



Hochskalierbare Datenverarbeitung in Echtzeit

Tempo machen mit Apache Kafka

Inhaltsverzeichnis

Das neue Paradigma:	03
Hochskalierbare Datenverarbeitung in Echtzeit	
Confluent Platform:	04
Der neue Weg: Apache Kafka® als Kern der Event-Streaming-Plattform	
Anwendungsfälle aus den Branchen:	06
Mehrwerte durch den Einsatz einer Event-Streaming-Plattform	
Fazit	08
Event-Streaming löst Geschäftsprobleme	

Hochskalierbare Datenverarbeitung in Echtzeit

TEMPO MACHEN MIT APACHE KAFKA®

Zukunftsweisende Datenstrategien fordern Event-Streaming-Plattformen, um Reaktionen in Echtzeit zu ermöglichen – Apache Kafka ebnet den Weg.

Wenn Unternehmen heute in unsicheren und sich schnell ändernden Marktumgebungen wettbewerbsfähig bleiben wollen, ist es überlebenswichtig, ebenso schnell auf Marktbewegungen wie Aktionen der Kunden zu reagieren. Dafür müssen Informationen in Echtzeit geteilt, verarbeitet und ausgewertet werden - bei jedem Klick, jedem Scan, jeder Sensormeldung und jeder Interaktion.

Die meisten bestehenden Architekturen sind jedoch nicht annähernd in der Lage, große Datenvolumen in Echtzeit zu verarbeiten. Damit agile Prozesse sowie Echtzeit-Entscheidungen möglich werden, ist eine Event-Streaming-Architektur zwingend erforderlich. Gleichzeitig dient sie als zentrales Nervensystem eines digitalen Unternehmens.

Hier ist insbesondere der Zeitwert als Wertprinzip für Daten zu fokussieren. Denn es geht nicht nur darum, wie viele Daten auf einmal verarbeitet werden können, sondern vor allem um die Geschwindigkeit, mit der sie verarbeitet werden.

Das neue Paradigma:

EVENT-STREAMING LÖST GESCHÄFTSPROBLEME

Für die Erhaltung der Wettbewerbsfähigkeit und die gestiegenen Anforderungen der Kunden müssen Unternehmen heute jederzeit in der Lage sein, ein vollständiges Bild der historischen Daten sowie einen lückenlosen Überblick über die Events, die in Echtzeit stattfinden, zu erhalten.

Heute werden Events - wie zum Beispiel die Bestellung eines Taxis - mit Unmengen an Echtzeitdaten ermöglicht und angereichert: Der Standort des Fahrzeugs, die geschätzten Verkehrszeiten, das laufende Tracking der Fahrt, die Bereitstellung von Fahrer- und Nutzerspezifika in Bezug auf Kontaktdaten usw.

Daraus resultiert sowohl eine Echtzeit-Kommunikation, als auch in diesem Fall das Bezahlen über eine Zahlungsinfrastruktur. Das ist im Prinzip bei jeder digitalen Aktion wie zum Beispiel einer Anbindung verteilter IoT-Anwendungen, Verwaltung von Bankkonten, Buchung eines Zugtickets oder Inventur von Unternehmen der Fall. Jedoch haben alle Aktionen eines gemeinsam: sie setzen sich aus Events zusammen oder sind Events.

NEUES EVENT-ZENTRIERTES DENKEN

Die Technologie des Event-Streaming löst als neues Paradigma das bedrohliche Geschäftsproblem, dass Kunden leichter zum digitalisierten Wettbewerb abwandern bzw. neue disruptive Anbieter eine schnell wachsende Konkurrenz bilden, an der man sich als Unternehmen messen lassen muss. Daten werden jetzt als ein ständig aktualisierter Strom von Events betrachtet, der es Unternehmen ermöglicht, eine völlig neue Kategorie von Event-gesteuerten Anwendungen sowie eine universelle Event-Pipeline zu schaffen.

Einer Gartner-Studie zufolge sollen bereits in den nächsten Jahren über 60 % aller Business-Transaktionen in digitalen Unternehmenslösungen Event-getrieben ablaufen.

Dafür ist es unerlässlich, eine Event-Streaming-Plattform in den Fokus der Datenstrategie zu stellen. Das heißt, Abkehr von den monolithischen Systemen und schwerfälligen Batch-Prozessen und hin zu einer agilen neuen Welt der entkoppelten Anwendungen bzw. Microservices.

