

STRFISH

If files then Starfish

Because files don't manage themselves

Trusted by the World's Most Demanding Environments



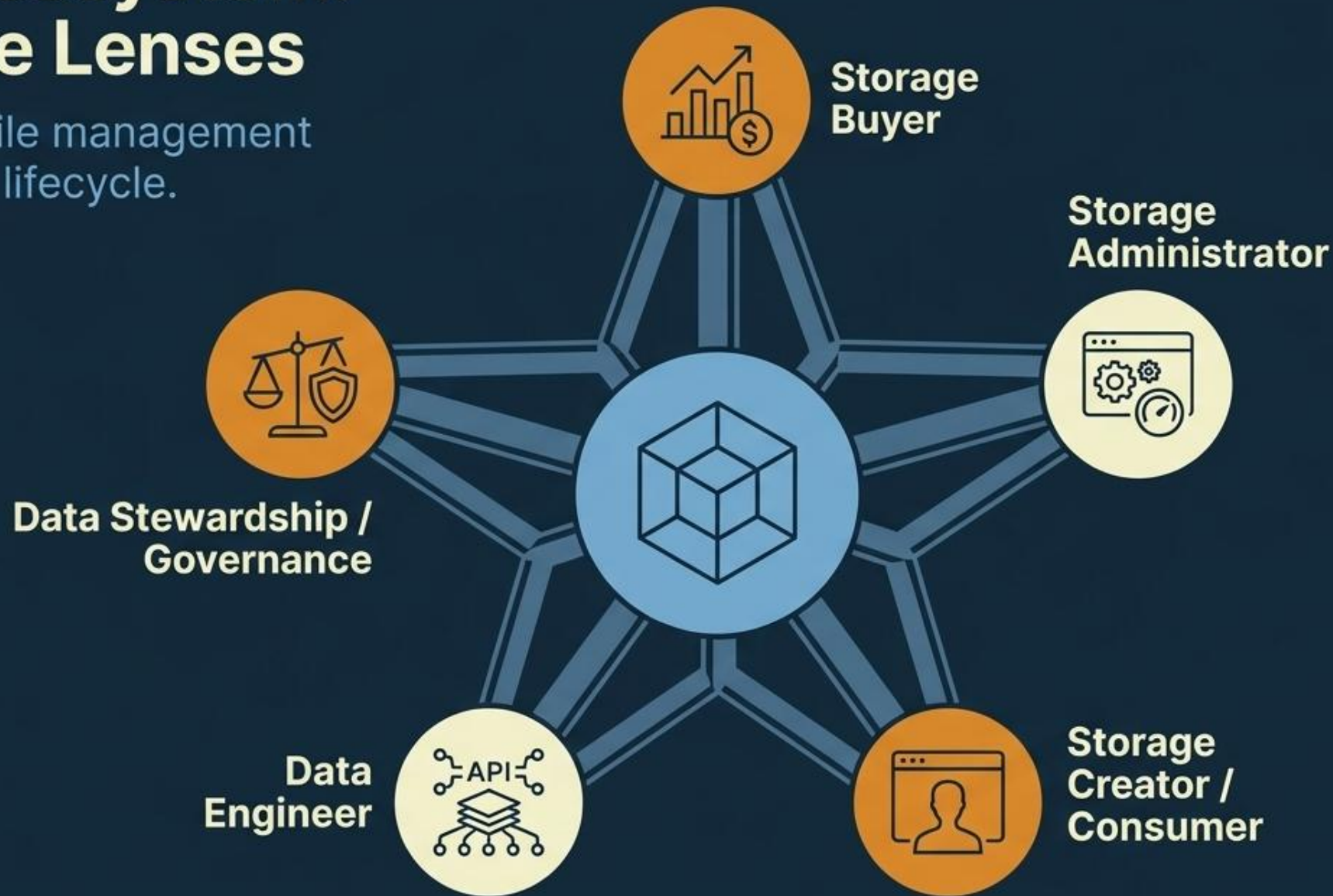
What does Starfish Do? That Depends. . .

*Who are you? What do
you care about?*

- Storage Buyer
- Storage / System Administrator
- Storage Consumer (a/k/a “the users”)
- Data Engineer
- Data Stewardship / Data Governance

The Starfish Ecosystem: One Truth, Five Lenses

Unifying petabyte-scale file management
across the organizational lifecycle.



