CLOUDERA

WHITEPAPER

The Keys to Building a Complete Kafka Framework

Cloudera delivers best-in-class streaming

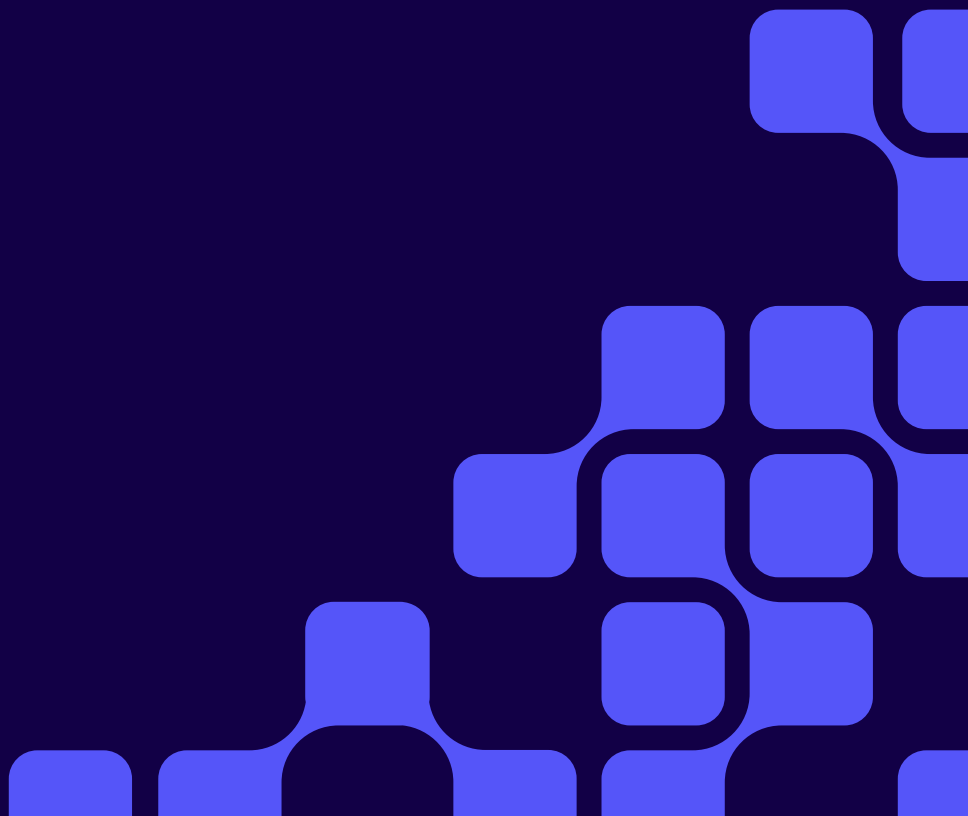
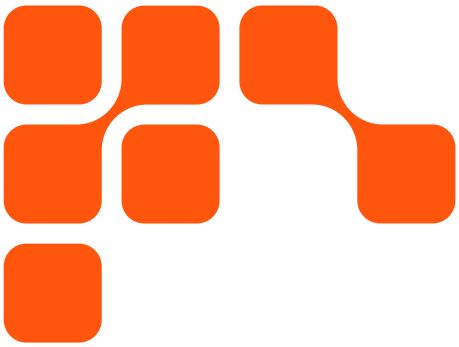


Table of Contents



Efficiency, Scale and Adaptability: You Can Have It All

To be successful, global businesses must continually transform and adapt to changing economic conditions and consumer expectations. Within each organization, both the business and technology communities share the responsibility of delivering innovative products and services that their customers, employees, and regulators expect in near real time. Apache Kafka is the key architectural component for a wide range of streaming data initiatives that enable enterprises to deliver on those responsibilities.

For large-scale deployments of hundreds to thousands of Kafka brokers, including a messaging solution in the streaming architecture isn't sufficient due to the data management challenges that arise at every stage, from data ingestion and preparation to real-time processing for predictive insights. Flow management, along with stream processing and analytics, are two additional components that must accompany stream messaging capabilities to implement a complete end-to-end streaming architecture.

While different vendors in the market claim to support Kafka, Cloudera stands out distinctively with its holistic and comprehensive offering. Cloudera's commitment to the open source community and its penchant for listening to the voice of the customer helps it deliver advanced innovations in the Kafka ecosystem of components. Also, as a trailblazer in the True Hybrid Platform market, Cloudera has been supporting and extending the Kafka ecosystem since the technology's inception. Cloudera packages Kafka for our customers and delivers a fully hybrid experience as part of its platform. Kafka can be deployed on premises or in the cloud with no refactoring required. Customers who do not need a platform and have their own Kubernetes environments can deploy Kafka independently from Cloudera as Kubernetes operators that install and manage the software lifecycle on their behalf.