Confluent Platform:

DER NEUE WEG: APACHE KAFKA ALS KERN DER EVENT-STREAMING-PLATTFORM

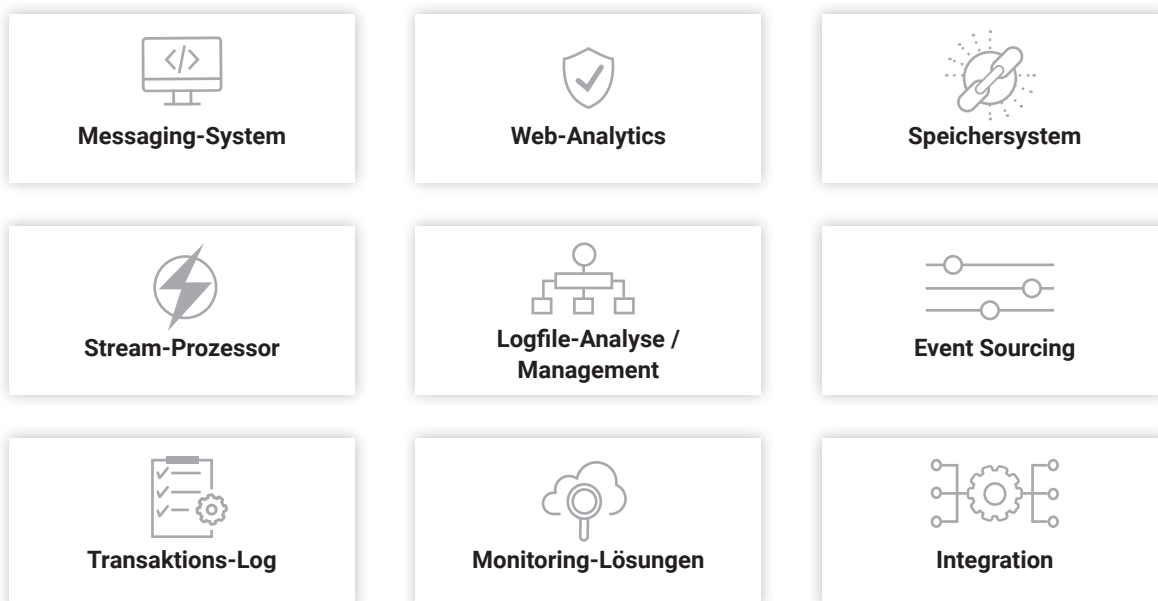
Die Open-Source-Lösung Apache Kafka wurde ursprünglich bei LinkedIn entwickelt und ermöglicht skalierbares Streaming von Daten in Echtzeit – kombiniert mit persistenter Speicherung. 2014 gründeten die Entwickler von Apache Kafka das Unternehmen Confluent aus LinkedIn heraus. Confluent treibt seither als Pionier der Event-Streaming-Plattform sowohl die Weiterentwicklung von Apache Kafka voran (80 % der Commits kommen aus dem Unternehmen), als auch die Vision eines zentralen Nervensystems im Unternehmen, ermöglicht durch eine vollständige Streaming-Plattform.

SZENARIEN FÜR DEN EINSATZ

Von Anfang an wurde Apache Kafka für einen hohen Schreib- und Lesedurchsatz konzipiert. Zusammen mit einer hohen Flexibilität, Skalierbarkeit und Fehlertoleranz eignet sich die Open-Source-Software für die verschiedensten Anwendungsfälle. Insbesondere für folgende Bereiche, die sich beliebig kombinieren lassen:

