VAST Data

Matinée Technologique

bertrand.ounanian@vastdata.com

About VAST Data

Building Infrastructure For HPC & AI Discovery

$100M+ Annual Recurring Revenue

Elite "Centaur" Status
2.5x Y/Y Growth

$3.7B
Series D | April 2021

100%
Recommended

Gartner peerinsights™
Source: VAST Customer Surveys

Cash Flow Positive

Founded Late 2015

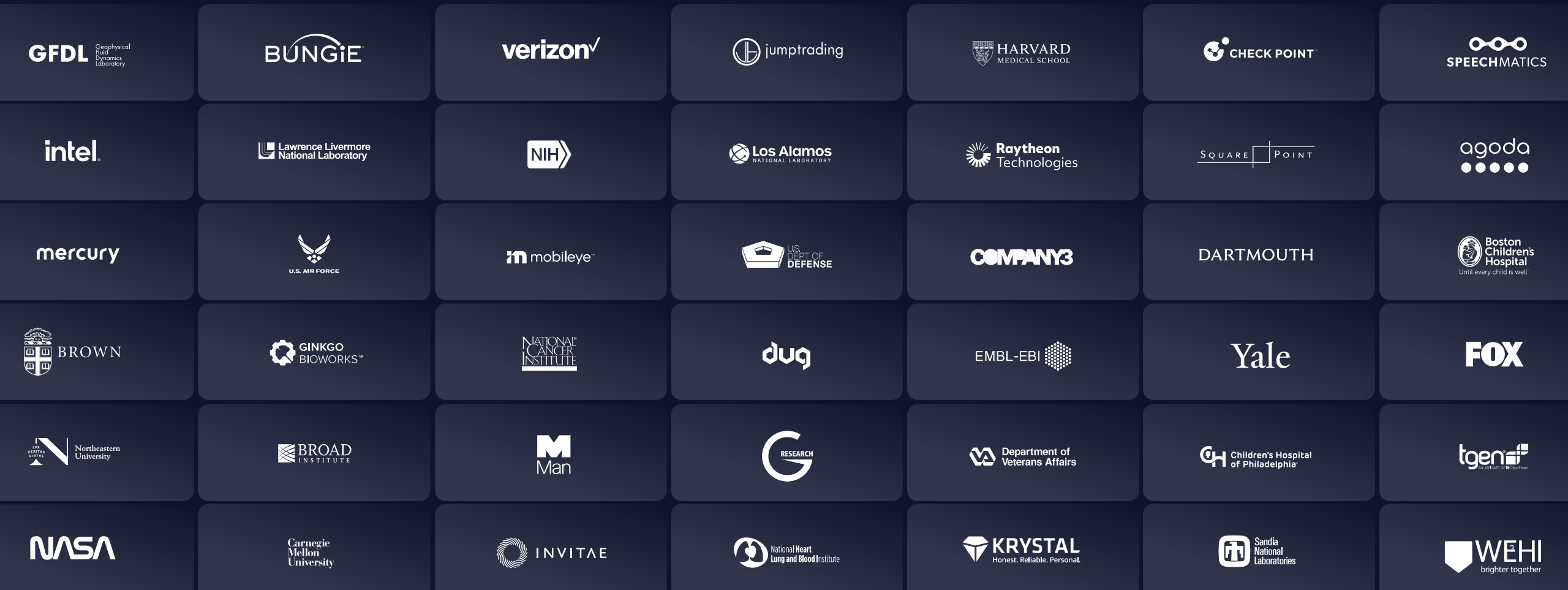
Over 700 People Globally

Product Of The Year

Goldman Sachs · 83NORTH · GENERAL ATLANTIC · NVIDIA · TIGERGLOBAL · Next47 · NORWEST · DELL Technologies CAPITAL

Over Eight Exabytes of VAST Data

The Fastest Growing In Infrastructure History



No More Tiers

Breaking Tradeoffs To Make Storage Simple

- Extinction-level Event for The HDD
- An End To 30 Years of Storage Tiering
- Unleash Big Data & AI Insights

Our Advantage

A Revolutionary Distributed System Architecture

UNIVERSAL STORAGE

STARTING FROM A RENAISSANCE IN HARDWARE

NVME OVER FABRICS FOR DISAGGREGATION

The latency of DAS, over switched commodity networks.