This paper describes how all three data-in-motion components are unified through a common data experience across on premises, hybrid, multi-cloud, and containerized kubernetes environments. Cloudera leverages best-in-class, open-source-based engineering to deliver the best Kafka ecosystem today that ensures a sustainable, scalable, and adaptable end-to-end streaming architecture.

The Three Components of a Unified End-to-End Streaming Architecture

- **Flow management** refers to the collection, distribution, and transformation of data across multiple points of producers and consumers.
- **Streams messaging** is the provisioning and distribution of messages between producers and consumers.
- **Stream processing and analytics** is the generation of real-time analytical insights from the data streaming between producers and consumers.

Cloudera's Data-in-Motion philosophy is a comprehensive, edge- to-cloud, real-time streaming suite of capabilities that collects, curates, and analyzes data so customers gain key insights for immediate actionable intelligence. It meets the challenges customers face implementing data in motion, such as real-time stream processing, streaming analytics, data provenance, and data ingestion from Internet of Things (IoT) devices and other sources.

Cloudera's approach to data in motion integrates all three components:

- Data capture and flow management at the edge
- Provisioning that data directly to or from your Kafka messaging backbone
- Stream processing and analytics

Read our solution brief, "Data-In-Motion Philosophy: A Blueprint for Enterprise-wide Streaming Data Architecture" to understand how it all comes together.

It Takes a Complete Streaming Platform

Business and technology teams seek to improve the flexibility, speed, efficiency, accuracy, and security of capturing, provisioning, distributing, and analyzing data that is streaming across their enterprise and with external parties. Thus, many organizations have pivoted from large monolithic database platforms to event-driven streaming architectures and microservices designs.

Apache Kafka has emerged as the central backbone of event-based architectures because it addresses the fundamental challenges of scalability and is highly optimized for both ad-hoc and sustained exchange of messages. However, Kafka doesn't address the challenges of how the data is ingested from multiple sources, or how real-time data streams are processed and analyzed with extremely low latency to produce automated responses and actionable insights for key decision makers.

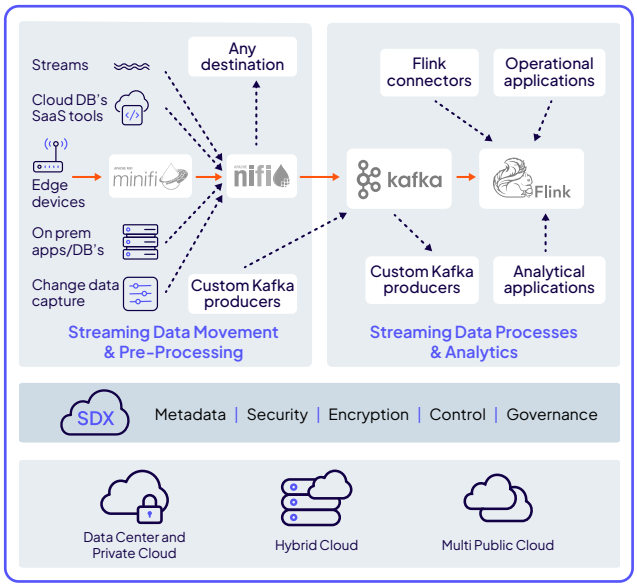
Cloudera integrates Kafka along with Apache NiFi and Apache Flink to deliver complete data-in-motion capabilities within Cloudera, the industry's only hybrid data platform that addresses the full data lifecycle.

The diagram in Figure 1 illustrates how Cloudera data in motion supports the entire streaming journey from data capture and flow management at the edge (1) to provisioning that data directly to or from your Kafka messaging backbone (2) and/or stream processing and analytics (3). It is tightly integrated with Cloudera Shared Data Experience (SDX), a common set of services that provide unified security, governance, lineage, and control (4) across data centers and multiple cloud environments (5).

In short, Cloudera data in motion provides the flow management and stream processing capabilities that IT teams need while enabling the data engineering and platform teams to deploy, manage, monitor, and replicate Kafka clusters with full end-to-end visibility. Cloudera enables safe and consistent deployments and migrations of such streaming capabilities across hybrid, private, and multi-cloud environments.

This paper describes how Cloudera provides the best Kafka ecosystem of components in the industry today.

Figure 1



The Complete Kafka Offering

As described earlier, Kafka is the key architectural component for a wide range of streaming data initiatives that enable enterprises to keep up with customer demand, provide better services, proactively manage risk, and more. Although there may be an assortment of streaming data approaches within the same organization, Kafka is the enterprise-wide common denominator because it provides:

- **High throughput and low latency —** Kafka supports millions of messages per second, with latencies as low as a few milliseconds.
- **Scalability —** Data teams can expand clusters elastically and transparently with no downtime.
- **Durability and reliability —** Messages are persisted on disk and replicated across clusters to prevent data loss.
- **Fault tolerance —** The platform is immune to machine failure in the Kafka cluster.
- **High concurrency —** Kafka can manage thousands of clients writing to and reading from it simultaneously.