- Veröffentlichung und Abonnieren von Datenströmen (Messaging-System). Selbst nach der Software-Erweiterung empfiehlt sich Apache Kafka nach wie vor für die direkte Nachrichtenübertragung via Warteschlangensystem oder für Broadcast-Nachrichtenaussendung.
- Weiterverarbeitung von Datenströmen. Dank der Stream-API qualifiziert sich Kafka Streams bestens für Applikationen, die in Echtzeit auf bestimmte Events reagieren sollen und zu diesem Zweck möglichst schnell und effektiv Datenströme weiterverarbeiten müssen.
- Wie Kafka Streams gehört auch Kafka Connect (Connect API) als Schnittstelle zum Laden / Exportieren von Daten aus / in Drittsysteme zum Apache Kafka-Framework. Dabei führt Kafka Connect sogenannte Konnektoren (Connectors) aus, welche die eigentliche Kommunikation mit dem Drittsystem wie zum Beispiel Datenbanken, Schlüsselwertspeichern, Suchindizes oder Dateisystemen übernehmen. Somit kann Kafka Connect ganze Datenbanken einlesen oder Messdaten von Anwendungsservern erfassen, um die Daten für die Stream-Verarbeitung mit geringer Latenz verfügbar zu machen.
- Speichern von Datenströmen. Die Software lässt sich ebenfalls als verteiltes, fehlertolerantes Speichersystem nutzen. Unerheblich davon, ob man 50 KB oder 50 TB an konsistenten Daten auf den Servern speichern möchte.

TYPISCHE USE CASES AUF EINEN BLICK:



GRUNDFUNKTIONEN

Zu den Grundfunktionen von Apache Kafka gehört die Optimierung der Übertragung und Verarbeitung von Datenströmen, die per Direktverbindung zwischen Datenempfänger und Datenquelle ausgetauscht werden. Kafka fungiert zwischen Sender (Producer) und Empfänger (Consumer) als Messaging- Instanz und bietet Lösungen für die Problemstellungen, die bei einer solchen Verbindung auftreten:

Kafka löst unter anderem das Problem, dass ein Zwischenspeichern von Daten verhindert wird, wenn der Empfänger nicht verfügbar ist (Netzwerkprobleme). Darüber hinaus macht die Kafka-Queue eine Überlastung des Consumers durch den Producer unmöglich. Dieses Problem tritt dann auf, wenn die Daten bei der Direktverbindung schneller gesendet werden als sie empfangen und verarbeitet werden können. Kafka ist sogar für die Fälle gewappnet, in denen das Zielsystem die Nachricht zwar erhalten hat, aber während des Verarbeitungsprozesses einen Systemabsturz erlebt. Dies geschieht, indem der Sender durch Apache Kafka eine Rückmeldung über den Fehlschlag erhält. Ein weiterer Punkt ist, dass auch später neue Anwendungen oder Microservices hinzugefügt werden können. Trotzdem können die alten Events als Consumer fungieren.

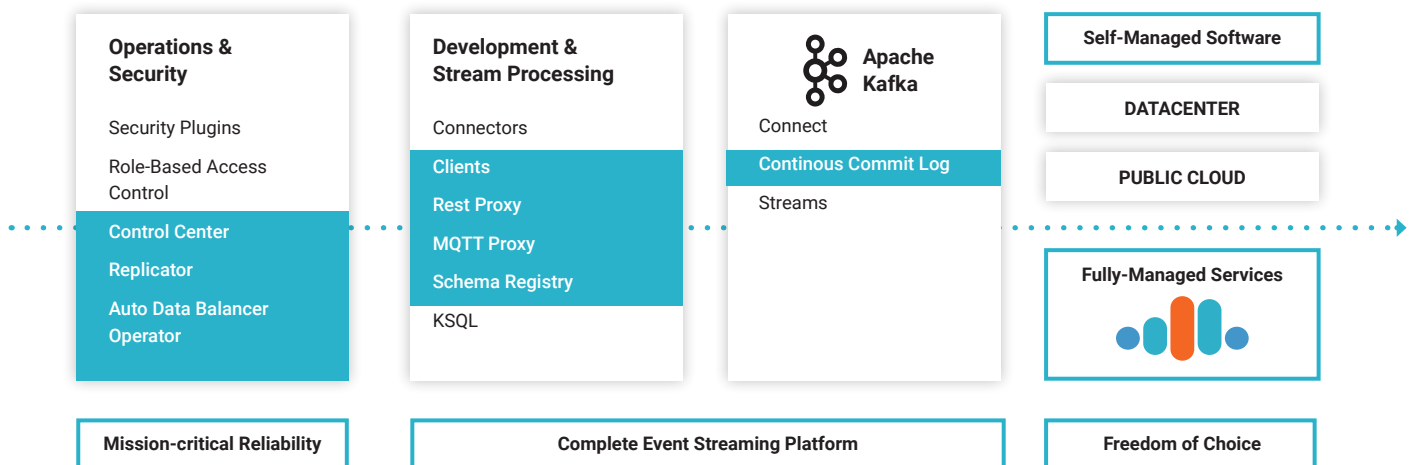
Im Vergleich zu den reinen Diensten für Nachrichten-Warteschlangen (Messaging Queue - MQ) wie zum Beispiel bei Datenbanken arbeitet die Software fehlertolerant. Dadurch schafft Apache Kafka die Basis für eine kontinuierliche Verarbeitung von Daten. In Verbindung mit einer sehr hohen Skalierbarkeit und der Fähigkeit, transportierte Informationen auf beliebig viele Systeme zu verteilen (verteiltes Transaktions-Log) ist Apache Kafka eine ideale Lösung für alle Services, die eine schnelle Speicherung und Verarbeitung von Daten sowie eine hohe Verfügbarkeit verlangen.

