UNIVERSAL STORAGE FOR

DATA ANALYTICS

ALL-FLASH DATA LAKES AT ARCHIVE ECONOMICS

As the modern data science stack continues to evolve, the motivations that previously drove organizations to adopt Direct Attached Storage (DAS) systems such as Hadoop Distributed File System (HDFS) have evolved. Enterprises continue to embrace real-time computing tools such as Apache Spark, Trino, and Elasticsearch, shifting away from batch-oriented tools of the Hadoop era.

To that effect, storage must be now architected for delivering real-time performance for any query load, while accommodating for the continuous evolution of data science tool-sets and the inevitable data growth. But most file and object storage systems (new and legacy) fail to scale performance affordably for your current applications, and are not architected to scale beyond a few petabytes.

ENTER VAST DATA

With it's unique Disaggregated and Shared-Everything (DASE) architecture, VAST has re-thought every aspect of storage system design to break all the tradeoffs of DAS and shared-nothing architectures, delivering a system that provides unprecedented price-performance and exabyte scale. The result is a highly scalable and affordable all-flash, file and object platform that allows you to run petabyte and exabyte scale analysis at less than half of the cost of traditional all-flash solutions, while being many times faster.

VAST'S KEY BENEFITS

High Performance At Any Query Load
Massively-parallel architecture delivers concurrency and real-time responsiveness at scale

Best-In-Class Price-Performance
Get all-flash performance for all data operations and searches at 1/2 the cost of traditional flash solutions

Flexible & Future-Proof
Use your preferred tools– Spark, Vertica, Trino, ElasticSearch, and easily adopt new tools that emerge

Limitless Scale
Scale capacity to 100s of petabytes and beyond without requiring complex capacity upgrades

No Data Silos
Access the same data via S3 & NFS eliminating the need for specialized storage or multiple copies of data

Turnkey & Simple
Turnkey appliance - with no client sw dependency, managed by VAST co-pilots (at no additional cost)

CUSTOMER SPOTLIGHT

Agoda Unlocks Real-Time Insights And Query Performance At Scale

"Customer information and travel supplier data are our most vital assets and we need a data science platform that can easily and cost-effectively scale with our growth," said Idan Zalzberg, Chief Data Officer, at Agoda. "To help provide our customers with the best value for their travel needs, we require a high-performance big data solution to run our machine learning algorithms, that's also infinitely scalable to meet our future needs."

A SMARTER WAY TO POWER YOUR QUERY ENGINES

The proliferation of ad hoc and interactive query engines means customers need predictable, real time performance for supporting thousands of queries simultaneously. Query engines need to be able to randomly read across massive data sets, which imposes significant I/O requirements, making flash infrastructure critical. This combined with the surge in Python programming and GPU-accelerated computing, means customers need to evolve beyond HDFS and traditional object storage solutions. VAST Data's unique approach to data lake storage features several such enhancements as seen below.

	HDFS	VAST DATA
Supports Apache Toolsets	Yes	Yes: **VAST S3 works with S3A**
Optimized for Python	No	Yes: **NFS+RDMA for DAS Speed**
Optimized for GPU	No	Yes: **GPUDirect Storage Enabled**
Multi-Protocol Namespace	No	Yes: **NFS & S3 MultiProtocol**
Distributed Metadata	No	Yes: **Scales to 10,000 Machines**
Optimized for Small Files	No	Yes: **NFS & S3 At Any I/O Size**
Built for Hyperscale Flash	No	Yes: **Radical Flash Efficiency**
Global Data Compression	No	Yes: **VAST's Similarity Reduction**

BREAKING TRADEOFFS WITH UNIVERSAL STORAGE

Most storage technologies just repackage the 20-year old shared-nothing, scale-out architecture, which is not designed for flash and struggle to reliably scale beyond a few petabytes. VAST Data's Universal Storage is an enterprise grade file and object storage system, that has reimagined every aspect of what has become typical in storage system design with DASE, ultimately delivering superior scalability, resilience, and Quality of Service (QoS) at a radically lower TCO for your rapidly evolving data analytics applications.

Scale-Out Architecture
DASE allows users to scale the performance, independently from the capacity of their system, allowing you to scale to 100s of petabytes and TB/s within the same namespace.

Massively-Parallel Scalability
Get real-time responsiveness for all your data. With no east-west cluster traffic, DASE enables virtually unlimited linear, predictable scale. VAST systems in production regularly exceed 1000 GB/s.

Multi-Protocol Support
Enable access to the same data via S3 & NFS simultaneously, eliminating the need to create multiple copies of the same data. Simply write via S3 and read that same data back via NFS or vice-versa.

Dedicated Quality of Service (QoS)
Pool compute servers to provide dedicated QoS for batch, interactive, and machine learning jobs on the same namespace

Superior Data Reduction
VAST's unique similarity based data reduction provides the industry's highest level of data reduction, allowing you to get better storage efficiency.

