The Aerospike Real-time Data Platform is a shared-nothing, multi-threaded, multimodal data platform designed to operate efficiently on a cluster of server nodes, exploiting modern hardware and network technologies to drive reliably fast performance at sub-millisecond speeds across petabytes of data.

The Aerospike Real-time Data Platform enables organizations to act in real time across billions of transactions using massive parallelism and a hybrid memory model to ensure the smallest possible server footprint. It ingests and acts on streaming data at the edge and can combine edge data with data from systems of record, third party sources, data warehouses, or data lakes for operational, transactional, or analytical workloads - all in real time.

**Highlights**

- Predictable real-time performance at petabyte scale
- Cluster node reduction up to 80%
- No-code integrations with Spark, Kafka, Presto/Trino, Starburst, Pulsar, JMS
- Fault and disaster tolerant
- Multi-site, multi-cloud clustering, and data replication
- Real-time transactions with strong

Figure 1: The Aerospike Real-time Data Platform architecture
Aerospike Database 6: real-time data engine

Aerospike Database 6 is the engine that drives the Aerospike Real-time Data Platform. It is responsible for all of the core cluster management and data services of the platform and is now a true multi-model database. It supports key-value and document database functions as well as graph database and time-series capabilities.

Real-time performance with tiered storage

Aerospike has a patented Hybrid Memory Architecture™ (HMA), which places data on solid-state drives (SSDs) and indexes in DRAM. Aerospike treats SSDs as raw devices, writing data in large blocks. Net-net: you can get the performance of DRAM for the price and reliability of SSDs. This approach has been extended to and optimized for All-Flash and persistent memory configurations, delivering true tiered storage density options.

Node reduction by up to 80 percent

In-memory DRAM can only store a fraction of the hot data that Aerospike Hybrid Memory Architecture™ SSDs store per node. Figure 3 illustrates this node reduction. It compares an in-memory system using AWS Memory Optimized instance nodes versus an Aerospike HMA system using Storage Optimized instances. The critical difference is in the addressable space: In-memory only has 122 GiB/node, whereas Storage Optimized nodes have 3.2 terabytes of addressable space per node - a whopping difference.

Multi-model data operations

Aerospike Database 6 is the first expansion of our flagship NoSQL database beyond its initial key-value data model to include full document database capabilities including both document storage, indexing and query. Key-value data operations are native and document database services are new in Aerospike Database 6. Proven graph database services and time series database services are deployed today in customer production environments. ANSI SQL operations have been added via Aerospilne SQL Powered by Starburst.
Modern Distributed Data Services

The data services of the Aerospike Real-time Data Platform are the internal processes of Aerospike Database 6 that enable global operations, enterprise grade availability, and sub-millisecond performance from gigabyte to terabyte scale. See figure 5.

Real-time Engine

Aerospike’s real-time engine delivers the maximum performance possible and can scale to millions of transactions per second at sub-millisecond latency. Aerospike’s real-time engine can seamlessly scale out across nodes, and can scale up on one node.

Query and Search

Aerospike utilizes technologies such as set indexes, secondary indexes and Aerospike Expressions to power real-time query performance on data at the edge, in a System of Record, and across data centers and clouds. Massively parallel secondary indexes enable complex querying and SQL queries. Aerospike Expressions enable highly efficient value-based

Dynamic Data Distribution

Aerospike provides dynamic data distribution via its shared-nothing cluster architecture, where all nodes are peers and there is no single point of failure. Using the Aerospike Smart Partitions™, data is distributed evenly across all nodes in the cluster for maximum performance and scale.

Change Data Notification

Aerospike’s Change Notification feature enables outside databases and message queues to be updated in parallel with writes from database clients. This is a key usability feature when Aerospike is used as the database of record.

Smart Cluster Management

Aerospike’s dynamic cluster management handles node membership, network fault detection, and ensures a single consistent view of current cluster members.

