Solution Brief:
Aerospike Connect for Spark

POWERING REAL-TIME ANALYTICS USING AEROSPIKE AND SPARK

Highlights

- Expanding the size of dataset that is processed in real-time by joining stream data from Spark with transactional data from Aerospike
- Lower TCO for real-time analytics by operating on larger datasets yet with a smaller cluster footprint
- Gain closed-loop business insights by operating on both transactional and stream datasets
- Rapidly develop using Spark libraries - no additional skill set required
- Accelerate business insights by enabling decisions in seconds as opposed to hours or days

Overview

In the world of Digital, AI, IoT and the Algorithmic economy – data and real-time insights have gained paramount importance within enterprises going through a digital transformation. Real-time, actionable insights drive better decisions which can greatly improve customer experience (CX), operational efficiency and support new business models. Actionable insights are derived from analysis and decisioning based on data from both the transactional and streaming data workloads. While stream data analysis is derived from high-velocity data, transactional data analysis is typically derived from low-velocity data from larger data stores. Traditionally, enterprises dealt with this challenge by using an assortment of disparate products / components with high complexity and cost.

A real-time transactional-analytical system needs to combine transactional and streaming data in a single high-performance database that can operate as fast as the inbound data streams in. It also needs to work with various analysis frameworks including artificial intelligence and machine learning (AI/ML) systems and toolsets.

Aerospike Connect for Spark for Real-Time Decisioning

Real-time decisioning systems must consider both streaming data and historical transaction data to provide valuable results. Most streaming engines operate in-memory and hence tend to operate on a limited dataset as the cost of loading all data in memory outweighs the results obtained.

The Aerospike Connect for Spark (Figure 1.) member of the Aerospike Connect product portfolio addresses this challenge. It combines transactional and historical data stored in the Aerospike database with streaming event data for consumption by machine learning and artificial intelligence engines using Apache Spark. This enables real-time decisioning and insights. From a deployment perspective there are two options:

- Customers can utilize their existing Spark instances in conjunction with the Aerospike Connect for Spark
- Customers who don't have existing Spark installations can utilize the embedded Spark instance of Aerospike Connect for Spark

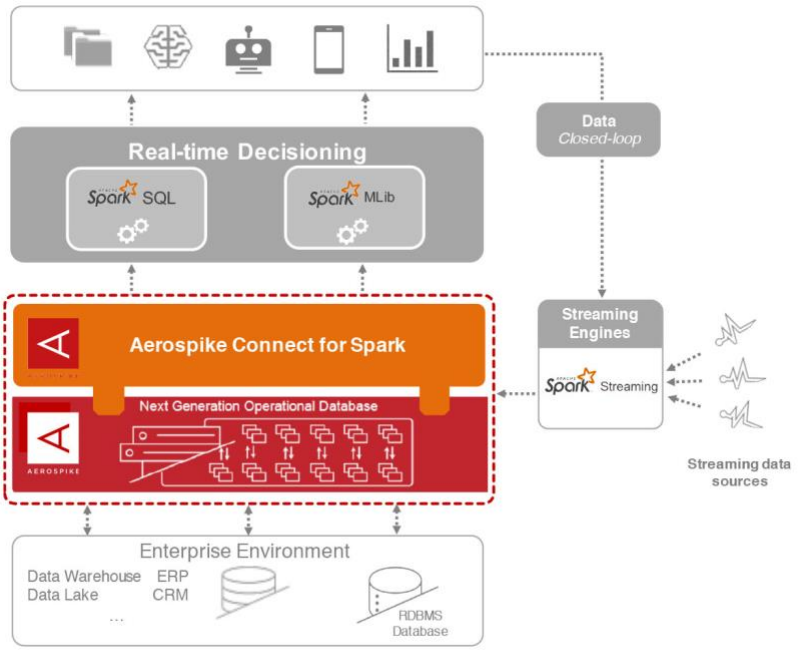

Figure 1. Aerospike Connect for Spark

Solution Brief:

Aerospike Connect for Spark

Existing Solutions Are Overly Complex

In the new age for digital enterprises, there are typically multiple systems that are combined for transactional + streaming analysis:

- **Streaming** - CEP engines/Spark and/or Kafka
- **Transaction store** - Hadoop and/or Cassandra + cache(s)
- **Analysis processing/predictive analysis** - Spark and/or a traditional OLAP engine

Aerospike Connect for Spark Capabilities

Aerospike Connect for Spark leverages Apache Spark and Aerospike to:

- Load Aerospike data into Spark streaming engine for processing
- Join stream data with Aerospike transactional data using keys of interest from the stream data – greatly expanding the size of the dataset that is processed in real-time.
- Utilize either customer's existing Spark instances or the embedded Spark instance for analysis
- Enable analysis and AI/ML processing on joined data
- Enable querying on the combined data by Spark SQL

Benefits

- Real-time analytics by operating on larger datasets yet with a smaller cluster footprint thus lower TCO
- Gain closed-loop business insights by operating on both transactional and stream datasets
- Rapidly develop using Spark libraries - no additional skill set required
- Flexibility to utilize either the embedded or customer's existing Spark instances for analysis
- Accelerate business insights by enabling decisions in seconds as opposed to hours or days

Typical Use Cases for Aerospike Connect for Kafka

Financial Services and FinTech: Real-time fraud prevention, trading risk computation, operational risk monitoring, real-time lending, robot advisory

e-Commerce and Retail / CPG: Personalized recommendations, real-time targeting, dynamic price optimization

Telco: Intelligent CSR management, churn reduction, real-time network strength monitoring

Industrial Internet: Detect issues in real-time, asset management and safety, predictive maintenance, etc.

Logistics: Detect issues in real-time, asset management and

AEROSPIKE

About Aerospike

Aerospike is trusted by leading enterprises around the world to help them build and deploy modern data architecture solutions with confidence. The Aerospike enterprise-grade non-relational database helps companies power mission critical, strategic operational applications that make digital transformation possible. Powered by a patented Hybrid Memory Architecture™ and autonomic cluster management, Aerospike is used by enterprises in the financial services, telecommunications, technology, retail, e-commerce, adtech, and online gaming industries and is well-suited for fraud prevention, digital payments, recommendation engines, real-time bidding and other applications that require extreme uptime, performance and scale. Aerospike customers include Adobe, Airtel, FlipKart, Kayak, Nielsen, and Snap. The company is headquartered in Mountain View, Calif.

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