QLC FLASH FOR COST SAVINGS

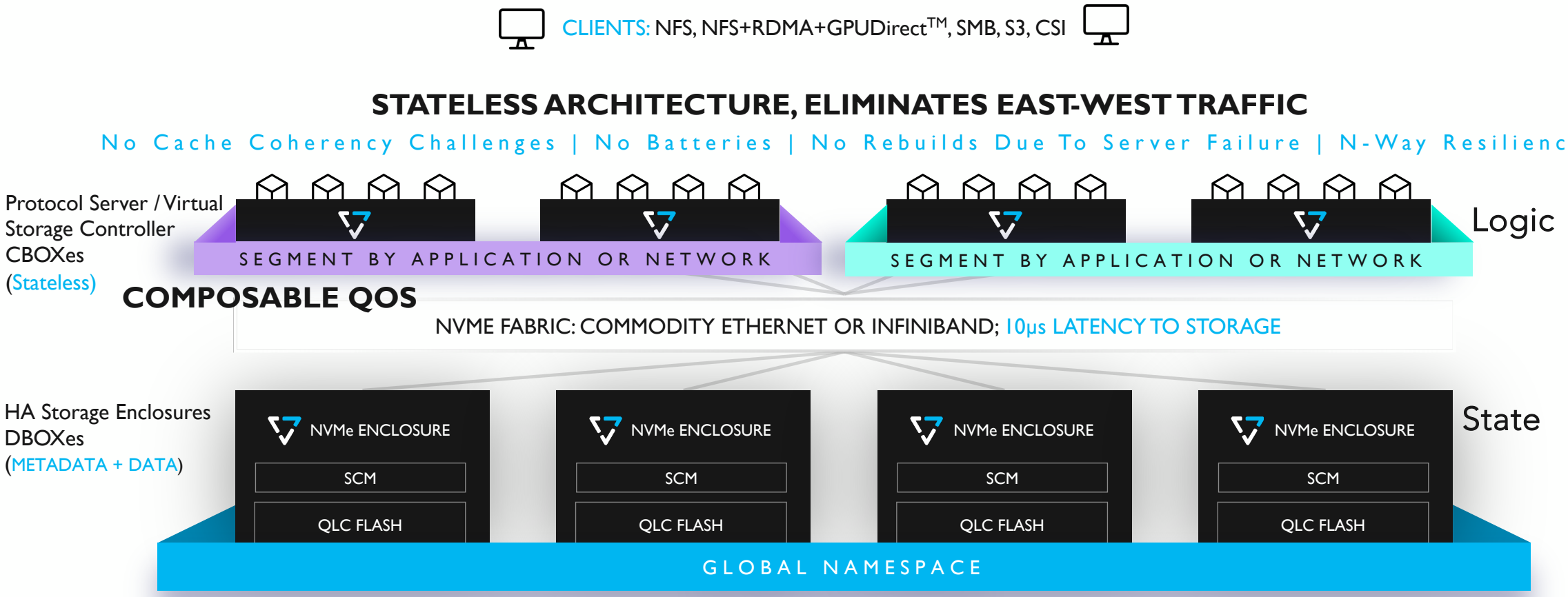
Low-cost, lower-endurance hyperscale flash.

STORAGE CLASS MEMORY LONGEVITY & EFFICIENCY

Enables write shaping to QLC and rich metadata.

DASE: Disaggregated, Shared-Everything Architecture

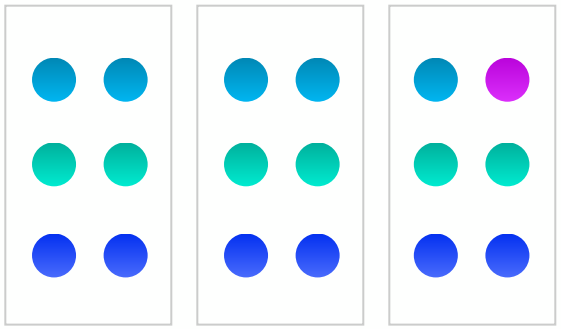
THE STORAGE ARCHITECTURE OF THE FUTURE - IDC, 2020



