

DÉFINIR LE VRAI COÛT DE LA VERSION OPEN SOURCE DE KAFKA ET DE MSK

Les bonnes équipes s'épuisent à gérer Kafka®. Les équipes performantes utilisent Confluent.

Introduction	2
L'approche de Confluent en matière de rentabilité	2
Comparaison entre Confluent et un environnement Kafka open source autogéré.	3
Les coûts directs et indirects liés à l'autogestion de Kafka	3
Réduisez le TCO de votre infrastructure de streaming de près de 60 %	4
Limiter l'empreinte de votre infrastructure.	5
Assurer l'efficacité de vos opérations à grande échelle	5
Réduire les risques et les temps d'arrêt	6
Accélérez votre délai de rentabilisation	7
Multipliez par 3 le ROI de vos dépenses liées à Kafka	8
Comparaison entre Confluent Cloud et MSK.	9
Infrastructure	10
Coûts opérationnels	10
Coûts immatériels	11
Conclusion.	11
À propos de Confluent	12

Introduction

Chez Confluent, nous sommes convaincus que les données en mouvement seront au centre de la prochaine vague d'innovation technologique au sein des entreprises modernes. Elles permettent en effet de connecter et de traiter des flux de données, et d'y réagir en temps réel pour alimenter des activités et des applications diverses. Au lieu de stocker les données au repos et de les interroger de manière périodique, les entreprises les plus performantes traitent leurs flux de données en temps réel et de manière continue afin d'automatiser leurs opérations et de créer des expériences numériques de qualité pour leurs clients.

Utilisé par plus de 70 % des entreprises du classement Fortune 500, Apache Kafka® s'est depuis longtemps imposé comme l'infrastructure de choix pour les entreprises qui souhaitent mettre leurs données en mouvement dans le cadre d'une stratégie numérique moderne. Ces entreprises redoublent aujourd'hui d'efforts pour rester compétitives dans un environnement dynamique et résolument numérique et, pour bon nombre d'entre elles, Kafka est devenu un composant critique de leur infrastructure de données, et l'un des facteurs qui conditionnent leur réussite.

La gestion en interne de Kafka demande toutefois un investissement non négligeable en matière d'infrastructure et de ressources d'ingénierie, et s'accompagne d'un risque important d'interruption de service. Dans un contexte économique incertain (qui engendre de multiples contraintes), les équipes en charge de Kafka doivent gérer le déploiement et la mise à l'échelle de l'infrastructure, ainsi que les opérations quotidiennes, avec un budget et des effectifs souvent réduits. Certaines décident de se tourner vers des services Kafka hébergés sur le cloud pour réduire les coûts et la charge opérationnelle. Malheureusement, héberger un système de données sur une infrastructure cloud ne suffit pas pour exploiter pleinement la valeur et le potentiel de cette technologie.

C'est la raison pour laquelle de nombreuses entreprises se tournent vers Confluent afin de réduire le coût total de possession (TCO) de leur infrastructure de streaming de près de 60 %, d'accélérer leur délai de rentabilisation de plus de 6 mois et de multiplier par 3 leur retour sur investissement (ROI), tout en mettant en place une architecture de données solide et évolutive accessible à l'ensemble de leurs équipes. Alors que de nombreuses équipes doivent aujourd'hui « faire plus avec moins », Confluent est la seule solution à leur fournir un service Kafka complet et véritablement cloud native pouvant être déployé, exploité et mis à l'échelle en quelques minutes, tout en leur permettant de transférer la charge et la complexité opérationnelles, ainsi que les risques associés, à des experts de Kafka.

L'un des objectifs de ce whitepaper est de présenter notre approche pour mesurer la rentabilité de notre plateforme de streaming de données. Il explique le modèle d'évaluation que nous utilisons, l'approche consultative que nous avons choisi d'adopter auprès de nos clients, et les leçons que nous avons tirées de cette évaluation concrète réalisée auprès d'entreprises de différents secteurs et niveaux de maturité. Afin de mettre en avant les avantages que Confluent peut offrir aux entreprises, nous avons également choisi de présenter plusieurs exemples où nous comparons :

- Confluent et un environnement Kafka open source autogéré
- Confluent Cloud et Amazon MSK

L'approche de Confluent en matière de rentabilité

Afin d'illustrer les avantages professionnels qu'offre notre plateforme de streaming de données, nous avons décidé de développer trois axes :

- **Réduction du TCO**—Améliorer l'efficacité des opérations tout en réduisant les coûts d'infrastructure et de maintenance, et les risques d'interruption d'activité
- **Accélération du délai de rentabilisation**—Déployer Kafka à grande échelle en seulement une semaine
- **Optimisation du ROI**—Garantir la rentabilité de vos projets grâce à un lancement rapide et à une charge opérationnelle réduite

La plupart des entreprises cherchent à réduire leurs coûts et à optimiser leurs performances, en particulier au niveau de l'architecture de données qui sous-tend leurs activités numériques. Dans le cadre de l'évaluation du TCO, nous nous sommes intéressés aux coûts liés à la gestion à grande échelle d'un système distribué complexe. Ces coûts peuvent être répartis en trois catégories :

1. Les **coûts d'infrastructure**, qui incluent les coûts de traitement, de stockage et de mise en réseau alloués aux clusters provisionnés. Si le cloud a optimisé l'utilisation des infrastructures, les coûts potentiels restent importants lorsque les entreprises décident de prendre en charge des cas d'usage liés aux données en mouvement et de mettre en place des clusters supplémentaires pour gérer une quantité croissante de données. Ces coûts sont encore plus élevés si l'entreprise choisit de gérer elle-même l'infrastructure sur site.