Cloudera supports hundreds of customers deploying Kafka, processing billions of messages per second. Cloudera provides the most comprehensive data-in-motion capabilities, with all of the necessary ecosystem components, and some productivity-boosting innovations for a complete Kafka implementation.

Cloudera provides simple deployment and robust troubleshooting and monitoring of Kafka, as well as shared compliance-ready security, governance, lineage, and control in one simple application across multiple on premises, hybrid, private, public, or multi-cloud environments.

Embedded in that ecosystem is a unified set of tooling to connect data sources, manage and reuse schemas, optimize clusters, and enable high availability and disaster recovery replication use cases. The following sections cover some of the key Kafka ecosystem components.

Kafka Streams

Kafka Streams is the built-in stream processing library of the Apache Kafka project. It provides real-time stream processing and analytics with high throughput and very low latency. It is a good fit for teams who are developing solely within a Kafka-to-Kafka pipeline, teams who don't need or want another cluster for stream processing and analytics in the future, and operational and resilience requirements are simple or handled elsewhere.

Kafka Streams enables common stream processing functions like filtering, joins, aggregations, and enrichments directly on the data stream. Good use cases include building lightweight microservices, performing straightforward Extract, Transform & Load (ETL) jobs, and simple stream analytics apps. For more sophisticated use cases, consider Apache Flink **(see Data-in-Motion Philosophy on page 12)**.

Because Kafka Streams is an integral part of Cloudera's Kafka ecosystem, customers have the additional capability to build microservices apps that address complex security, governance, and audit requirements **(see Security and Governance Are First-Class Citizens on page 9).**

Kafka Connect

Kafka also includes a connectivity framework called Kafka Connect (KConnect). It is a good fit for simple use cases, but it is not ideal for use cases that require complex logic for error handling, filtering, or transformation of data. The complexity of managing network plug-ins and lack of support for schema evolution make Kafka Connect less effective than Apache NiFi for many use cases that require more fine-grained control of data. Apache NiFi is an incredibly versatile and efficient data movement tool with a no-code visual user interface and an extensive library of pre-built processors to support ingress and egress to and from Kafka.

Cloudera also supports NiFi stateless for KConnect. NiFi Stateless is a runtime option for where flows built in NiFi can be executed in the KConnect framework.

Any security and governance restrictions and complexities of a standalone Kafka Connect are mitigated when it is implemented as a part of Cloudera **(see Shared Data Experience on page 9)**.

Kafka Cruise Control

Kafka Cruise Control enables management and load balancing for large Kafka installations. It is the solution for platform teams that need premium management services that address hard problems such as frequent hardware or virtual machine failures, cluster expansion and reduction, and load skew among brokers. It solves these challenges by balancing clusters intelligently and with automated anomaly detection and remediation.

While it automatically balances partitions based on user-defined goals, Kafka Cruise Control also detects and actively addresses anomalies. For example, if there is a broker failure, Kafka Cruise Control will fix the cluster by removing the failed brokers. In the case of a disk failure, all of the offline replicas will be moved to healthy brokers.

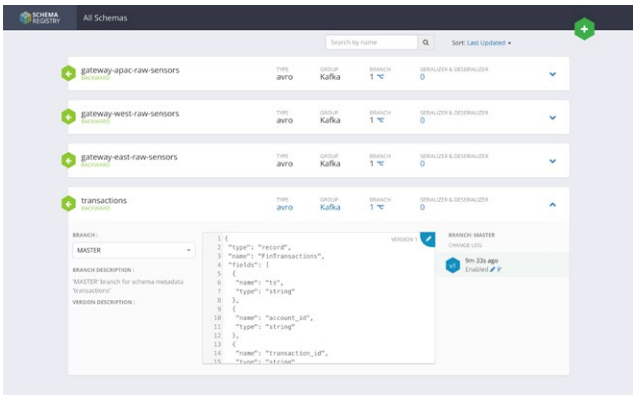
Schema Registry

Schema Registry is an important component of the Cloudera Kafka ecosystem because it enables teams to safely mitigate interruptions that occur due to schema mismatches. It manages, shares, and supports the evolution of all producer and consumer schemas across the Kafka landscape. Users can also avoid having to attach a schema to every piece of data.

As part of Cloudera's streams messaging capabilities, Schema Registry provides a shared repository of schemas that enables applications to flexibly interact with each other across the Kafka landscape by using the same schemas from end to end. This capability is particularly useful for managing data flows with schema-based routing. An example of this would be parsing a syslog event to extract the event type, and then routing it to a downstream Kafka topic based on that type.

