

REPORT

Non solo messaging: Confluent apre la strada a un approccio più completo a Kafka e alle piattaforme di dati in streaming.

Giovanni Cervellati

EXECUTIVE SNAPSHOT

FIGURA 1

Executive Snapshot: Confluent apre la strada a un approccio più completo a Kafka e alle piattaforme di data streaming

Il presente Report di IDC illustra i principali risultati del Kafka Summit svoltosi a Londra e ospitato da Confluent il 25 e 26 aprile 2022. I dati vengono visti come una sequenza praticamente infinita di eventi, che riguardano qualsiasi possibile tipo di azione, evento o cambiamento identificato o registrato dai software o dalle applicazioni. Kafka è una piattaforma di event streaming open source e la piattaforma Confluent, interamente gestita e basata su Apache Kafka open source, è in grado di gestire servizi di data in motion in tempo reale e permette di semplificare notevolmente il percorso degli utenti di Kafka, grazie a connectors predefiniti per diverse origini di dati, e per i relativi elementi di gestione, governance e sicurezza dei dati.

Punti chiave

- Il data streaming viene generato ininterrottamente da fonti di dati, che in genere inviano contemporaneamente record di dati, di piccole dimensioni (alcuni kilobyte) in tempo reale. Le piattaforme di data streaming vengono quindi utilizzate per raccogliere, elaborare, archiviare e integrare tali dati in tempo reale su larga scala. Kafka è un prodotto software open source noto come piattaforma di event streaming.
- I casi d'uso tradizionali per le piattaforme di data streaming comprendono le applicazioni Web, gli acquisti online, le attività dei giocatori in-game e le informazioni provenienti dai social network e dalla messaging. L'attenzione si sta spostando sulla valorizzazione degli event streaming come parte dell'ecosistema di dati. L'utilizzo di Kafka si è diffuso, da una semplice piattaforma utilizzata dagli sviluppatori per la messaging è diventata una piattaforma universale più completa per applicazioni e data pipeline.
- L'idea dei sostenitori di Kafka e di Confluent è che le piattaforme di data streaming diventeranno un "sistema nervoso centrale" di dati, collegando e analizzando tutto ciò che è acceso, tutto ciò che è spento, cosa sta cambiando, cosa invece rimane statico e cosa sta succedendo in real time.
- La piattaforma interamente gestita di Confluent, basata su Apache Kafka open source, può semplificare in modo significativo i percorsi degli utenti di Kafka, grazie ai connectors predefiniti per varie fonti dati e ai relativi elementi di gestione, governance e sicurezza dei dati. La sua capacità di archiviazione, virtualmente infinita, potrebbe comportare un nuovo modo di intendere l'archiviazione dei dati in futuro.

Fonte: IDC, 2022

IN QUESTO REPORT:

Il presente Report di IDC illustra i principali risultati del Kafka Summit ospitato da Confluent, che si è svolto a Londra il 25 e il 26 aprile 2022. Durante il Summit sono stati presentati i casi d'uso e le best practices, si è discusso sul futuro delle tecnologie di streaming e c'è stato spazio anche per delle sessioni pratiche, che hanno fornito un ampio panorama degli sviluppi attuali e futuri di Kafka (e di Confluent). Tra i partecipanti al Kafka Summit vi erano sviluppatori, architetti e ingegneri dei dati e molti altri desiderosi di conoscere le tecnologie di data streaming.

Il data streaming: Dal Messaging al Sistema Nervoso Centrale

Il data streaming viene generato ininterrottamente da fonti di dati, che in genere inviano contemporaneamente record di dati, di piccole dimensioni (alcuni kilobyte) in tempo reale. Le piattaforme di data streaming vengono quindi utilizzate per raccogliere, elaborare, archiviare e integrare tali dati in tempo reale su larga scala.

I data streaming includono un'ampia varietà di dati come i file di registro generati dai clienti che utilizzano applicazioni mobili o applicativi web, gli acquisti online, le attività dei giocatori in-game e le informazioni provenienti dai social network, i trading floor finanziari o i servizi geospaziali. L'implementazione di questi servizi aiuta le aziende a rimanere competitive in un mercato più ampio come quello digitale. L'integrazione di dati in tempo reale nel data fabric per scelte aziendali a bassa latenza ha un impatto enorme sui servizi quali le relazioni con i clienti delle società di telecomunicazioni e la gestione del percorso del cliente.

