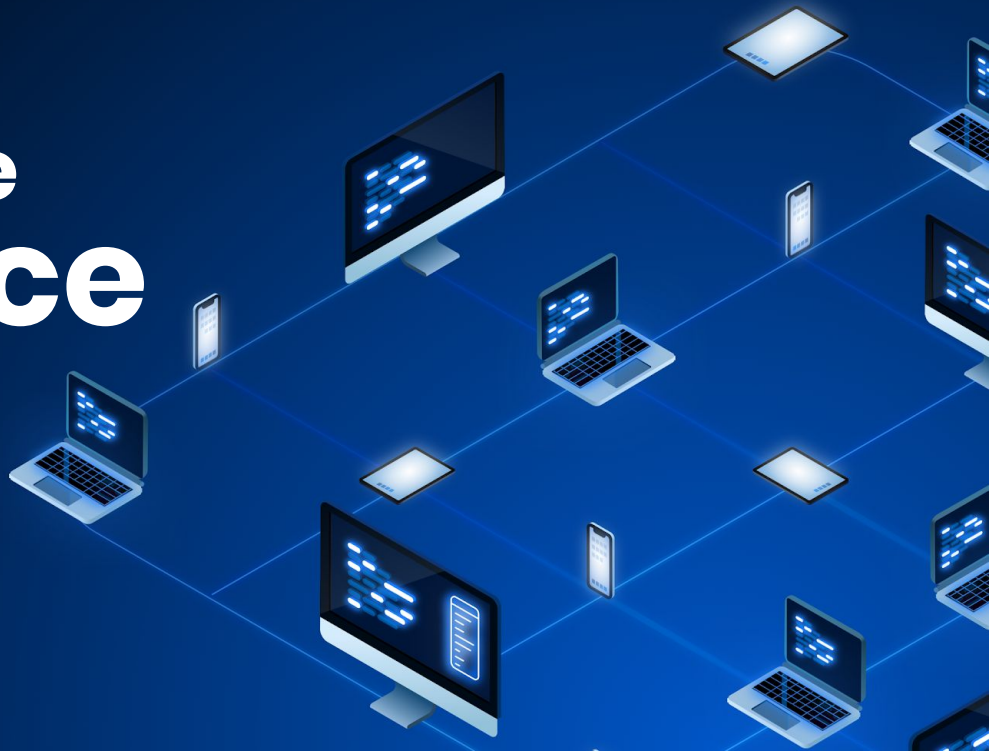
We're in the Midst of a Major Transformation

"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