

## ESTUDIO DE MERCADO

# No es solo un mensaje: Confluent señala el camino hacia un enfoque más completo de Kafka y de las *streaming data platforms*

Giovanni Cervellati

## RESUMEN EJECUTIVO

### Imagen 1.

## Resumen ejecutivo: Confluent señala el camino hacia un enfoque más completo de Kafka y de las *streaming data platforms*

Este estudio de mercado de IDC analiza las principales conclusiones del Kafka Summit en Londres, que fue organizado por Confluent los días 25 y 26 de abril de 2022. Los datos se ven como una secuencia virtualmente infinita de eventos que se refieren a cualquier tipo posible de acción, incidente o cambio, que es identificado o registrado por el *software* o las aplicaciones. Kafka es un *open-source streaming event platform*, y esta plataforma de Confluent está totalmente gestionada. Se construyó sobre el *open-source* de Apache Kafka, puede manejar servicios de *data in motion* en tiempo real y simplificar en gran medida la experiencia de los usuarios de Kafka; gracias a los *connectors* prediseñados ante varias fuentes de datos, y a sus elementos de gestión de datos, gobernanza y seguridad.

### Puntos clave

- El *data streaming* son datos que se generan continuamente por fuentes, que suelen enviar registros de estos datos simultáneamente y a un tamaño reducido (del orden de los *kilobytes*) en tiempo real. Por lo tanto, las *streaming data platforms* se utilizan para recoger, procesar, almacenar e integrar estos datos en tiempo real a escala. Kafka es un *software open-source* que se conoce como *streaming event platform*.
- Los casos de uso tradicionales de *streaming data platforms* incluyen las aplicaciones web, las compras en el comercio electrónico, la actividad en los juegos y la información de las redes sociales y los mensajes. Este enfoque está cambiando hacia el beneficio de los *streaming events* como parte del ecosistema de datos. Kafka ha pasado de ser solo una plataforma que se utilizaba para mensajería por los desarrolladores, a ser una plataforma universal más completa para aplicaciones y *data pipelines*.
- La visión de los partidarios de Kafka, como Confluent, es que *streaming data platforms* evolucionarán hasta convertirse en un «sistema nervioso central» de datos, que conectará y analizará todo lo que esté encendido o apagado, cambiando o parado y, además, en tiempo real.
- La plataforma totalmente gestionada de Confluent, fue construida sobre el *open-source* de Apache Kafka y puede simplificar, en gran medida, las experiencias de los usuarios de Kafka, gracias a sus *connectors* prediseñados ante varias fuentes de datos, y a sus elementos de gestión de datos, gobernanza y seguridad. Su capacidad de almacenamiento prácticamente infinita puede dar lugar a una nueva forma de ver el almacenamiento de datos en el futuro.

Fuente: IDC, 2022

Mayo de 2022, IDC #EUR149117722

## EN ESTE ESTUDIO DE MERCADO

Este estudio de mercado de IDC analiza las principales conclusiones de Kafka Summit en Londres, dicho evento fue organizado por Confluent los días 25 y 26 de abril de 2022. En la cumbre se mostraron casos de uso y las mejores prácticas, así mismo, se debatió el futuro de las tecnologías de *streaming* y contó con sesiones prácticas, que proporcionaron una amplia visión del desarrollo actual y futuro de Kafka (y Confluent). Entre los asistentes al Kafka Summit se encontraban desarrolladores, arquitectos e ingenieros de datos y otras personas interesadas en aprender sobre las tecnologías del *streaming data*.

### **Streaming data: del mensaje al sistema nervioso central**

Los *streaming data* es información que se genera continuamente por fuentes de datos, que suelen enviar registros de estos datos simultáneamente y a tamaño reducido (del orden de los *kilobytes*) en tiempo real. Por lo tanto, *streaming data platforms* se utilizan para recoger, procesar, almacenar e integrar estos datos en tiempo real a escala.