Kafka è un prodotto software open source noto come piattaforma di eventi di streaming: i dati vengono visti come una sequenza praticamente infinita di eventi, che riguardano qualsiasi possibile tipo di azione, evento o cambiamento identificato o registrato dai software o dalle applicazioni. Questo, ovviamente, racchiude i dati. Negli ultimi anni, quest'idea affascinante ha spesso sostituito la vecchia mentalità batch: le aziende ritenevano che l'elaborazione del flusso fosse necessaria perché le esigenze aziendali impongono maggiori requisiti in tempo reale a sviluppatori e architetti. Confluent è un'organizzazione che utilizza Kafka come nucleo della propria piattaforma, interamente gestita. L'implementazione di questi servizi aiuta le aziende a rimanere competitive in un mercato più ampio come quello digitale. L'integrazione di dati in tempo reale nel data fabric per scelte aziendali a bassa latenza ha un impatto enorme sui servizi quali le relazioni con i clienti delle società di telecomunicazioni e la gestione del percorso del cliente.

Come dimostra uno studio del 2022 di IDC sui data streaming, le dimensioni delle imprese più affidabili dal punto di vista digitale stanno ripensando le proprie architetture software con un contesto in tempo reale all'avanguardia. Praticamente nessuna delle "best in class" ha dichiarato di avere in corso o di aver pianificato delle strategie di data streaming. Inoltre, oltre l'80% delle aziende intervistate afferma che nei prossimi 12-18 mesi, effettuerà delle ricerche o investirà in nuove funzionalità di streaming. I casi d'uso più comuni sono quelli in cui la rapidità d'azione è una necessità estrema; tra questi vi sono la gestione della sicurezza e dei rischi, l'esperienza del cliente e i dati finanziari in tempo reale.

Il passaggio dai trasferimenti di dati in batch al data streaming risulta essere in linea con le altre innovazioni nell'attuale ecosistema di dati: le distribuzioni on-premise si stanno spostando nel cloud, i monoliti si stanno evolvendo in microservizi e così via. Tutto ciò implica il concetto di decentramento, che è una parte logica e intrinseca alla struttura stessa dello streaming/evento.

Kafka, lungi dall'essere una semplice piattaforma utilizzata dagli sviluppatori per la messaggistica, è diventata una piattaforma universale, più completa, per applicazioni e pipeline di dati. Sono disponibili altre piattaforme di data streaming, ma attualmente Kafka è tra le più utilizzate.

L'idea dei sostenitori di Kafka e di Confluent è che le piattaforme di data streaming diventeranno un "sistema nervoso centrale" di dati, collegando e analizzando tutto ciò che è acceso, tutto ciò che è spento, cosa sta cambiando, cosa invece rimane statico e cosa sta succedendo in real time.

Inoltre, immaginare che una piattaforma di data streaming venga messa al centro di un ecosistema di dati favorisce nuovi innovativi modi di strutturare i progetti di dati: l'acquisizione diretta in streaming per il training di modelli predittivi (anche direttamente dai dati all'edge), la distribuzione di modelli e di analisi in tempo reale con conseguente live-action, il potenziamento del controllo in tempo reale di possibili violazioni dei dati e della sicurezza. Inoltre è importante prendere in considerazione due caratteristiche di Kafka: la grande capacità di archiviazione e il linguaggio KSQL che consente l'elaborazione dei dati in modalità streaming utilizzando un linguaggio simile al SQL. Tali caratteristiche rappresentano un nuovo modo di vedere Kafka: eseguire l'analisi batch su Kafka senza archiviare i dati in un grande data storage e leggerli direttamente, potrebbe consentirci di gestire, in futuro, tutti i nostri dati senza un data lake.

Le piattaforme di data streaming su vasta scala vanno oltre Kafka

I dati in tempo reale sono così diventati più una necessità che un vantaggio. Ma anche una piattaforma open source, ad alta codifica, orientata agli sviluppatori come Kafka, per quanto molto popolare, presenta dei limiti. Quando la quantità di dati inizia a scalare, sono necessari sforzi di gestione eccessivi. Di solito le aziende che non hanno dimestichezza con l'elaborazione di eventi in streaming iniziano con una fase di sperimentazione (che è più adatta per Kafka). Quando però il progetto viene attuato come mission-critical e le aziende scalano per collegare Kafka a una o più linee commerciali, è probabile che Kafka non sia più sufficiente.