2. Les **coûts opérationnels**, qui couvrent le temps, les ressources d'ingénierie et les services tiers nécessaires au développement et à la gestion d'une plateforme de streaming de données. Outre les ingénieurs chargés de créer et maintenir les outils permettant de passer en production de manière fiable et sécurisée, cette catégorie inclut aussi ceux qui gèrent la plateforme au quotidien. Elle peut également intégrer les dépenses pour des services professionnels, formations, plans d'assistance tiers, etc.
3. Les **coûts immatériels** (et associés aux risques professionnels) incluent les coûts liés aux temps d'arrêt non planifiés des clusters et aux incidents de sécurité. Ces coûts peuvent être difficiles à quantifier et leur étendue se révèle généralement en cas de problème. Sans parler du fait que, dans ce cas, des ressources précieuses doivent être réquisitionnées pour gérer les failles et les temps d'arrêt.

Comparaison entre Confluent et un environnement Kafka open source autogéré

Les coûts directs et indirects liés à l'autogestion de Kafka

Même si Kafka est devenu la plateforme de choix pour le streaming de données, sa gestion et sa mise à l'échelle en interne impliquent une charge, des coûts et des risques non négligeables. Voici quelques-uns des défis liés à l'autogestion de Kafka :

- **Charge opérationnelle et utilisation de ressources**—Gérer et mettre à l'échelle la plateforme afin de répondre à une demande toujours plus importante
- **Sécurité et gouvernance**—Garantir que les données de streaming sont aussi sécurisées et fiables que les données au repos, alors que de plus en plus d'entreprises utilisent Kafka
- **Connectivité et traitement en temps réel**—Tirer parti de données historiques de qualité pour alimenter des applications et des expériences modernes, basées sur le cloud
- **Disponibilité globale**—Assurer une haute disponibilité au sein d'environnement variés, avec un risque d'interruption minimal

Plus les cas d'usage et la quantité de données critiques liées à Kafka se multiplient, plus ces défis deviennent importants pour votre entreprise. Quel est leur impact sur votre activité ? Même si Kafka peut être téléchargé, personnalisé et utilisé gratuitement, sa gestion en interne implique des coûts directs et indirects non négligeables, notamment :

1. En termes de coûts d'infrastructure, le surprovisionnement dû à la fluctuation de la demande peut entraîner **des dépenses pouvant atteindre plusieurs centaines de milliers de dollars en cas d'infrastructures surdimensionnées et sub-optimales**. Avec Apache Kafka, les fonctionnalités de stockage et de traitement sont liées. Il existe en outre une limite concrète à la quantité de données que vous pouvez stocker sur un broker Apache Kafka. Une fois que vous avez atteint cette limite, vous devez provisionner des brokers supplémentaires et payer pour des ressources parfois inutiles.
2. En matière de coûts opérationnels, vous devez réquisitionner des ingénieurs pour créer et maintenir des outils liés à l'infrastructure de bas niveau au lieu de tirer parti de leurs connaissances sur des projets stratégiques à réelle valeur ajoutée pour votre entreprise. Il est en outre de plus en plus compliqué de recruter et de fidéliser des spécialistes de Kafka car il s'agit de compétences très recherchées. Tout cela peut engendrer des coûts de **3 à 5 millions de dollars en frais de développement et de fonctionnement de la plateforme**.
3. Enfin, pour ce qui est des risques liés à l'activité de votre entreprise, vous devez vous assurer que vos cas d'usage de Kafka sont suffisamment résilients et sécurisés pour éviter tout temps d'arrêt ou toute faille potentielle. Ces risques augmentent avec le nombre de cas d'usage et d'applications/systèmes/environnements utilisant Kafka au sein de votre entreprise. Les temps d'arrêt et les failles de sécurité peuvent également avoir de graves conséquences à long terme, y compris des **dommages à la réputation de votre entreprise, une baisse du niveau de satisfaction client, une perte de revenus et de données, et d'autres effets négatifs liés à des manquements réglementaires**.

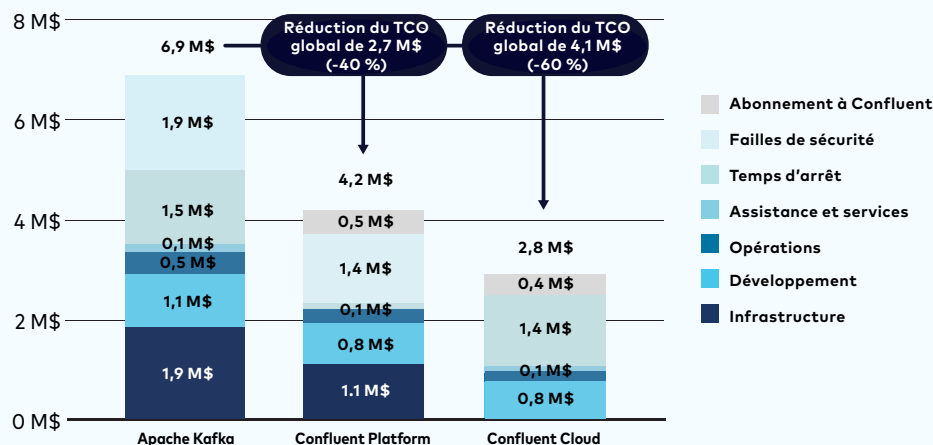
Comparons maintenant les performances de Confluent par rapport à celles de la version autogérée d'Apache Kafka.

Réduisez le TCO de votre infrastructure de streaming de près de 60 %

Nous avons commencé par estimer la répartition des coûts de fonctionnement d'un client gérant en interne une version open source de Kafka, et les éléments influençant ces coûts.

