Aerospike® SQL Powered by Starburst

PRODUCT BRIEF

Delivering massively parallel SQL analytics on petabytes of data

Because SQL is the lingua-franca of data scientists, data analysts, and business intelligence users, more and more data teams are looking for integrated ANSI SQL query processing in their real-time data pipelines. The development of Aerospike SQL Powered by Starburst provides a tightly integrated SQL processing engine powered by Starburst Enterprise, the enterprise-grade implementation of the open source Trino project (formerly PrestoSQL). The combination of the Aerospike Real-time Data Platform and the Starburst SQL engine offers a rapid and cost effective way for SQL developers to generate valuable insights in your dashboards, ad-hoc queries, or Python applications.
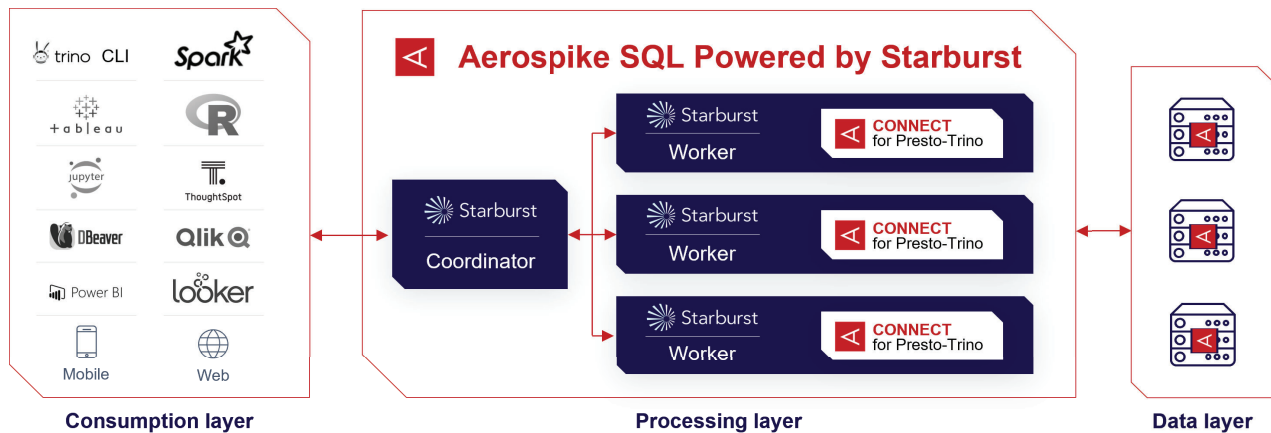

Figure 1: High speed SQL access to Aerospike real-time data

Aerospike SQL's no-code integration of Starburst Enterprise with Aerospike Real-time Data Platform drastically reduces the time and expense of running SQL analytics at gigabyte-to-petabyte scale. Aerospike SQL comes with best in class support from both Aerospike and Starburst, so that you can focus on generating valuable insights at scale from the data stored in Aerospike using ANSI SQL to drive your critical business decisions.

Key Features

Aerospike SQL is a complete gigabyte-to-petabyte scale SQL solution for real-time enterprise analytics and data applications:

- Run ANSI SQL queries to perform in-place, on-demand analytics on massive amounts of data.

- No-code integration of Starburst Enterprise and the Aerospike Database with enterprise grade support.

- Federate queries across multiple Aerospike clusters.

- Integrate Aerospike into an ecosystem of multiple data-storage technologies.

- Query Aerospike by using familiar BI tools, such as Tableau, Qlik and Looker.

- Accelerate queries by using Aerospike's massive parallelism, predicate pushdown, and secondary indexes.

- Leverage the Trino Cost-Based Optimization via row count for query optimization.

- Query records with different schemas within the same set in Aerospike.

- Enterprise grade security:

    - TLS to secure connection between Trino and the Aerospike clusters.

    - LDAP for authenticating Trino users with the Aerospike database.

    - PKI support for passwordless authentication over TLS.

    - Avoid excessive usage via server quotas.

- Deploy in the cloud (AWS and GCP) or on premises.

