



Meghan McClelland
MSST 2026



What We Do at Versity:

Manage large data collections at the lowest possible cost

- A Modern Software-Defined Storage Platform
- For Mass Storage & Large Archive Systems

Versity Global Installed Base

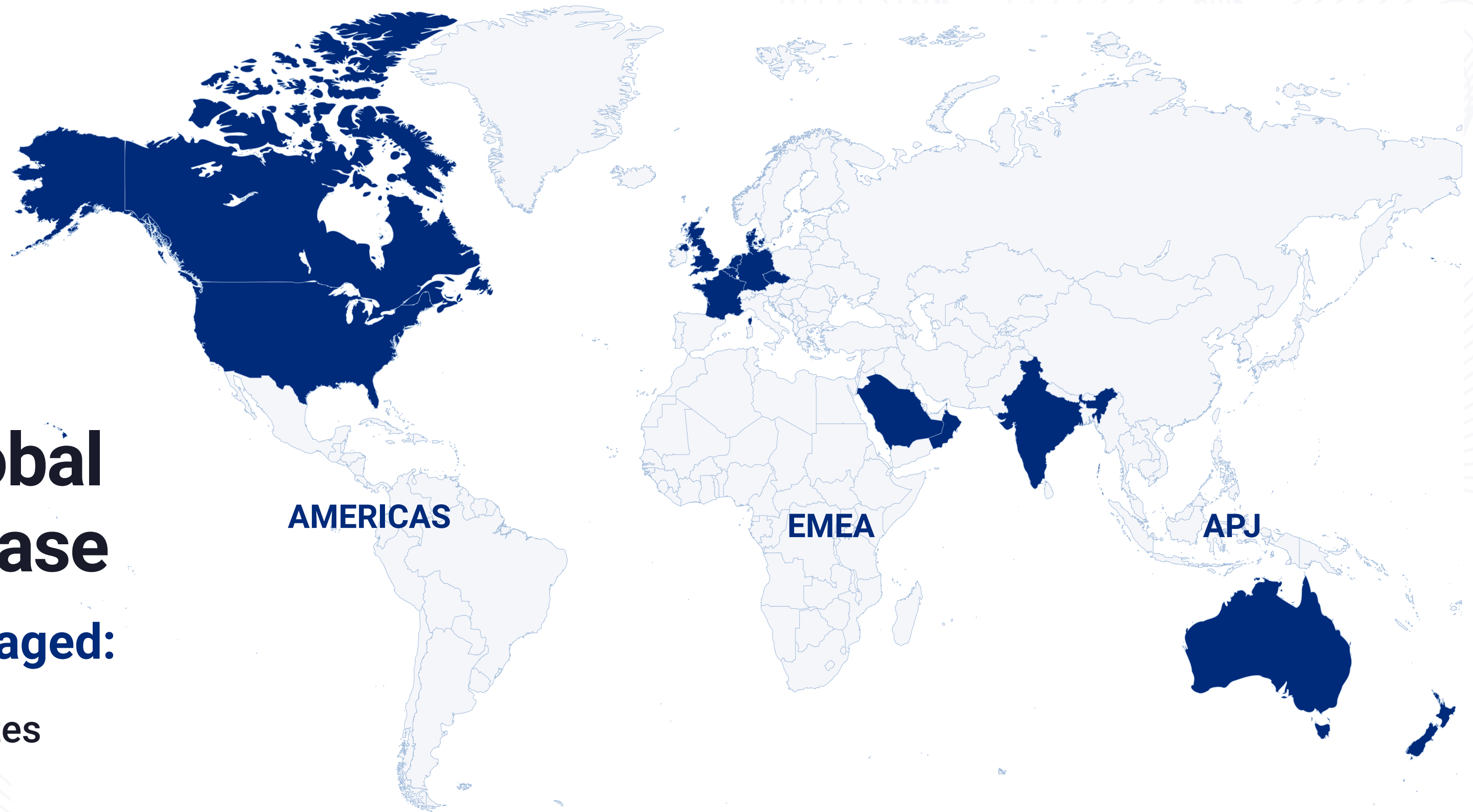
Total Data Managed:



Over 4 Exabytes



Over 100 Sites



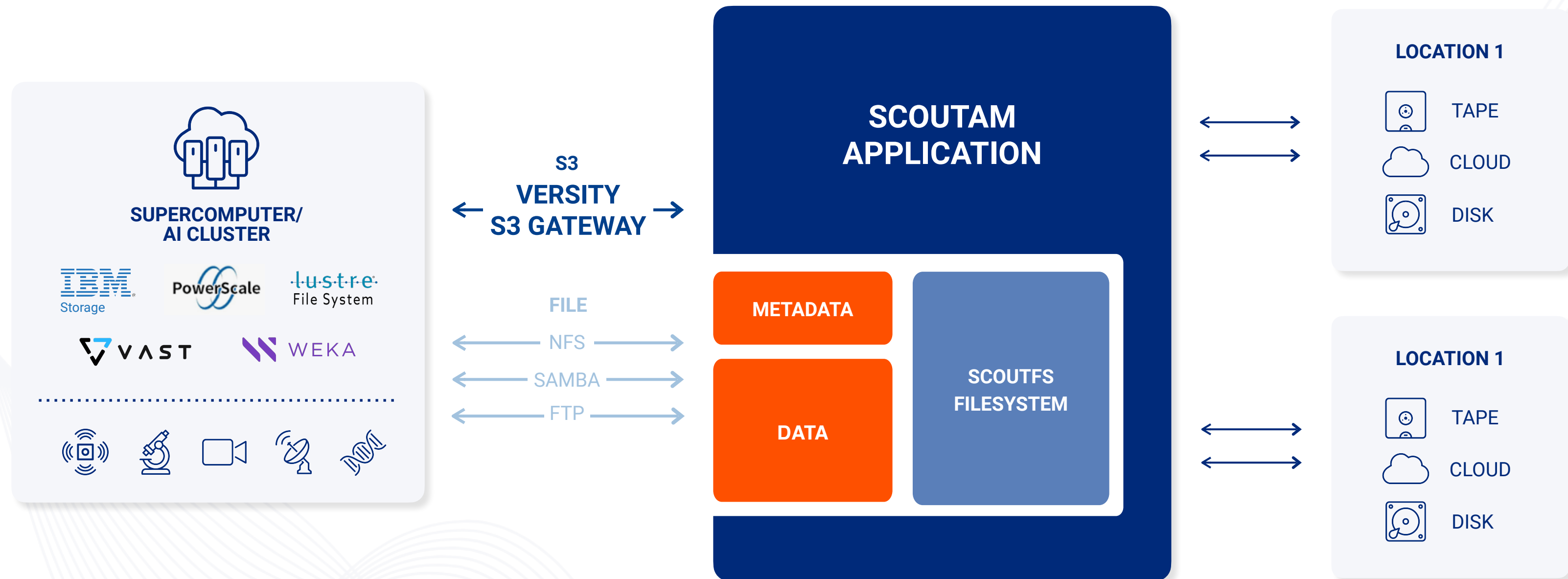
Versity Customers

- Have extremely large data collections that are too big for backup
- Require long-term preservation
- Actively read & write large data sets
- Need: cost efficiency, versatility, control, and performance
- Want: SIMPLICITY