Los *streaming data* incluyen una amplia variedad de datos, como los archivos de registro que se generan por los clientes que utilizan aplicaciones móviles o web, las compras del comercio electrónico, la actividad de los jugadores en los juegos y la información de las redes sociales, los mercados de valores o los servicios geoespaciales. La implantación de estos servicios ayuda a las empresas a seguir siendo competitivas en el mercado digital general. La integración de los datos en tiempo real en el tejido de datos para la toma de decisiones empresariales de baja latencia tiene un gran impacto en servicios como la gestión de la relación con el cliente y la experiencia de este en las empresas de telecomunicaciones.

Kafka es un producto *software* de *open-source* que se conoce como *streaming event platform*: los datos se ven como una secuencia virtualmente infinita de eventos, que se refieren a cualquier tipo posible de acción, incidente o cambio que sea identificado o registrado por un *software* o por aplicaciones. Esto, por supuesto, incluye los datos. En los últimos años, este fascinante concepto ha sustituido a menudo a la antigua mentalidad de los lotes: las organizaciones consideraban que *stream processing* era necesario a medida que las demandas empresariales imponían más requisitos en tiempo real a los desarrolladores y arquitectos. Confluent es una organización que potencia a Kafka como núcleo de su propia plataforma; la cual está totalmente gestionada. La implantación de estos servicios ayuda a las empresas a seguir siendo competitivas en el mercado digital general. La integración de los datos en tiempo real tiene un gran impacto en servicios como la gestión de la relación con el cliente y la experiencia de este en las empresas de telecomunicaciones.

Como muestra una encuesta de IDC sobre *streaming data* en 2022, las empresas más maduras, desde el punto de vista digital, se están replanteando sus arquitecturas de *software* con el contexto en tiempo real. Ninguno declaró tener estrategias de *streaming data*, ni en marcha ni planificadas. Asimismo, más del 80 % de los encuestados afirma que investigará o invertirá en nuevas oportunidades de *streaming* en los próximos 12 o 18 meses. Los casos de uso más comunes son aquellos en los que la acción rápida es una necesidad imperiosa; entre ellos, la gestión de la seguridad o amenazas, la experiencia del cliente y los datos financieros en tiempo real.

El cambio de las transferencias de datos por lotes a *streaming data* está en consonancia con todas las demás innovaciones del ecosistema de datos actual: los despliegues *on-prem* se están trasladando a *cloud*, los monolitos están evolucionando hacia los microservicios, etc. Todo esto implica el concepto de descentralización, que es una parte lógica e inherente a la propia estructura *streaming* o *event*.

Kafka ha pasado de ser solo una plataforma que se utiliza para mensajería por los desarrolladores, a ser una plataforma universal más completa para aplicaciones y *data pipelines*. Existen otras en *streaming*, pero Kafka es una de las que más se utiliza actualmente.

La visión de los partidarios de Kafka, como Confluent, es que *streaming data platforms* evolucionarán hasta convertirse en un «sistema nervioso central» de datos, que conectará y analizará todo que esté encendido o apagado, esté cambiando o parado y lo que esté sucediendo en tiempo real. Además, la visión de *streaming data platforms* como elemento central de un ecosistema de datos impulsa nuevas formas innovadoras de estructurar los proyectos de datos: el consumo directo de *streaming* para el entrenamiento de modelos predictivos (incluso directamente desde los datos en el límite), el despliegue de modelos y análisis en tiempo real con la consiguiente acción en directo, mejora la seguridad o control de la violación de datos en tiempo real.

Por último, debemos tener en cuenta dos características de Kafka: sus amplias capacidades de almacenamiento y el lenguaje KSQL que permite *streaming data processing* al utilizar un dialecto similar a SQL. Estas características proporcionan una nueva forma de ver a Kafka: ejecutar análisis por lotes directamente en Kafka y leer todos los datos desde este podría permitirnos gestionar todos nuestros datos sin tener que almacenarlos.

## Las *streaming data platforms* a gran escala son más que Kafka

Por estos motivos, los datos en tiempo real se han convertido más en una necesidad que en una ventaja. Sin embargo, una plataforma *open-source* orientada a los desarrolladores y de alta codificación como Kafka, aunque es muy popular, tiene sus límites. Cuando la cantidad de datos empieza a incrementar, requiere una gran gestión.