En analysant les opportunités d'optimisation de ces coûts, en partenariat avec nos clients, nous avons réalisé les observations suivantes :

TCO sur 3 ans : comparaison entre Apache Kafka et Confluent



Confluent offre à ses clients des fonctionnalités essentielles qui leur permettent de rentabiliser les différents coûts associés à leur infrastructure de streaming de données. Les économies en matière de TCO global sont de l'ordre de 60 % avec Confluent Cloud et de 40 % avec Confluent Platform.

Fonctionnalités de Confluent

Limiter l'empreinte de votre infrastructure

Assurer l'efficacité de vos opérations à grande échelle

Réduire les risques et les temps d'arrêt

Fonctionnalités de Confluent	Limiter l'empreinte de votre infrastructure	Assurer l'efficacité de vos opérations à grande échelle	Réduire les risques et les temps d'arrêt
Évolutivité flexible, sans serveur	●	●	
Stockage infini/hierarchisé	●	●	
Autoéquilibrage des clusters	●	●	
Clusters multi-régions	●	●	●
Fonctionnalité Cluster Linking	●	●	
ksqlDB/Stream Designer	●	●	
Contrôle d'accès basé sur les rôles, logs d'audit, fonction Secret Protection	●	●	●
UI Cloud/Control Center		●	●
Health+	●	●	●
Plus de 120 connecteurs préconfigurés	●	●	●
Outils d'infrastructure en tant que code		●	●
Validation de schémas, gouvernance de flux		●	●
Expertise basée sur les <i>committers</i>		●	●

Limiter l'empreinte de votre infrastructure

L'autogestion de Kafka sur une infrastructure privée peut engendrer des coûts d'infrastructure importants, liés notamment aux tâches de traitement, de stockage et de mise en réseau. Le fait d'héberger Kafka dans le cloud ne résout pas pleinement le problème puisque cela implique de payer des frais à un fournisseur cloud. Par exemple, si vous souhaitez exécuter Kafka sur AWS, vous devrez payer des instances EC2 pour exécuter vos brokers — et il vous reviendra d'en réduire ou d'en augmenter le nombre en fonction de la demande. Si vous utilisez un service Kubernetes comme EKS, vous devrez payer des nœuds et le service lui-même (c'est-à-dire les *masters* Kubernetes), sans oublier les coûts de mise en réseau. Le transfert de données sur des instances EC2 n'est en outre pas gratuit et, en fonction de vos paramètres réseau (VPC, lien privé ou internet public), vous devrez peut-être payer pour envoyer et recevoir des données.

Confluent Cloud a repensé Kafka pour éliminer les coûts d'infrastructure privée grâce à un service entièrement géré qui tire pleinement parti de la puissance du cloud. Grâce à son fonctionnement sans serveur, vos développeurs pourront créer et lancer des applications sans avoir à se préoccuper de l'architecture sous-jacente. Vous pourrez également profiter de la flexibilité de Confluent pour augmenter vos capacités afin de répondre à la charge souvent élevée liée aux données en mouvement, ou les réduire pour éviter le surprovisionnement. Avec notre modèle de tarification à l'utilisation, vous ne payez que ce que vous utilisez et l'infrastructure sous-jacente s'adapte dynamiquement à vos besoins sans impacter les applications en cours d'utilisation.

En termes de stockage, les opérateurs de plateformes Kafka doivent habituellement trouver un compromis entre offrir aux développeurs de longues périodes de rétention des données et limiter l'augmentation des coûts d'infrastructure. En effet, avec la version open source de Kafka, plus la période de rétention est longue, plus les besoins en matière de traitement et de stockage sont importants. La fonction Infinite Storage de Confluent a mis fin à ce dilemme. Elle permet de payer uniquement pour les données effectivement envoyées vers Confluent Cloud, tandis que le système s'adapte automatiquement au débit et à la période de rétention souhaitée. Plus besoin non plus de préprovisionner des ressources de traitement et de stockage, ou d'attendre avant que la mise à l'échelle soit effective.

En ce qui concerne Confluent Platform, sa fonction de stockage hiérarchisé permet à Kafka de définir deux entités de stockage : les disques locaux et les systèmes de stockage d'objets tels qu'Amazon S3 ou Google Cloud Storage (GCS). Cela vous permet de transférer les données issues d'anciens topics depuis des brokers de stockage coûteux vers des systèmes de stockage d'objets plus rentables afin de réduire considérablement les coûts. Cette fonctionnalité est particulièrement utile pour les cas d'usage impliquant des données historiques, par exemple le stockage de données de transactions financières pour des raisons réglementaires.

La fonctionnalité d'autoéquilibrage des clusters automatise totalement la réassignation des partitions, assurant une distribution cohérente de vos données sur chaque cluster afin d'optimiser les performances et l'utilisation des ressources. Enfin, les clusters multi-régions réduisent les besoins d'infrastructure liés au déploiement de plusieurs data centers. Ils tirent parti du moteur de réplication interne de Kafka pour répliquer de manière durable des partitions sur plusieurs centres de données à l'aide d'un seul cluster étendu, ce qui évite d'avoir à déployer un outil de réplication séparé basé sur Kafka Connect.

Assurer l'efficacité de vos opérations à grande échelle

L'utilisation de Kafka en production nécessite également un investissement considérable et à long terme en matière de ressources d'ingénierie. Les équipes DevOps et celles en charge de la plateforme doivent configurer, déployer, sécuriser, mettre à niveau et à l'échelle les clusters, et assurer leur disponibilité. Cela implique des tâches manuelles et complexes, qui prennent beaucoup de temps (un temps qui aurait pu être consacré à des projets à forte valeur ajoutée pour l'entreprise). D'autre part, ces tâches requièrent souvent la mise en place d'outils d'automatisation pour simplifier leur gestion et s'assurer que la plateforme fonctionne de manière fiable, mais le développement de tels éléments en interne implique des risques et nécessite une fois de plus de monopoliser des ressources d'ingénierie.