STORAGE EFFICIENCY

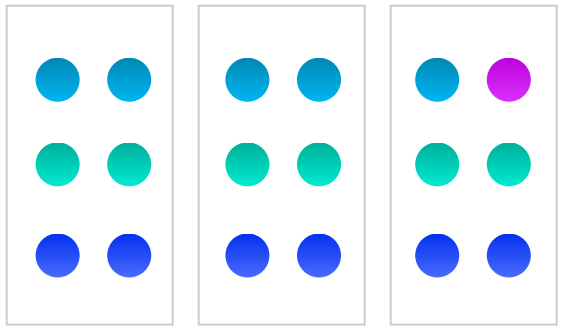
SIMILARITY IS GAME-CHANGING

COMPRESSION



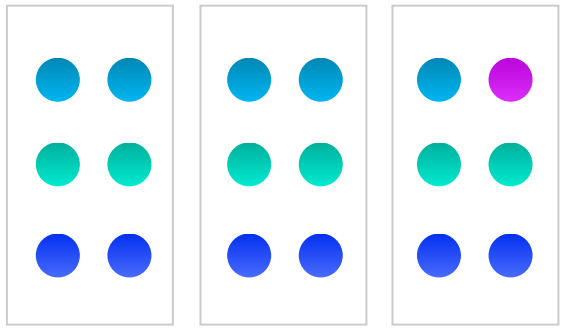
Fine Grained, But Local

DEDUPLICATION



Global, But Coarse

VAST DATA SIMILARITY REDUCTION



Global & Fine Grained

EXAMPLE SAVINGS FROM SIMILARITY

3:1	Pre-Reduced Backups	**3:1**	Pre-Compressed Log Files	**2:1**	Life Science Data	**3:1**	HPC Data	**3:1**	Animation Data	**8:1**	Uncompressed Time-series Data

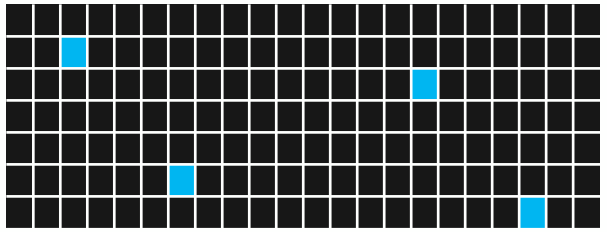
STORAGE EFFICIENCY

TRANSFORMING THE CALCULUS OF FLASH OWNERSHIP

VAST DATA FORESIGHT

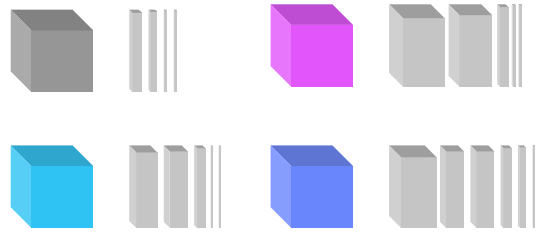
QLC + 10-Year Longevity

LOCALLY DECODABLE ERASURE CODES



2.7% Overhead Data Protection

SIMILARITY-BASED DATA REDUCTION



Global, Fine-Grained Compression

VAST Data Platform Features

Data Science Pushdowns
Apache Spark & Trino Predicate Pushdown

The VAST Data Catalog
Query unstructured data at any scale, in real-time

Similarity-based Data Reduction
Unprecedented storage efficiency

VAST DB API
Transact rows natively to the VAST Database

No-Overhead Snapshots
A write-in-free-space approach to non-disruptive snaps

Locally-decodable Data Protection
Unrivaled resilience, only 2.7% overhead

NFS V3.0, 4.1
w/ RDMA, GPUDirect

Native Replication
N-1 and 1-N Topologies, extremely low levels of RPO (15s)

Data Aware Compression
Floating Point Data, Integers

SMB 2.1 (Resilient SMB)
Stateful file services; stateless servers

Backup To Cloud/S3
Snapshot to an S3 endpoint of choice

Quality of Service
Logical QOS with min/max by path; CNode Pools

S3-Compatible API, W HTTPS
Also featuring object versioning, lock, etc.

Load Balancing
Internal DNS services

Quota Management
Directory, User & Group Quotas

Multi-Protocol Namespace
Easily traverse between file & object

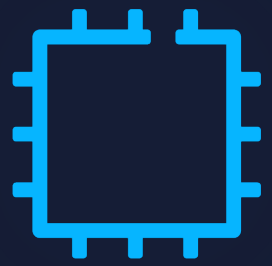
Uplink: The VAST Cloud Portal
Fleet management with ML-based capacity prediction