Por lo general, las empresas que son nuevas en *streaming event processing* comenzarán con una fase de experimentación (que es la más adecuada para Kafka). No obstante, a medida que implementen este proyecto como de misión crítica y escalen para conectar Kafka a una o más líneas de negocio, entonces Kafka probablemente no será suficiente.

Como acabamos de mencionar, operar con Kafka por cuenta propia puede ser agotador, especialmente porque carece de algunas características importantes de gestión de las que una organización puede querer liberarse. A medida que la organización se amplía, se necesitará más personal para poder controlarla. Esto supondrá una carga para los que la utilizan, así como un aumento de los costes. La implementación de Kafka como plataforma central de datos, de hecho, requerirá que las organizaciones habiliten la conectividad a todas las fuentes de datos, aplicaciones de datos y aplicaciones finales o de límite. Al hacerlo, también tendrán que pagar por soluciones de integración para recoger datos y «eventos» de los sistemas heredados. Además, Kafka no es *cloud native*, y moverlo aquí puede no resultar eficiente.

Esta excesiva y creciente carga de gestión para las organizaciones que utilizan Kafka impulsó la búsqueda de soluciones nuevas, más fáciles y mejor conectadas. Confluent es una de esas soluciones. Se creó a partir del *open-source* Apache Kafka y puede desplegarse como una oferta SaaS *cloud native* o como *software* autogestionado. La oferta *cloud native* puede ejecutarse en una infraestructura de *cloud* múltiple o híbrida. Está construido para gestionar *data in motion* en tiempo real y, hasta ahora, sus usuarios más significativos han sido de los sectores financiero, minorista y de comercio electrónico, de fabricación, de medios de comunicación y entretenimiento.

Los costes de los ETC, el gasto en infraestructura y los riesgos empresariales que están relacionados con un posible tiempo de inactividad no planificado o un fallo de Kafka que son evitables con una plataforma gestionada. Los *connectors* prediseñados, autogestionados y más fáciles de usar con Kafka (y con la mayoría de las posibles fuentes de datos y sistemas); la gestión y la observabilidad de los datos; las capacidades de descubrimiento de datos; y la seguridad y el gobierno de los datos son características que podrían habilitarse mediante el uso de una plataforma de datos a gran escala. Es más, el uso de Kafka en un entorno de *cloud* híbrida se ve obstaculizado por las dificultades para conectarse a varios entornos en tiempo real, algo para lo que plataformas como Confluent están exactamente construidas. Por último, las notables capacidades de almacenamiento de Kafka se vuelven prácticamente ilimitadas con Confluent.

Teniendo en cuenta todo esto, sería difícil implementar *data streams* de calidad y de confianza que se requieren para alcanzar la arquitectura del «sistema nervioso central». La implantación de una solución de *streaming data* más completa facilitará la conectividad en tiempo real para el *streaming* de todos los datos de una empresa, tanto para su ingesta como para su inyección allí donde sea necesario.

## PUNTO DE VISTA DE IDC

---

Los datos en tiempo real se han convertido, sin duda, en un tema clave para la mayoría de las organizaciones. Se consideraban en un principio solo como plataformas de mensajes, ahora, *streaming data platforms* están ganando terreno en más campos. Dado que la seguridad, la gobernanza y la visibilidad se han convertido en áreas cruciales para los datos, depender de la sincronización de lotes para las respuestas de datos resultará insuficiente para muchos.

El desarrollo de soluciones de *streaming data* en tiempo real acabará convirtiéndose en una parte vital de muchas organizaciones, comenzando inicialmente con algunos casos de uso y expandiéndose a más líneas de negocio. El creciente interés en torno a las *streaming data platforms* es la prueba de su creciente importancia para las industrias, ya que solo un pequeño sector de empresas declara no estar interesado en implementarlas.

Las *streaming data platforms* a gran escala ya van un paso más allá, permitiendo a algunas organizaciones construir toda su infraestructura de datos en torno al *data en streaming*. Además, el uso de estas plataformas para los controles de seguridad en tiempo real y las tareas de seguridad de los datos tiene un enorme impacto y potencial para su futuro desarrollo y difusión.