The screenshot in Figure 2, below shows how you would use the Schema Registry UI to create schema groups and schema metadata, and how to add schema versions.

Figure 2



Stream Processing and Analytics: Choose the Right Tool For the Job

Companies deploying real-time streaming applications encounter many challenges and there are several ways to address them.

The best stream processing and analytics engine depends on the use cases and team makeup as well as various technological, operational, and organizational factors.

To understand the best fit for purpose across the stream processing and analytics engines, including Kafka Streams, Spark Structured Streaming, and Flink, read the whitepaper, "Choose the Right Stream Processing Engine for Your Data Needs."

Cloudera's Expertise Guarantees Your Success

With more experience across more production customers for more use cases, Cloudera Professional Services and Training (PS&T) is the leader in Kafka end-to-end services and support. For example:

- Cloudera PS&T helped one of the largest truck manufacturers in North America build a vehicle telematics pipeline using a streaming platform built with Kafka. Sensors continuously send data from over 150,000 trucks in North America every 2 minutes. The manufacturer processes data in real time and provides vehicle details, including speed and idling, fuel use, low tire pressure, and more.
- One of the largest energy companies in the world engaged Cloudera PS&T to build a mission-critical commodity trading platform using Kafka. Combined with other complementary Cloudera technologies, this platform processes extremely high throughput and low latency messages to generate trader alerts based on proprietary machine learning algorithms on streaming data.

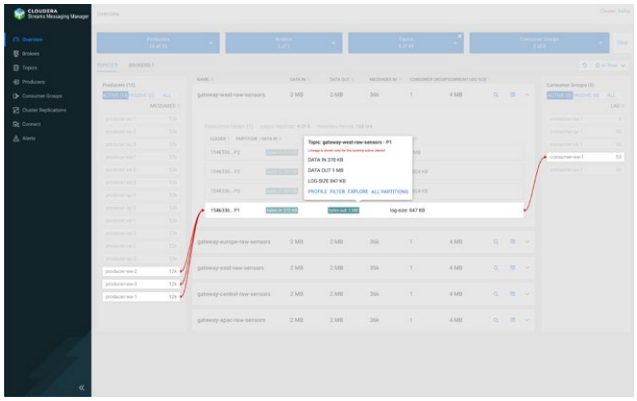
Cloudera PS&T is considered a trusted advisor by customers around the globe and across a range of industries.

Streams Messaging Manager

Perhaps the most critical component of the Cloudera Kafka ecosystem is Cloudera Streams Messaging Manager, a single monitoring and management dashboard that provides end-to-end visibility into how data moves across Kafka clusters between producers, brokers, topics, and consumers. It is a complete Kafka toolset that addresses the unique needs of DevOps, application development, platform operations, governance, and security teams.

As an example, the image in Figure 3 (below) shows interactive visualizations that enable visibility into how data flows across Kafka clusters.

Figure 3



An example of why Streams Messaging Manager is important to the Cloudera Kafka ecosystem is that you're able to optimize Kafka environments based on key performance insights gathered from various brokers and topics. Through tight integration with the Schema Registry service **(see page 6)** data from topics can be automatically deserialized when inspecting them through Streams Messaging Manager.

Streams Messaging Manager is a differentiating innovation that enables Kafka customers to gain complete visibility into their Kafka clusters. Read more about this in the whitepaper titled, "Manage, Monitor and Replicate Apache Kafka Across the Enterprise and Cloud."

+1000

Successful customer engagements

+300

Professional services consultants and architects globally

+150

Kafka experts

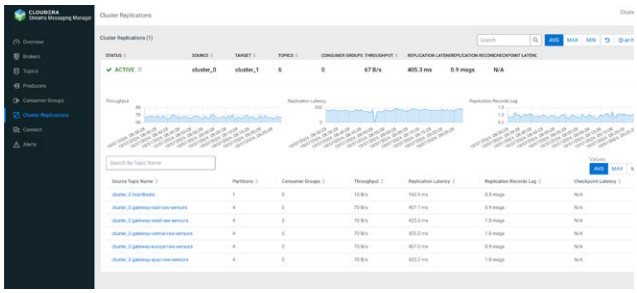
Streams Replication Manager

While Cloudera is focused on delivering key management and monitoring capabilities for IT users, we are also deliberate about ensuring business continuity and high availability for streaming architectures.

Streams Replication Manager is an enterprise-grade replication solution that enables fault-tolerant, scalable, and robust cross-cluster Kafka topic replication for several business-critical replication use cases such as high availability, disaster recovery, cloud migrations, geographic proximity, and others (see left sidebar **Get More out of Streams Replication Manager**).

