Starburst MinIO Solution Brief

In the modern enterprise, data is the driver of both strategy and execution. Accordingly, the enterprise is built around data, whether they make aircraft, shoes or services. The data-centric enterprise, however, can be siloed, with each organization focused on their own needs and with their own solution stack. This can be difficult to manage as the desire to get the right answer from the data often outweighs corporate mandates over which architecture to adopt.

Starburst and MinIO have a unique approach to delivering analytical performance at scale that enables the enterprise to have a consistent analytics infrastructure - across clouds, both public and private as well as on-prem. This solution delivers exceptional speed coupled with modern disaggregation best practices that ensure viability to exabyte scale.

A Modern Data Architecture

The modern data stack is cloud-native, performant, simple, secure and scalable. These attributes should be present in any stack - but are of particular importance when it comes to creating a data mesh architecture to complement or replace older datalakes/warehouses.

Let's look at each briefly:

Starburst

Starburst Enterprise, based on open source Trino (formerly PrestoSQL) is the fastest SQL-based MPP query engine. It presents a common query tool, abstracted from the systems that store data. This provides the ability to run analytics on data wherever it sits. As a result, no data movement or copies are required. Starburst runs on the cloud, on-prem or wherever it is needed.

Starburst articulates this approach as creating a Data Mesh. A Data Mesh approach moves the flow and requirements of data management out to the domain experts to create data products to share with the rest of the organization.

Starburst eliminates the need to maintain a traditional data warehouse and separates storage from compute, allowing organizations to leverage high-performance, low-cost storage without sacrificing insights. In addition to the economic benefits, Trino improves productivity and time to insight on actionable data, enabling decision makers to impact the business with faster, smarter, data-driven outcomes.

MinIO

MinIO is a high-performance, Kubernetes-native object store. Optimized for the multi-cloud and the S3 API, MinIO delivers AWS S3-like infrastructure on public clouds, private clouds, Kubernetes distributions and the edge.

Enterprises use MinIO to deliver against ML/AI, analytics, backup and archival workloads - all from a single platform. Remarkably simple to install and manage, MinIO offers a rich suite of enterprise features targeting security, resiliency, data protection, scalability and identity management.

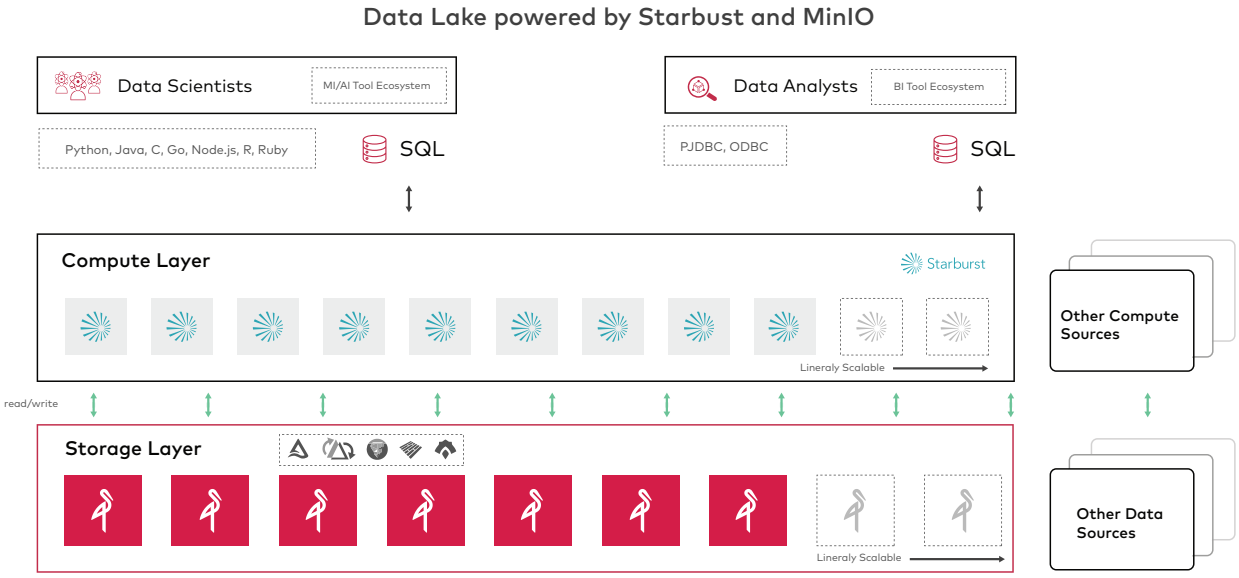
MinIO is the only object storage solution to have a major footprint in the public cloud (1M+ hosts across AWS, GCP and Azure), the private cloud, the Kubernetes distributions (Tanzu, OpenShift, Ezmeral, Rancher/SUSE) and the edge.

MinIO is guided by a few core principles:

- **Cloud Native:** MinIO was built from scratch in the last four years and is native to the technologies and architectures that define the cloud. These include containerization, orchestration with Kubernetes, microservices and multi-tenancy.
- **Performance.** With its focus on high performance, MinIO enables enterprises to support multiple use cases with the same platform. MinIO object storage is used as the primary storage for cloud native applications that require higher throughput and lower latency than traditional object storage can provide.
- **Scalability**. At MinIO scaling starts with a single cluster which can be seamlessly expanded to create a global namespace, spanning multiple data centers if needed. Expansion of the namespace is possible by adding more clusters, more racks until the goal is achieved.
- **Simplicity.** Minimalism is a guiding design principle at MinIO. Simplicity reduces opportunities for errors, improves uptime, delivers reliability while serving as the foundation for performance.