Automation Plugins
K8S CSI, Manilla and Cinder Plugins for Openstack

The Rise of a Data Platform

Three Elements of a Data Platform

Compute Engine



Runtime + Events + K8S

Database Engine

Streams + Analytics

Storage Engine

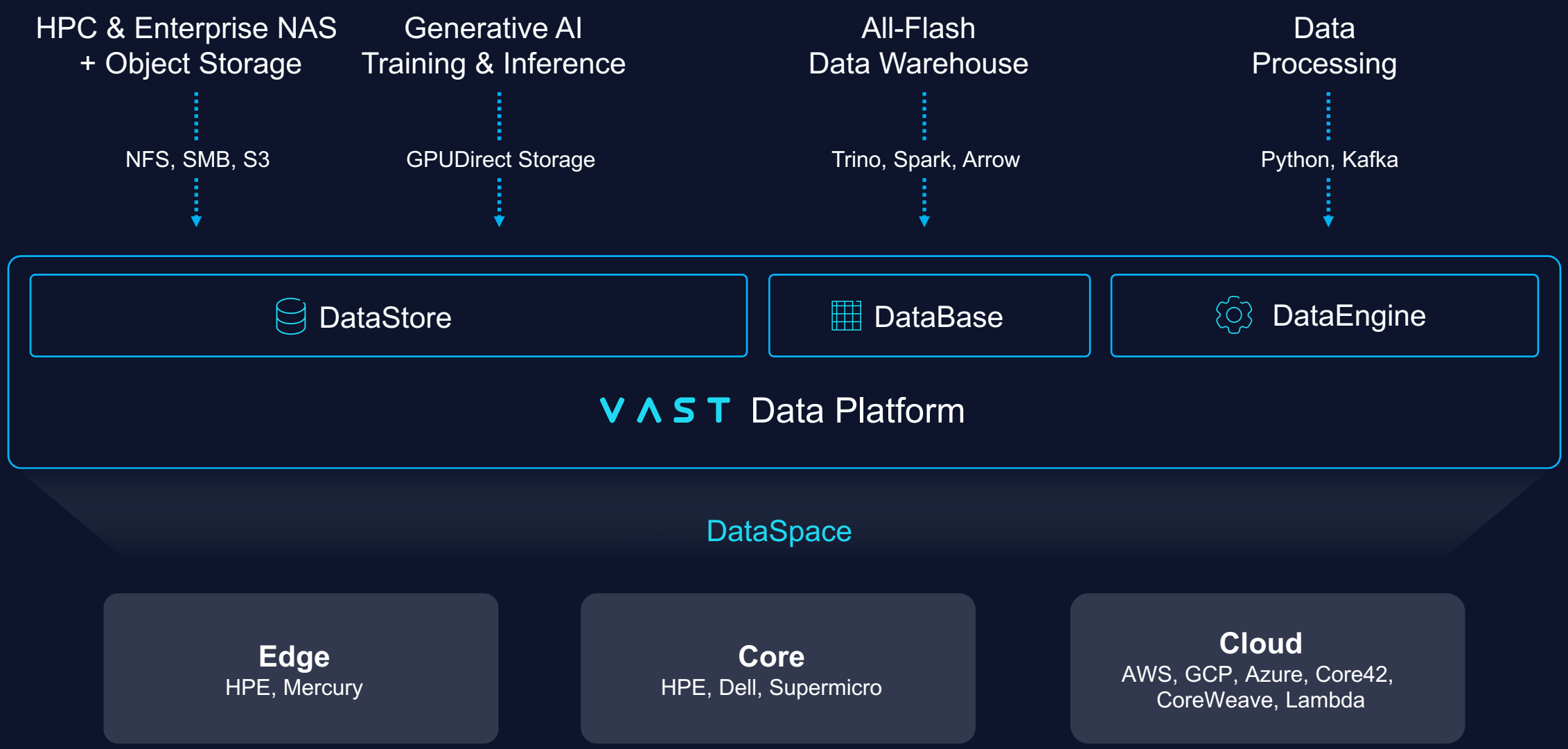
Raw Data + Data Protection

VAST

Data **Platform**

DataSpace

Edge Core Cloud


HPC & Enterprise NAS
+ Object Storage
Generative AI
Training & Inference
All-Flash
Data Warehouse
Data
Processing
NFS, SMB, S3
GPUDirect Storage
Trino, Spark, Arrow
Python, Kafka
DataStore
DataBase
DataEngine
VAST Data Platform
DataSpace
Edge
HPE, Mercury
Core
HPE, Dell, Supermicro
Cloud
AWS, GCP, Azure, Core42,
CoreWeave, Lambda

Case Studies

The Lustre Problem

We did a Lustre upgrade last year that actually broke a couple of our applications and it was one of those things where it's like, man, it's like weird. You can't reproduce it.

TACC
Tommy Minyard, Ph.D.
Director of Advanced Computing Systems
Texas Advanced Computing Center (TACC)

Lawrence Livermore National Laboratory

“VAST Data is very interesting to us… it’s used for enterprise NAS & in place of Parallel File Systems.

Success Story

- Support for /home, /scratch, /training – mounted on 22 different clusters
- Rapid access to all data, support for the DOE’s hardest challenges
- Embarrassingly parallel architecture for at-scale computing
- 3:1 data reduction on petabytes of file data

VAST

merci