

## NOTE DE MARCHÉ

# Pas seulement une messagerie : Confluent ouvre la voie à une approche plus complète de Kafka et des plateformes de Streaming de Données

Giovanni Cervellati

## RESUMÉ

---

### FIGURE 1

---

#### Résumé : Confluent ouvre la voie à une approche plus complète de Kafka et des plateformes de Streaming de Données.

Cette note de marché IDC aborde les principales conclusions du Kafka Summit de Londres, organisé par Confluent les 25 et 26 avril 2022. Les données sont considérées comme une séquence virtuellement infinie d'événements, qui font référence à tout type possible d'action, d'incident ou de changement identifié ou enregistré par des logiciels ou des applications. Kafka est une technologie de streaming événementielle open source et Confluent est une plateforme entièrement gérée qui a été construite sur Apache Kafka open source. Elle permet de gérer des services de données en temps réel et de simplifier considérablement le parcours des utilisateurs de Kafka, grâce à des connecteurs préconstruits vers diverses sources de données et à ses éléments de gestion, de gouvernance et de sécurité.

#### Points essentiels à retenir

- Les flux de données sont générés en permanence par différentes sources de données, qui envoient simultanément et en temps réel des enregistrements de données de petites tailles (de l'ordre de quelques kilo-octets). Les plateformes de streaming de données sont utilisées pour collecter, traiter, stocker et intégrer ces données en temps réel à l'échelle. Kafka est un logiciel open source connu sous le nom de Plateforme de Streaming de données.
- Les cas d'usages traditionnels des plateformes de streaming de données comprennent les applications web, les achats en ligne, l'activité des joueurs *in game* et les informations provenant des réseaux sociaux et des messageries. Dans le cadre de l'écosystème de données, l'exploitation du streaming d'événements devient de plus en plus importante. Kafka était d'abord une plateforme utilisée par les développeurs pour la messagerie. Elle est devenue une plateforme universelle plus complète pour les applications et les pipelines de données.
- La vision des promoteurs de Kafka tels que Confluent est que les plateformes de streaming de données évolueront vers un « système nerveux central » de données, connectant et analysant ce qui est *On*, ce qui est *Off*, ce qui change, ce qui est statique, et ce qui se passe en temps réel.
- La plateforme entièrement gérée Confluent, construite sur Apache Kafka open source, peut considérablement simplifier le parcours des utilisateurs de Kafka, grâce à ses connecteurs préconstruits vers diverses sources de données et à ses éléments de gestion, de gouvernance et de sécurité. Sa capacité de stockage virtuellement infinie pourrait entraîner une nouvelle façon d'envisager le stockage des données à l'avenir.
- Les flux de données sont générés en permanence par différentes sources de données, qui envoient simultanément et en temps réel des enregistrements de données de petites tailles (de l'ordre de quelques kilo-octets). Les plateformes de streaming de données sont utilisées pour collecter, traiter, stocker et intégrer ces données en temps réel à l'échelle. Kafka est un logiciel open source connu sous le nom de Plateforme de Streaming de données.

## DANS CETTE NOTE DE MARCHÉ

---

Cette note de marché IDC aborde les principales conclusions du Kafka Summit de Londres, organisé par Confluent les 25 et 26 avril 2022. Le sommet a présenté des cas d'usages et les meilleures pratiques. Des discussions sur l'avenir des technologies de streaming et les sessions pratiques ont fourni une vue d'ensemble des développements actuels et futurs de Kafka (et de Confluent). Les participants au Kafka Summit comprenaient des développeurs, des architectes et des ingénieurs de données, ainsi que d'autres personnes désireuses de s'informer sur les technologies de streaming de données.

### Streaming Data: de la messagerie au système nerveux central

Les flux de données sont générés en permanence par différentes sources de données, qui envoient simultanément et en temps réel des enregistrements de données de petites tailles (de l'ordre de quelques kilo-octets). Les plateformes de streaming de données sont utilisées pour collecter, traiter, stocker et intégrer ces données en temps réel à l'échelle.