« Grâce à Confluent, nous avons pu réduire nos ressources internes dédiées à Kafka d'environ 50 %, ce qui a permis à nos ingénieurs et développeurs de se consacrer au développement de fonctionnalités innovantes et utiles pour nos clients. Depuis que nous faisons confiance à Confluent, nous n'avons également plus à nous préoccuper des interruptions de service. »

— JOSH LITTLESUN, DIRECTOR OF TECHNOLOGY CHEZ ROADSIDE

Avec Confluent Cloud, vous pourrez confier la gestion de vos clusters aux plus grands experts mondiaux, bénéficiant de plus d'un million d'heures d'expérience sur Kafka, et ainsi vous assurer que vos ingénieurs n'auront pas à se préoccuper de la gestion d'infrastructures de bas niveau et pourront se concentrer sur des tâches de développement qui offriront un avantage compétitif à votre entreprise.

Confluent Cloud vous permet :

- De déployer vos clusters Kafka en quelques minutes et de pouvoir les mettre à l'échelle de manière flexible pour répondre à tous vos besoins en matière de streaming de données
- De connecter vos sources et vos récepteurs de données critiques afin de créer des pipelines de données de streaming à l'aide de connecteurs entièrement gérés, sans charge opérationnelle et sans monopoliser vos développeurs
- De traiter facilement vos données de streaming avec une interface SQL simple et interactive grâce à notre technologie ksqlDB entièrement gérée

Confluent Platform peut également vous aider à réduire le temps et les ressources que vous consacrez à la création de composants permettant de faire de Kafka une plateforme de streaming de données professionnelle, et à la gestion quotidienne de cette plateforme.

La conception, la création, les tests et la maintenance de connecteurs Kafka peuvent prendre jusqu'à 6 mois d'ingénierie chacun, et la connexion à des systèmes propriétaires peut impliquer des coûts de licence non négligeables. L'extraction de données importantes concernant la santé des clusters implique d'identifier au préalable quelles données permettent de surveiller et de gérer efficacement ceux-ci. C'est pourquoi Confluent Platform inclut plus de 120 connecteurs préconfigurés et éprouvés, ainsi que la fonctionnalité Health+ qui propose une surveillance et des alertes basées sur le cloud, afin que les développeurs puissent se concentrer sur ce qui compte le plus pour l'entreprise. D'autre part, Confluent pour Kubernetes simplifie l'utilisation de Confluent Platform en tant que système cloud native sur Kubernetes afin de réduire davantage la charge opérationnelle grâce à des fonctionnalités de configuration, de déploiement, de mise à jour et de mise à l'échelle automatiques.

Confluent s'appuie également sur une assistance, des services professionnels et des formations pour réduire encore le temps nécessaire au développement et à l'exploitation d'une plateforme. Riches de plus d'1 million d'heures d'expérience technique sur Kafka, les équipes de Confluent aident leurs clients à maîtriser rapidement les bases de Kafka, à l'intégrer dans leur architecture existante, à s'assurer que la configuration de la plateforme correspond à leurs meilleures pratiques, et à réduire drastiquement le temps nécessaire à la maintenance et au dépannage.

Réduire les risques et les temps d'arrêt

Il est important de ne pas sous-estimer l'impact de risques tels que les temps d'arrêt, la dégradation des performances et les failles de sécurité. Selon une [étude réalisée par le cabinet Forrester](#) (document en anglais), les temps d'arrêt planifiés coûtent aux entreprises 1,5 million de dollars par trimestre. Et les interruptions non planifiées coûtent 35 % plus cher par minute. Une autre étude récente, réalisée par IBM, a révélé qu'une faille impliquant des données coûte en moyenne 4,35 milliards de dollars à l'entreprise concernée, un montant astronomique qui inclut les coûts de détection et de résolution, les amendes et frais d'indemnisation, et les dommages à la réputation de l'entreprise.

Si certaines entreprises indiquent ne rencontrer quasiment aucun temps d'arrêt au sein de leurs environnements autogérés, elles doivent malgré tout faire face à des problèmes de performances, qui peuvent à leur tour engendrer une baisse de la productivité et de la satisfaction des clients.

Nous avons par exemple travaillé avec une entreprise employant 2 000 personnes au total, dont 25 % étaient des ingénieurs logiciels. Parmi ces derniers, 25 % travaillaient sur des projets liés à Kafka. Cela signifie qu'environ 125 employés étaient impactés en cas de problèmes de performances de Kafka.

En calculant le coût total de ces ressources et l'impact en matière de productivité, nous avons également pu déterminer que le coût total par incident était de 230 000 dollars, sans parler de la frustration et de la charge que cela représentait pour les ressources d'ingénierie.

Celles-ci doivent disposer d'un haut niveau d'expertise et s'investir considérablement pour garantir une haute disponibilité de la plateforme, des besoins qui augmentent d'ailleurs avec le temps. Confluent Cloud allège cette charge grâce à son niveau de disponibilité

de 99,99 % garanti par SLA, couplé à l'absence d'interruption lors des mises à jour ou des opérations de maintenance planifiées. Les clusters multi-régions de Confluent Platform permettent d'autre part aux opérateurs d'améliorer la durabilité des données avec un RPO (objectif de temps de reprise) et un RTO (objectif de délai de restauration) proches de zéro, tout en automatisant la reprise après sinistre. Enfin, l'assistance 24h/24, 7 j/7 basée sur les *committers* de Confluent permet d'éviter les incidents et de garantir une résolution rapide en cas de problème.