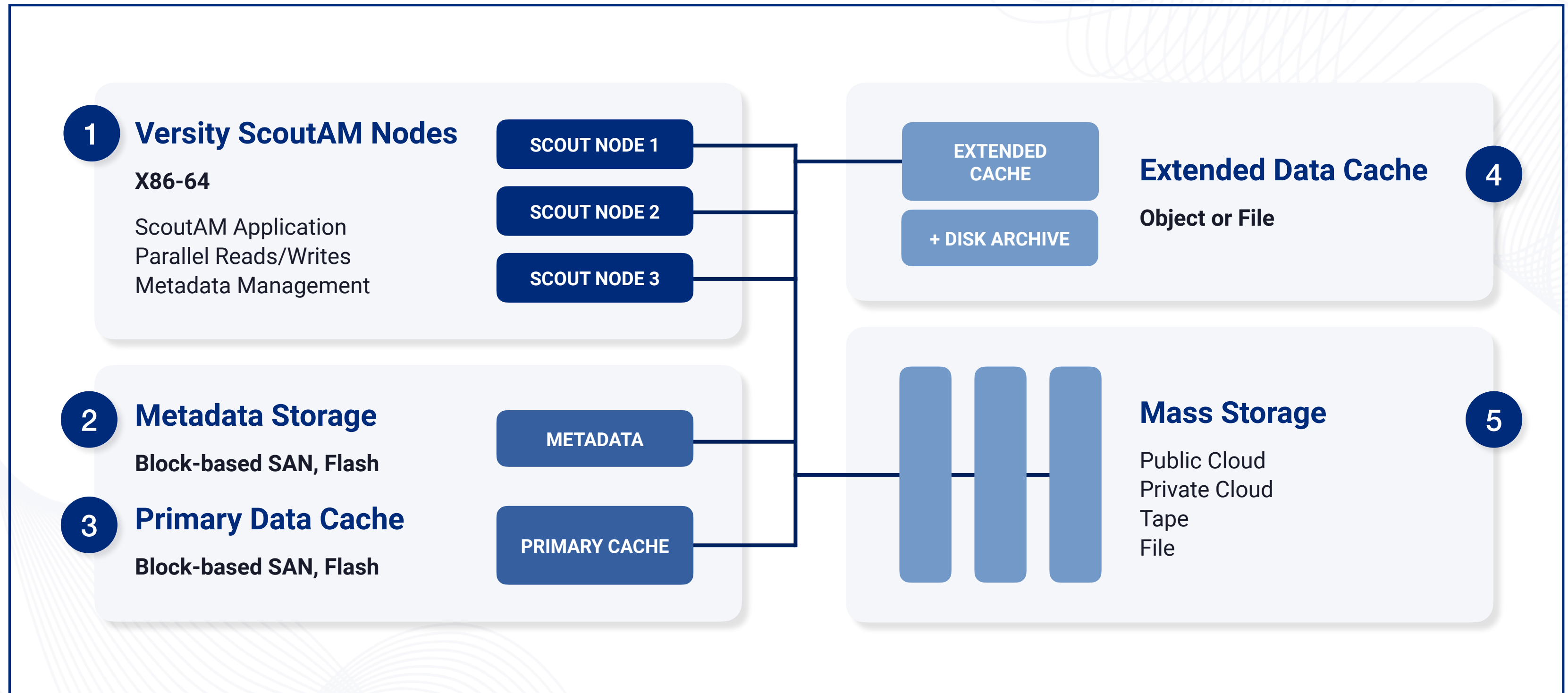


Versity: Archive with Intelligence

A Modern Data Management Platform



Versity Reference Architecture



TACC Solution Architecture

Data

Horizon Supercomputer
400 petaflops

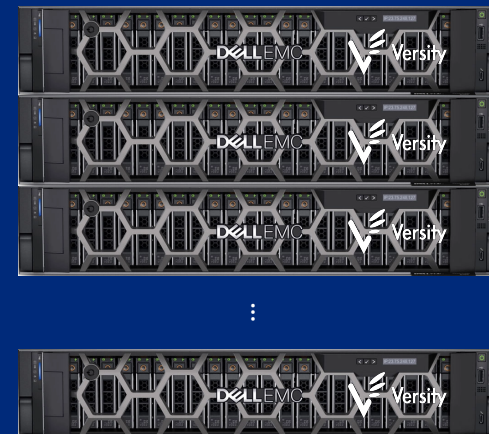
AI Data from Biosciences, Computer
Vision and NLP Applications