Les données en streaming comprennent une grande variété de données telles que les fichiers journaux générés par les clients utilisant des applications mobiles ou web, les achats en ligne, l'activité des joueurs *in game* et les informations provenant des réseaux sociaux, des salles de marché financières ou des services géospatiaux. La mise en œuvre de ces services aide les entreprises à rester compétitives sur le marché numérique. L'intégration de données en temps réel dans le tissu de données pour des décisions commerciales à faible latence a un impact considérable sur des services tels que la relation client et la gestion du parcours client des opérateurs de télécommunications.

Kafka est un logiciel open source connu comme une plateforme de streaming de données : les données sont considérées comme une séquence virtuellement infinie d'événements, qui font référence à tout type possible d'action, d'incident ou de changement identifié ou enregistré par des logiciels ou des applications. Ces dernières années, ce concept fascinant a souvent remplacé l'ancienne méthode du traitement par lots : Les organisations ont reconnu que le traitement en flux était nécessaire pour permettre aux développeurs et aux architectes de répondre aux demandes des entreprises pour les capacités en temps réel.

Comme le montre une enquête IDC 2022 *Streaming Data Survey*, les entreprises les plus matures sur le plan numérique repensent leurs architectures logicielles en mettant le contexte en temps réel au premier plan. Pratiquement aucune n'a déclaré avoir des stratégies de données en streaming ni en place, ni planifiées. En outre, plus de 80 % de toutes les personnes interrogées déclarent qu'elles vont rechercher ou investir dans de nouvelles capacités de streaming au cours des 12 ou 18 prochains mois. Les cas d'usages les plus courants sont ceux pour lesquels la rapidité d'action est une nécessité absolue ; il s'agit notamment de la gestion de la sécurité ou des menaces, de l'expérience client et des données financières en temps réel.

Le passage des transferts de données par lots au streaming de données s'aligne sur toutes les autres innovations de l'écosystème de données actuel : les déploiements sur site se déplacent vers le cloud, les monolithes se transforment en micro-services, etc. Tout cela implique le concept de décentralisation, qui est une partie logique et inhérente à la structure streaming/événementielle.

Kafka est passé du statut de simple plateforme utilisée par les développeurs pour la messagerie à celui de plateforme universelle plus complète pour les applications et les pipelines de données. Il existe d'autres plateformes de streaming de données, mais Kafka est l'une des plus utilisées aujourd'hui.

La vision des promoteurs de Kafka tels que Confluent est que les plateformes de streaming de données évolueront vers un « système nerveux central » de données, connectant et analysant ce qui est *On*, ce qui est *Off*, ce qui change, ce qui est statique, et ce qui se passe en temps réel.

De plus, le fait d'envisager une plateforme de streaming de données comme centrale à un écosystème de données offre des possibilités innovantes de structurer les projets de données : l'ingestion directe de streaming pour la formation des modèles prédictifs (même directement à partir de données périphériques/edge), le déploiement de modèles et d'analyses en temps réel avec actions conséquentes et immédiates, l'optimisation du contrôle en direct des violations de la sécurité/de données éventuelle. Enfin, nous devrions considérer deux caractéristiques de Kafka : ses vastes capacités de stockage et le langage KSQL, qui permet de traiter les données en streaming à l'aide d'un dialecte de type SQL. Ces caractéristiques offrent une nouvelle façon d'envisager Kafka : l'exécution d'analyses par lots directement dans Kafka, sans les stocker dans un grand entrepôt de données et en lisant toutes les données, pourrait nous permettre de gérer toutes nos données sans *data lake* à l'avenir.

## Les plateformes de streaming de données à grande échelle surpassent Kafka

Pour ces raisons, les données en temps réel sont devenues plus une nécessité qu'un atout. Mais une plateforme open source à haut niveau de codage, orientée vers les développeurs, telle que Kafka - bien que très populaire - a ses limites. Lorsque la quantité de données commence à augmenter, elle nécessite une gestion trop importante.