Apache Kafka kann für die erfolgreiche Integration verschiedener Legacy- oder moderner Anwendungen genutzt werden. Das ist heute eins der wichtigsten Kriterien in Projekten. Dabei ist es unerheblich, ob es sich um ein neues Projekt oder die Ablösung von alten ETL-/ESB-Lösungen handelt. Letztlich hängt es von dem zu betrachtenden Szenario ab.

Durch „Confluent Platform“ steht den Unternehmen eine Enterprise-Streaming-Plattform zur Verfügung, wahlweise on-premise, in der Cloud oder auch als komplett gemanagte Streaming-Lösung. Ein wesentlicher Vorteil liegt in der nahtlosen Integration in bereits bestehende Landschaften. Gepaart mit der Möglichkeit, neue Technologien und Verfahren wie Microservices, DevOps und Containerlösungen, neue Applikationsentwicklungen und regulatorische Anforderungen gleichzeitig bedienen zu können, resultiert daraus eine hohes Einsparungspotenzial bei den laufenden Betriebskosten.

CONFLUENT PLATFORM

Support, Services, Training and Partners



Anwendungsfälle aus den Branchen:

MEHRWERTE DURCH DEN EINSATZ EINER EVENT-STREAMING-PLATTFORM

Was haben ING, Lyft, Audi, Bosch oder JP Morgan Chase gemeinsam? Der wirtschaftliche Erfolg dieser Unternehmen lässt sich nicht zuletzt darauf begründen, dass entscheidende Erkenntnisse schnell aus großen Datenmengen gezogen werden können.

ANWENDUNGSFALL: AUTOMOTIVE

Connected Car Services sollen beispielsweise Fahrzeugdaten für Predictive Maintenance oder zur Abwicklung von Bestellungen und Auslieferung neuer Fahrzeug-Features in Echtzeit (z. B. Motorleistungs-Upgrade) liefern. Diese Event-Daten werden an verschiedenste „Consumer“ wie zum Beispiel Analytics-Anwendungen, Buchhaltung oder andere Plattformen gestreamt.

Für eine **verbesserte Fahrzeugsicherheit und mehr Fahrerlebnis** können Fahrzeuge heute hinsichtlich ihres technischen Zustands, der Motoreffizienz, der Kraftstoff- und Emissionswerte uvm. kontinuierlich überwacht werden. Zusätzlich lassen sich diese Informationen mit Services wie zum Beispiel Verkehrsinformationen und GPS-Services kombinieren. Kafka Connect erlaubt eine schier unbegrenzte Zuschaltung dieser Dienste über benutzerdefinierte Konnektoren. Auf diese Weise kann die Software leicht auf Hunderte oder sogar Tausende von Endgeräten skalieren.

Erfolgreiche Customer-360-Projekte: Mit der wachsenden Zahl verschiedenster Sensoren, Geräte und Interaktionsmöglichkeiten wird es zunehmend schwieriger, einen präzisen Überblick zu den Aktivitäten beispielsweise eines Fahrers zu erhalten. Eine Streaming-Plattform ermöglicht mit einer übergreifenden Betrachtung der Ergebnisse ein ganzheitliches Bild der Aktivitäten des Fahrers.

