

MARKET NOTE

Not Just Messaging: Confluent Points the Way to a More Comprehensive Approach to Kafka and Streaming Data Platforms

Giovanni Cervellati

EXECUTIVE SNAPSHOT

FIGURE 1

Executive Snapshot: Confluent Points the Way to a More Comprehensive Approach to Kafka and Streaming Data Platforms

This IDC Market Note discusses the main takeaways from the Kafka Summit in London, hosted by Confluent on April 25 and 26, 2022. Data is seen as a virtually infinite sequence of events, which refer to any possible type of action, incident, or change that is identified or recorded by software or applications. Kafka is an open source streaming event platform, and fully managed platform Confluent, built on open source Apache Kafka, can handle real-time data-in-motion services and greatly simplify Kafka users' journey, thanks to prebuilt connectors to various data sources, and its data management, governance, and security elements.

Key Takeaways

- Streaming data is data that is generated continuously by data sources, which typically send in data records simultaneously and in small sizes (order of kilobytes) in real time. Therefore, streaming data platforms are used to collect, process, store, and integrate these live data at scale. Kafka is an open source software product known as a streaming event platform.
- Traditional use cases for streaming data platforms include web applications, ecommerce purchases, in-game player activity, and information from social networks and messaging. The focus is shifting toward leveraging streaming events as part of the data ecosystem. Kafka usage has spread from being just a platform used by developers for messaging to being a more comprehensive universal platform for applications and data pipelines.
- The vision of Kafka supporters such as Confluent is that streaming data platforms will evolve into a "central nervous system" of data, connecting and analyzing what is on, what is off, what is changing, what is static, and what is happening in real time.
- Fully managed platform Confluent, built on open source Apache Kafka, can greatly simplify Kafka users' journeys, thanks to its pre-built connectors to various data sources, and its data management, governance, and security elements. Its virtually infinite storage capability may result in a new way of looking at data storage in the future.

Source: IDC, 2022

IN THIS MARKET NOTE

This IDC Market Note discusses the main takeaways from the Kafka Summit hosted by Confluent, which took place in London on April 25 and 26, 2022. The summit showcased use cases and best practices, discussed the future of streaming technologies, and contained practical sessions, which provided a broad view of current and future Kafka (and Confluent) developments. Kafka Summit attendees included developers, architects, and data engineers and others keen to learn about streaming data technologies.

Streaming Data: From Messaging to the Central Nervous System

Streaming data is data that is generated continuously by data sources, which typically send in data records simultaneously and in small sizes (order of kilobytes) in real time. Therefore, streaming data platforms are used to collect, process, store, and integrate these live data at scale.

Streaming data includes a wide variety of data such as log files generated by customers using mobile or web applications, ecommerce purchases, in-game player activity, and information from social networks, financial trading floors, or geospatial services. Implementing these services helps companies stay competitive in the wider digital marketplace. Weaving real-time data into the data fabric for low-latency business decisions has a massive impact on services such as telcos' customer relationship and customer journey management.

Kafka is an open source software product known as a streaming event platform: data is seen as a virtually infinite sequence of events, which refer to any possible type of action, incident, or change that is identified or recorded by software or applications. This, of course, encompasses data. In recent years, this fascinating concept has often replaced the old batch mindset: organizations felt stream processing is needed as business demands impose more real-time requirements on developers and architects. Confluent is an organization that leverages Kafka as the core of its own platform, which is fully managed. Implementing these services helps companies stay competitive in the wider digital marketplace. Weaving real-time data into the data fabric for low-latency business decisions has a massive impact on services such as telcos' customer relationship and customer journey management.

As an IDC 2022 *Streaming Data Survey* shows, the most digitally mature enterprise sizes are rethinking their software architectures with real-time context at the forefront. Virtually none of the "best in class" declared to have streaming data strategies neither in place nor planned. Moreover, more than 80% of all respondents say they will research or invest in new streaming capabilities in the next 12 to 18 months. The most common use cases are those in which fast action is an ultimate necessity; these include security/threat management, customer experience, and real-time financial data.

The shift from batch data transfers to data streaming aligns with all other innovations in today's data ecosystem: on-prem deployments are moving to the cloud, monoliths are evolving into microservices, and so on. All this implies the concept of decentralization, which is a logical and inherent part of the streaming/event structure itself.

Kafka has shifted from being just a platform used by developers for messaging to being a more comprehensive universal platform for applications and data pipelines. There are other streaming data platforms available, but Kafka is among the most widely used today.

The vision of Kafka supporters such as Confluent is that streaming data platforms will evolve into a "central nervous system" of the data, connecting and analyzing what is on, what is off, what is changing, what is static, and what is happening in real time.

Moreover, envisioning a streaming data platform as central to a data ecosystem drives new innovative ways of structuring data projects: direct streaming ingestion for predictive model training (even directly from data at the edge), the deployment of models and analytics in real-time with consequent live-action, enhancement of possible security/data breach control live. Finally, we ought to consider two

features of Kafka: its ample storage capabilities and the KSQL language enabling data processing in a streaming fashion using an SQL-like dialect. These features provide a new way of looking at Kafka: executing batch analytics directly on Kafka without storing it in an ample data storage and reading all data from it could enable us to manage all our data without a data lake in the future.

