

Aerospike® for Streaming Data Pipelines

SOLUTION BRIEF

Persisting and processing streaming data in real time

Overview

The proliferation of mobile devices, social media, e-commerce, and sensors has fueled the tremendous growth of streaming data. Once confined to a narrow set of industries, streaming data is now pervasive worldwide, and market researchers anticipate significant increases for the foreseeable future. According to one analysis, the global streaming analytics market will grow from \$12.5 billion in 2020 to \$38.6 billion in 2025.

Aerospike and streaming data pipelines

Aerospike has developed critical technologies in our Aerospike Real-time Data Platform to help firms process, persist and protect petabytes of streaming data on the smallest possible server footprint. Aerospike's highly performant, scalable interface to streaming and messaging platforms enables its real-time data platform to ingest massive data volumes captured at the edge to power operational and transactional applications.

The Aerospike Real-time Data Platform architecture features massive parallelism, support for modern hardware architectures, intelligent workload processing, strong data consistency, and self-managing features. In addition, Aerospike's support for Trino (PrestoSQL) and Spark enables SQL programmers to access streamed data stored in Aerospike for immediate query and analysis.

Highlights

- Powerful Change Data Notifications informs external systems and CDC applications when records are modified, inserted, or deleted from a data stream.
- Provides a high-performance persistence and real-time processing engine for streaming data pipelines.
- Seamless integration between Aerospike and streaming data pipelines components such as Kafka, Spark, Pulsar and Trino.
- Enables event-driven and microservices architectures.

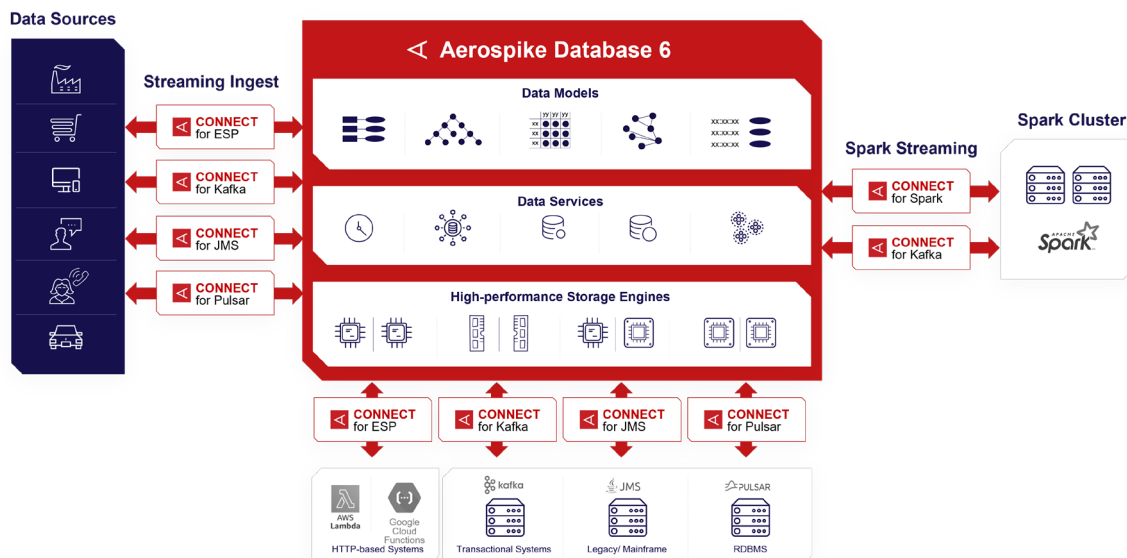


Figure 1: Aerospike connectivity to streaming data pipeline components

Aerospike for streaming data pipelines

Key features

- Stream change notifications from Aerospike using the Cross Data Center Replication (XDR) protocol to streaming data applications, change data capture (CDC) use cases and streaming analytics.
- Supports popular message formats such as Apache Avro, binary (MessagePack) and flat JSON, making application integration easier.
- Ability to route messages based on namespaces, sets, bin values, and static topic names from Aerospike to the downstream application via Kafka, Pulsar, and Spark Streaming.
- Transform messages in real-time (sub-millisecond) going into and out of Aerospike to downstream applications via Kafka, Pulsar, and Spark.
- Rapid, parallelized ingest of data into Aerospike from streaming applications with a variety of data models including document/JSON, key-value, time series, and geospatial.

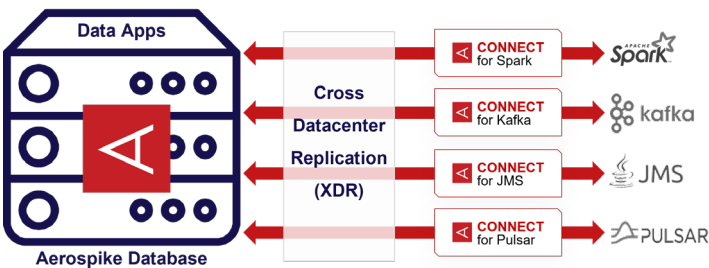


Figure 2: Aerospike interacting with common streaming components

Benefits

- Enterprise wide connectivity between Aerospike and streaming data pipelines components such as Kafka, Spark Streaming, Trino/Starburst, Pulsar, and more.
- Enables event driven architecture & microservices architectures
- Make Aerospike a part of your high throughput and low latency streaming pipeline
- Lower TCO by leveraging the cost efficiencies of Aerospike and Kafka
- Deploy in the cloud (AWS, GCP, Azure) or on premises

Typical use cases

Financial services and FinTech

Bridge legacy systems such as mainframes to modern services such as fraud detection, risk management, and customer 360 using best of breed streaming data pipeline components.

e-Commerce and retail / CPG

Faster and more reliable use of behavioral, social, recommendation data as well as clickstream integration with product data.

Industrial internet

Edge and device data synchronization with back-end systems.

Telecommunications

Stream data from Aerospike to back-end core systems. Persist the data, train AI/ML models, and perform real-time analysis.

Data Lake / EDW Integration for real-time Analytics

Integrating operational data with data lakes and data warehouses for the broadest set of real-time analytics use cases.

AdTech

Real-time clickstream data synchronization and integration.

<EROSPIKE

Aerospike unleashes the power of real-time data to meet the demands of The Right Now Economy. Global innovators and builders choose the Aerospike real-time, multi-model, NoSQL data platform for its predictable sub-millisecond performance at unlimited scale with dramatically reduced infrastructure costs. With support for strong consistency and globally distributed, multi-cloud environments, Aerospike is an essential part of the modern data stack for Adobe, Airtel, Criteo, DBS Bank, Experian, PayPal, Snap, Sony Interactive Entertainment, The Trade Desk, and Wayfair. A global company, Aerospike is headquartered in Mountain View, California, with 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.