Come accennato in precedenza, utilizzare Kafka da solo può essere faticoso, soprattutto perché privo di alcune importanti caratteristiche di gestione di cui un'azienda farebbe volentieri a meno. Man mano che l'organizzazione cresce, la necessità di personale, per la sua gestione, cresce. Tutto ciò implica un onere per coloro che lo utilizzano e farà lievitare i costi. L'utilizzo di Kafka come piattaforma centrale per i dati, infatti, richiederà alle aziende di abilitare la connettività a tutte le origini dati, le applicazioni dati e le applicazioni end/edge e, in questo modo, dovranno anche occupare risorse per delle soluzioni di integrazione per l'acquisizione di dati e "eventi" da legacy systems. Kafka inoltre non è cloud native e spostarlo nel cloud potrebbe non essere efficiente.

Costi di gestione eccessivi e sempre maggiori, per le aziende che utilizzano Kafka, hanno portato alla ricerca di nuove soluzioni, più semplici e meglio collegate. Confluent è una di queste soluzioni. Basato su Apache Kafka open source, può essere distribuito come offerta SaaS cloud native o come software autogestito. L'offerta cloud native può essere eseguita su un'infrastruttura multicloud o cloud ibrida. È costruito per gestire i dati in tempo reale nei servizi in motion e, finora, gli utenti più importanti provengono dal settore finanziario, dal retail ed e-commerce, da quello manifatturiero e dal media & entertainment.

I costi degli ETP, le spese per l'infrastruttura e i rischi aziendali relativi a possibili tempi di inattività non pianificati o alle violazioni di Kafka possono essere evitati grazie a una piattaforma gestita. I connectors predefiniti, autogestiti e più semplici da usare, per Kafka (e per la maggior parte delle potenziali origini e sistemi di dati); la gestione dei dati e l'osservabilità dei dati; la capacità di rilevamento dei dati; la sicurezza e la governance dei dati sono tutte funzionalità che potrebbero essere abilitate utilizzando una piattaforma dati su vasta scala. L'utilizzo di Kafka in un ambiente cloud ibrido viene inoltre ostacolato dalle difficoltà di connessione a diversi ambienti in tempo reale, attorno ai quali vengono proprio costruite piattaforme come Confluent. Per concludere, le notevoli capacità di archiviazione in Kafka, grazie a Confluent diventano praticamente illimitate.

Alla luce di tutto ciò, sarebbe difficile distribuire flussi di dati di qualità comprovata necessari per ottenere la già citata architettura del "sistema nervoso centrale". L'implementazione di una soluzione di data streaming più completa renderà molto più semplice abilitare la connettività in tempo reale per il data streaming di un'azienda, sia ingerendoli che immettendoli ove necessario.

IL PUNTO DI VISTA DI IDC

È indubbio che i dati in tempo reale sono diventati un argomento chiave per la maggior parte delle aziende. Le piattaforme di dati in streaming, che all'inizio erano viste solo come piattaforme di messaging, stanno guadagnando terreno in più campi. Poiché la sicurezza, la governance e la visibilità si sono sviluppate in aree cruciali per i dati, fare affidamento sulla tempistica batch per le risposte dei dati si rivelerà insufficiente per molti.

Lo sviluppo di soluzioni di data streaming in tempo reale alla fine, diventerà una parte vitale per molte aziende, partendo inizialmente con alcuni casi d'uso ed espandendosi a più linee commerciali. Il sempre maggior interesse per le piattaforme di data streaming è la prova della loro crescente importanza per diversi settori, con solo un piccolo segmento di aziende che dichiara di non essere interessato a implementarle. Le piattaforme di data streaming su vasta scala

rappresentano già un passo avanti e consentono ad alcune aziende di costruire intere infrastrutture di dati attorno ai data streaming. L'utilizzo di piattaforme di data streaming per controlli di sicurezza in tempo reale e attività di sicurezza dei dati, inoltre, ha un enorme impatto e rappresenta un grande potenziale per lo sviluppo e la diffusione futuri.

