AEROSPIKE

Aerospike Connect for Event Stream Processing

SOLUTION BRIEF

Seamlessly connect Aerospike to HTTP-based systems to build efficient streaming pipelines

Overview

Event stream processing (ESP) systems commonly rely on Change Data Capture (CDC) data patterns which inform external systems when records are inserted, modified, or deleted in a database. These CDC patterns (in Aerospike CDC is implemented as "Change Notification" but unlike CDC, may include several changes) are published as they occur via the Aerospike Cross-Datacenter Replication (XDR) feature.

Highlights

SUPPORT FOR SERVERLESS DESIGN
Supports serverless processing design patterns, where CDCs are streamed to either a Lambda function in AWS or to Google Cloud Functions.

EASY INTEGRATION
Easily integrate the Aerospike Real-time Data Platform with HTTP-based systems either in the cloud or on-premises.

FLEXIBLE DEPLOYMENT
Offers flexible deployment options either in the cloud or on-premises.

Connecting Aerospike with HTTP based enterprise systems through Connect for Event Stream Processing

As a member of the Connect product line, Aerospike Connect for ESP converts change notifications over XDR into HTTP requests and streams them to downstream consumers.

Aerospike is a hyperscale, multi-model, real-time distributed data platform that forms a core building block of any enterprise data architecture. In addition, to avoid data siloing, it is imperative to support free data movement between Aerospike and other enterprise systems, such as those relying upon the HTTP protocol. Aerospike Connect for ESP provides an efficient way to move data out of Aerospike to other enterprise and cloud systems, as seen on Figure 1.

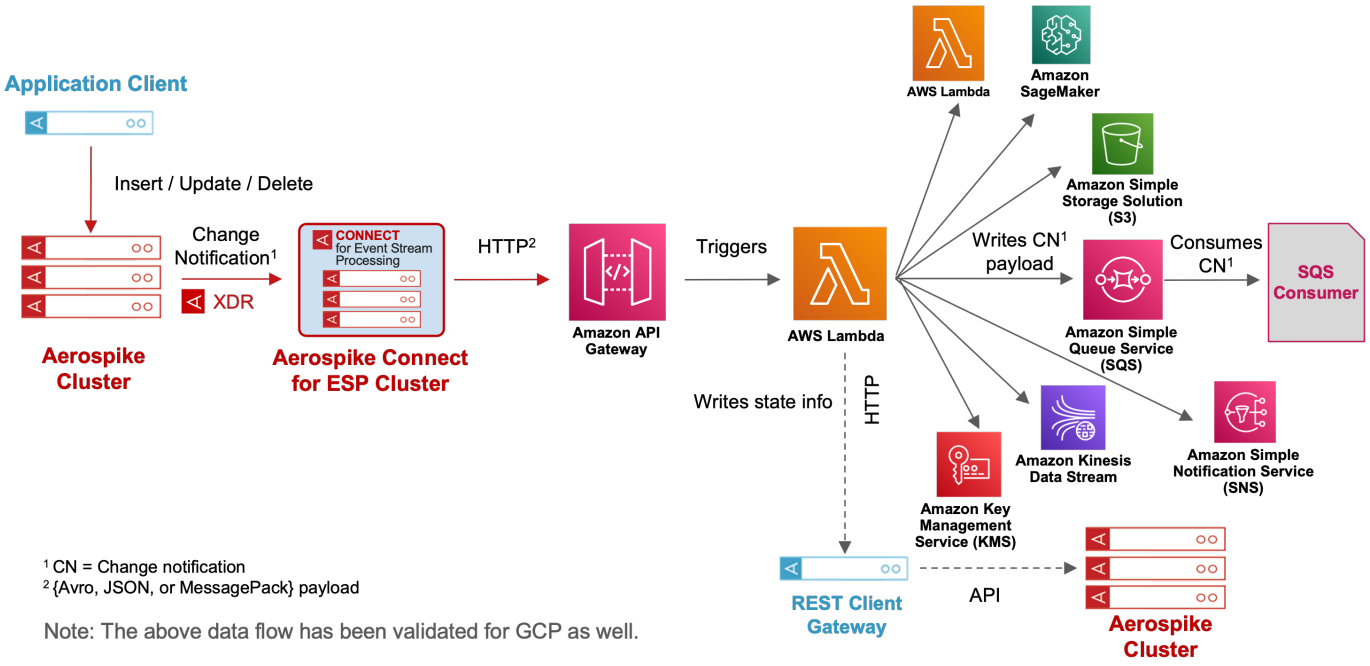

Figure 1: Reference architecture for Amazon SQS with Aerospike Connect for Event Stream Processing data pipline.

Aerospike Connect for ESP functionality

Connect for ESP converts change notifications over Aerospike's XDR into HTTP requests and streams them to downstream consumers.

- **HTTP/1 or HTTP/2 conversion:** Connect for ESP converts Change Notification over XDR into HTTP/1 or HTTP/2 requests, which potentially opens up connections to several HTTP-based systems such as ElasticSearch, AWS Lambda, Splunk, etc.
- **Payload serialization:** Serializes the change notification payload into a text format such as JSON, or a binary format, such as Avro or MessagePack, for efficient data exchange.
- **Trigger systems:** Can be used to trigger AWS Lambda via API Gateway or Application Load Balancer, or Google Functions, to process change notifications and make it available to the broader big data ecosystem in AWS or GCP.
- **LUT shipping:** Ships the LUT (last-update-time) of the record to enable downstream applications to build their own custom logic for ordering messages.
- **Delivery guarantee:** Extends XDR's at-least-once delivery guarantee to ensure zero message loss.
- **Transmission control:** Can be used in conjunction with XDR filter expressions to filter out records before transmission for greater control.
- **Flexible deployment:** Offers flexible deployment options either in the cloud or on-prem.

Benefits

Support for serverless design

Supports serverless processing design patterns, where CDCs are streamed to either a Lambda function in AWS or to Google Cloud Functions.

Supports efficient data exchange between systems

Supports JSON and binary formats such as Avro and MessagePack.

Easy Integration

Easily integrate The Aerospike Real-time Data Platform with HTTP-based systems either in the cloud or on-premises.

Simplifies Operations

Removes the overhead of operationalizing Kafka or Pulsar production clusters for use cases that do not warrant it.

Use Cases:

The basic idea is to connect the Aerospike database with any system that can accept Aerospike Change Notification (which is, again, Aerospike's version of Change Data Capture) as HTTP POST requests. Here are a few select examples:

Search

Export shortened Aerospike Change Notification messages to Elasticsearch using its Document REST API. This pattern can be extended to make Aerospike data available for analysis in Splunk using their HTTP Event Collector.

Backup

Back up Aerospike data to S3 (via AWS Lambda) or Google Cloud Storage (via Google Cloud Functions) for archival purposes.

AI/ML data processing

Stream Aerospike data to AWS SageMaker (via AWS Lambda) or Google AI (via Google Cloud Functions) to accelerate your cloud native AI/ML pipelines.

Compliance

Encrypt Aerospike data for compliance using AWS KMS via Lambda, before ingesting it into your data lake.

Web application development

Develop a web application in the language of your choice to ingest and process Aerospike Change Notification.

Autoscale streaming connectors

Autoscale Aerospike streaming connectors for Kafka, JMS, or Pulsar to enable you to build a cloud-native streaming pipeline.

AEROSPIKE

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 up to petabyte scale with five-nines uptime and 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.

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