Petabytes of Astrophysics
Simulation Data

Climate Modeling and Weather
Simulation Data

Versity ScoutAM Archive Solution

13 X SCOUTAM NODES
DELL R760



5 X DELL ME
1 X METADATA
ME5024
7.6 TB
4 X DATA
ME5084
4 PB

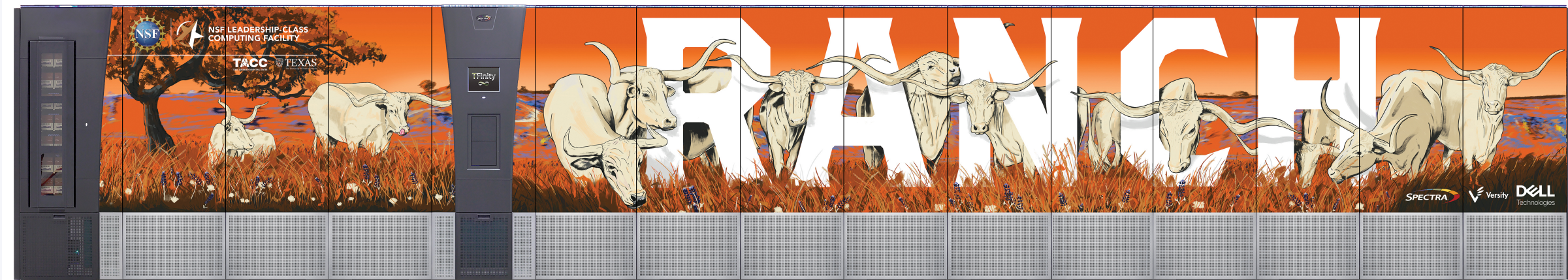


EXTENDED CACHE
DELL ECS
16PB



Mass Storage

SPECTRA LOGIC
16 X LTO 9 DRIVES
8 X LTO 10 DRIVES



DKRZ Workflow Architecture

HPC

Levante
3,000 compute nodes + GPUs

Data

AI/ML Research Data

Earth-System Model Data

Climate Modeling Data
ICON, ECHAM, FESOM

lustre
File System

130 PB

HPSS and Stronglink Data

Versity ScoutAM Archive Solution



SCOUTAM NODES X 7
SUPERMICRO SYS-222C



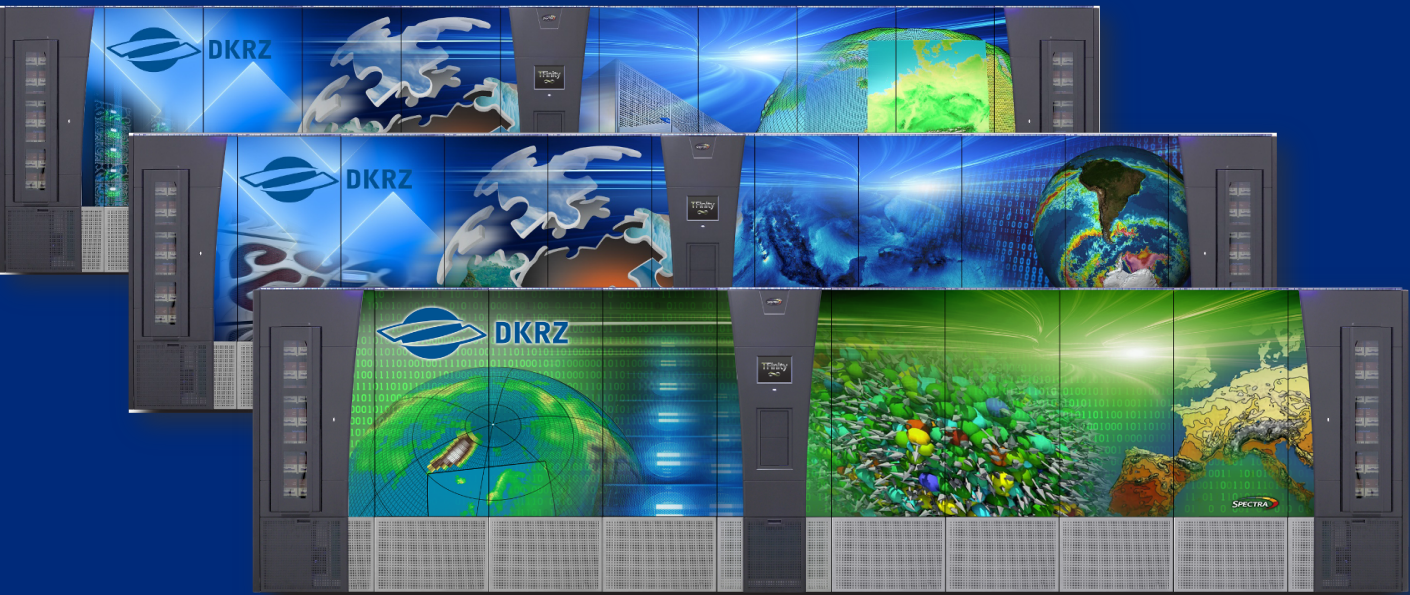
8 SEAGATE
EXOS X

DATA
1.5 PB

METADATA
120 TB



EXTENDED CACHE
SEAGATE EOSX 4006
3.4 PB



SPECTRA LOGIC
2 X 11 FRAME TFINITY
1 10 FRAME TFINITY

1 EXABYTE

Mass Storage

REMOTE COPY



SCOUTAM NODES X 2
SUPERMICRO SYS-222C



SEAGATE
EOSX 4006

DATA
240 TB

METADATA
30 TB

MASS STORAGE

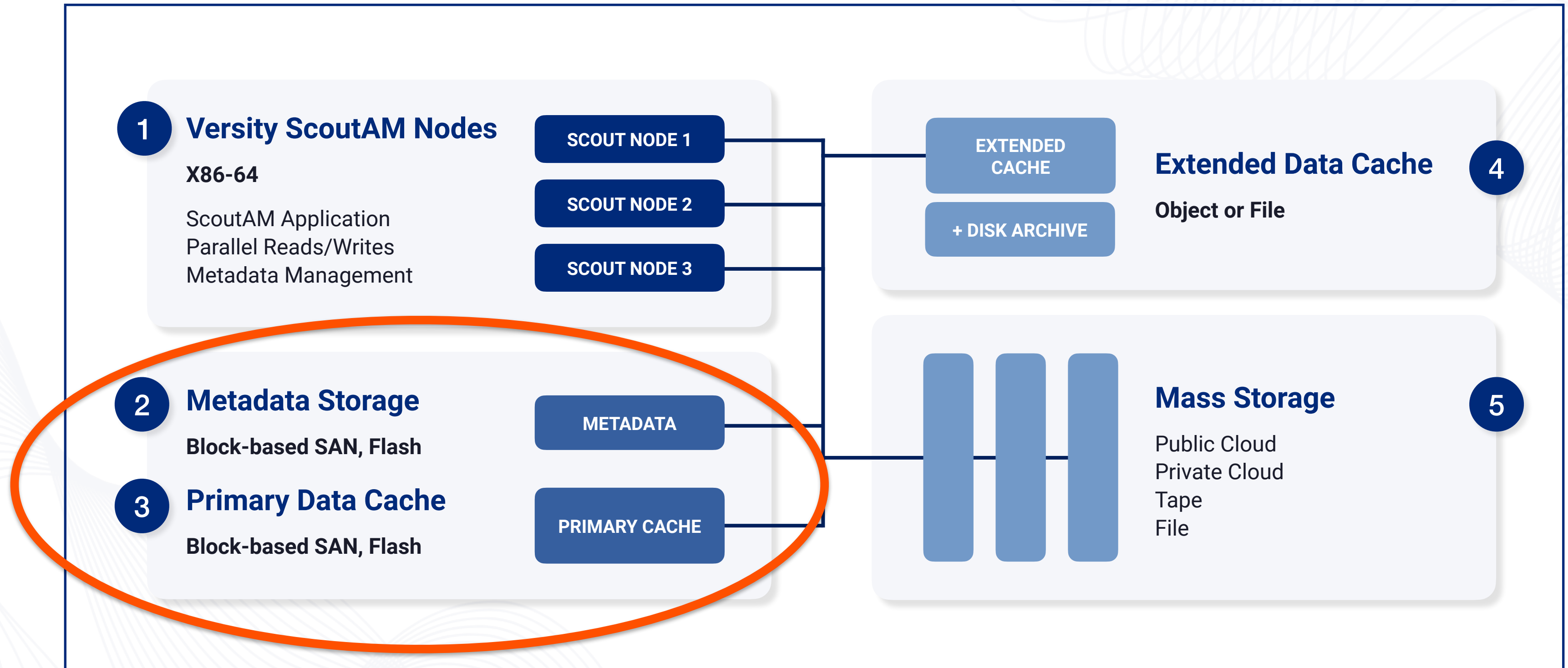


STORAGETEK
SL8500

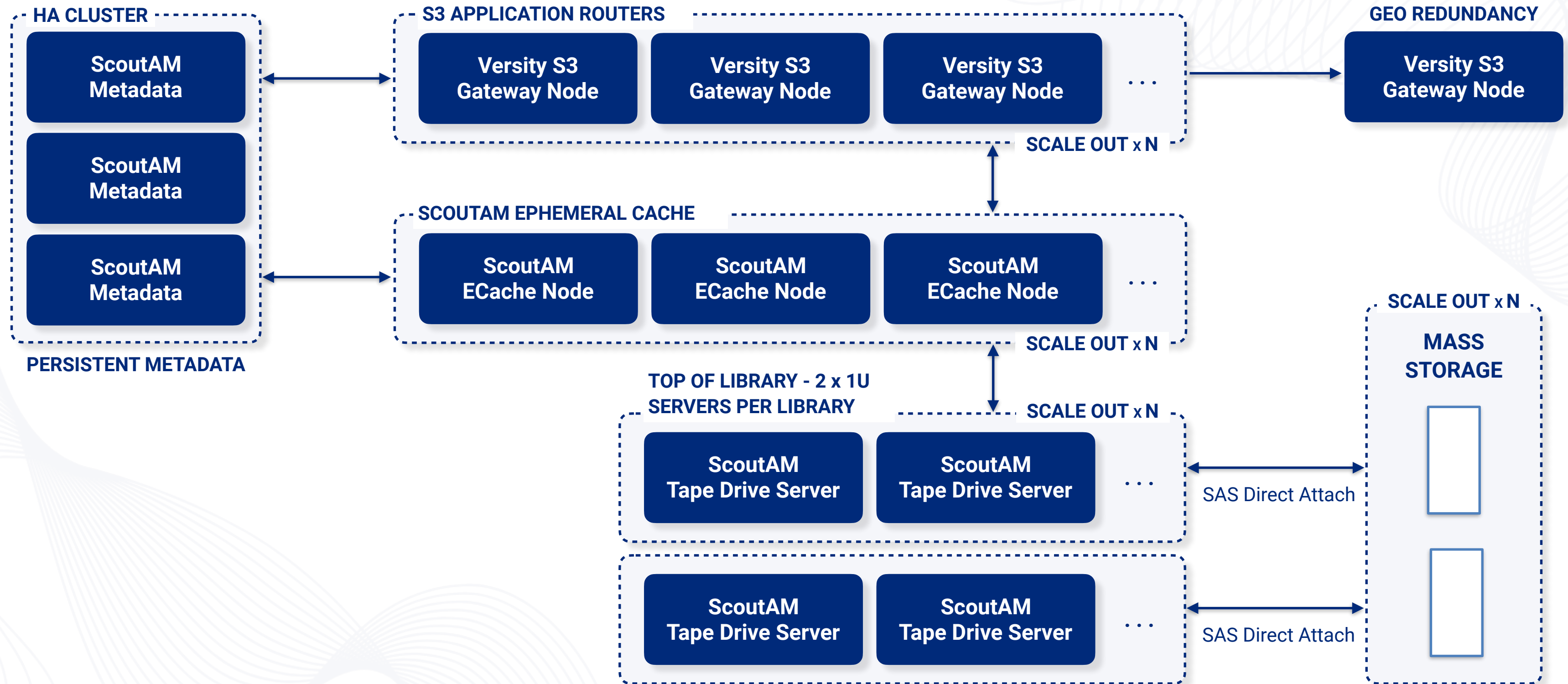


SPECTRA
CUBE

Versity Reference Architecture



Versity Hyperscale Solution



RapidFS: Next Generation Filesystem

Open-Source High Performance Distributed NVMe Filesystem

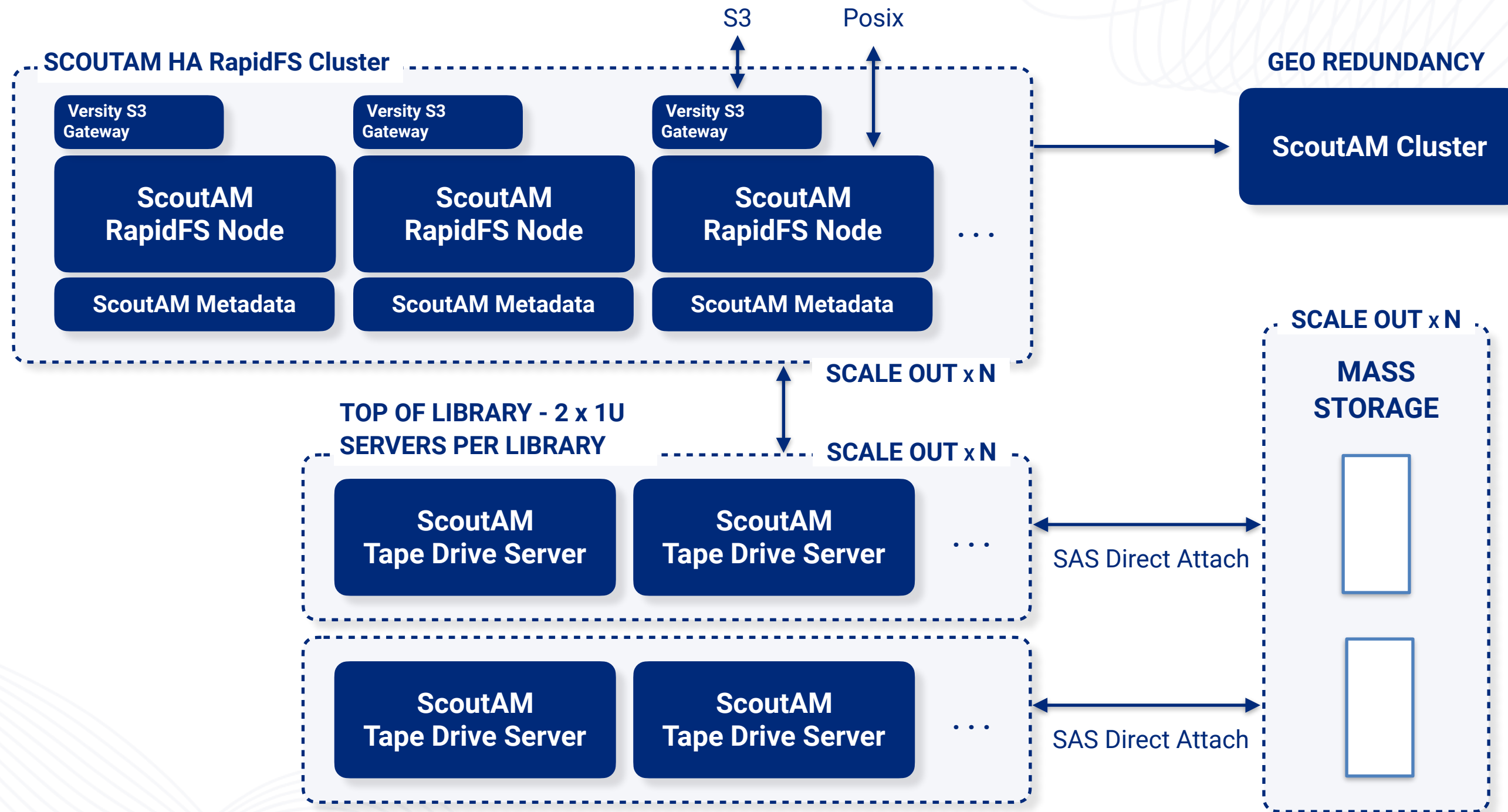
High performance distributed filesystem using commodity hardware

- Standard Ethernet (100 GbE+) and RDMA networks
- Internal NVMe devices with parity across nodes
- Native file or Object clients
- Consistent caches across nodes
 - Fine-grained locking