Storage Buyer

*A Petabyte saved is a
Petabyte earned.*

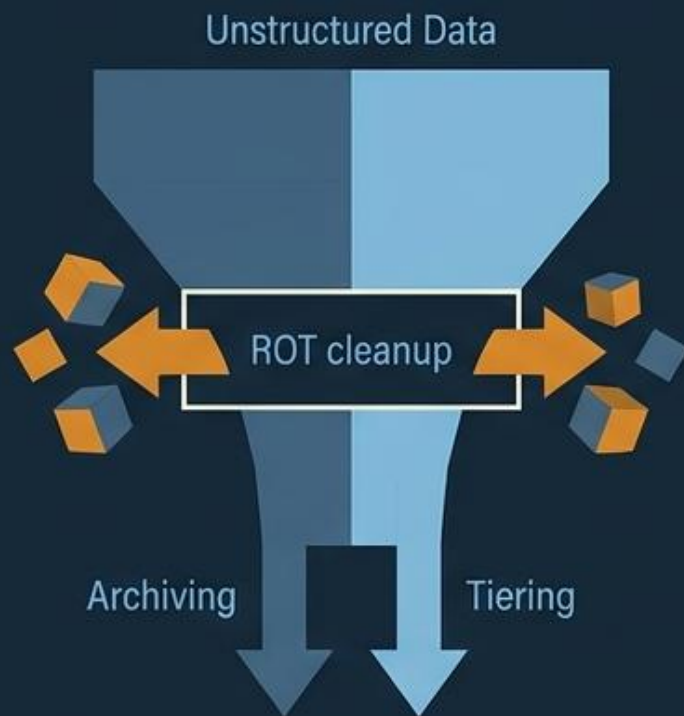
- Capacity Reclamation
 - Archiving
 - Tiering
 - ROT cleanup
- Capacity consumption metrics and trend analysis
- Chargeback / showback

Storage Buyer

A petabyte saved is a petabyte earned.



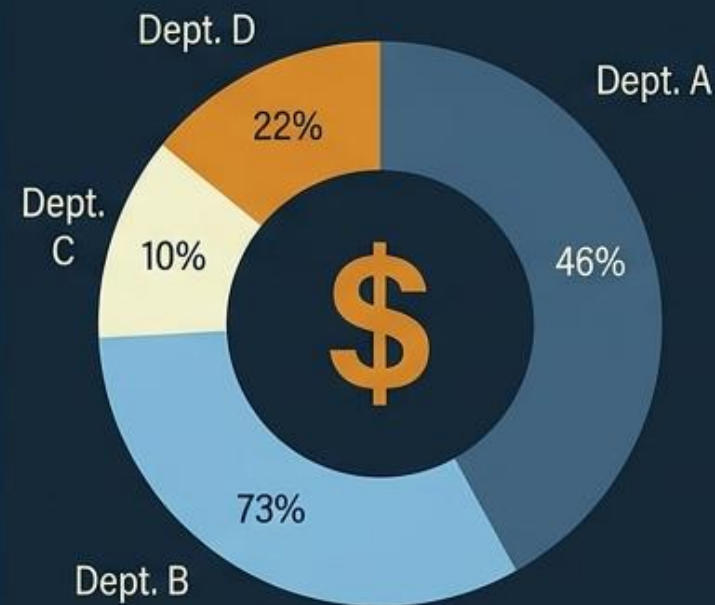
Capacity Reclamation



Capacity Consumption



Chargeback & Showback



Storage Administrator

Regain the capabilities you used to have --
before you had way too many files.

- Starfish works like a **FIND** statement + **CSV** + **SCRIPT**
 - Make a discovery take action upon it
- Move data using any common data movement paradigm
 - Backup / Data Protection
 - Archiving
 - Replication
 - Migration
 - Workflow
- Analytics - spot problems and opportunities
- Automation - fix problems, automate

Storage Consumer

Give your users the tools. Now you can hold them accountable.

- AI Chatbot
- Data classification (tagging on files and directories)
 - Ability to manage their own storage
- Discovery tools
- Reporting / analytics
- Collaboration tools

