Teradata Solution Brief

The modern data stack is performant, disaggregated and cloud native. Each component – analytics/data processing, networking and storage – are best of breed and designed for demanding, petascale workloads.

This is the environment in which Teradata and MinIO have collaborated, creating a data stack that is designed for customers looking to achieve performance at scale in their private cloud deployments.

The integrated solution has two use cases: adding Object Storage as a backend option and adding a migration path for Hadoop installations.

Native Object Storage and MinIO

Native Object Storage (NOS) allows Teradata customers to natively read data where it lives – which, in the modern enterprise, is increasingly oriented around object storage. NOS will write SQL to access data in S3 buckets enabling Vantage to read directly from the object store. This achieves several key objectives for Teradata customers seeking to move past the constraints of traditional file and block systems:

- Facilitates ad hoc data exploration of unknown data sets
- Enables users to join relational and semistructured data sets
- Enables queryable archives by offloading relational data, yet queryable using the same SQL as when the data was in Vantage
- Simplifies loading semi-structured data into relational tables with a single SQL command
- Accelerates data sharing across multiple systems and 3rd parties

The integration supports key data formats such as JSON, CSV and Parquet with a roadmap to support AVRO, ORC and TXT in the future.

While NOS will support both public and private clouds, the integration with MinIO is targeted at Teradata customers who wish to remain on premises with their data yet have full S3 compatibility.
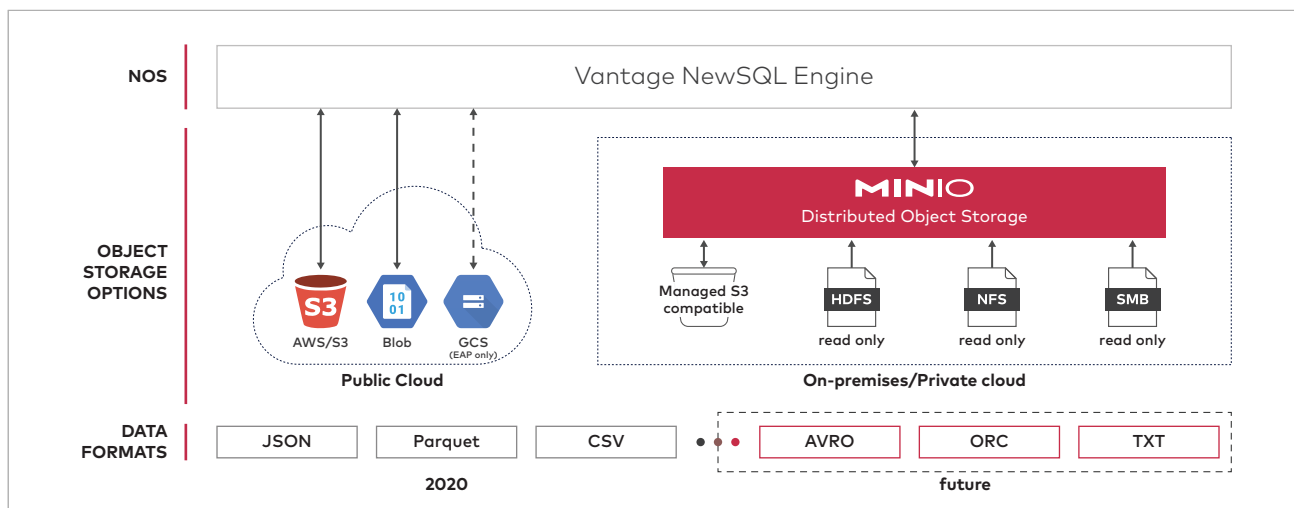
Hadoop Migration

Many enterprise customers want to move away from their existing Hadoop infrastructures. This approach will depend on the entreprises requirements, but many will seek to read data from existing Hadoop deployments while capturing new data in a modern object store. To achieve this Teradata customers would utilize MinIO's HDFS to S3 Gateway to pull data out of HDFS and make it available to Vantage for analytics. This approach allows the older data to persist on HDFS while new data is populated to MinIO. Vantage can access the data from either location seamlessly using the S3 API.
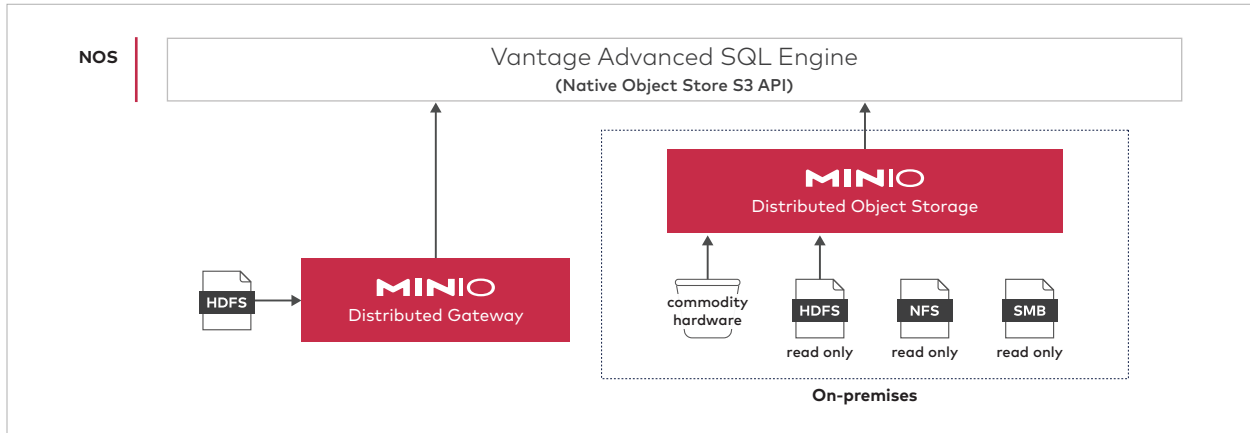

About MinIO

MinIO is a high performance, distributed object storage system. It is software-defined, and runs on standard hardware. Because MinIO is purpose-built to serve only objects, a single-layer architecture achieves all of the necessary functionality without compromise. The result is a cloud native object server that is simultaneously performant, scalable and lightweight.

In S3 Benchmark testing with an 8-node, 100GBe, NVMe setup MinIO achieved GET throughput of 55GB/s and PUT throughput of 35GB/s. These are the fastest object storage numbers ever published.

Summary

Companies that combine MinIO and Teradata will achieve performance at scale across a range of workloads and use cases. From advanced analytics to Hadoop migrations the two companies superbly complement each other to deliver scale, data resiliency, economic value and industry-leading performance.
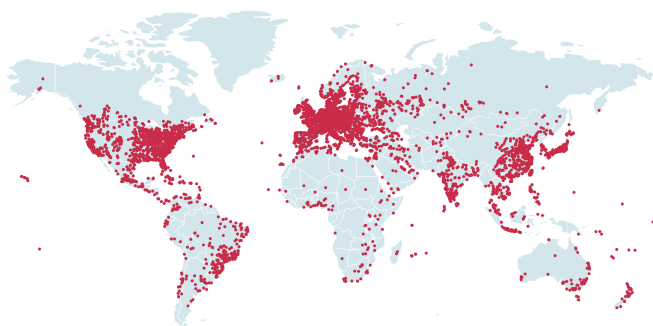
18.4K+
GITHUB STARS

267M+
DOCKER PULLS

5742
SLACK MEMBERS

480
CONTRIBUTORS

*numbers in map as of October 2019