- Intelligent Block Indexing

Motivation

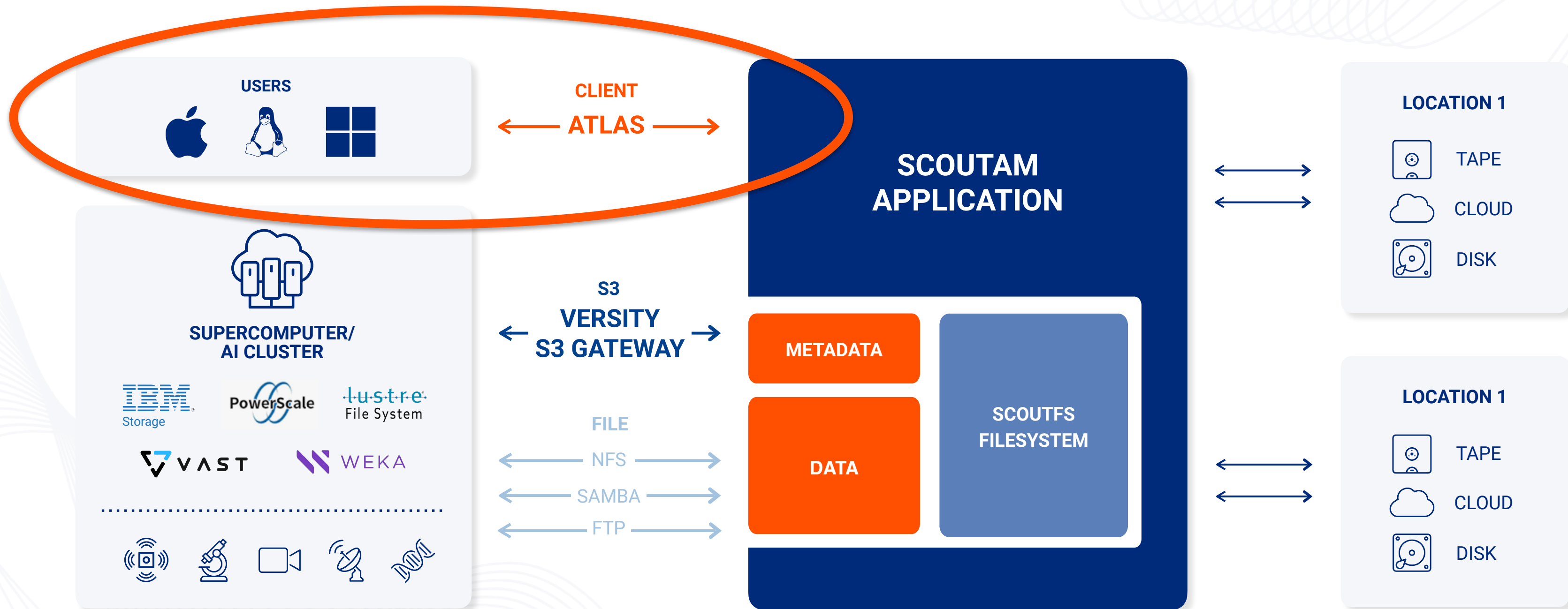
- Truly open-source distributed NVMe platform in upstream kernel
- Eliminate shared disk controllers and SAN
- NVMe in front of new tape generations (1+ GB/sec/drive)

Versity Hyperscale Solution - RapidFS



Versity: Archive with Intelligence

A Modern Data Management Platform



Atlas

Visibility, search, and self-service control - delivered natively within ScoutAM.

UNLOCK

Automatic Metadata and State

- Live archive status and information
- Automatic Metadata
- User Defined Metadata
- All from single source of truth

SEARCH

Find what matters, across billions of files

- FAST Indexed queries
 - Across billions of objects
- Full SQL query support
- No Duplicative infrastructure

EXECUTE

Act on what you find, without leaving the archive

- Initiate archive operations directly from search results
- RBAC restricts access to authorized only
- Admins retain full visibility and control

UNLOCK

Rich metadata extraction and tagging

AUTOMATIC EXTRACTION

What ScoutAM extracts for free

- **Scientific formats:** NetCDF, FITS, DICOM, GRIB, GRIB2, BUFR, HDF5 and more - parsed at ingest. Climate variables, dimensions, time ranges. Medical imaging fields.
- **General formats:** Image EXIF and GPS. Document metadata. All searchable without any user effort.

USER-DEFINED TAGS

Only what you can know

- Tag, classify, and enrich billions of files in one command.
- Create tags that reflect real workflows for detailed reporting and deeper insight.