En général, les entreprises qui découvrent le traitement des événements en continu commencent par une phase d'expérimentation (ce qui convient le mieux à Kafka). Cependant, lorsqu'elles mettent en œuvre ce projet en tant que mission critique et qu'elles passent à l'échelle pour connecter Kafka à un ou plusieurs secteurs d'activité, alors Kafka ne sera probablement pas suffisant.

Comme nous venons de le mentionner, l'exploitation autonome de Kafka peut s'avérer éprouvante, d'autant plus qu'il lui manque certaines fonctions de gestion importantes dont une organisation pourrait vouloir se libérer. Au fur et à mesure que l'organisation se développe, elle aura besoin de plus de personnel pour gérer Kafka. Cela représentera une charge pour ceux qui l'utilisent et augmentera les coûts. En fait, la mise en œuvre de Kafka en tant que plateforme centrale pour les données exigera des organisations qu'elles permettent la connectivité à toutes les sources de données, les applications de données et les applications *end/edge*. Ce faisant, elles devront également payer pour des solutions d'intégration permettant de capturer des données et des « événements » à partir des systèmes existants. De plus, Kafka n'est pas *cloud-native* et le déplacer vers le cloud peut ne pas être efficace.

Cette gestion excessive et croissante pour les organisations utilisant Kafka a poussé à la recherche de nouvelles solutions plus faciles et mieux connectées. Confluent est l'une de ces solutions. Construite sur Apache Kafka open source, elle peut être déployée comme une offre SaaS *cloud-native* ou comme un logiciel autogéré. L'offre *cloud-native* peut être exécutée sur une infrastructure multicloud ou hybride. Elle est conçue pour les services de données en mouvement en temps réel et, jusqu'à présent, ses principaux utilisateurs sont issus des secteurs de la finance, de la vente au détail et du commerce électronique, de la fabrication, des médias et du divertissement.

Les coûts d'ETP, les dépenses d'infrastructure et les risques commerciaux liés à un éventuel arrêt non planifié ou à une violation de Kafka sont évitables avec une plateforme gérée. Des connecteurs préconstruits, autogérés et plus faciles à utiliser pour Kafka (et la plupart des sources de données et systèmes potentiels), la gestion et l'observabilité des données, les capacités de découverte des données, ainsi que la sécurité et la gouvernance des données sont autant de fonctionnalités qui pourraient être activées en utilisant une plateforme de données complète. De plus, l'utilisation de Kafka dans un environnement de cloud hybride est entravée par les difficultés de connexion à divers environnements en temps réel ; les plateformes telles que Confluent sont précisément construites pour contourner cela. Enfin, les capacités de stockage sur Kafka deviennent pratiquement illimitées avec Confluent.

Compte tenu de tout cela, il serait difficile de mettre en œuvre les flux de données de qualité et fiables qui sont nécessaires pour atteindre l'architecture de « système nerveux central ». La mise en œuvre d'une

solution de streaming de données plus complète facilitera grandement la connectivité en temps réel pour le streaming de toutes les données d'une entreprise, à la fois en les ingérant et en les injectant là où c'est nécessaire.

## LE POINT DE VUE DE L'IDC

---

Les données en temps réel sont devenues un sujet clé pour la plupart des organisations. Considérées à l'origine comme de simples plateformes de messagerie, les plateformes de données gagnent du terrain dans de plus en plus de domaines. La sécurité, la gouvernance et la visibilité étant devenues des domaines critiques pour les données, s'appuyer sur le traitement par lots pour les réponses aux données s'avérera insuffisant pour beaucoup.

Le développement de solutions de streaming de données en temps réel finira par devenir un élément vital pour de nombreuses organisations, en commençant par certains cas d'usages et en s'étendant à d'autres secteurs d'activité. L'intérêt croissant pour les plateformes de données en continu est la preuve de leur importance grandissante pour les industries, seul un petit segment d'entreprises déclarant ne pas être intéressé par leur mise en œuvre. Des plateformes de streaming de données à grande échelle ont déjà franchi une étape supplémentaire, permettant à certaines organisations de construire l'intégralité de leurs infrastructures de données autour des données en streaming. En outre, l'utilisation des plateformes de streaming de données pour les contrôles de sécurité en temps réel et les tâches de sécurisation des données a un impact et un potentiel énormes pour leur futur développement et leur diffusion.