Data Engineers / Preservationists / Archivists

The single pain of glass
for all things file storage

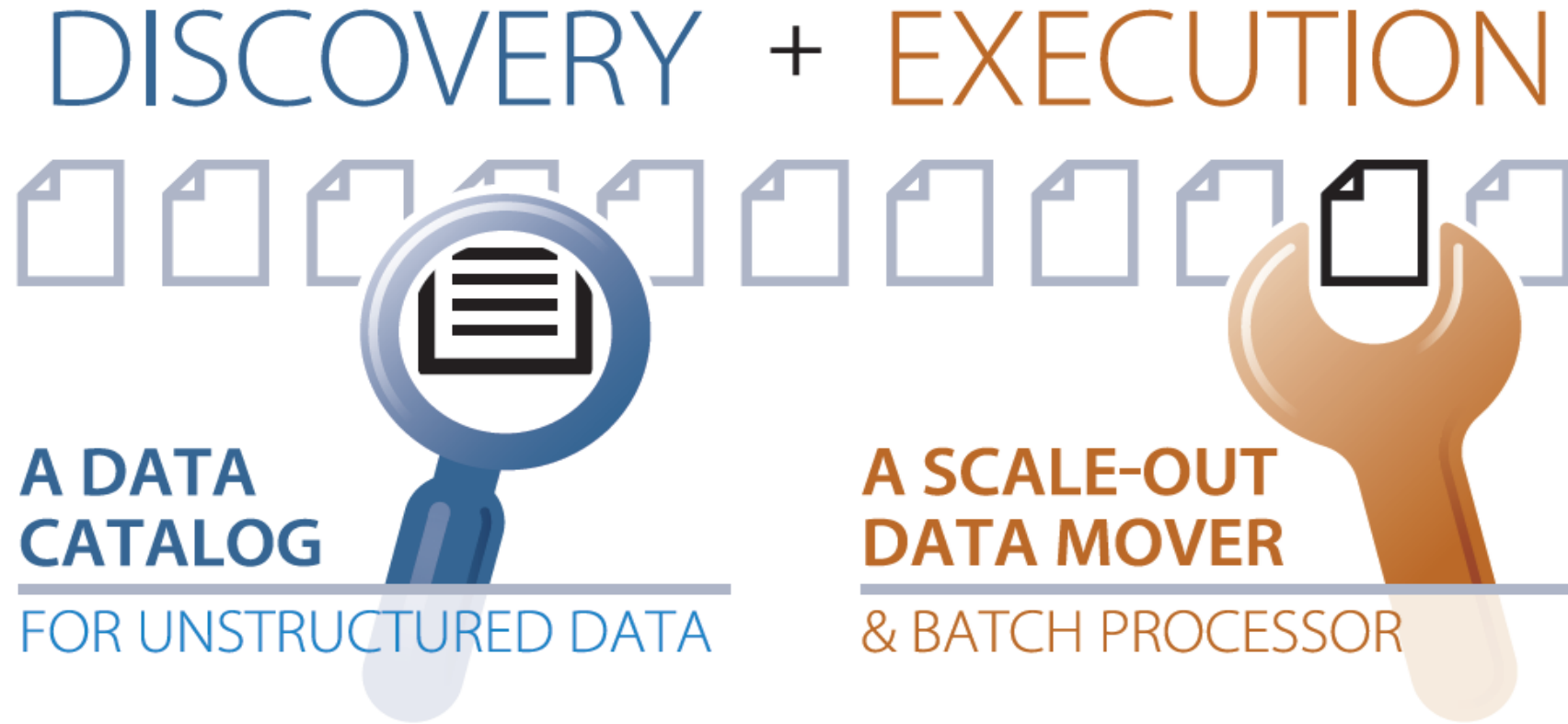
- A global map of all file content
- A single API for all file storage
- MCPs for Agentic Automation
- Hooks into 3rd party Workflow Engines
 - N8N, Snakemate, AirFlow, NextFlow, etc.

Data Stewardship / Governance

Relate file storage to business objectives, liabilities, responsibilities, best practices

- Data stewardship
- AI swim lanes
 - Metadata defined boundaries
 - Authorization, Data Quality Attestation, Sandboxes, etc.
- Policy definition and execution
- Preservation / protection of assets
- Disposition of liabilities

How is Starfish So Diverse? It's the basics. . .