En matière de risques, la sécurité de vos données est bien entendu essentielle. Il peut être coûteux et risqué de gérer ce sujet vous-même avec la version open source de Kafka. Confluent offre des fonctionnalités professionnelles de sécurité, de conformité et de gouvernance qui vous permettront de vous assurer que vos données seront toujours en sécurité.



« En termes de coûts directs, Confluent a réduit nos dépenses liées à l'infrastructure en limitant les licences et le matériel. Nous avons réduit le coût annuel de 69 % par rapport à nos prévisions, et avons évité de consacrer plusieurs mois à des réunions et à des projections. Nous avons donc également réduit nos coûts indirects. »

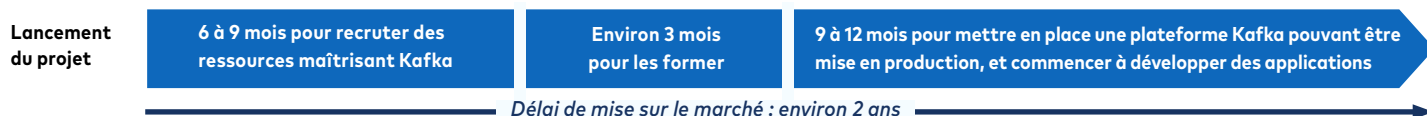
— JUSTIN DEMPSEY, SENIOR MANAGER, SOFTWARE DEVELOPMENT, CLOUD & INFORMATION SERVICES

Accélérez votre délai de rentabilisation

En plus de vous aider à réduire le TCO de votre infrastructure Kafka, Confluent peut vous permettre d'accélérer le délai de rentabilisation de vos cas d'usage (existants et nouveaux) liés au streaming de données. Si de nombreuses infrastructures de données modernes sont basées sur Kafka, sa version open source n'inclut pas nativement toutes les fonctionnalités dont votre entreprise a besoin. Voici quelques-unes des contraintes liées à la mise en place de cas d'usage impliquant des données en mouvement :

1. Configurer et déployer manuellement des brokers Kafka et ZooKeeper
2. Créer, tester et maintenir des connecteurs afin d'intégrer vos systèmes de données et applications critiques
3. Intégrer des outils GUI et dégager des indicateurs pertinents pour contrôler la santé des clusters et alerter en cas de problème
4. Créer des fonctionnalités de sécurité personnalisées en matière de contrôle d'accès, de cryptage et d'audit
5. Mettre en place des outils et une stratégie de reprise après sinistre pour protéger votre entreprise contre l'interruption d'activité et la perte de données

Développement traditionnel avec Kafka



Pour pouvoir déployer des applications et pipelines de streaming fiables, vous devez au préalable mettre en place les composants de base de Kafka. De nombreuses entreprises doivent recruter des ingénieurs maîtrisant Kafka pour gérer ce processus, ce qui rallonge encore le délai. L'ensemble du processus peut prendre jusqu'à deux ans (avant de pouvoir mettre en production à grande échelle), ce qui rallonge considérablement le délai de rentabilisation et empêche les entreprises de se montrer réactives face à leurs concurrents.

Développement avec Confluent



Afin d'aider les entreprises à exploiter pleinement le potentiel du streaming de données, Confluent offre un service flexible, évolutif et disponible dans le monde entier qui leur permet d'accélérer drastiquement le processus décrit plus haut. Plutôt que de monopoliser leurs équipes de développement pour mettre en place les outils nécessaires au fonctionnement de Kafka, nos clients peuvent tirer parti de notre plateforme complète pour passer en production rapidement et de manière fiable et sécurisée.

En tant que service cloud native entièrement géré et déployable en quelques minutes sur les principaux clouds publics, Confluent Cloud vous permet de vous lancer gratuitement et de limiter vos dépenses grâce à sa tarification à l'utilisation, qui évolue avec votre entreprise. Confluent Cloud vous libère d'une charge considérable : vos clusters sont provisionnés instantanément et leur maintenance assurée sans effort. Vous pouvez commencer à streamer des données dès votre inscription, en quelques minutes au lieu de plusieurs semaines voire plusieurs mois, sans charge opérationnelle importante.

Confluent Platform est une solution autogérée qui inclut des outils d'infrastructure en tant que code permettant de programmer automatiquement les clusters, plus de 120 connecteurs préconfigurés pour les sources et récepteurs les plus populaires, des outils de gestion et de surveillance spécifiquement conçus pour Kafka, des fonctionnalités de sécurité et de reprise après sinistre de niveau professionnel, et bien plus encore.

Grâce à son expertise basée sur les committers, Confluent permet également d'accélérer le délai de rentabilisation de Kafka. Pour les utilisateurs novices de Kafka, certains manquements en matière de conception architecturale et de configuration peuvent entraîner une augmentation des pannes, de la dette technique et des délais de réalisation des projets. Confluent offre un ensemble de services techniques et d'assistance qui vous aideront à prendre les meilleures décisions de manière à déployer des clusters évolutifs, sécurisés et résilients. D'autre part, notre offre de formation permet à nos clients de bénéficier rapidement d'une expertise Kafka en interne afin que leurs développeurs puissent commencer à créer des applications en temps réel tirant parti des données en mouvement.

Multipliez par 3 le ROI de vos dépenses liées à Kafka

Pour calculer le retour sur investissement, il est possible de tenir compte des prévisions de bénéfices liés à la mise en place d'un cas d'usage lié au streaming de données. Afin que ce calcul soit réaliste, il doit inclure une plage de résultats, ainsi que des limites inférieures et supérieures (simulation Monte Carlo). Nous essayons systématiquement de déterminer une fourchette de modélisation autour de l'option la plus conservatrice. C'est la meilleure façon de prévoir des événements potentiels malgré l'imprévisibilité de certains facteurs liés au contexte actuel.