Il mondo del data streaming tuttavia, non ha acquisito una conoscenza molto diffusa oltre al classico utilizzo per applicazioni e attività come la messaggistica. In questo senso i software come Confluent rappresentano un ottimo punto di partenza. Per velocizzare il riconoscimento dei numerosi casi d'uso e della sua centralità nelle piattaforme dati però, potrebbe essere utile allontanarsi dalla sua natura troppo orientata allo sviluppatore e optare per un percorso di adozione più semplice.

In concreto, il notevole miglioramento delle funzionalità senza codice consente a più utenti aziendali e di analisi di sfruttare i vantaggi dei dati in tempo reale senza doversi affidare agli sviluppatori, che di solito sono piuttosto lontani dalla mentalità aziendale. Infine, un altro problema potrebbe essere quello dell'integrazione di piattaforme di data streaming come "sistema nervoso centrale" nelle aziende che dispongono già di troppi software e piattaforme nei loro ecosistemi di dati e di analisi. Caratteristiche quali il catalogo dei dati, la provenienza dei dati o la gestione dei dati master hanno un enorme impatto sui moderni data stack. Essere in grado di sfruttare appieno tutte queste funzionalità, eliminando legacy systems equivalenti, rappresenta un passo finale per una strategia vincente.

SCOPRI DI PIÙ

In breve

Il presente Report di IDC illustra i principali risultati del Kafka Summit svoltosi a Londra e ospitato da Confluent il 25 e 26 aprile 2022. Lo sviluppo di soluzioni di data streaming in tempo reale diventerà una parte vitale per molte aziende. Il sempre maggior interesse per le piattaforme di data streaming è la prova della loro crescente importanza per diversi settori. La piattaforma, interamente gestita, Confluent, basata su Apache Kafka open source, è in grado di gestire servizi di data in motion in tempo reale e permette di semplificare notevolmente il percorso degli utenti di Kafka, grazie a connectors predefiniti per diverse origini di dati, e per i relativi elementi di gestione, governance e sicurezza dei dati. In conclusione, la sua capacità di archiviazione, virtualmente infinita, potrebbe comportare un nuovo modo di intendere l'archiviazione dei dati in futuro.

Chi è IDC

International Data Corporation (IDC) è la prima società mondiale specializzata in ricerche di mercato, servizi di consulenza e organizzazione di eventi nei settori ICT e dell'innovazione digitale. IDC aiuta i professionisti dell'IT, i dirigenti d'azienda e la comunità degli investitori a prendere decisioni basate sui fatti sugli acquisti di tecnologia e sulla strategia aziendale. Oltre 1.100 analisti mettono a disposizione a livello globale, regionale e locale le proprie competenze sulle opportunità e le tendenze tecnologiche e industriali in oltre 110 paesi in tutto il mondo. Da 50 anni, IDC offre visioni strategiche che permettono ai clienti di raggiungere i propri obiettivi aziendali chiave. IDC è una società controllata di IDG, la società leader mondiale nel campo dei media, della ricerca e degli eventi tecnologici.

IDC Europa Centrale GmbH

IDC Europa Centrale GmbH – Germania e Svizzera

Hanauer Landstraße 182 D

60314 Francoforte sul Meno, Germania

+49 (0)69 90502-0

Twitter: https://twitter.com/idc_deutschland

www.idc.de

Avviso sul Copyright

Il presente documento di ricerca IDC è stato pubblicato come parte di un servizio permanente di IDC a scopo informativo, che fornisce ricerche scritte, interazioni con gli analisti, tele briefing e conferenze. Per saperne di più sull'abbonamento IDC e sui servizi di consulenza, visita www.idc.com. Per visualizzare l'elenco degli uffici IDC nel mondo, visita www.idc.com/offices. Contatta il numero verde IDC all'800.343.4952, int. 7988 (o +1.508.988.7988) o sales@idc.com per informazioni sull'applicazione, sul prezzo del presente documento, l'acquisto di un servizio IDC o per informazioni su copie aggiuntive o diritti web.

Copyright 2022 IDC. Vietata la riproduzione non autorizzata. Tutti i diritti riservati.