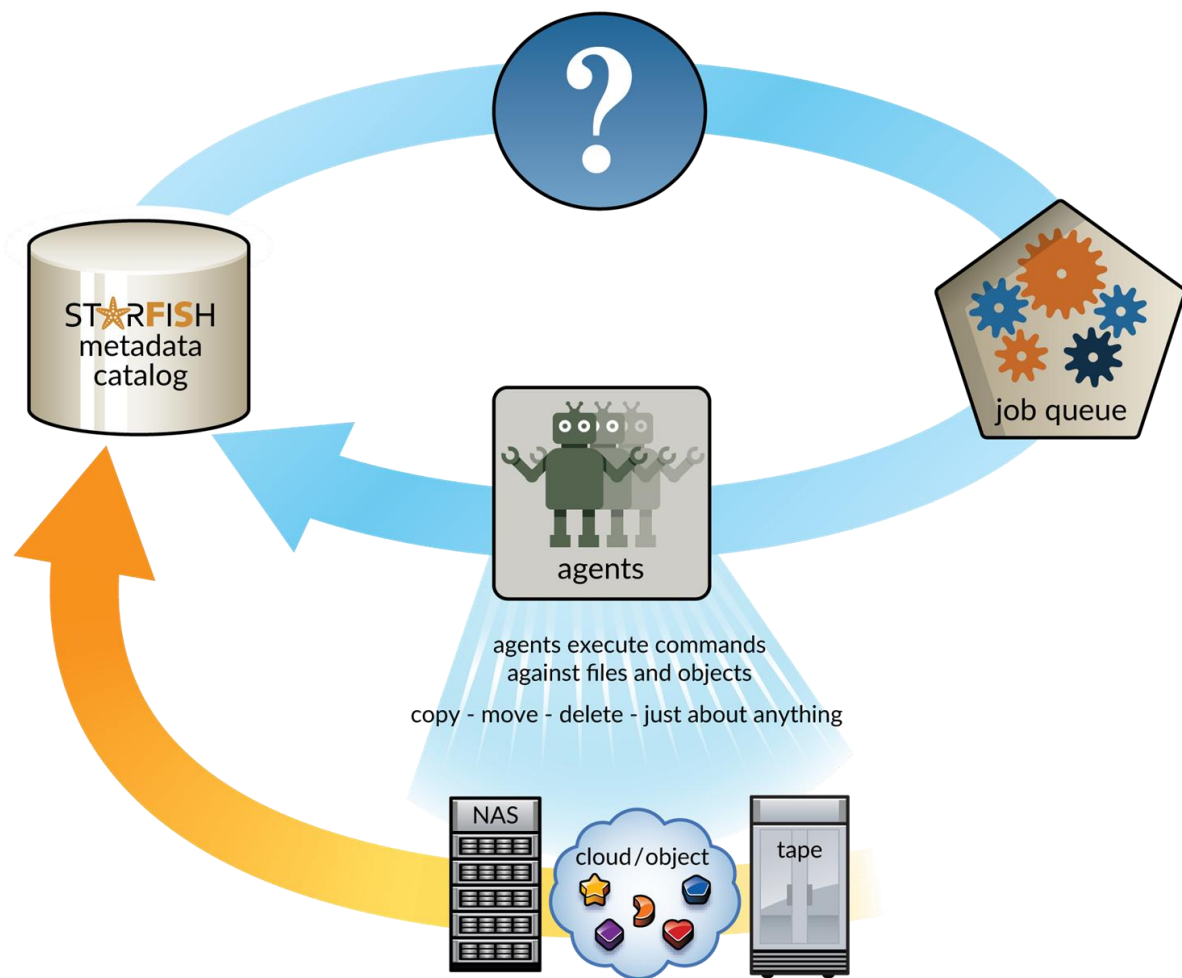


The **Data Catalog** enumerates and describes all of the files and directories in your storage environment

The **Automation Engine** moves and processes files based on discoveries from the data catalog

Discovery + Execution – Feedback loop

The output of a query is the
input of a batch process



Query database
Process the result set
Update the database

Query database
Process the result set
Update the database

Query database . . .

What's a Data Catalog?

It's more than just a database of files

- A system of metadata for data sets / data resources that describes and governs the discovery and usage
- How does the data relate to the business?
- Data Governance:
 - Who “owns” the data and can make business decisions about who uses it and what for?
 - What are the policies for protection, access, life cycle management, etc.
- Data Stewardship
 - Who is responsible for data quality, care and feeding?

Differentiators – Starfish Data Catalog

Storage Vendor Agnostic

Starfish works across all enterprise
NAS, HPC file systems, and S3-style object stores



