

White Paper

# Trino with VAST

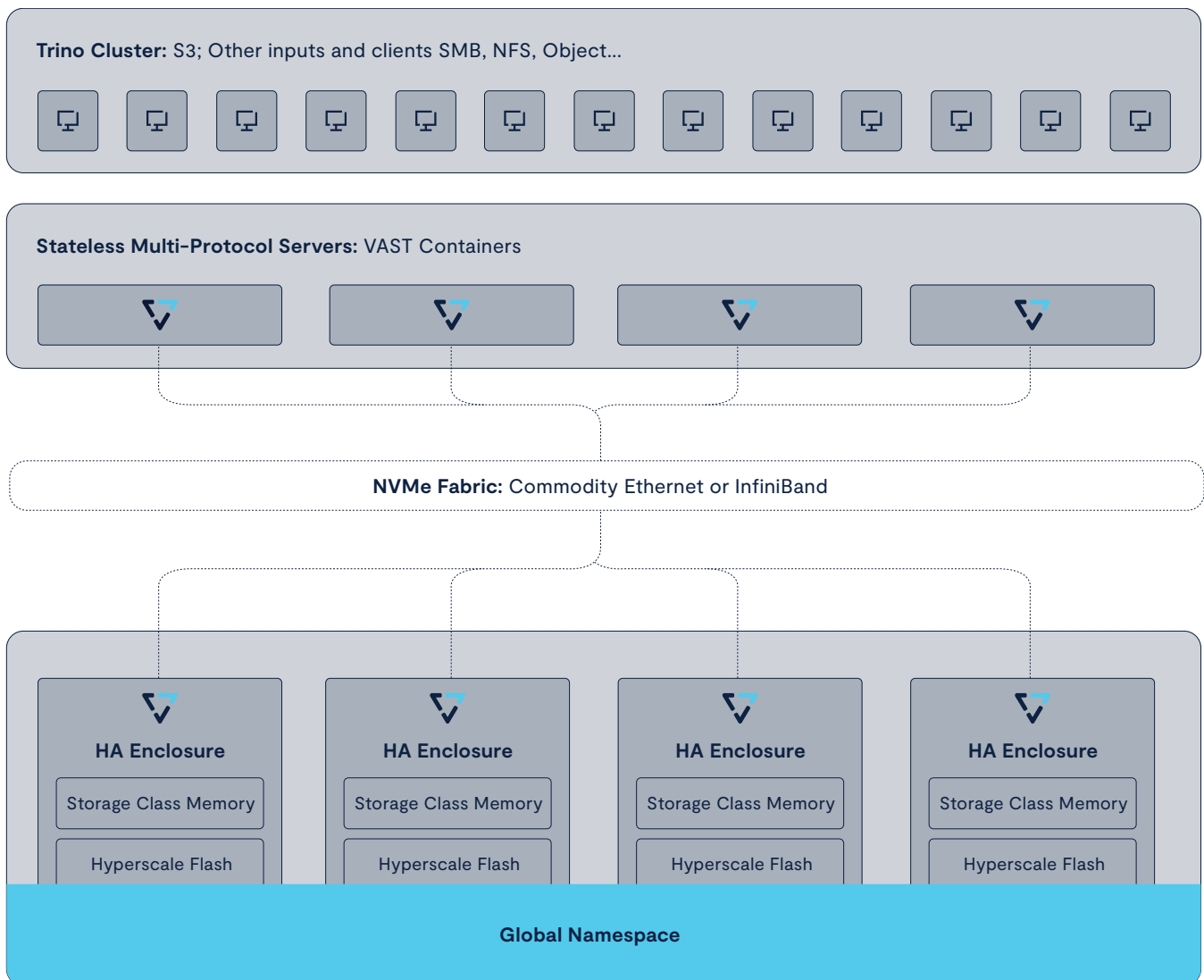
Trino is an advanced SQL-anywhere query engine capable of coordinating disparate data sources so that you may operate on them in a single namespace. Trino enables novel ways to generate value from data: Execute SQL queries on sensor data streaming through Kafka; run analytical queries spanning tables over multiple OLAP databases; tightly integrate your data warehouse with JSON files landing in your data lake in real time. Trino is intended to be both flexible and performant, well suited for ad hoc queries, real time use and reporting. With VAST housing your data, Trino becomes even more powerful.



# VAST

VAST's DASE architecture and mission to make all-flash affordable converge to create the ideal solution for data-lake/lake-house architectures, where Trino can be leveraged to yield data warehouse-like response times on petabyte scale storage. Current generations of object storage products in public and private clouds conceptualize S3 as an archive-class target. On VAST, it is a high-performance target with an enormous amount of enterprise capability at archival costs.