sst_global_2025.nc

/users/adhya/sst_global_2025.nc

Metadata

Tags

File Info

Archive Info

application/x-netcdf · Extracted 2026-05-22 19:56

Calendar

gregorian

Dimensions

lat

180

time

90

Dim Lat

180

Dim Time

90

Conventions

CF-1.8

lon

360

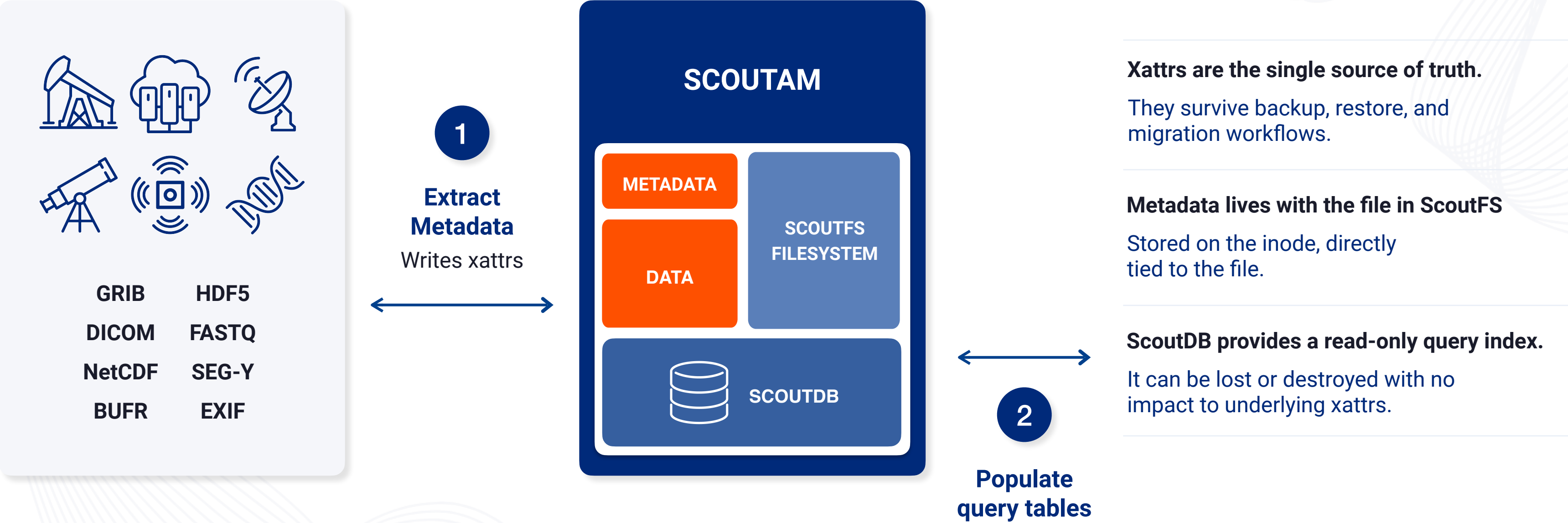
Dim Lon

360

History

Created 2026-05-22T23:52:13.711131+00:00

Extraction Pipeline



SEARCH

Find the right data in seconds.



Hundreds of billions of objects

Fast indexed queries across the entire archive, no matter the scale. **Full SQL support.**



No duplicate infrastructure

Search uses ScoutAM architecture, so no need for an external metadata database that must sync with storage



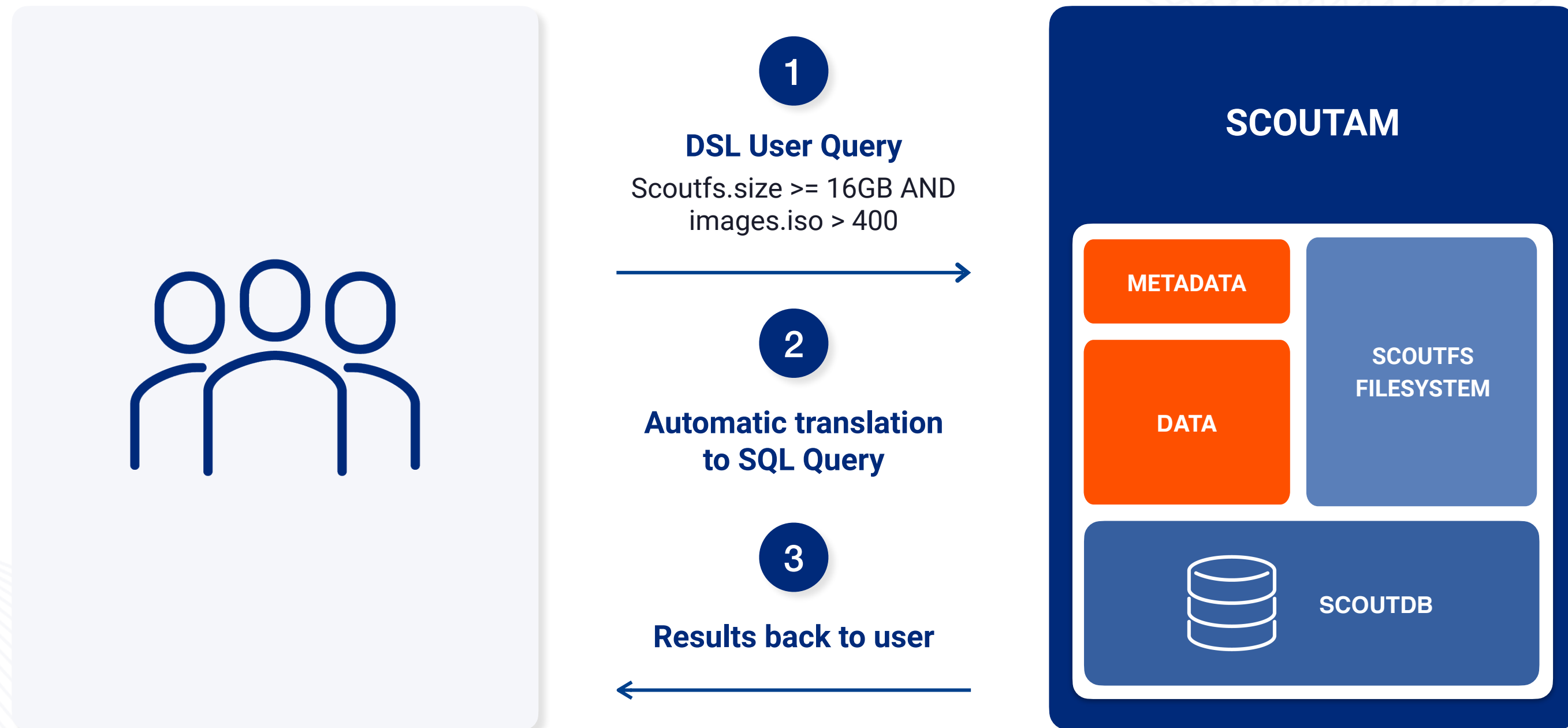
Single source of truth

Results reflect the same metadata plane as used by ScoutAM.

```
acli
[archive:/archive/users/nds] > search --tags images --query "posix.owner=525"
/archive/users/nds/images/IMG_6844.HEIC
/archive/users/nds/images/IMG_0020.jpeg
/archive/users/nds/images/IMG_0022.jpeg
/archive/users/nds/images/IMG_0024.jpeg
/archive/users/nds/images/IMG_5148.jpeg
/archive/users/nds/images/IMG_7161.jpeg
/archive/users/nds/images/IMG_7162.jpeg
/archive/users/nds/images/IMG_0030.jpeg
/archive/users/nds/images/IMG_6843.jpeg
/archive/users/nds/images/IMG_7156.jpeg
/archive/users/nds/images/IMG_6495.HEIC

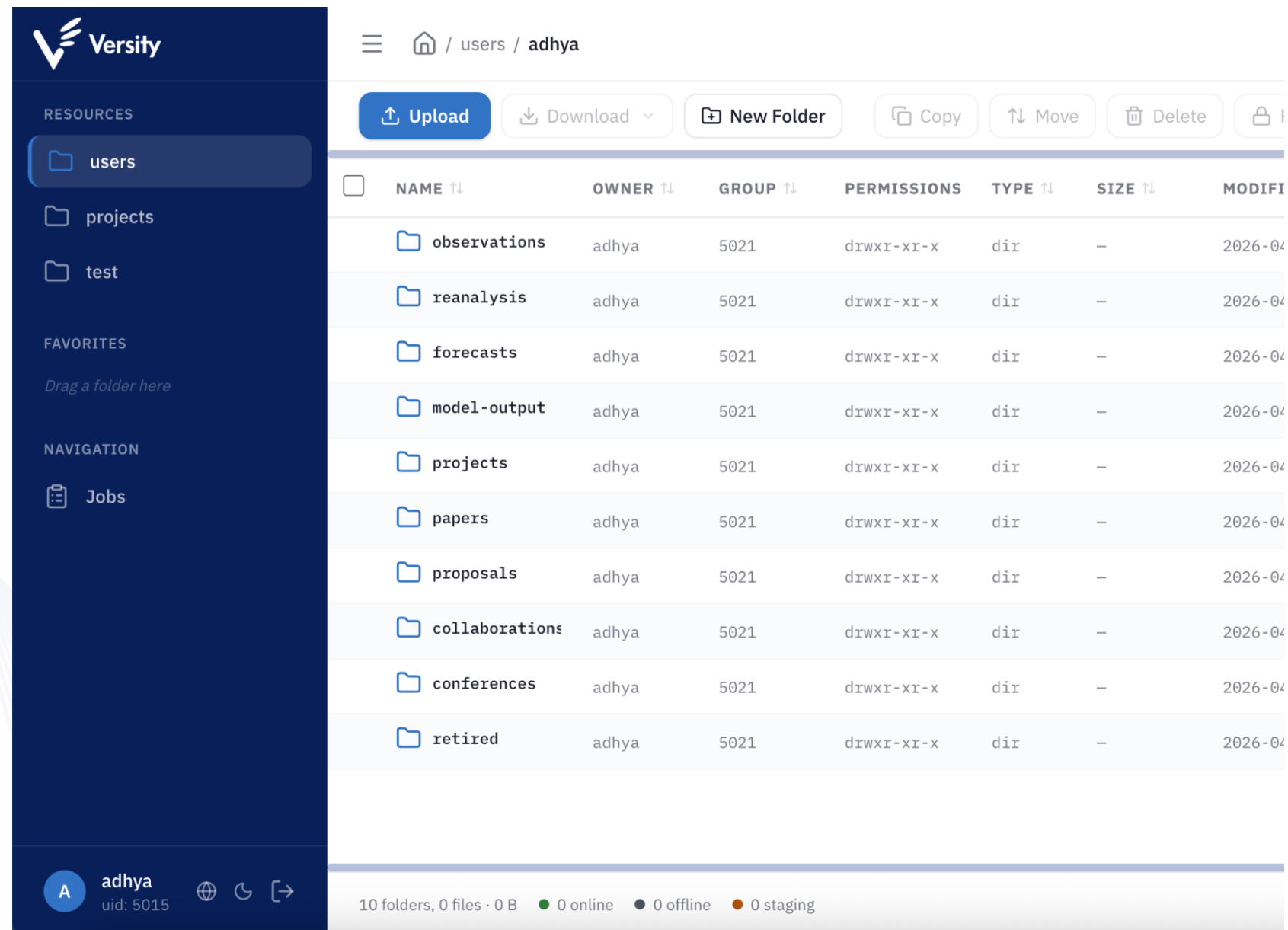
Found 11 matches
[archive:/archive/users/nds] > █
```

