

Low latency stateful stream processing powered by Apache Flink

CLUDERA STREAMING ANALYTICS

Why Cludera Streaming Analytics?

Next-generation streaming engine - Can process hundred billion events per day, consuming both offline and live data, and continuously monitoring and processing events in real-time

Stream processing simplified - Building streaming applications is no longer for the Java and Scala developers but also for developers who prefer SQL.

Multi-Cloud and Hybrid Cloud Support - With the power of CDP, you can choose a flexible deployment model for your streaming architecture that spans across the edge, on-premises, and multi-cloud environments

Unified Streaming and Batch - The stream processing engine can join and enrich from a variety of sources including streaming sources like Apache Kafka as well as batch sources like Hive, Kudu, and JDBC.

Cludera DataFlow (CDF) is a comprehensive edge-to-cloud streaming data platform. It addresses the key data management challenges with streaming and IoT data for all types of enterprises. Cludera Streaming Analytics (CSA) is part of the CDF platform that addresses all types of stream processing and analytics use cases. All of this is available within the Cludera Data Platform (CDP) to support multi-cloud and hybrid cloud architectures.

Cludera Streaming Analytics

Cludera Streaming Analytics (CSA) comprises of a stream processing engine powered by Apache Flink. It helps businesses democratize streaming analytics across the firm and improve detection and response to critical events that deliver business outcomes. CSA makes streaming analytics more accessible to business analysts and data engineers by simplifying streaming application development using comprehensive APIs and SQL to handle event processing. CSA is integrated with Apache NiFi and Apache Kafka and brings the ability to increase new data sources without the heavy development cycles and scale real-time analytics as data increases.



Streaming Analytics Challenges

Stateless - Unable to process broader insight across different time periods (windows), especially when the use case needs to combine historical events with current events

Lack of Comprehensive APIs - Unable to process both batch data and streaming data to unify the information for more accurate event detection and alerts

Lack of SQL access - Business Analyst are unable to leverage their SQL skills to access real-time data

Limited or No Monitoring - Lack visibility to pinpoint streaming analytics failures

Increase in Ingest Data - unable to scale as data ingestion increases and cannot process the events fast enough to guide business decision making

Apache Flink

Apache Flink is an open-source, next-generation distributed stream processing system built to handle any data movement, from offline data to live data, but highly scalable to process massive amounts of data for real-time analytics. Some of the more transformative businesses use Flink to separate the massive data noise from detecting the most important events that help the business automate more accurate decision making.

What makes Apache Flink unique from other stream processing solutions is that it maintains a state locally (in-memory or on-disk) to provide a checkpoint or a consistent snapshot of all the tasks. Apache Flink can easily help developers identify the latest task and correlate past and current events for more accurate event detection.

Though it was built primarily for Java and Scala developers, Flink is now available to business analysts, who are already SQL developers and want to start building streaming analytics for their business needs. Flink is also integrated easily with all major cloud vendors such as Amazon Web Services, Azure, and Google Cloud, delivering the business analysts the option to build real-time analytics quickly.

SQL Stream Builder

SQL Stream Builder allows developers, analysts, and data scientists to write streaming applications using industry-standard SQL. It provides an interactive experience - so the development process is quick, easy, and productive. It offers syntax checking, error reporting, schema detection, query creation, sampling results, and creating outputs. It provides an advanced materialized view engine to interface with applications, tooling, and services via REST. SQL Stream Builder continuously runs SQL via Apache Flink. This truly democratizes access to real-time data to many more developers across the enterprise.

Key Features

- Simple query interface with SQL Stream Builder
- Easier to consume and build streaming data using Materialized Views
- Rich integration with Apache Kafka for streaming sources and sinks, including exactly-once processing
- Event-driven application architecture
- Guaranteed consistency by checkpoints
- Support for multi-cloud and hybrid cloud models
- Unified data governance and security delivered through Cloudera SDX integration
- Available on CDP Private Cloud Base and Data Hub
- Join and enrich streams of data with batch data via SQL
- Available on CDP Private Cloud Base and Data Hub (coming soon)