Cloudera Stream Processing with IBM

Advanced messaging and
stream processing

Why Cloudera Stream Processing with IBM?

100 percent open source
Invest in your architecture and scale with confidence knowing that there's no vendor lock-in.

Data lineage and provenance
The only platform to offer end-to-end data lineage tracking and provenance across Kafka, MiNiFi, NiFi and so on.

Choice of multiple streaming analytic engines
Cloudera DataFlow (CDF) supports Kafka Streams, Spark Streaming and Apache Flink for real-time insights and predictive analytics.

Hundreds of Kafka customers
Cloudera has hundreds of happy customers getting excellent support on their advanced Kafka deployments.

Edge-to-enterprise comprehensive platform
Kafka is tightly integrated with NiFi, MiNiFi and the rest of the enterprise services for security and governance.

Cloudera Stream Processing with IBM

Cloudera Stream Processing with IBM (CSP) offers the latest certified, secure and governed Apache Kafka that provides the messaging backbone for real-time streaming use cases that require low-latency ingestion and durability along with decentralized management across producers and consumers. CSP also includes support for Kafka Streams for real-time analytics, Schema Registry for centralized schema management, Cloudera Manager for cluster management and monitoring, and Apache Sentry for rich access control and security.

CSP is a key part of Cloudera DataFlow with IBM (CDF), a comprehensive edge-to-enterprise streaming data platform that addresses the key data management challenges with streaming and Internet of Things (IoT) data for essentially all types of enterprises. As part of the CDF platform CSP addresses key ingestion and stream processing use cases   in streaming architectures at scale.

Cloudera Streams Management

Cloudera Streams Management (CSM) includes the key management and monitoring capabilities you'll need to maintain business continuity of your Kafka implementation. It's comprised of two key components:

– Cloudera Streams Messaging Manager (SMM)—for monitoring and management of enterprise Kafka
– Cloudera Streams Replication Manager (SRM)—for disaster recovery and replication of enterprise Kafka clusters

Kafka is used as the core stream processing engine across most of the streaming architectures across enterprises today. But, it also has its own set of challenges within platform operations, DevOps and the security and governance teams. CSM acts as the perfect complement for CSP to provide the required visibility for Kafka users.
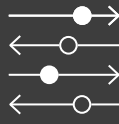
Cloudera DataFlow
platform with IBM

Edge management

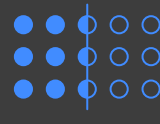Edge data collection, routing and monitoring

– MiNiFi
– Edge Flow Manager
– NiFi Registry

Flow management

Enterprise data ingestion, transformation and enrichment

– Apache NiFi
– NiFi registry

Stream processing

Real-time stream processing at IoT scale

– Apache Kafka
– Schema Registry

Stream management

– Streams Messaging Manager
– Streams Replication Manager

Stream analytics

Predictive analytics and real-time insights

– Kafka Streams
– Apache Flink
– Spark Streaming

Enterprise services

Provisining, management and monitoring

– Unified security
– Edge-to-enterprise governance
– Single sign-on

Cloudera Streams Messaging Manager (SMM)

Cure Kafka blindness with a single monitoring and management dashboard that lets you:

– Troubleshoot your Kafka environment to identify bottlenecks, throughputs, consumer patterns, traffic flow and so on.
– Visualize end-to-end flows and complete data lineage of message streams from producers to topics to consumers.
– Optimize your Kafka environment based on the key performance insights gathered from various brokers and topics.

Cloudera Streams Replication Manager (SRM)

Next-generation Kafka replication powered by MirrorMaker 2.0 (MM2) allows you to:

– Leverage significantly improved Kafka replication capabilities with MirrorMaker 2.0.
– Extend your Kafka architecture to handle active-active clusters and disaster recovery scenarios with SRM.
– Install an MM2 cluster easily, provide lifecycle management actions on the cluster like start and stop; and manage and monitor the replication flows across clusters.

Key benefits

IoT-level scale of streaming architectures
– Process millions of messages per second with
  Apache Kafka.
– Achieve high scale between diverse producers and
  consumers with Kafka's decoupled architecture.
– Connect from virtually any data producer to consumer
  using Kafka's publish-subscribe model.

Messaging efficiency and data governance
– Reuse schemas, define the relationship between schemas,
  and manage schema versions with Schema Registry.
– Avoid attaching a schema with every message and
  boost message transport efficiency.
– Leverage the integration of Schema Registry across
  Kafka and Apache NiFi by using the same schemas
  from end-to-end.

Kafka management, monitoring and replication
– Take advantage of Cloudera Streams Management
  to manage and monitor enterprise Kafka clusters.
– Cure "Kafka blindness" by getting visibility into all your
  Kafka clusters with Streams Messaging Manager.
– Manage enterprise Kafka data effectively for active-active
  cluster replication and disaster recovery use cases.

For more information

To learn more about Cloudera Stream Processing
with IBM, visit the IBM and Cloudera webpage
or contact an IBM data management expert.

IBM