- Best in class support from Starburst and Aerospike

How Aerospike SQL Powered by Starburst works

This is the process flow that describes the Aerospike/Starburst integration

1. client submits SQL query to Trino coordinator.

2. Trino coordinator constructs query plan, distributes portions among workers.

3. Aerospike connectors send parallel partition scan requests and push down predicates to Aerospike DB. Presto connector loads scanned data from Aerospike partitions into Presto splits.

5. The coordinator fetches results from the workers and returns them to the client application.

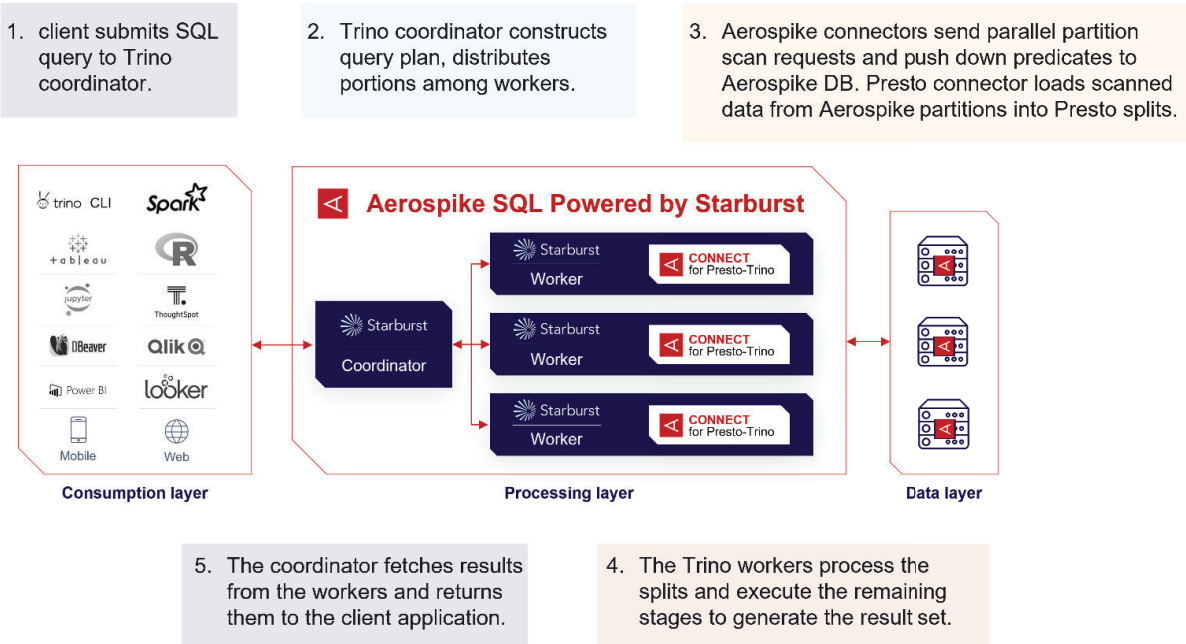4. The Trino workers process the splits and execute the remaining stages to generate the result set.

Figure 2: Aerospike SQL process flow

Use Cases

Aerospike SQL is a complete gigabyte-to-petabyte scale SQL solution for real-time enterprise analytics and data applications.:

Data Analysts

- Run **ad-hoc queries**. For example, "Count the number of users that have clicked the new banner ad", "What are some categories of ads they've seen?", etc.
- Run **BI dashboards and applications** on massive datasets with standard enterprise BI front end platforms (Tableau, Looker, Qlik, etc.)

Compliance Teams

- Use SQL queries for **audit and compliance** applications

Data Engineers and Data Scientists

- **Run complex queries programmatically** using Python and Jupyter notebook.
- Data Engineers and Data Scientists can perform **exploratory data analysis (EDA)** to identify features, patterns, anomalies in the data.
- Develop **complex data models** using Aerospike Collection data types (CDT) i.e. maps and lists and query them performantly using Trino JSON Functions.

2525 E Charleston Road, Mountain View, CA, 94043 | (408) 462-2376 | aerospike.com