Votre investissement dans Confluent peut être facilement multiplié par 3, via des économies sur les coûts d'infrastructure, le nombre d'ETP nécessaires et les coûts d'installation et d'exploitation.

	Confluent Cloud	Confluent Platform	Variation maximale par rapport à Kafka
Réduction en matière d'infrastructure	1,85 M\$	0,80 M\$	100%
Réduction en matière de développement	0,30 M\$	0,25 M\$	30%
Réduction en matière d'opérations	0,23 M\$	0,15 M\$	50%
Réduction en matière d'assistance et de services	0,14 M\$	0,14 M\$	100%
Coûts d'abonnement	-0,37 M\$	-0,47 M\$	N/A
ROI sur les abonnements	x5.9	x3.7	(Économies-abonnements)/ Abonnements

Le ROI est encore plus important si l'on prend en compte les coûts et bénéfices difficiles à quantifier tels que la réduction des temps d'arrêt, des risques en matière de sécurité et de la dégradation des performances, ou l'accélération de la mise sur le marché.

Il existe en outre d'autres avantages qui ne sont pas forcément pris en compte dans le calcul du TCO : amélioration de l'efficacité des développeurs et de l'agilité de l'entreprise, réduction des risques et, avantage non négligeable, libération de ressources clés auparavant mobilisées sur le paramétrage et l'exploitation de clusters Kafka. En déchargeant les meilleurs éléments de cette charge, ceux-ci sont en mesure de se concentrer sur les défis et problèmes véritablement importants pour l'entreprise.

Grâce à notre expérience inégalée au service de plus de 4 000 clients (de différents secteurs et avec différents niveaux de maturité) que nous avons aidés à exploiter le potentiel des données en mouvement, nous sommes les mieux placés pour vous aider à mettre en œuvre différents cas d'usage liés aux données de streaming, du concept au déploiement à grande échelle. Cela implique de préparer votre entreprise pour l'avenir en modernisant votre environnement technique à l'aide d'une architecture basée sur des microservices, afin d'intégrer plus facilement les autres services de l'entreprise qui pourraient adopter la plateforme ultérieurement. Cela implique également de créer une dynamique de collaboration au sein de vos équipes, de mettre en place une feuille de route prévoyant la mise en œuvre de cas d'usage en temps réel, et de favoriser un meilleur niveau de compréhension et d'adoption de la technologie.

Comparaison entre Confluent Cloud et MSK

Il existe de nombreux services Kafka prétendument autogérés, mais le concept d'autogestion semble être subjectif. Dans certains cas, il s'agit d'une expérience véritablement cloud native où les fonctionnalités produits et les automatisations facilitent considérablement les tâches opérationnelles et permettent aux équipes de se concentrer sur leurs applications. C'est le cas de Confluent Cloud, qui offre une expérience complète véritablement cloud native, de la planification au déploiement à grande échelle.

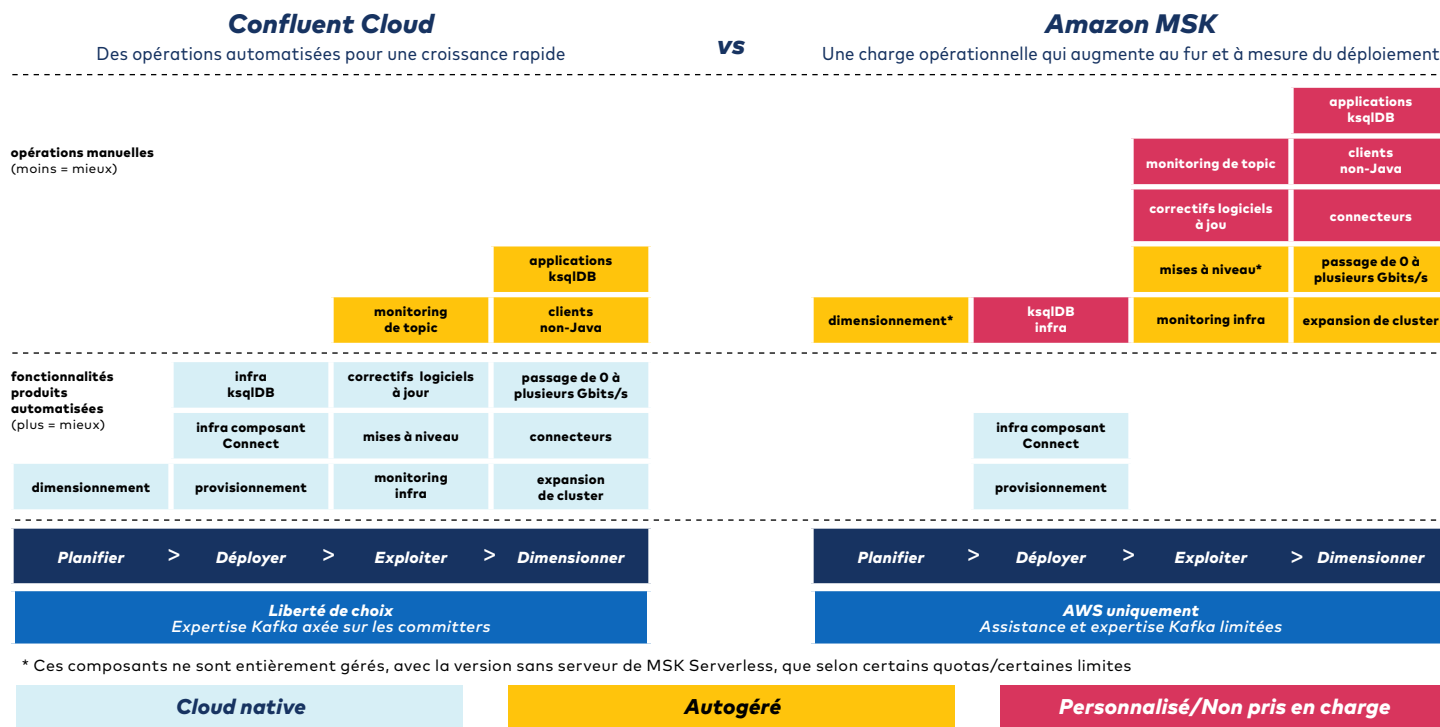
Il peut également s'agir de services pour lesquels certaines fonctionnalités sont absentes ou non prises en charge par le fournisseur, ce qui oblige les équipes à réaliser des développements personnalisés sans aucune assistance.

