K2D (Kafka to Druid) stack for real-time analytics applications

"Operating multi-tenant services requires fine-grained visibility down to the individual user, tenant, or application behavior, where most traditional monitoring stacks fail to scale or become cost-prohibitive."

- Xavier Léauté and Zohreh Karimi, Lead Engineers at Confluent

How leading organizations leverage Confluent/Kafka together with Imply/Druid

The new enterprise imperative

In order to succeed in today's increasingly digital world, modern organizations need real-time, granular insights into their business and operations. These insights include a real-time understanding of operational attributes such as uptime, performance, security, or usage-based billing of the applications and services they deliver, as well as a real-time understanding of business attributes such as product usage or user behavior. In addition, it's often not enough for a modern organization to gain these insights internally; increasingly organizations must share these real-time operational and business insights externally with the customers and partners they serve.

### Real-time analytics applications based on K2D

To gain these real-time operational and business insights, and to share them with customers and partners, modern organizations are building real-time analytics applications based on the K2D (Kafka® to Druid®) architecture.

As examples, **Netflix** built an internal-facing observability application to minimize any delay from when videos start to play, **Salesforce.com** built an internal-facing edge intelligence application to understand the health of their services, and **Confluent** built an internal-facing observability application to understand attributes such as performance, uptime, and billing for Confluent Cloud. To extend insights to customers, **Citrix** built their external-facing Citrix Analytics Service to share security and performance insights with their customers, **Confluent** built the external-facing Health+ to share Confluent Cloud monitoring and health insights with their customers, and **Imply** built their external-facing Imply Clarity to share monitoring and health insights with their customers.

These organizations, and many more, chose the K2D architecture because data warehouses, existing databases and legacy monitoring stacks fail to deliver on key technical requirements. Those existing solutions do not enable internal and external users to rapidly interact with data at massive data volumes. They struggle with delivering insights from modern data infrastructures centered around data in motion. And, they are not built to deliver compelling price/performance when the application serves a large number of end users and concurrent queries.

Widespread adoption of the K2D architecture

Successfully building real-time analytics applications starts with having a central nervous system where data in motion can move freely throughout the organization. That data in motion must then feed into a real-time analytics database built for analytics in motion. This combination of data in motion with analytics in motion provides the foundation for the development of real-time analytics applications.
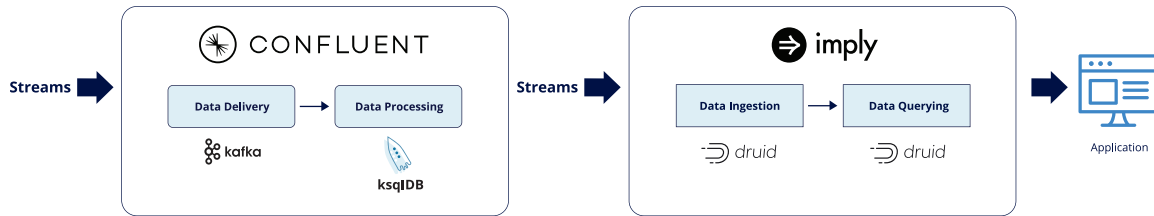
K2D Stack: Kafka serves as the event streaming platform, Druid serves as the real-time analytics database

In the K2D stack, Kafka serves as the central repository of streams and provides high throughput event delivery. Druid enables events to be explored immediately after they occur and enables analysis of real-time events coupled with historical events. Druid can ingest data at a rate of millions of events per second, functioning as a natural complement to an event streaming platform such as Kafka. Together, Kafka and Druid make building real-time analytics applications a reality.

Case Study: Confluent Health+

Confluent Health+ provides Confluent Cloud customers with the visibility needed to ensure the health of their data-in-motion infrastructure and to minimize business disruption. Health+ offers intelligent alerts, cloud-based monitoring and visualizations, and a streamlined support experience.

"We built an observability platform powered by Kafka and Druid. This solution ingests over 3.5 million events per second and handles hundreds of queries on top of that. And this gives us real-time insights into the operations of thousands of these Kafka clusters within Confluent Cloud."

- Jay Kreps, CEO, Confluent

"Leveraging Druid as part of our stack means we don't shy away from high-cardinality data which means we can find the needle in the haystack. As a result, our teams can detect problems before they emerge and quickly troubleshoot issues to improve the overall customer experience. The flexibility we have with Druid also means we can expose the same data we use internally also to our customers, giving them detailed insights into how their own applications are behaving."

- Xavier Leaute and Zohreh Karimi, Lead Engineers at Confluent

Case Study: Imply Clarity

Imply Clarity is a visual analysis tool that delivers real-time monitoring and performance tuning. It's designed to catch problems before they occur and then quickly visualize, explore and drill down into the root cause.

"Across our company, we build based on Confluent Cloud / Kafka in order to harness the power of our data in motion. By leveraging Imply technology with Confluent Cloud / Kafka, we build internal observability applications, technical workshops, and our external-facing Imply Clarity offering."

- Fangjin Yang, CEO, Imply

"Because of the high volume of ingested real-time events required to fuel the experience, we leverage Apache Kafka as our event streaming platform. This gives us a platform built for high throughput and reliable event delivery. Ultimately, we're able to provide an exceptional customer experience by identifying and resolving issues in real-time."

- Gabriel Tavridis, Sr. Director of Product Management at Imply

About Confluent

Confluent® is pioneering a fundamentally new category of data infrastructure focused on data in motion. Confluent's cloud-native offering is the foundational platform for data in motion—designed to be the intelligent connective tissue enabling real-time data, from multiple sources, to constantly stream across the organization. With Confluent, organizations can meet the new business imperative of delivering rich, digital front-end customer experiences and transitioning to sophisticated, real-time, software-driven backend operations.

About Imply

Imply®, founded by the original creators of Apache Druid®, develops an innovative database purpose-built for modern analytics applications. Imply is driving a new era in data analytics, called Analytics in Motion, where interactive queries, real-time and historical data at unlimited scale, combine with the best price/performance, to realize the full potential of data.