

Data Movement Via Real-Time CDC to Hadoop, Kafka, Cloud Infrastructures

Striim provides a non-intrusive, end-to-end solution for moving change data from transactional systems to Hadoop, Kafka, as well as other data targets.

WHY CDC FOR DATA MOVEMENT TO BIG DATA?

Change Data Capture (CDC) provides a mechanism for monitoring or recording database transactions to help companies understand what's happening within their enterprise. The output is a record of all activity in the form of inserts, updates, and deletes against that database.

This CDC activity captures every event from operational systems. Unlike obtaining data in batches where the dataset represents a snapshot, streaming CDC provides a complete real-time account of activity that takes place in the enterprise. If batch is like a photograph, CDC is like a video.

Unfortunately, many businesses are not considering database change events as part of their historical dataset, resulting in only a partial picture of their enterprise. If companies could tap into their change streams, they could get a complete, real-time view into their business.

BENEFITS

- Real-time with low latency
- Non-intrusive streaming CDC
- Complete end-to-end solution
- Streaming processing/enrichment
- Drag-and-drop UI and SQL-like language
- Easy-to-understand



Companies that realize this opportunity are looking to move change data via real-time CDC to Hadoop, Kafka, Amazon, or other data targets for analysis and reduced latency. However, this is no easy task. There is no simple off-the-shelf solution for this, and options provided by database vendors are often expensive and complex to implement.

What's more, existing solutions typically only allow a simple, one-to-one delivery of change events to records in HDFS files, Kafka message queues, etc. If companies want to perform any transformation, filtering, aggregation or enrichment of the data as it streams, they need additional software products.

SHORTFALLS IN ANALYZING BIG DATA

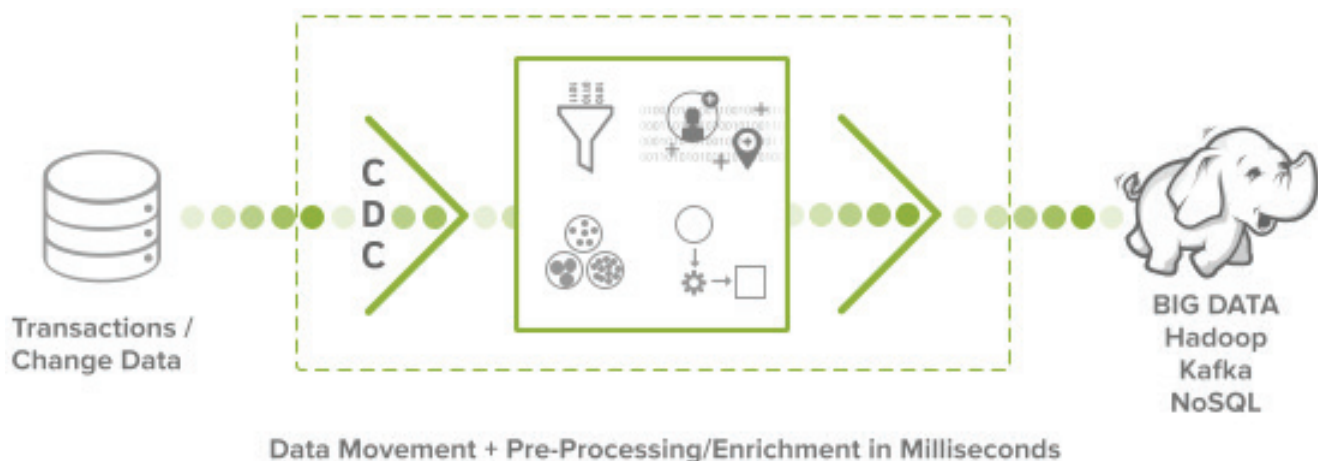
The introduction of Big Data into the enterprise has introduced a “store now, process/analyze later” approach to harvesting valuable insights.

To analyze data, companies often store snapshots of the data in Hadoop – which leads to extreme latency and expensive queries against production data. If they are using snapshots and deltas using timestamps, they are likely to be missing intermediate operations that may be essential to record from an auditing or compliance perspective.

Many companies would prefer a non-intrusive CDC solution. CDC is the only approach accepted by most DBAs, but it is expensive and complex to implement. It also may require multiple software products to achieve anything beyond simple transformations.

STRIIM PLATFORM: REAL-TIME CDC TO HADOOP, KAFKA, CLOUD

Striim offers a simple, cost-effective CDC-to-Big Data/Cloud delivery solution managed through a single application that provides enterprise-quality security, scalability and fault tolerance. For more advanced applications, it is easy to add in data transformations, filtering, aggregation and enrichment of streaming data through simple SQL-like queries as it flows from the database to Hadoop, Kafka, Amazon, or another of a broad range of data targets.



By enabling up-front processing of data as it's streaming, as well as filtering and enrichment before the data lands on disk or on a message queue, the data storage footprint is reduced, and stored records are more actionable. Populating stream caches with historical data from Hadoop allows real-time correlation of deep context with real-time streaming data.

BENEFITS OF THE STRIIM PLATFORM

In comparison to other CDC mechanisms, the Striim platform does not require other software products or complex configuration. Striim uses a simple drag-and-drop UI to configure both simple and complex data flows. Any data manipulation is carried out using easy-to-understand SQL-like queries. Data flows can be built up incrementally, and streaming data can be interrogated at any point.

For more information, or to schedule a free trial, please contact us at info@striim.com or at **+1 (650) 241-0680**.