Best in Class Discovery, Reporting & Analytics

Scale

- Starfish works in the largest and most complex file environments

Version history

- Measure changes over time

Metadata

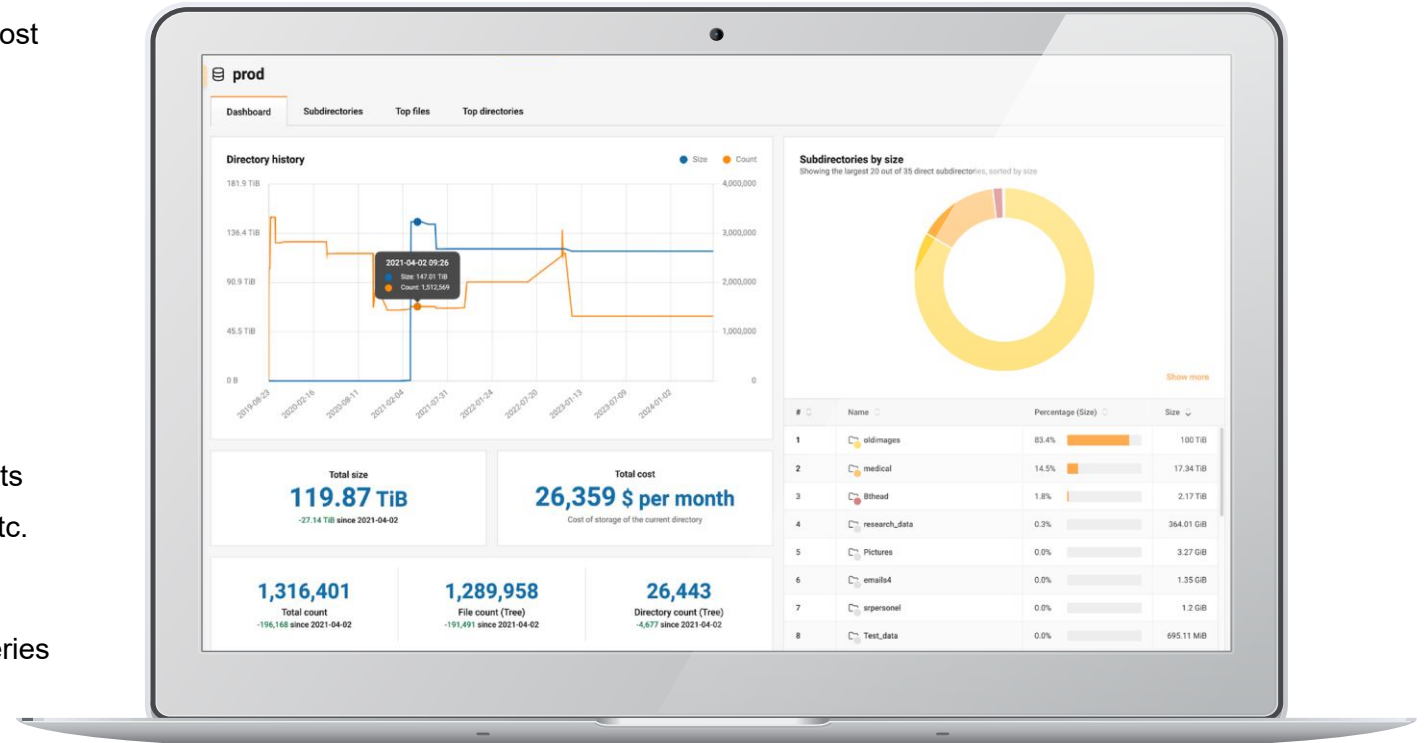
- Metadata enables more specific analysis

SQL Openness

- Customized dashboards and reports
- Expose to Tableau, Microsoft BI, etc.

Actionable

- Take action based on your discoveries



Empower Users to Manage Their Own Files

Starfish ZONES

a logical grouping of branch directories from anywhere in the file storage environment.

- assign users by User ID and Group ID
- assign a name and any other properties to the zone
- define sets of tags that can be used within the zone
- define roles for users in the context of the zone



Use Cases for Zones

- Zones allow users to classify and manage their own files.
- Zones are the organizing principle for feeding data into LLM pipelines and controlling who has access to the tuned model.
- Zones can match to billing codes for chargeback or cost-accounting
- Zones are boundaries for data ownership and stewardship
- Zones provide swim lanes for agentic AI

Starfish is more than just a data catalog

- Integrated data warehouse for analytics
- Metadata constructs
 - Zones
 - Collections / Data Sets
- Batch Processor – do anything to files that you can express in code
- Data movers – move data any which way
- S3 export gateway – export query results as an S3 bucket
- API / MCPs
- Webhooks for workflow engines

A Few Last Words

- Starfish is free to try. We are very generous with PoC terms. Bring it in. Figure out how it will pay for itself before you buy anything.
- Starfish has acquired StorJ technology for private and community cloud storage.
 - Salvage unused storage capacity
 - Turn it into a resilient storage cloud across geographies
- Our parent organization will be announcing a new product for salvaging and pooling GPU resources at Supercomputing in Chicago.