Entre les deux, il existe des services qui sont vendus comme entièrement gérés mais qui nécessitent de réaliser de nombreuses opérations manuelles et de disposer d'outils supplémentaires. Bon nombre de ces services promettent d'alléger la charge opérationnelle liée à Kafka mais, dans les faits, la plupart d'entre eux (y compris Amazon MSK) sont seulement capables de répondre aux besoins initiaux de votre stratégie de streaming de données à l'aide de fonctionnalités faciles à déployer. Il leur manque de véritables fonctionnalités d'automatisation qui permettraient aux équipes d'abandonner totalement la gestion quotidienne de Kafka pour se concentrer sur le développement d'applications. Même si Amazon propose désormais une offre sans serveur (MSK Serverless) qui répond à certaines de ces problématiques opérationnelles, cette offre reste limitée et il lui manque plusieurs composants pour être vraiment utilisable telle quelle en production.

Idéalement, chaque étape de votre stratégie de streaming événementiel doit être cloud native par nature, afin que votre équipe ne soit pas surchargée par des tâches liées à Kafka qui devraient être automatisées par votre service géré.

C'est pourquoi, lorsque l'on compare différentes solutions, il est important de prendre en compte des considérations techniques basées sur le cycle de vie des projets : PLANIFICATION > DÉPLOIEMENT > EXPLOITATION > MISE À L'ÉCHELLE.

Le diagramme ci-dessous compare le fonctionnement de Confluent Cloud sur l'ensemble de ce cycle de vie par rapport à celui d'Amazon MSK, et illustre clairement les différences qu'il existe entre une solution véritablement cloud native et une solution hébergée sur le cloud.



MSK n'étant ni entièrement géré, ni véritablement cloud native, les coûts et la charge opérationnelle qui y sont associés ne font qu'augmenter, en particulier après la phase de déploiement du cycle de vie des projets de streaming de données. Cela a un impact sur les différentes catégories de coûts liées à Confluent Cloud.

Infrastructure

Étant donné que MSK utilise un modèle de facturation par nœud, l'infrastructure occasionne des frais en permanence, que ce soit pour des opérations de traitement ou de stockage, et pour chaque cluster, même ceux qui offrent un débit limité pendant la phase de développement. Vous devez également allouer du temps et des ressources techniques à la réalisation de nombreux tests de performance afin de déterminer les types et les quantités de brokers. Si MSK Serverless intègre une fonctionnalité de dimensionnement basé sur le débit capable de prendre en charge jusqu'à 200 Mbits/s en écriture et 400 Mbits/s en lecture, celle-ci est limitée à 120 partitions maximum, ce qui est trop limité pour pouvoir mettre en œuvre un cas d'usage de production critique.

En termes de provisionnement, Confluent Cloud et Amazon MSK (que ce soit en version provisionnée ou sans serveur) permettent tous deux d'accéder à la demande à des fonctionnalités de provisionnement rapide dans le cadre d'une expérience cloud native. Ce qui distingue toutefois Confluent Cloud, c'est que vous payez à l'utilisation et pas en fonction de votre infrastructure, ce qui vous permet d'optimiser les coûts de votre stratégie Kafka.

Autre différence : Confluent Cloud utilise un modèle basé sur le débit (contrairement au modèle basé sur les nœuds de MSK) qui facilite le dimensionnement de votre infrastructure Kafka sous-jacente. Ce type de modèle vous permet de réduire vos coûts en évitant le surprovisionnement et en ne payant que pour la capacité de stockage que vous utilisez.

SecurityScorecard

« Depuis que nous avons créé Horus, notre plateforme mondiale de scan IPv4 basée sur Confluent, nous avons économisé plus d'un million de dollars par rapport à la version open source de Kafka ou de MSK. Nous sommes désormais une entreprise résiliente qui offre à ses clients une expérience de qualité et un haut niveau de disponibilité. Si nous y sommes parvenus, c'est grâce à Confluent, qui nous offre une infrastructure à la fois fiable et sécurisée pour tirer parti du streaming de données. »

— MATT CODDINGTON, SENIOR DIRECTOR OF DEVOPS ENGINEERING

Coûts opérationnels

Avec un service hébergé sur le cloud comme Amazon MSK, vous devez exploiter des paramètres bas niveau pour contrôler, diagnostiquer et corriger les problématiques liées aux performances des clusters.

MSK permet également à ses clients de mettre à niveau leur système en utilisant leur console ou une API, mais ce processus implique une surveillance manuelle active de l'environnement. MSK nécessite également des interventions manuelles pour rééquilibrer et réassigner les partitions, ce qui constitue une charge non négligeable en cas de *scale out*. En outre, MSK ne prend pas en charge le *scale in*. MSK Serverless offre des fonctionnalités de mise à l'échelle flexible sans avoir à redimensionner les clusters. Toutefois, comme déjà mentionné, il existe une limite matérielle de 120 partitions par cluster et une limite de rétention de 1 jour qui rend cette option inadaptée aux cas d'usage en production.