Streams Replication Manager is built on the innovations that Cloudera brought to MirrorMaker, the original Kafka open-source messaging replication tool. Cloudera addressed some of the severe shortcomings of the original by unveiling MirrorMaker 2, which infuses the concepts of clusters, global configuration, and global management APIs.

Figure 4



SQL Stream Builder

Cloudera SQL Stream Builder enables developers, data analysts, and data scientists to write streaming applications using just 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 with its powerful interface and APIs.

SQL Stream Builder continuously runs SQL via Apache Flink. Developers don't need to understand the Java and Scala programming languages or the Flink API. The SQL job inherits and leverages the robust nature of Apache Flink. It can be restarted and retain its state, and it has massive scalability and a robust run-time framework.

Learn more about SQL Stream Builder.

Get More out of Streams Replication Manager

Replication is often associated with disaster recovery and high availability use cases. However, Streams Replication Manager enables additional business-critical use cases, some of which are described below.

- **Aggregation for analytics:** Aggregate data from multiple streaming pipelines and across multiple data centers to run batch analytics jobs that provide a holistic view of data from across the enterprise.
- **Data deployment after analytics:** This is the opposite of the aggregation use case in which the data generated by the analytics application in one cluster is broadcast to multiple clusters across data centers for end-user consumption.
- **Isolation:** Due to performance or security requirements, data must often be replicated between different environments to isolate access. In many deployments, the ingestion cluster is isolated from the consumption clusters.
- **Geographic proximity:** Replication moves data closer to the access location in geographically-distributed access patterns that require low latency.
- **Cloud migration:** As more enterprises manage on premises and cloud environments, they can use Kafka replication to migrate data to the public or private cloud and back.
- **Legal and compliance:** Much like the isolation use case, companies with legal and compliance requirements can use policy-driven replication to limit what data is accessible in a cluster.

For more about innovation in replication, read the white paper, "Manage, Monitor and Replicate Apache Kafka Across the Enterprise."

Security and Governance are First-Class Citizens

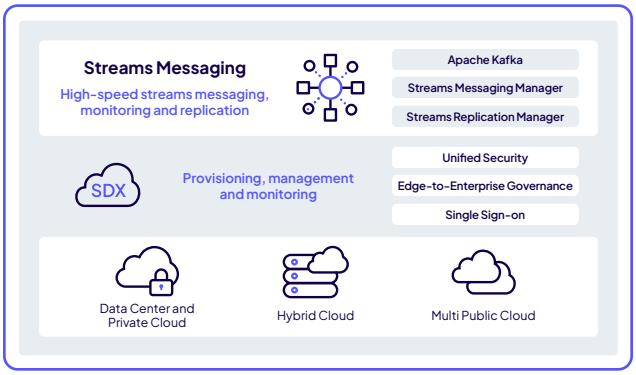
The previous pages described the strong capabilities of each key component that makes up the Kafka ecosystem within the Cloudera framework. Below, we describe why and how security and governance are first-class citizens in our platform.

Cloudera Shared Data Experience (SDX)

Cloudera SDX is a key differentiator from other platform providers. It enables the seamless integration of all parts of the Kafka ecosystem and safe, efficient, and consistent experience of deployments and migrations across all data environments: on premises, hybrid, private, or multi-cloud (see Figure 5 below).

With Cloudera SDX, data security, control policies, governance, and lineage are set once and automatically enforced on every data platform and across all components of the streaming architecture. Below, we briefly describe some of the key components of Cloudera SDX.

Figure 5

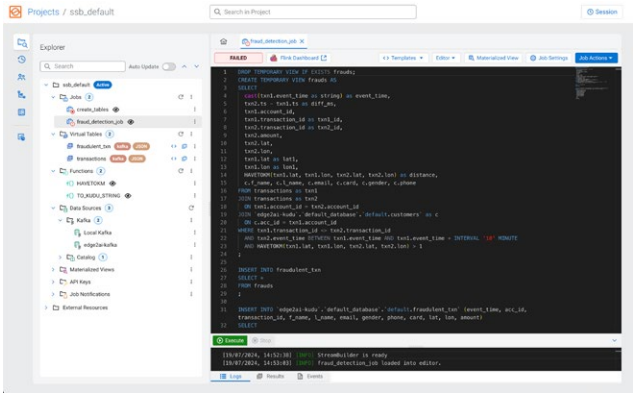


Apache Ranger

Apache Ranger provides a single pane of glass for enterprise-wide security with centralized, granular, and consistent access control. It addresses the four main pillars of security that are needed to support sensitive and regulated data use cases: identity, access, data protection, and visibility.

A sample of capabilities includes data encryption, dynamic row filtering, dynamic column masking, attribute-based and fine-grained access control, and the ability to enforce logical and physical separation of administrative duties across all infrastructures.