- **Zugriff auf Kundendaten in Echtzeit:** Wenn Benutzerdaten durch eine Streaming-Plattform laufen, können sie in einem oder mehreren Systemen gespeichert werden.
- **Keine Silos:** Nur wenn umgedacht, Silo-Denken abgeschafft und die Architektur daran angepasst wird, kann eine effektive Single Source of Truth geschaffen werden.
- **Reaktionsvermögen:** Event-getriebene Anwendungen ermöglichen automatisierte, hochskalierbare Reaktionen auf vom Fahrer oder Fahrumfeld ausgelöste Events.
- **Geschäftserfolg:** Streaming-Plattformen ermöglichen eine Umsatzsteigerung und binden durch neu ermöglichte Use Cases ihre Kunden und Fahrer an die Marke.
- **Marktforschung:** Wachstumsfelder für die Zukunft kontextsensitiv identifizieren, neue Produkte entwickeln sowie Bedrohungen rechtzeitig erkennen.

ANWENDUNGSFALL: PRODUKTION

Da die Entstehung der Daten in Echtzeit mitverfolgt und somit adjustiert werden können, lassen sich im Bereich Manufacturing mit Apache Kafka nicht nur Wartungsaufwände reduzieren, sondern auch Endprodukte optimieren.

- Streaming von Daten der einzelnen Maschinen, Produktionsstandorte und Abläufe. Dafür werden Microservices sowie Machine Learning integriert.
- Unternehmen sind somit in der Lage, Bauteile und Komponenten auszutauschen, bevor sie defekt sind. Gleichzeitig können aufgrund der Daten Wartungszyklen verlängert werden.
- Erhebliche Kostenersparnis abhängig von der Branche und dem Anwendungsfall (z. B. in der Fertigung), Umsatzsteigerung (z. B. an Verkaufsautomaten) oder Erhöhung der Kundenzufriedenheit (z. B. Vorhersage von Telko-Netzwerkausfällen).

ANWENDUNGSFALL: BANK / FINANZWESEN

Im Finanzwesen wird es aufgrund der verschiedensten Kundenzugänge, Geräte und weiteren Interaktionsmöglichkeiten für einen Anbieter immer schwieriger, einen umfassenden Überblick zu den Aktivitäten des Kunden zu erhalten.

Eine Streaming-Plattform unterstützt Finanzdienstleister insbesondere bei folgenden Herausforderungen:

- **Betrugserkennung:** Geldwäsche, Zahlungsbetrug, unternehmensinterne Diebstähle, Identitätsdiebstahl und andere Formen des Betrugs sind eine gewaltige Herausforderung für Organisationen, die versuchen, nur mit Batch-Verarbeitung und Silostrukturen Angreifer zu entlarven. Event-Streaming reduziert erheblich den Ressourcenaufwand sowie den Zeitabstand zwischen Angriff und Erkennung.

Hier ist es hilfreich, eine Streaming-Plattform als zentrale Quelle für Echtzeit-Daten zu nutzen, um entsprechend zu reagieren und betrügerisches Verhalten zu verhindern. Mit der „Confluent Plattform“ können Teams aus dem Bereich Betrugsanalyse oder auch Data Scientists stets in vollem Umfang auf die aktuellsten Daten zugreifen und diese als Input für ihre Aufgaben oder Modelle des maschinellen Lernens nutzen.

- **Kosteneinsparung (OpEx) durch Mainframe Offloading:** Kritische Anwendungen und Daten werden von Finanzdienstleistern meist auf Mainframes betrieben und verwaltet. Diese gelten zwar als sehr zuverlässig, sind aber auch mit vergleichsweise recht hohen Betriebskosten verbunden. Daher modernisieren derzeit immer mehr Finanzdienstleister ihre Architektur durch den Umstieg von monolithischen Anwendungen auf Microservices, Cloud-Migration und die Einbeziehung offener Systeme.

Im Zuge dieses Modernisierungsprozesses ist ein Zugriff auf alle relevanten Daten hilfreich. „Confluent Plattform“ kann zur Umleitung von Datenströmen genutzt werden, die normalerweise über den Mainframe an eine Event-getriebene Architektur geleitet worden wären. Dies entlastet nicht nur den Mainframe, sondern spart vor allem Geld. Dadurch müssten die Datenflüsse keine lähmenden Batch-Prozesse mehr durchlaufen und könnten auf sie bzw. die Events in Echtzeit zugreifen.

- **Customer 360:** Rund zwei Drittel der Bankkunden erwarten mehr Zugriffsmöglichkeiten auf Online-Interaktionen und möchten ihr Banking überall, jederzeit und auf allen Geräten in Echtzeit abwickeln. Des Weiteren wollen sie Kredite ohne persönliche Treffen erhalten. Das veraltete Batch-System ist damit hoffnungslos überfordert.