Confluent Cloud permet un monitoring proactif (et une réduction des coûts associés à celui-ci) sans avoir besoin d'exporter et d'analyser des paramètres relatifs à l'infrastructure. En ce qui concerne le monitoring des topics, Confluent Cloud fournit gratuitement des paramètres préagrégées, tandis que MSK applique une facturation à la consommation et nécessite une agrégation manuelle. D'autre part, Confluent Cloud offre des fonctionnalités flexibles de mise à l'échelle qui réduisent la charge opérationnelle et les coûts grâce à une tarification « jusqu'à zéro » pour les clusters basiques.

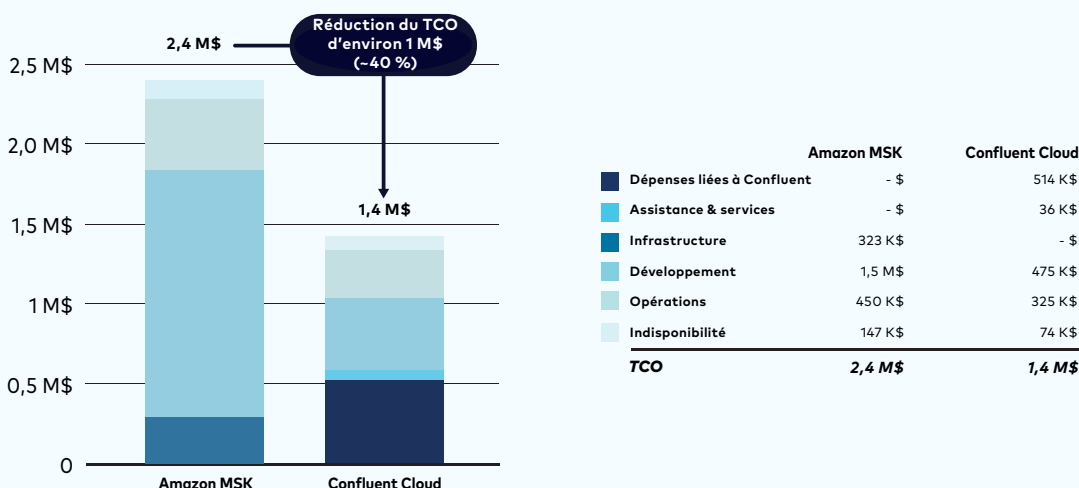
Coûts immatériels

Différents facteurs peuvent augmenter le risque d'interruption d'activité avec MSK. Cela s'explique notamment par le fait qu'il est possible de sous-estimer les besoins de stockage relatifs à un broker, ce qui peut provoquer une panne et des coûts associés en cas de pic de demande. D'autre part, la répartition manuelle des données implique des risques d'erreurs qui augmentent le risque d'interruption d'activité.

L'absence de mise à niveau progressive et de correctifs, le nombre limité de versions prises en charge et les exclusions de SLA en matière de pannes liées à Kafka sont autant d'éléments qui font qu'un problème peut persister pendant de longues périodes.

À l'inverse, Confluent Cloud vous permet de mettre facilement à jour vos clusters afin de les adapter à vos besoins en matière de streaming, de manière à ne pas sous-estimer les besoins en stockage. Grâce à la répartition automatique des données, Confluent Cloud optimise en permanence le placement de celles-ci sur les clusters Kafka afin d'équilibrer la charge, éliminant tout risque d'erreur manuelle qui pourrait provoquer des temps d'arrêt ou une dégradation des performances. Enfin, les mises à jour progressives vers la dernière version stable de Kafka (incluant des correctifs stratégiques) avant même le lancement prévu par Apache, ainsi que le niveau de disponibilité de 99,99 % garanti par SLA, sécurisent votre investissement et évitent les interruptions d'activité.

Dans le cadre d'une évaluation qui consistait à déterminer la différence de coût sur trois ans, nous avons modélisé des économies cumulées d'un million de \$ avec Confluent Cloud par rapport à Amazon MSK, ce qui représentait une réduction de 40 % du TCO lié au streaming de données.



Conclusion

Quels que soient votre expérience et votre niveau de maturité par rapport à Kafka (et au streaming de données), Confluent peut vous offrir des économies qui vous permettront d'amortir plusieurs fois le coût de l'abonnement. Si l'on ajoute la réduction du délai de rentabilisation des nouveaux cas d'usage, il devient évident que le ROI de Confluent est largement supérieur à celui de la technologie open source ou d'autres services Kafka hébergés sur le cloud. Vous devrez toutefois également veiller à mettre en place une stratégie efficace basée sur les données en mouvement, qui apportera une réelle valeur ajoutée à votre entreprise.

Si vous avez besoin que nous réalisons une évaluation du TCO spécifiquement adaptée à votre cas d'usage de Kafka, n'hésitez pas à [nous contacter](#) pour engager le dialogue.

Et, si vous souhaitez tester gratuitement Confluent Cloud, notre plateforme cloud native, rendez-vous sur notre [page d'inscription](#).

À propos de Confluent

Confluent est le pionnier d'une catégorie nouvelle d'infrastructure axée sur les données en mouvement. La solution cloud native de Confluent, qui a popularisé les données en mouvement, est conçue pour fournir un tissu conjonctif intelligent permettant aux données en temps réel provenant de sources multiples d'être diffusées en permanence dans l'ensemble de l'entreprise. Les clients de Confluent peuvent ainsi satisfaire à de nouveaux impératifs : offrir à leurs clients une expérience numérique de premier ordre au niveau front end, tout en mettant en place des opérations back end de niveau avancé, pilotées en temps réel par logiciel. Pour en savoir plus, rendez-vous sur notre site www.confluent.io/fr-fr/.