Figure 6



Delivering Iot-Enabled Predictive Maintenance, Vehicle Diagnostics and Management, and Route Optimization to Fleet and Truck Owners

Navistar is a leading manufacturer of commercial trucks, buses, defense vehicles, and engines, widely known for its International® Truck and IC Bus® brands.

To help fleet and vehicle owners move from a reactive approach to a more proactive, predictive maintenance model, Navistar needed to analyze a wider range of data in real-time, including vehicle sensor data. However, its traditional data warehouses couldn't support the growing volume of fast-moving, high-volume telematics data. |"As we collected more data, the analytic process slowed to a near halt on our legacy systems," said Ashish Bayas, Chief Technology Officer, for Navistar.

By choosing Cloudera, Navistar:

- Built an IoT-enabled remote diagnostics platform, called OnCommand® Connection, on Cloudera with Cloudera SDX
- Reduced maintenance costs by more than 30%
- Minimized downtime via early detection with OnCommand® Connection
- Enabled Navistar to ensure school buses run safely and on time

Read the complete customer success story, "Navistar: reducing maintenance costs more than 30 percent for connected vehicles."

In another example of seamless integration across Cloudera, gateway-based Single-Sign On (SSO) access is provided through Apache Knox while Apache Atlas enables full end-to-end data classification and auditing in order to perform analytics on regulated data and to redact sensitive data when needed.

Apache Knox

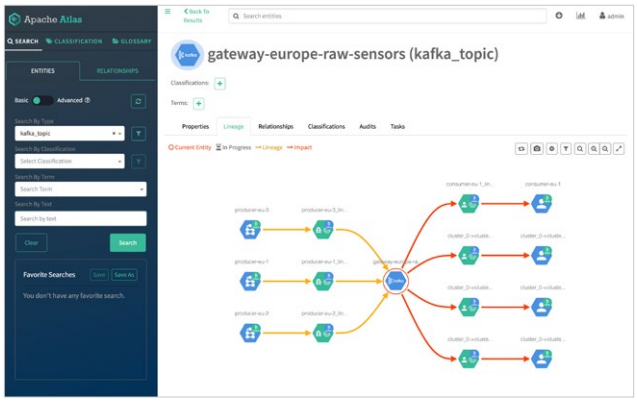
Apache Knox is a gateway-based SSO solution that simplifies security controls with seamless, secure user access to cluster data and the permissions needed to execute jobs while maintaining compliance with enterprise security policies. It provides a single entry point for all user interfaces across the Kafka ecosystem, alleviating the need to remember which node uses which services, for example. From a security administration point of view, Knox reduces the number of ports that need to be opened.

Knox is one of the reasons why all teams across the enterprise can safely and easily access their applications regardless of the data environment in which it is located.

Apache Atlas

Earlier, we described how Apache Atlas enhances Ranger security and data protection functions with full end-to-end data classification and audit capabilities. However, Atlas' enterprise-grade auditing, lineage, and governance capabilities are critical across the entire Kafka ecosystem. For example, Atlas provides access to the metadata and metrics about every Kafka topic and can produce complete data lineage and audit trails, even across multiple Kafka hops.

Figure 7



Cloudera Enables Multi-Cloud Support

Enterprises struggle to take their streaming data to the cloud because they often need to retain their on premises footprint for reasons like data sensitivity, data gravity, or security. In this scenario, they need to adopt a hybrid cloud architecture. For enterprises facing such challenges and with complex environments, Cloudera is an excellent platform to embrace multi-cloud and hybrid strategies.

With Cloudera, customers can extend the same holistic enterprise-wide on premises streaming experience to the cloud as well. For example, customers can quickly provision Kafka and Streams Messaging Manager in a public cloud in just a few minutes and can continue to take advantage of Cloudera's unified data security, governance, lineage, and control. Similarly, the other components of Cloudera data in motion, like Flow Management and Stream Processing & Analytics, are also available on Cloudera, giving customers complete control of how they deploy their streaming architecture across all environments.

This model enables development teams to leverage the same tools and platform across multiple environments and avoid the struggle of managing data with multiple tools. This also enables DevOps teams to easily spin up clusters of the streaming component of choice based on the specific use case they are handling. Additionally, Cloudera SDX offers a seamless security and governance experience across all components and environments Security and governance professionals can also feel confident in how data is managed and protected across their entire data estate.

The Complete Edge-to-Cloud Streaming Data Platform

This paper has focused primarily on the streams messaging aspects of the Kafka ecosystem and how to secure, monitor, balance, and replicate large-scale Kafka environments across on premises, hybrid, private, and public cloud environments. The data-in-motion reference architecture diagram in Figure 8 below puts this into perspective.

As a result of supporting customers through their data journeys, Cloudera has learned that it is not enough to have the best messaging solution at the heart of an end-to-end streaming architecture. As represented in the diagram **on page 10**, flow management and stream processing and analytics, must be unified with streams messaging capabilities. These three components, if properly integrated, will ensure a sustainable, scalable, and adaptable end-to-end streaming architecture and are the basis of Cloudera's data-in-motion philosophy.