Cependant, le monde du streaming de données n'a pas acquis une connaissance très répandue en dehors de l'utilisation classique pour des applications et des tâches telles que la messagerie. Des logiciels tels que Confluent constituent un bon point de départ à cet égard. Toutefois, ce qui pourrait accélérer la reconnaissance de ses nombreux cas d'usages et de sa place centrale dans les plateformes de données, c'est l'abandon de sa nature excessivement orientée vers les développeurs au profit d'une voie d'adoption plus facile.

En particulier, l'amélioration considérable des fonctionnalités no-code permettra à un plus grand nombre d'utilisateurs commerciaux et analytiques de tirer parti des avantages des données en temps réel sans avoir à faire appel à des développeurs, qui sont généralement assez éloignés de la mentalité commerciale. Enfin, un autre problème peut consister à intégrer des plateformes de streaming de données en tant que « système nerveux central » dans les organisations qui ont déjà trop de logiciels et de plateformes en place dans leurs écosystèmes de données et d'analyse. Des capacités telles que le catalogue de données, la traçabilité des données ou la gestion des données de référence ont un impact énorme sur les piles de données modernes ; être capable d'exploiter pleinement toutes ces fonctionnalités tout en se débarrassant des systèmes hérités équivalents sera une étape finale pour une stratégie gagnante.

## EN SAVOIR PLUS

---

### Synopsis

Cette note de marché IDC aborde les principales conclusions du Kafka Summit de Londres, organisé par Confluent les 25 et 26 avril 2022. Le développement des solutions de streaming de données en temps réel va les aider à devenir un élément vital pour de nombreuses organisations ; l'intérêt croissant pour les plateformes de streaming de données est la preuve de leur importance grandissante pour les industries. La plateforme entièrement gérée Confluent, construite sur Apache Kafka open source, peut gérer des services de données en mouvement en temps réel et simplifier considérablement le parcours des utilisateurs de Kafka, grâce à ses connecteurs préconstruits vers diverses sources de données ainsi qu'à ses éléments de gestion, de gouvernance et de sécurité des données. Sa capacité de stockage virtuellement infinie pourrait entraîner une nouvelle façon d'envisager le stockage des données à l'avenir.

## À propos de IDC

International Data Corporation (IDC) est le premier fournisseur mondial de renseignements sur le marché, de services consultatifs et d'événements pour les marchés des technologies de l'information, des télécommunications et des technologies grand public. IDC aide les professionnels de l'informatique, les dirigeants d'entreprise et la communauté des investisseurs à prendre des décisions fondées sur des faits concernant les achats de technologie et la stratégie commerciale. Plus de 1 100 analystes d'IDC fournissent une expertise mondiale, régionale et locale sur les opportunités et les tendances technologiques et industrielles dans plus de 110 pays au monde. Depuis 50 ans, IDC fournit des informations stratégiques pour aider ses clients à atteindre leurs principaux objectifs commerciaux. IDC est une filiale d'IDG, le leader mondial des médias, de la recherche et des événements technologiques.

## 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)

---

### Avis de droit d'auteur

Ce document de recherche IDC a été publié dans le cadre d'un service d'intelligence continue IDC, fournissant des recherches écrites, des interactions d'analystes, des télébriefings et des conférences. Visitez [www.idc.com](http://www.idc.com) pour en savoir plus sur IDC, ses services d'abonnement et de conseil. Pour consulter la liste des bureaux IDC dans le monde, visitez [www.idc.com/offices](http://www.idc.com/offices). Veuillez contacter la hotline IDC au 800.343.4952, ext. 7988 (ou +1.508.988.7988) ou [sales@idc.com](mailto:sales@idc.com) pour des informations sur l'application du prix de ce document à l'achat d'un service IDC ou pour des informations sur des copies supplémentaires ou des droits web.

Copyright 2022 IDC. Toute reproduction est interdite sauf autorisation. Tous droits réservés.