Full-Scale Data Streaming Platforms Are more than Kafka

For these reasons, real-time data has become more of a necessity than a plus. But a developer-oriented, open source high-coding platform such as Kafka – although very popular – has its limits. As the amount of data starts to scale, it requires too much management.

Generally, businesses that are new to streaming event processing will start with an experimentation phase (which is most suitable for Kafka). However, as they implement this project as mission critical and scale to connect Kafka to one or more lines of business, then Kafka will probably not be sufficient.

As we just mentioned, operating Kafka on your own can be strenuous, especially as it lacks some important management features from which an organization may want to be freed. As the organization scales, it would need more personnel to be able to handle it. This will place a burden on those who use it as well as increase costs. Implementing Kafka as a central platform for data, in fact, will require organizations to enable connectivity to all data sources, data applications, and end/edge-applications – in doing so, they will also have to pay for integration solutions for capturing data and "event" from legacy systems. Moreover, Kafka is not cloud native, and moving it to the cloud may not be efficient.

This excessive and growing management toll on organizations using Kafka drove the search for new, easier, and better-connected solutions. Confluent is one such solution. Built on open source Apache Kafka, it can be deployed as a cloud-native SaaS offering or as self-managed software. The cloud-native offering can be run on a multicloud or hybrid cloud infrastructure. It is built to handle real-time data in motion services, and so far, its most significant users have been from the financial, retail and ecommerce, manufacturing, and media and entertainment sectors.

FTE costs, infrastructure spend, and business risks related to a possible unplanned downtime or breach of Kafka are avoidable with a managed platform. Prebuilt, self-managed, and easier-to-use connectors to Kafka (and most potential data sources and systems); data management and data observability; data discovery capabilities; and data security and governance are features that could be enabled by using a full-scale data platform. Moreover, using Kafka in a hybrid cloud environment is hindered by difficulties in connecting to various environments in real time, which platforms such as Confluent are exactly built around. Finally, notable storage capabilities in Kafka become virtually almost-limitless with Confluent.

Considering all of this, it would be tough to implement the trusted quality data streams that are required to reach the already mentioned "central nervous system" architecture. Implementing a more comprehensive streaming data solution will make it much easier to enable real-time connectivity for streaming all of a company's data, both ingesting them and injecting them wherever needed.

IDC'S POINT OF VIEW

Real-time data has surely become a key topic for most organizations. Seen originally just as messaging platforms, streaming data platforms are gaining ground in more fields. As security, governance and visibility have developed into crucial areas for data, relying on batch timing for data responses will prove to be insufficient for many.

The development of real-time streaming data solutions will eventually become a vital part of many organizations, initially starting with some use cases and expanding to more lines of business. Growing interest around streaming data platforms is the proof of their increasing importance for industries, with just a tiny segment of companies declaring to not be interested in implementing them. Full-scale data

streaming platforms are already a step further, enabling some organizations to build their entire data infrastructures around streaming data. Furthermore, the use of streaming data platforms for real-time security checks and data safety tasks has an enormous impact and potential for its future development and spread.

However, the world of streaming data has not gained a very widespread knowledge other than classical usage for applications and tasks such as messaging. Software such as Confluent are a good starting point in this regard. However, what may accelerate this acknowledgment of its many use cases and its centrality in data platforms may come from moving away from its excessively developer-oriented nature and moving to an easier adoption path.

In particular, greatly improving no-code features will enable more business and analytics users to leverage the advantages of real-time data without having to rely on developers, who are generally quite far from the business mindset. Finally, another issue may be integrating streaming data platforms as the "central nervous system" in organizations that already have too many software and platforms in place in their data and analytics ecosystems. Capabilities such as data catalog, data lineage, or master data management have a huge impact on modern data stacks; being able to leverage all these functionalities fully while getting rid of the equivalent legacy systems will be a final step for a winning strategy.

LEARN MORE

Synopsis

This IDC Market Note discusses the main takeaways from the Kafka Summit in London, hosted by Confluent, on April 25 and 26, 2022. The development of real-time streaming data solutions will help them become a vital part of many organizations; growing interest in streaming data platforms is proof of their increasing importance to industries. Fully managed platform Confluent, built on open source Apache Kafka, can handle real-time data-in-motion services and greatly simplify Kafka users' journeys, thanks to its prebuilt connectors to various data sources as well as its data management, governance, and security elements. Finally, its virtually infinite storage capability may result in a new way of looking at data storage in the future.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC Central Europe GmbH

IDC Central Europe GmbH - Deutschland & Schweiz
Hanauer Landstraße 182 D
60314 Frankfurt am Main, Deutschland
+49 (0)69 90502-0
Twitter: https://twitter.com/idc_deutschland
www.idc.de

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or web rights.

Copyright 2022 IDC. Reproduction is forbidden unless authorized. All rights reserved.