## Trino/VAST Architecture



## True Multi-Protocol Support

VAST is a multi-protocol storage system that exports a single namespace to all protocols: S3, NFS and SMB. This is important because data lakes work best when there are no limits to how data is written or read. Applications can work on the same data – read and write – over different protocols without creating copies or being forced to silo data. And Trino can work on your data no matter how it was written or where it is.

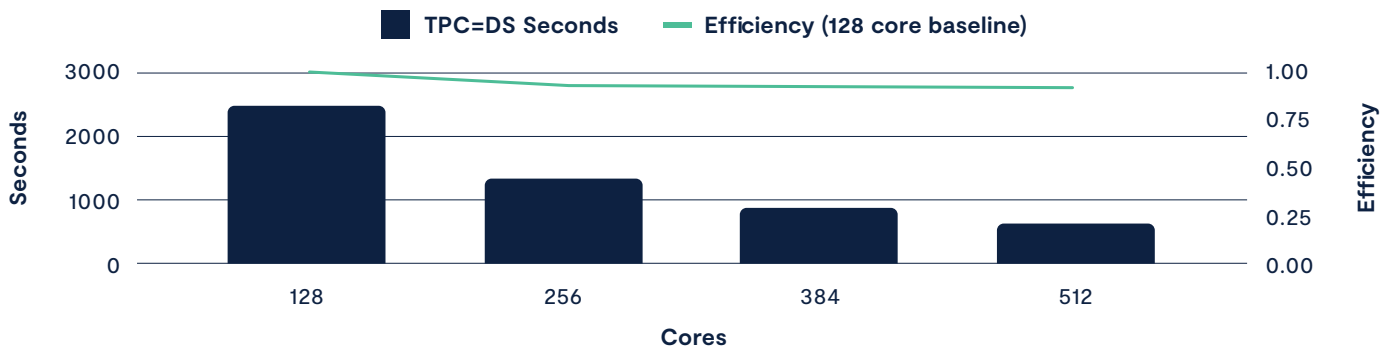
**Multi-Protocol Access:** NFS, NFS+RDMA+GPUDirect™, SMB, S3, K8S CSI



## Performance

VAST's DASE architecture can service over 100,000 S3 GETs per second and supply read throughputs of up to 35GB/s on an entry level system. Larger systems will achieve larger throughputs. VAST can keep your Trino cluster 100% utilized, without bottlenecking, supplying results in timeframes associated with data warehouses. Lakehouse architectures like Deltalake, Iceberg and Hudi – all supported in Trino – can truly live up to the “house” side of the name, adding support for access control, transactional database behaviors and a host of other advanced capabilities.

### Trino Scaling Efficiency on VAST



An entry-level 3x1 VAST system keeps 512 cores of Trino working without storage bottlenecks.  $Efficiency = ( [Time\ to\ execute\ on\ 128\ cores] / [n\_cores] ) / [time\ to\ execute\ on\ n\ cores]$ .

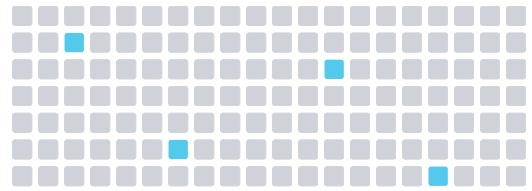
Scale Trino to large numbers of cores, even on a small VAST system. Optimize your query execution environment, don't worry about your storage.

## Scaling and Manageability

VAST's all-flash [DASE](#) architecture allows you to scale performance and capacity independently. Expansion and maintenance can be executed with minimal impact on operations. It realizes the dream of a load-following, scalable data-lake.

## Data Reduction and Resilience

VAST reduces data in three ways: deduplication, similarity deduplication and compression. We present a single data reduction domain. No matter how large the namespace, data is deduplicated and compressed across all of it. This is ideal for large storage systems that might contain redundant data, store a lot of compressible data or obviate the multiple copies of data that applications require for housing separate projections of the same information. All data is erasure coded with an overhead of only 2.7%.



## Flexibility

VAST presents a peerless high-performance S3 interface alongside NFS and SMB. No matter how data is written Trino can read and operate on it without storage bottlenecks. Second-to-none data reduction eliminates many data redundancy problems and transparently deals with easily compressible formats (e.g. JSON). VAST's DASE architecture supplies a scalable solution that, together with our [Gemini](#) licensing model, can load-follow your data-lake/lake-house needs without interruption and without [hardware refreshes](#).

<b>Advanced compression and deduplication</b>	Flash performance at archival costs
<b>High-performance S3</b>	Fast S3 capability means reduced time to decision and real-time reporting
<b>True multi-protocol access</b>	Access all of your data via Trino regardless of the protocols used to write data
<b>DASE Storage Architecture</b>	Scale capacity and performance with ease, and enjoy 99.99999% availability



For more information on Universal Storage and how it can help you solve your application problems, reach out to us at [hello@vastdata.com](mailto:hello@vastdata.com).