Sin embargo, el mundo del *streaming data* no se ha extendido mucho más allá del uso clásico para aplicaciones y tareas como los mensajes. Programas como Confluent son un buen punto de partida en este sentido. Sin embargo, lo que puede acelerar este reconocimiento de los muchos casos de uso y la centralidad en las plataformas de datos puede venir de alejarse de la naturaleza excesivamente orientada a los desarrolladores y pasar a una vía de adopción más fácil.

En particular, la mejora de las funciones sin código permitirá a un mayor número de usuarios empresariales y analíticos aprovechar las ventajas de los datos en tiempo real sin tener que depender de los desarrolladores, que suelen estar bastante alejados de la mentalidad empresarial. Por último, otro problema puede ser la integración de las plataformas de *streaming data* como «sistema nervioso central» en las organizaciones que ya cuentan con demasiados programas y plataformas en sus ecosistemas de datos y análisis. Aptitudes como el catálogo de datos, el linaje de datos o la gestión de datos maestros tienen un gran impacto en los pilares de datos modernas; ser capaz de aprovechar todas estas funcionalidades plenamente mientras te deshaces de los sistemas heredados equivalentes será un paso final para una estrategia ganadora.

## MÁS INFORMACIÓN

---

### Sinopsis

Este estudio de mercado de IDC analiza las principales conclusiones de Kafka Summit en Londres, que fue organizada por Confluent los días 25 y 26 de abril de 2022. El desarrollo de soluciones de *streaming data* en tiempo real ayudará a que se conviertan en una parte vital de muchas organizaciones; el creciente interés por las plataformas de *streaming data* es una prueba de la creciente importancia para las industrias. La plataforma totalmente gestionada de Confluent, que fue construida sobre el *open-source* de Apache Kafka, puede utilizar servicios de *data-in-motion* en tiempo real y simplificar en gran medida la experiencia de los usuarios de Kafka, gracias a sus *connectors* preconstruidos a varias fuentes de datos, así como a sus elementos de gestión de datos, gobernanza y seguridad. Por último, su capacidad de almacenamiento prácticamente infinita puede dar lugar a una nueva forma de ver el almacenamiento de datos en el futuro.

## Sobre IDC

*International Data Corporation* (IDC, por sus siglas en inglés) es el principal proveedor mundial de inteligencia de mercado, servicios de asesoramiento y eventos para los mercados de las tecnologías de la información, las telecomunicaciones y la tecnología de consumo. IDC ayuda a los profesionales de informática, a los ejecutivos de las empresas y a la comunidad inversora a tomar decisiones con base en hechos sobre las compras de tecnología y la estrategia empresarial. Más de 1 100 analistas de IDC proporcionan conocimientos globales, regionales y locales sobre las oportunidades y tendencias tecnológicas e industriales en más de 110 países de todo el mundo. Durante 50 años, IDC ha proporcionado conocimientos estratégicos para ayudar a nuestros clientes a alcanzar sus objetivos empresariales clave. IDC es una filial de IDG, la empresa líder mundial en medios de comunicación, investigación y eventos tecnológicos.

## 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](https://twitter.com/idc_deutschland) [www.idc.de](http://www.idc.de)

### Aviso de Copyright

Este documento de estudio de IDC se publicó como parte de un servicio de inteligencia continua de IDC, que proporciona investigación escrita, interacciones con analistas, tele-informes y conferencias. Visita [www.idc.com](http://www.idc.com) para obtener más información sobre los servicios de suscripción y consulta de IDC. Para ver una lista de las oficinas de IDC en todo el mundo, visita [www.idc.com/offices](http://www.idc.com/offices). Por favor, contacta con la línea directa de IDC mediante el número 800.343.4952, ext. 7988 (o +1.508.988.7988) o [sales@idc.com](mailto:sales@idc.com) para obtener información sobre la aplicación del precio de este documento a la compra de un servicio de IDC o para obtener información sobre copias adicionales o derechos en la web.

Copyright 2022 IDC. Queda prohibida la reproducción, salvo autorización. Todos los derechos reservados.