Search Pipeline



EXECUTE

Cross Platform Client



ATLAS EXPLORER

Moves data

Drag, drop, stage, release. Operates the way users already work.

See archive status

Online, offline, archived, staging viewable at a glance, on every file.

Batch Operations from GUI or CLI

Select thousands of files, stage in one click or pipe to `accli` for scripted ops.

macOS | Linux | Windows

Native client for every desk in your organization.

EXECUTE

Powerful CLI

ACLI

Runs on Linux, Mac, Windows

One binary. Every operating system your team touches.

Full query capability

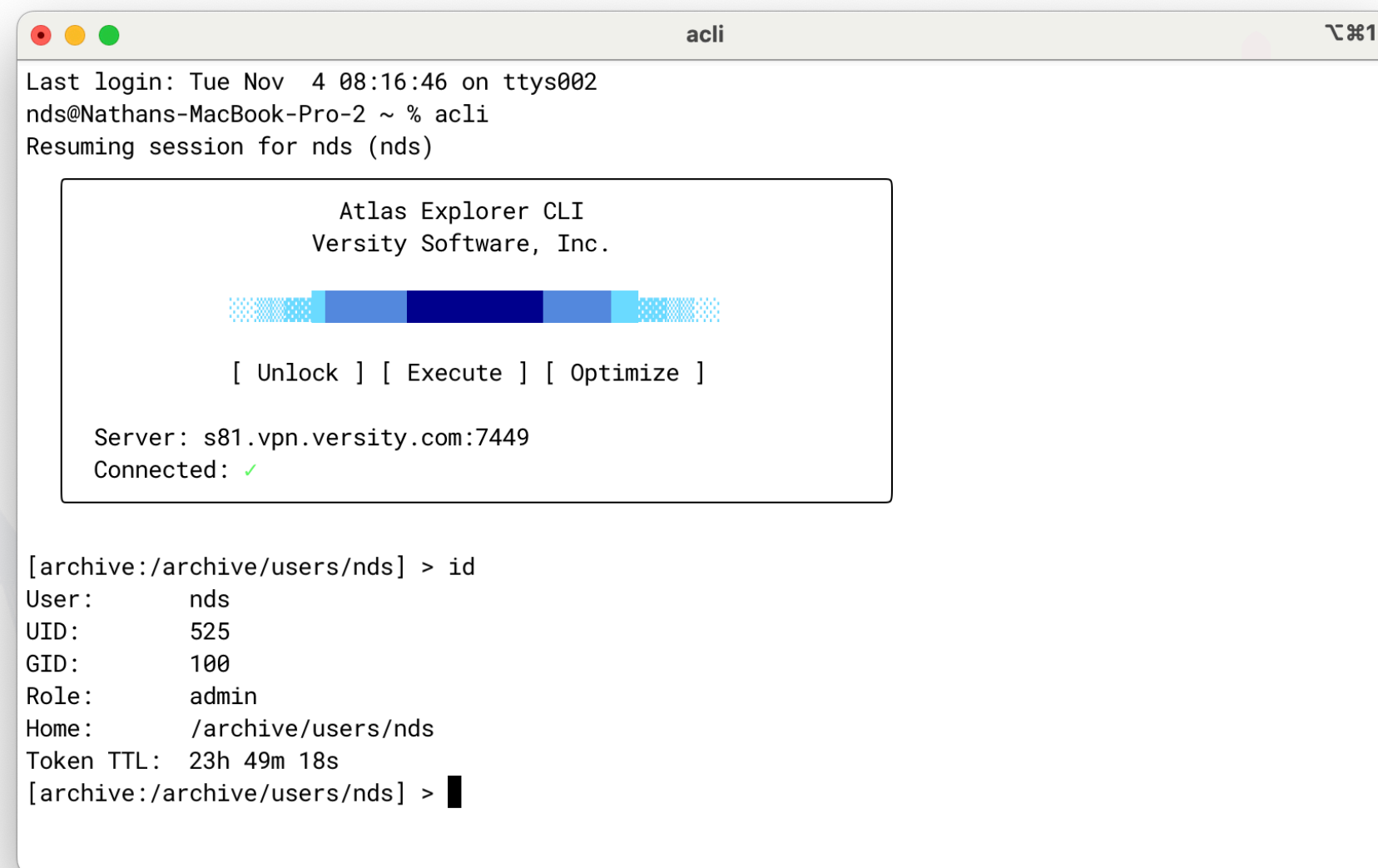
SQL across hundreds of billions of objects.

Script complex tasks

Tag a million files by project in one go. Migrate an entire directory to tape with a single command.

JSON & CSV out of the box

Human-readable for the terminal. Machine-parseable for your pipelines.



```
acli
Last login: Tue Nov  4 08:16:46 on ttys002
nds@Nathans-MacBook-Pro-2 ~ % acli
Resuming session for nds (nds)

Atlas Explorer CLI
Versity Software, Inc.

[ Unlock ] [ Execute ] [ Optimize ]

Server: s81.vpn.versity.com:7449
Connected: ✓

[archive:/archive/users/nds] > id
User:      nds
UID:       525
GID:       100
Role:      admin
Home:      /archive/users/nds
Token TTL: 23h 49m 18s
[archive:/archive/users/nds] > █
```

EXECUTE

Your archive, plus everything around it.

```
acli
[archive:/archive/users/nds] > resources
NAME          TYPE      SLOTS  NODES          DESCRIPTION          STATUS
-----
/scratch      posix     32/32   s81.vpn.versity.com  Lustre scratch filesystem used for HP...  available
/archive      scoutfs   16/16   s81.vpn.versity.com  Longterm data archive  available
[archive:/archive/users/nds] > cd /scratch/nds
[scratch:/scratch/nds] > ls
-rw-r--r-- 0 0      0      33554432 Oct 20 19:04 -o--- 0 file.8
-rw-r--r-- 0 0      0      163840 Oct 20 20:30 -o--- 0 file.9
-rw-r--r-- 0 525    100    2784 Oct 24 21:22 -o--- 0 pres_temp_4D.nc
-rw-r--r-- 0 525    100    3187 Oct 27 20:33 -o--- 0 stats.pdf
drwxr-xr-x 0 525    100    4096 Nov 02 19:00 ----- 0 test
[scratch:/scratch/nds] >
```

INTEGRATE WITH THIRD-PARTY STORAGE

Connect what you already have.

Lustre, IBM Spectrum Scale, VAST, PowerScale, and other storage platforms become Atlas-managed resources.

One interface for every tier.

Browse, search, and move data across primary and archive storage from the same GUI or CLI.