No More Migration
Add enclosures and compute servers of multiple generations into a single cluster and namespace, eliminating the need for forklift upgrades needed or migration of data from the old to new clusters.

ANALYTICS USE-CASES POWERED BY UNIVERSAL STORAGE

Implement High-Throughput Data Ingestion

Ingest massive streams of data in real time, including log data, using tools such as Apache Kafka and Apache Nifi.

Integrate Data Warehouses

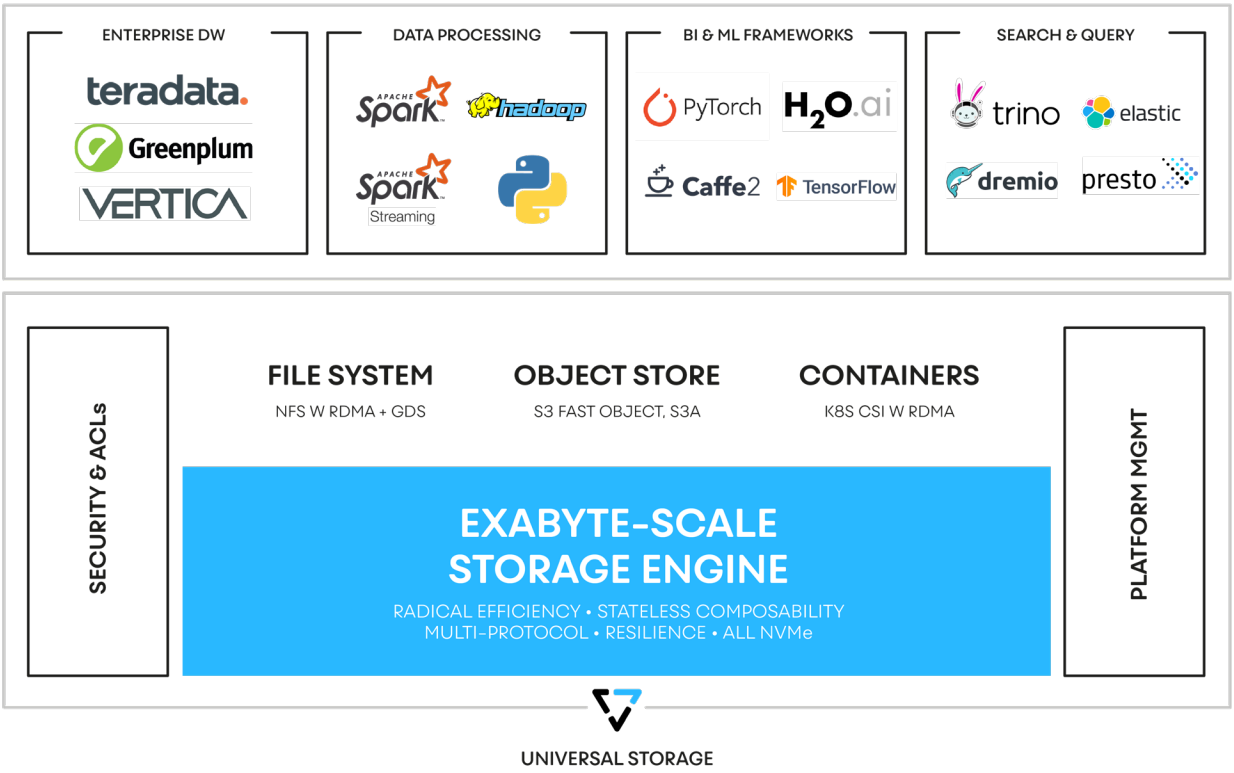
Unify data warehouses, data lakes, and other siloed data on a single platform, enabling your users to run fast, scalable analytics on your structured data as well as unstructured data.

Power High-Performance Interactive Analytics

Enable data scientist & data analysts to process data and run ad-hoc queries at scale, using their choice of analytic tools, including Apache Spark, ElasticSearch, Trino, Dremio, and Vertica.

Accelerate AI and ML adoption

Consolidate all your data on a single tier of flash, allowing data scientists to accelerate advanced data processing using their preferred ML framework, such as Apache Spark MLlib, TensorFlow, Pytorch, and Apache MXNet.



FEATURED PARTNERS

VERTICA

Deliver a unified analytics platform powered by an all-flash data lake to enable real-time access to all your data at scale.

trino

Accelerate time to insights with a flash-optimized VAST data lake to power Trino's distributed SQL query engine at scale.

dremio

Deliver a unified analytics platform bringing the performance and functionality of a data warehouse to a flash-powered VAST data lake.

splunk>

Future-proof your Splunk storage infrastructure with a flash-optimized VAST data lake and enable sub-millisecond read latency for all Splunk data at scale.