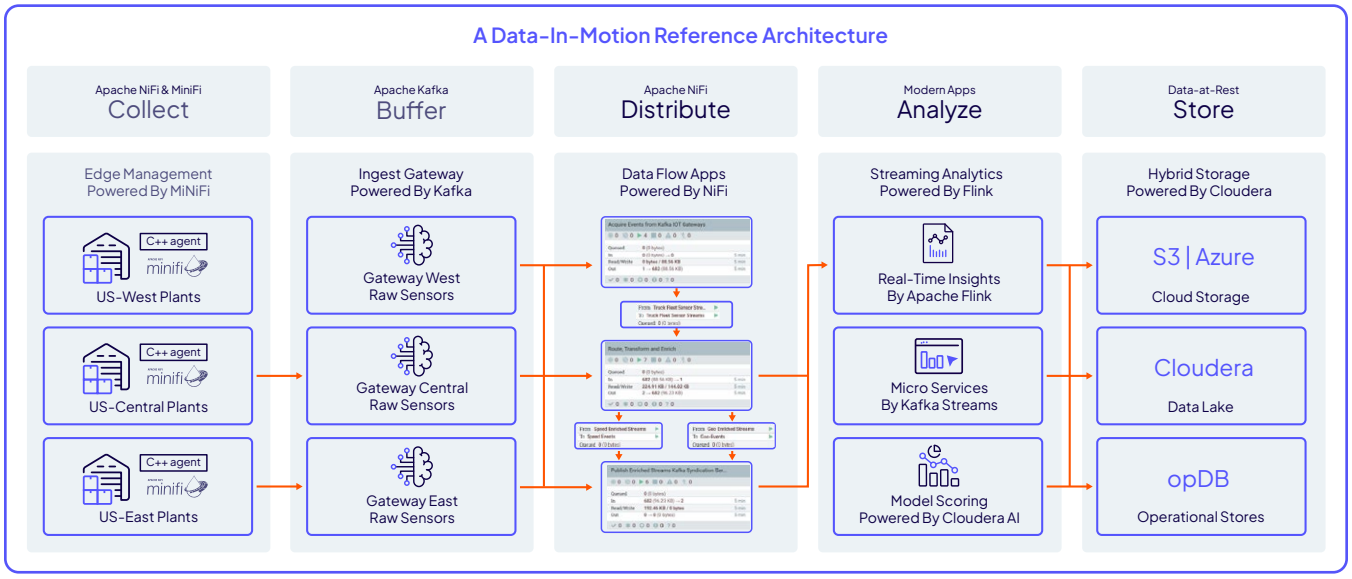
Real Life Data-in-Motion

Cloudera demonstrates its data-in-motion philosophy in real-life scenarios.

A global medical device manufacturer successfully modernized their messaging architecture to support a new line of implantable medical devices that generate more data more often and at a higher resolution than previous products.

- **Flow management:** Due to the private nature of medical data, the data flow was complex, requiring in-motion and at-rest encryption. NiFi's no-code user interface enabled business users to drive the initiative, reducing their reliance on the technology team.
- **Streams messaging:** Messaging volume jumped from quarterly reporting of device status to real-time health monitoring. Kafka enabled the business to scale that volume across multiple on premises and cloud environments.
- **Stream processing and analytics:** The company had to transition from batch to real-time data processing. Flink handled both processing models, and complex event processing is planned for the near future. The company only needs to adopt and support one type of stream processing and analytics engine.

Figure 8



CLOUDERA

Data-in-Motion Philosophy

Cloudera's philosophy is that best-in-class compute engines are required to adequately address the unique challenges of data flow management, streams messaging, and stream processing and analytics. As described in the table below, Cloudera delivers on that vision by supporting best-in-class data streaming compute engines while providing a high level of abstraction so that data teams can focus on the true business logic of building streaming data pipelines.

For additional insight into the below, read our solution brief, "Data-In-Motion Philosophy: A Blueprint for Enterprise-wide Streaming Data Architecture".