No External Gateways. No Movers.

Native lifecycle transitions across flash, object, and tape. No extra middleware. No vendor lock-in.

Self-service at **exabyte scale**

Built for the People Who Actually Use the Archive

RESEARCHERS



Easily your data across billions of files. Tag it in the language of your science. Retrieve it the moment you need it.

- Search your entire archival collection using the attributes your team actually uses
- Self-service stage and batch stage from search results
- Clear status and retrieval progress at every step

STORAGE ADMINS



Replace fragile scripts with policy-driven automation. Empower researchers within the guardrails you set.

- Real-time operational visibility across jobs, placement, and copy state
- Role-based permissions that are explicit, auditable, and easy to manage
- Policy-driven archive workflows without third-party movers

IT LEADERSHIP



Metadata intelligence, search, tagging, and self-service archive - delivered natively with the archive platform.

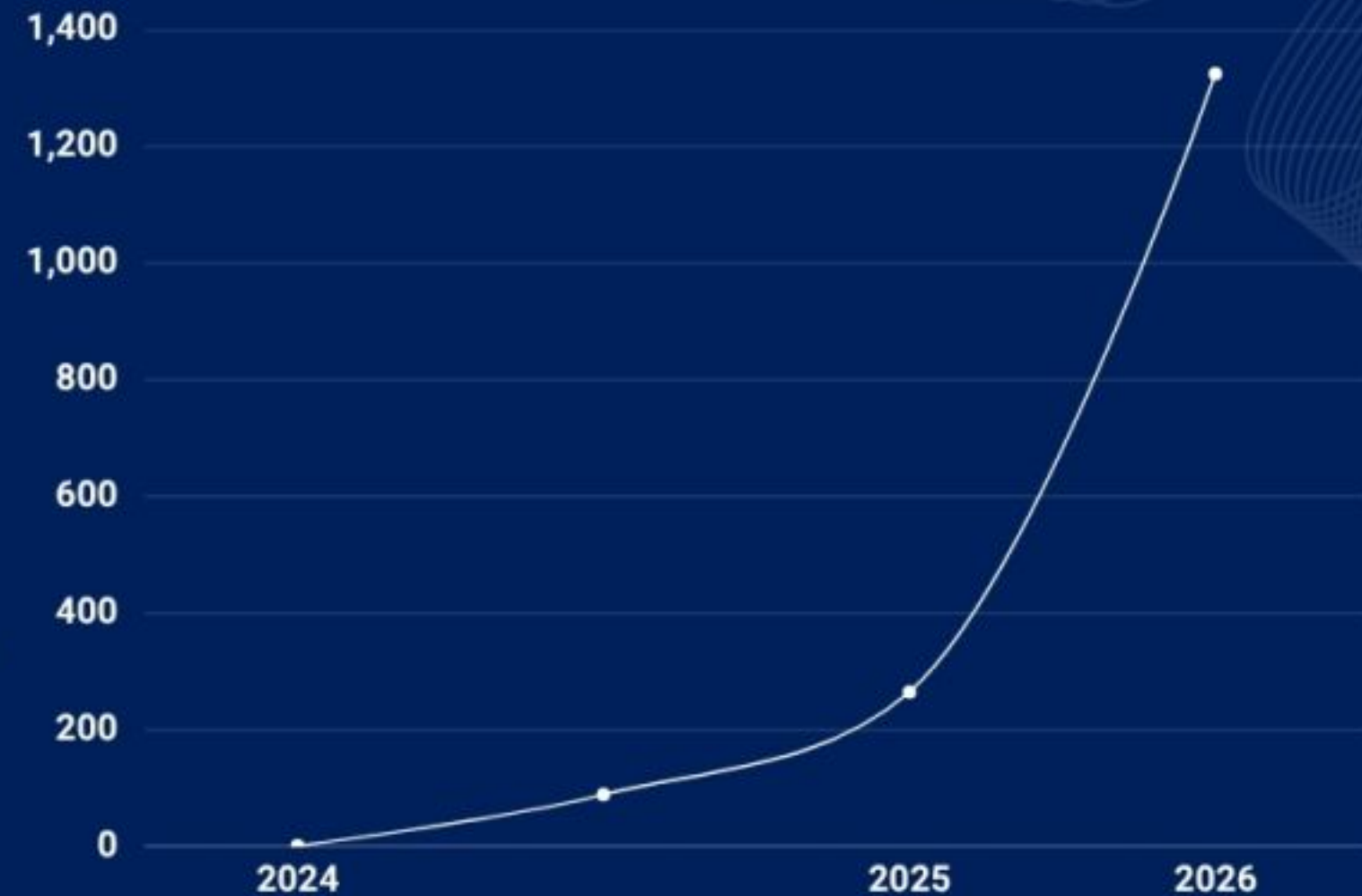
- Cost analytics reporting you can bring to a budget review
- Bundled with ScoutAM - no additional infrastructure or third-party dependencies or standalone licenses
- Open formats that eliminate long-term vendor lock-in risk

Versity S3 Gateway

This growth speaks for itself.

Organizations are moving to the Versity S3 Gateway for open, scalable object-to-archive workflows that go beyond traditional gateway models.

Github Stars
★



Versity S3 Gateway

A High-Performance S3 to Storage System tool

Scalable

Deployed in a cluster

Each instance is stateless,
requests are independently
handled by any instance

Performance linearly scales with
each instance

Minimal CPU and memory
overhead

Modular Backend Support

ScoutAM/ScoutFS

General Filesystem

S3 Proxy

Computational Devices (NVMe)

Azure

Memory

Open Source

Apache 2.0 license

Available on Github

Contributions welcome

Easy to use and deploy

Versity S3 Gateway

What's New

- **Web-based management UI** - browser interface for managing buckets, objects, and users without CLI tools
- **Plugin backend system** - connect any external storage system to the gateway at runtime without modifying the core codebase
- **Virtual host-style bucket addressing** - access buckets as subdomains (e.g., bucket.gateway.example.com) in addition to path-style URLs
- **Helm chart for Kubernetes** - deploy versitygw on Kubernetes via a published OCI Helm chart, with support for persistence, HA, web UI, and admin API + community COSI driver
- **Public bucket access** - serve unauthenticated reads from designated buckets
- **Signed chunked encoding with trailer** - improved compatibility with AWS SDK clients that use this upload mode
- **Multi-address listener** - bind S3, admin, and web UI ports to different network addresses independently
- **IPv6 support** - listen on IPv6 sockets

A lot of bug fixes, performance improvements, and compatibility testing

Versity S3 Gateway GUI

Versity

ADMIN

Dashboard

Users

Buckets

Explorer

RESOURCES

Documentation

Bug Reports

Releases

GitHub

admin1
Admin

Sign Out

VersityGW Dashboard

Total Users
1

Total Buckets
5

System Status
● Connected

Quick Actions

Manage Users
Create, edit, and delete user accounts

Manage Buckets
View and manage bucket ownership

Connection Info

Endpointhttp://127.0.0.1:7070

Regionus-east-1

Access Keyadmin1

Status● Authenticated

Recent Users

View all

Access Key	Role	Project ID	User ID	Group ID
admin1	Admin	0	0	0

Versity

ADMIN

Dashboard

Users

Buckets

Explorer

RESOURCES

Documentation

Bug Reports

Releases

GitHub

admin1
Admin

Sign Out

VersityGW Buckets

Buckets

+ Create Bucket

Note: Create buckets, view existing buckets, and transfer ownership between users.

Search buckets...