Für ein verbessertes Kundenerlebnis und der damit verbundenen gestärkten Marktposition sammelt die Streaming-Plattform alle Daten, die das Kundenverhalten abbilden und veröffentlicht sie in Echtzeit in den großen Daten-Pipelines, wodurch eine „Single Source of Truth“ für nachgelagerte Systeme entsteht.

Ein 360°-Kunden-Überblick ermöglicht eine Unterstützung von Omnichannel-Banking-Initiativen und Kundenerlebnisse in Echtzeit!

ANWENDUNGSFALL: RETAIL

Event-Streaming bietet auch für den Handel hinsichtlich der Optimierungen bei der Übertragung, Verarbeitung und Auswertung von Datenströmen in Echtzeit viele wertvolle Vorteile.

- **Personalwesen:** Schnelleres Onboarding von Verkäufern.
- **Marketing / Vertrieb:** Schnellere Aufstellung von Produktlisten.
- **Einkauf:** Rationalisierung der Lieferkette und Modernisierung der bestehenden Infrastruktur.
- **Marketing / Vertrieb:** Abstimmung von Kundendaten, um kontextspezifische und persönliche Angebote bzw. Rabatte zu erstellen, bevor der Kunde die Filiale verlässt oder das Geschäft betritt.
- **Marketing / Vertrieb:** Nutzung von Echtzeitinformationen (z. B. standortbezogene Daten, Zahlungsdaten), aber auch historische Daten (z. B. Informationen aus der CRM- oder Loyalty-Plattform), um jedem einzelnen Kunden das beste Angebot zu unterbreiten.

FAZIT

Apache Kafka ist eine Event Streaming-Plattform, die Messaging, Storage, Stream Processing und Integration vereint. Als verteiltes System ist es hochverfügbar, hochskalierbar und fehlertolerant und somit für Big-Data-Anwendungen bestens geeignet.

Apache Kafka verbessert herkömmliche Message Broker durch Weiterentwicklungen beim Datendurchsatz, bei der Partitionierung, Replikation, Latenz und letztlich Zuverlässigkeit.

Weiterhin lässt sich Apache Kafka in vielfältigen Bereichen einsetzen: Messaging, Echtzeit-Tracken der Aktivitäten von und auf Webseiten, Überwachen der operativen Kennzahlen von verteilten Anwendungen, Aggregation der Log-Dateien von zahlreichen Servern, Event Sourcing beim Protokollieren und Anfordern von Statusänderungen in einer Datenbank, Übergeben von Log-Dateien, in denen verteilte Systeme ihre Daten synchronisieren sowie Wiederherstellen von Daten nach dem Ausfall von Systemen.

Für eine nachhaltige Erhaltung der Wettbewerbsfähigkeit müssen Unternehmen heute auf Events jeder Art reagieren können – und zwar in Echtzeit! Dies erfordert ein Umdenken sowohl in der Unternehmensbetrachtung als auch in der Architektur. 60 % der Fortune 100 setzen hier schon auf Apache Kafka und immer mehr Firmen komplementieren diesen Ansatz mit den Komponenten von Confluent Plattform, um geschäftskritische Datenströme ausfallsicher und skalierbar zu managen und auswerten zu können. Ergänzt durch Trainings-, Zertifizierungs- und Service-Angebote schaffen Unternehmen ganz neue Geschäftsfelder und festigen so ihre Stellung im Markt.

Confluent, gegründet von den Entwicklern der Open-Source-Lösung Apache Kafka®, bietet die führende Streaming-Plattform, welche es Unternehmen ermöglicht, den Wert ihrer Daten zu maximieren. Die Plattform von Confluent erlaubt es, Daten aus isolierten Systemen in eine Echtzeit-Datenpipeline zu übertragen, wo sofort darauf reagiert werden kann.

Heute konzentriert sich Confluent auf die Weiterentwicklung einer Streaming-Plattform sowohl on-premise als auch in der Cloud, die es anderen Unternehmen ermöglicht, schnell und einfach auf Echtzeit-Streams mit Unternehmensdaten zuzugreifen. Zu den Kunden von Confluent zählen Unternehmen wie ING, Target, Audi, LinkedIn, BOSCH, Deutsche Bahn oder Nordea.

confluent.io
info-de@confluent.io