

# Streaming Integration for Hadoop

## Ingest Real-Time, Pre-Processed Data for Operational Intelligence

Striim® is a software product that continuously moves real-time data from a wide range of sources into Hadoop, Kafka, relational and NoSQL databases — on-prem or in the cloud — with in-line transformation and enrichment capabilities. Brought to you by the core team behind GoldenGate Software, Striim offers a non-intrusive, quick-to-deploy solution for streaming integration so your Hadoop solution can support a broader set of operational use cases.

With the following capabilities Striim enables a smart data architecture that supports use case-driven analytics in enterprise data lakes:

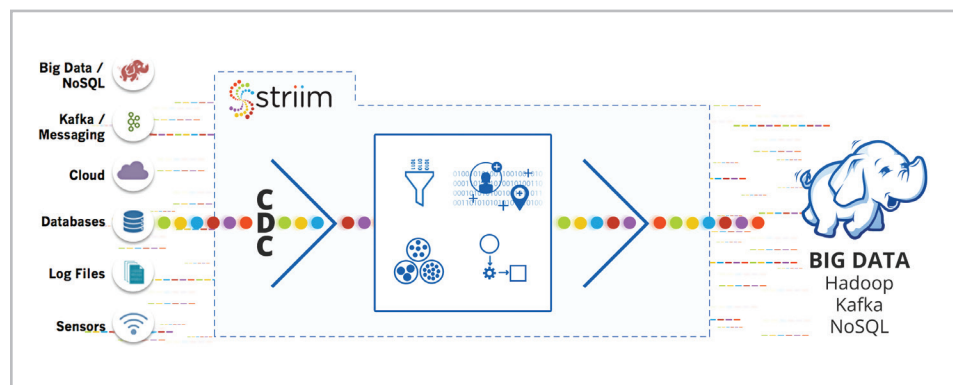
- **Ingests large volumes of real-time data** from databases, log files, messaging systems, and sensors
- **Collects change data non-intrusively** from major enterprise databases such as Oracle, SQL Server, MySQL, HPE NonStop, MariaDB, and Amazon RDS
- **Delivers data in milliseconds** to Hadoop (HDFS, HBase, Hive, Kudu), Kafka, Cassandra, MongoDB, relational databases, cloud environments, and other targets
- **Supports mission-critical environments** with end-to-end security, reliability, HA, and scalability

## BENEFITS

- Use low-latency data for operational use cases
- Accelerate time to insight with a continuous flow of transformed data
- Ensure scalability, security, and reliability for business-critical solutions
- Achieve fast-time to market with wizards-based UI and SQL-based language

## KEY FEATURES

- Enterprise-grade and fast-to-deploy streaming integration
- Real-time integration of structured and unstructured data
- In-flight filtering, aggregation, transformation, and enrichment
- Integration with existing technologies and open source solutions



Striim offers real-time data ingestion from a diverse set of sources, including databases using low-impact change data capture.

# Get More Operational Value from Hadoop Investments

Striim enables businesses to get the maximum value from high-velocity, high-volume data by delivering it to Hadoop environments in real time and in the right format for operational use cases.

## Real-Time, Low-Impact Change Data Capture

Striim ingests real-time data from transactional databases, log files, message queues, and sensors. For enterprise databases, including Oracle, Microsoft SQL Server, MySQL, and HPE NonStop, Striim offers a non-intrusive change data capture (CDC) feature to ensure real-time data integration has minimal impact on source systems and optimizes the network utilization by moving only the change data.

## In-Flight Data Processing

As data volumes continue to grow, having the ability to filter out and aggregate the data before analytics becomes a key way to manage the limited storage resources. Striim enables in-flight data filtering and aggregation before it delivers to Hadoop to reduce data storage footprint. By performing in-line transformation (such as denormalization) and enrichment with static or dynamically changing data in memory, Striim feeds large data volumes in the right format without introducing latency.

## Enterprise-Grade Solution

Striim is designed to meet the needs of mission-critical environments with end-to-end security and reliability — including out-of-the-box exactly once processing — high-performance, and scalability. Users can focus on the application logic knowing that from ingestion to alerting and delivery, the platform is bulletproof to support the business as required.

## Fast Time to Market

Intuitive development experience with drag-and-drop UI along with prebuilt data flows for multiple Hadoop targets from popular sources allow fast deployment. Striim uses an SQL-based language that requires no special skills to develop or modify streaming applications.

## Operationalizing Machine Learning

Striim can pre-process and extract features suitable for machine learning before continually delivering training files to Hadoop. Once data scientists build their models using Hadoop technologies, these can be brought into Striim, using the new open processor component, so real-time insights can guide operational decision making and truly transform the business. Striim can also monitor model fitness and trigger retraining of models for full automation.

## Differences from ETL

Compared to traditional ETL offerings that use bulk data extracts, Striim enables continuous ingestion of structured, semi-structured, and unstructured data in real time, delivering granular data flow for richer analytics. By performing in-memory transformations on data-in-motion using SQL-based continuous queries, Striim avoids adding latency and enables real-time delivery. While ETL solutions are optimized for database sources and targets, Striim provides native integration and optimized delivery for Hadoop, Kafka, databases, and files, on-premises or in the cloud. Striim also offers intelligence and data visualization capabilities within the same platform, without requiring additional licenses.