THE THREE TENETS OF A UNIFIED END-TO-END STREAMING ARCHITECTURE			
Tenet	**Compute Engine**	**Why**	**Importance of Cloudera**
Flow management, broadly speaking, refers to the collection, distribution, and transformation of data across multiple points of producers and consumers.	APACHE nifi	Apache NiFi is a real-time integrated data logistics and simple event processing platform. It is best-in-class because it inherently addresses the three important aspects of flow management: extensible tooling, ease of use, and data provenance.	As with the entire Kafka ecosystem, Cloudera provides a unified platform to handle the complexities of connecting, managing, and integrating these best-in-class engines through a high level of abstraction. This means that your teams can focus on the true business logic that goes into building an end-to-end data pipeline because Cloudera seamlessly renders that logic across the respective engines. This shields the user from that complexity.
Streams messaging is the provisioning and distribution of messages between producers and consumers.	kafka	Kafka has emerged as the single central backbone of streaming architectures for large organizations because it addresses the fundamental challenges of scalability and is highly optimized for both ad-hoc and sustained exchange of messages.	
Stream processing and analytics is how you generate real-time analytical insights from the data streaming between producers and consumers.	Flink	Apache Flink is a distributed processing engine and a scalable data analytics framework that can process millions of data points or complex events very easily and deliver predictive insights in real-time. It is best-in-class because it gives you loads of technological and operational control to address some of the more sophisticated analytic use cases.	

Cloudera is Superior in the Kafka Space

In this paper, we described how Cloudera delivers the best Kafka ecosystem by not only delivering the streams messaging aspects of securing, monitoring, balancing, and replicating large-scale Kafka environments but also by incorporating best-in-class flow management and stream processing and analytics engines to ensure a sustainable, scalable and adaptable end-to-end streaming architecture.

The following summarizes the key reasons why Cloudera delivers the beat Kafka ecosystem for enterprise streaming workloads:

Commitment to the open source community — Cloudera is dedicated to the Kafka ecosystem and continues to be actively involved with the Kafka open-source community through deep engineering relationships with other Kafka committers. This relationship has led to critical innovations and product improvements, many of which have been described in this paper.

Respect for the voice of the customer — Cloudera is constantly listening to and incorporating customer feedback into the development and enhancements of our products. Streams Messaging Manager, which we created directly as a response to customer feedback regarding visibility into their Kafka workloads, is a prime example of this commitment.

Kafka innovations — Kafka is one of the many open-source projects Cloudera is committed to across the platform. However, it plays a crucial role in our streaming services, and therefore we have been introducing groundbreaking innovations in this area for our customers. Other than Streams Messaging Manager, Streams Replication Manager is another example of best-in-class engineering and innovation. This service is based on Mirrormaker 2, which is a much-needed innovation that our engineers delivered for the open-source community.

Security and governance are table stakes — Cloudera SDX is a true differentiator compared to other offerings. Customers need a unified security and governance layer across all components and environments. SDX enables customers to understand end-to-end lineage of streaming data from the edge to the cloud across ingestion, messaging, and stream processing components, which is critical for companies that are struggling with compliance and regulations.

Multi-cloud and hybrid cloud support — Cloudera provides a true hybrid architecture, and customers can quickly provision streaming platform components in private and public clouds while leveraging the unified data security, governance, lineage, and control across all environments.

Global customer support — Cloudera supports massive Kafka deployments for hundreds of customers delivering near-real-time insights across multiple industry verticals. Beyond software support, Cloudera's global team of professional services and training staff make it easy for our customers to implement true end-to-end streaming architectures.

One platform — Cloudera delivers a true edge-to-cloud streaming data platform. Instead of adopting Kafka as simply a point solution, Cloudera addresses the data management challenges of the enterprise across all aspects of their data-in-motion journey with one integrated platform.

Learn more about Kubernetes Operator and Cloudera Streaming

Learn more about Cloudera

Learn more about Cloudera Stream Processing

Positive Impact of a Complete Streaming Platform

One of the world's largest steel mining businesses worked with Cloudera experts to significantly increase steel production by expanding their current streams messaging Kafka ecosystem with MiNiFi edge management, NiFi flow management, and a Cloudera data lake for high capacity data storage.

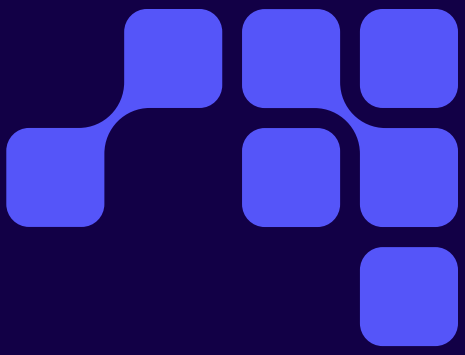
The business impact includes:

+6.5%
Increase in productivity

100K TONS
of new steel

0.7 SECONDS
for data to go through the solution

About Cloudera



Cloudera is the only true hybrid platform for data, analytics, and AI. With 100× more data under management than other cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable, trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to be able to use their data to solve the impossible — today and in the future.

To learn more, visit Cloudera.com and follow us on LinkedIn and X. Cloudera and associated marks are trademarks or registered trademarks of Cloudera, Inc. All other company and product names may be trademarks of their respective owners.

CLOUDERA

Cloudera, Inc. | 5470 Great America Pkwy, Santa Clara, CA 95054 USA | cloudera.com