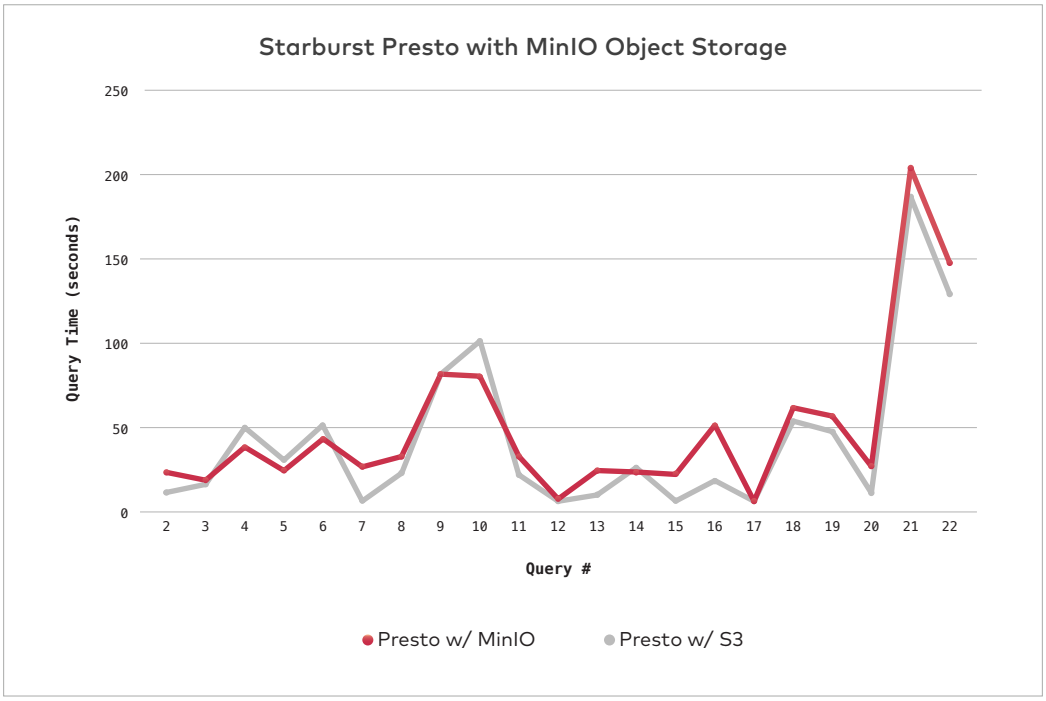
Better Together

The uniqueness of the Starburst-MinIO partnership is in the extensibility of the architecture. Both products can be deployed almost anywhere meaning that analytical insight can be achieved anywhere. Architecturally it looks like this:



What is immediately apparent is the modern principle of disaggregating storage and compute, eliminating costly overprovisioning. Further, Starburst presents the lingua franca of data science, SQL, to data managers, data scientists and analysts.

MinIO is optimized for the techniques employed by Starburst, such as pushdown, predicate filtering, dynamic filtering and Parquet acceleration. The result is an architecture that should, in most cases, choke the hardware or network on which it is run. This can be seen in the benchmarks run by the two companies against AWS. Here we find that MinIO's performance is superior to AWS across query types.



Use Cases

The highly extensible architecture serves a number of use cases, with the following representing the most frequently adopted:

HDFS Migration

HDFS deployment numbers are in free fall, but that doesn't change the fact that for the better part of a decade they were where the enterprise put their mission-critical enterprise data. Nonetheless, no one designs a new Hadoop instance, they either build new with Starburst and MinIO or they migrate their workload onto the modern alternative. Either way - HDFS replacement is the #1 use case for the combined solution.

AI/ML Datalake

As AI/ML moved from next-gen to part of the solution stack, so did the data infrastructure associated with it. Net new AI/ML builds are carefully curated, cloud-native componentry that are capable of simultaneously delivering both speed and scale. Again, pairing the front-end query performance of Starburst with the scale and throughput characteristics of MinIO presents a superb architecture for these use cases.

Perhaps most importantly - these net new datalakes can run almost anywhere - public cloud, private cloud, Kubernetes distributions.

Application Oriented Data Products

Data products require analytical interfaces and data storage. They should be easy to build, deploy and use - driving adoption across the enterprise. With Starburst and MinIO this is achieved effortlessly through the Starburst Data Products Console, the MinIO Console and the MinIO Operator. Using this suite of products, even non-technical types can quickly build, deploy and manage sophisticated data infrastructure to support a wide variety of data products.

Summary

Starburst and MinIO offer an end-to-end, cloud-native data architecture that is powerful but simple. The solution is scalable, secure, software-defined and open, ensuring superior economics and the extensibility to run on any cloud or commodity hardware.

To learn more download them and run them yourselves to see. You can get MinIO at https://min.io/download#/kubernetes and Starburst at https://www.starburst.io/platform/starburst-enterprise/download/ .