All Owners

Bucket Name	Owner	Actions
object-lock-bucket-1	admin1	↔ Owner
test-bucket-1	user1	↔ Owner
test-bucket-2	user1	↔ Owner
test-bucket-3	user1	↔ Owner
versioned-bucket-1	admin1	↔ Owner

Shared Library Plugin

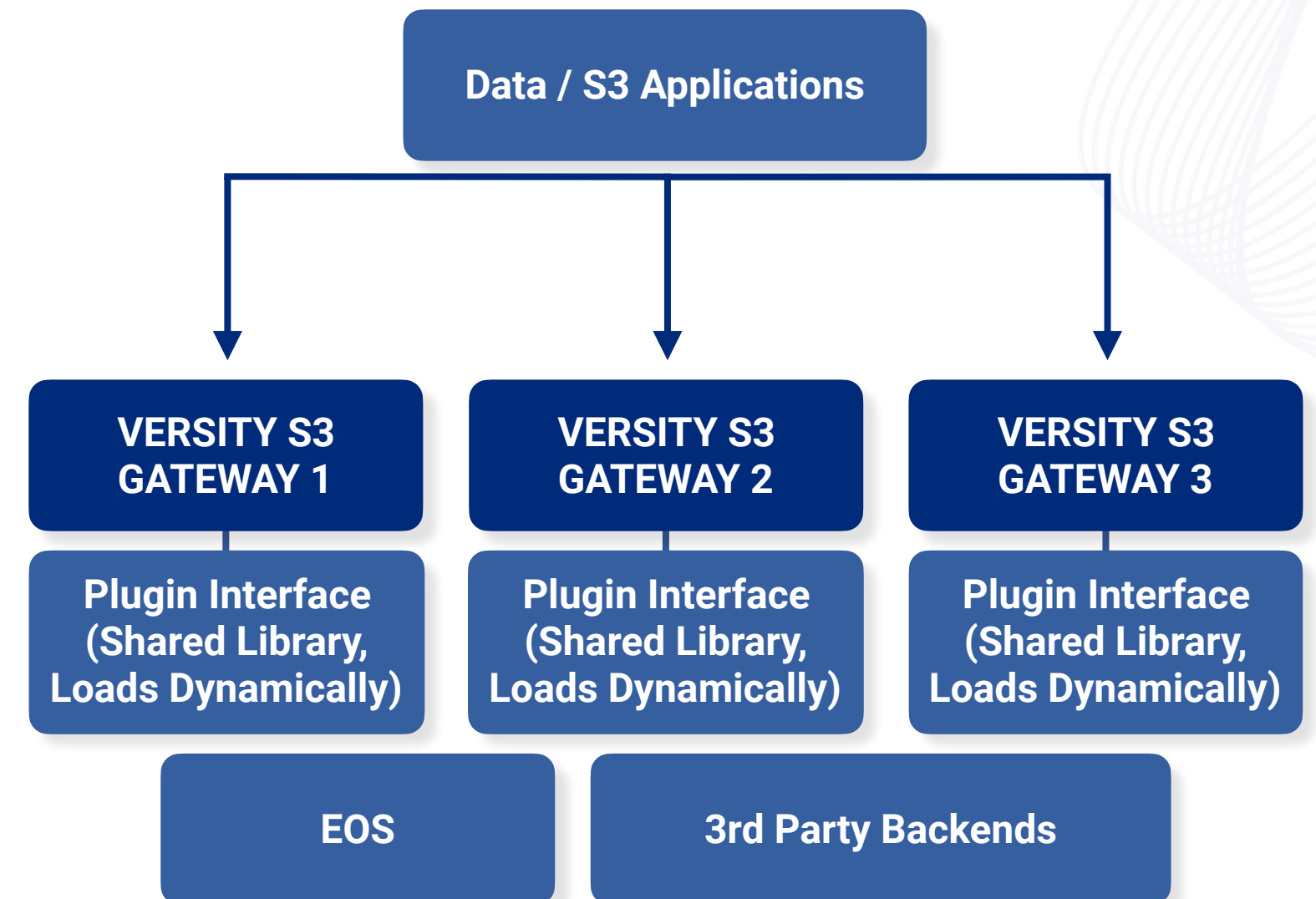
New Plugin Framework: Any backend can be loaded as a shared library

- Any backend can be loaded as a shared library
- Backends written independently, dynamically loaded
- Developers focus only on storage logic; gateway handles protocol, auth, metrics
- Skips POSIX layer → faster, more direct integration
- Enables rapid adoption of new storage systems without forking or patching

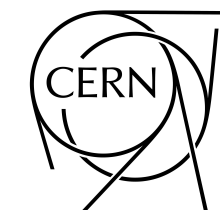
EOS Plugin

- CERN implemented the **EOS backend** (“eoss3”) using the plugin model
- Maps S3 operations to EOS APIs (gRPC + HTTP) for metadata and data handling
- Provides S3-compatible access to EOS, CERN’s large-scale distributed storage for high-energy physics
- Demonstrates the model for other institutions to follow when adding their own storage backends

Versity S3 Gateway Shared Library Plugin Architecture



Community Contribution - Thank You



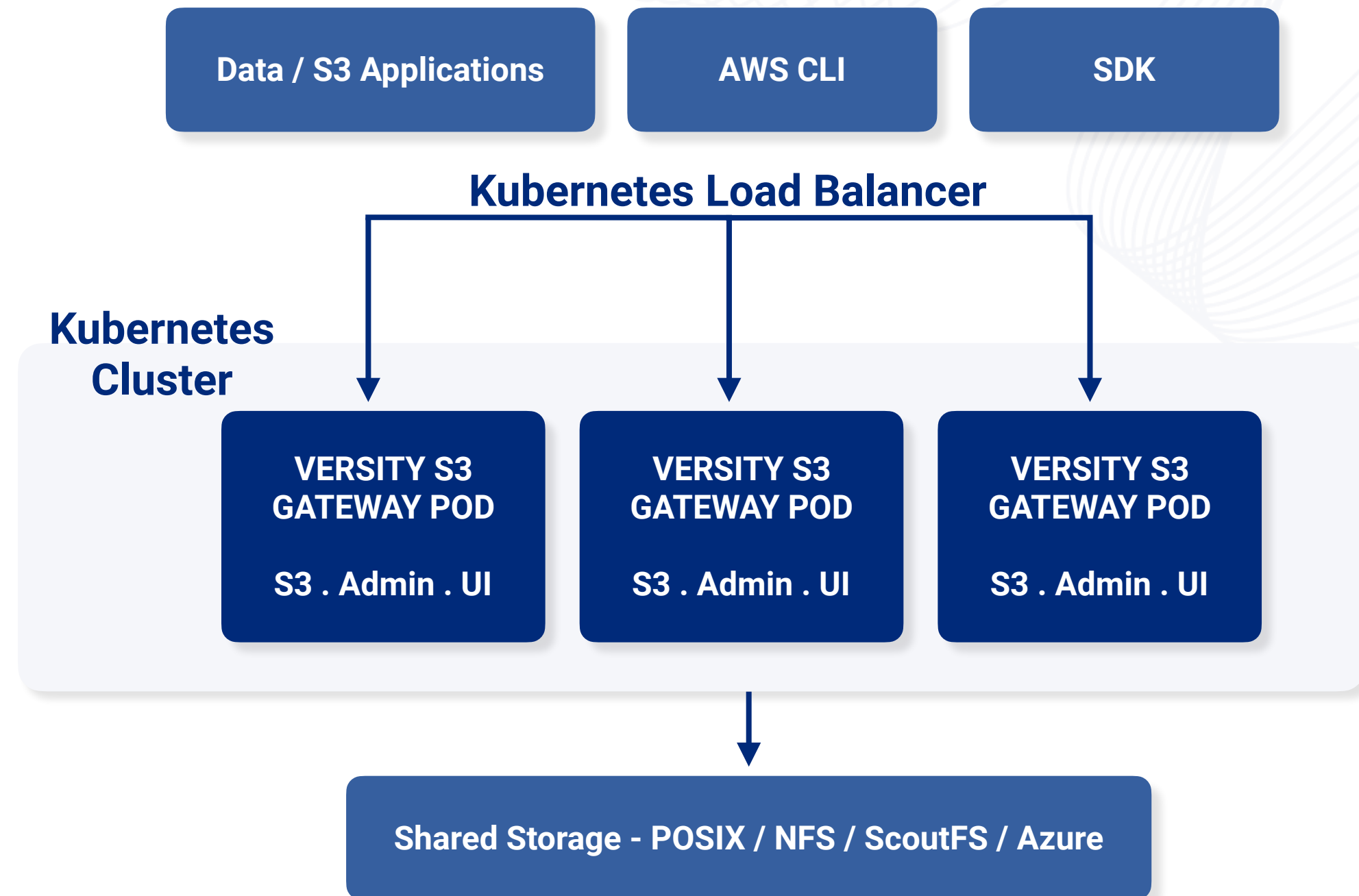
VersityGW on Kubernetes

Official Helm Chart that packages Versity S3 Gateway for Kubernetes - the standard way to install production software on any Kubernetes Cluster

- One command replaces 5–6 hand-written config files
- Scales horizontally - add replicas as demand grows
- Stateless design means any pod serves any request
- Fits existing Kubernetes, CI/CD, and GitOps workflows
- Cosign-signed chart for supply chain security

\$ helm install versitygw oci://ghcr.io/versity/versitygw/charts/versitygw

How the Versity S3 Gateway run on Kubernetes



Thank You

sales@versity.com



@versitysoftware