Figure 5: Aerospike distributed data services

Figure 6: Some Aerospike customer milestones
Aerospike Connect

No-code integrations with your data stack

Aerospike has created a series of connector products that provide tight integration with important components of a real-time data stack. Aerospike Connect products provide a direct and easy way to create real-time data pipelines with Spark, Kafka, Pulsar and more. These integrations require no special coding by developers and provide seamless SQL access and massively parallel data ingest for maximum throughput.

![Aerospike Connect Diagram]

**Figure 8: Aerospike Connect no-code integrations for streaming and SQL access**

**SQL Access Connectors**

- **CONNECT for Spark**
  Aerospike Connect for Spark supports APIs that leverage Structured Spark Streaming and Spark SQL. This provides very low latency for both reads and writes, enabling AI/ML use cases to run against data stored in Aerospike, while simultaneously ingesting new data at high rates.

- **CONNECT for Presto/Trino**
  Aerospike Connect for Presto enables data analysts to use ANSI SQL to query data stored in Aerospike via Trino (formerly PrestoSQL). Trino is a highly parallel distributed SQL query engine that is capable of running 100s of concurrent queries, and scaling to 1000s of workers.

**Streaming Connectors**

- **CONNECT for ESP**
  Connect for Event Stream Processing (ESP) integrates Aerospike with event stream processing systems that commonly rely on Change Data Capture (CDC) patterns to inform external systems when records are inserted, modified, or deleted in a database.

- **CONNECT for Kafka**
  Aerospike Connect for Kafka provides real-time data exchange between the Aerospike database and other data pipeline components that use Kafka to transport and consume data streams.

- **CONNECT for JMS**
  Aerospike Connect for JMS provides a simple and easy way to move data in and out of Aerospike to other enterprise systems via the JMS enterprise messaging system.

- **CONNECT for Pulsar**
  Aerospike Connect for Pulsar offers both an inbound connector to ingest Pulsar data into Aerospike and an outbound connector to ship change notifications from Aerospike to a Pulsar consumer for streaming data pipelines.
Aerospike SQL Powered by Starburst

Delivering massively parallel SQL analytics on petabytes of data

The Aerospike Real-time Data Platform has been enhanced by the addition of the massively parallel Trino engine provided by the integration of Starburst Enterprise. This enables SQL users to access data assets in Aerospike clusters for ad hoc analysis, interactive BI dashboards and data applications, audit and compliance applications, exploratory data analysis (EDA), and programmatic creation of powerful massively parallel SQL queries against real-time data.

Aerospike Real-time Data Platform benefits at a glance

- **High-speed processing and ingest**: Query response times in the microseconds to milliseconds range even for extremely large datasets.
- **Highly available**: Designed for the always-on business: enabling five-nines uptime for years on end.
- **Consistency**: Strongly consistent CRUD (create, read, update, delete) operations for accurate data, always in real time.
- **Plug compatibility**: No-code integrations for all data pipeline and event stream components.
- **Massively scalable**: Achieve real-time analytics and applications on datasets of gigabytes, terabytes, petabytes.
- **Predictable performance**: Linear scaling of workloads with low variability response times and low variance; even at massive scale.

The Aerospike Real-time Data Platform enables organizations to act instantly across billions of transactions while reducing server footprint by up to 80 percent. The Aerospike multi-cloud platform powers real-time applications with predictable sub-millisecond performance from terabytes to petabytes of data with five nines uptime with globally distributed, strongly consistent data. Applications built on the Aerospike Real-time Data Platform fight fraud, provide recommendations that dramatically increase shopping cart size, enable global digital payments, and deliver hyper-personalized user experiences to tens of millions of customers. Customers such as Airtel, Experian, Nielsen, PayPal, Snap, Wayfair and Yahoo rely on Aerospike as their data foundation for the future. Headquartered in Mountain View, California, the company also has offices in London, Bangalore and Tel Aviv.

©2022 Aerospike, Inc. All rights reserved. Aerospike and the Aerospike logo are trademarks or registered trademarks of Aerospike. All other names and trademarks are for identification purposes and are the property of their respective owners.