

PICNIC.APP

Picnic modernise son architecture de données avec Confluent pour l'analyse en continu



Siège

Amsterdam-Duivendrecht, Hollande du Nord, Pays-Bas

Industrie

Commerce de détail/Épicerie en ligne

Défi

Aborder les problèmes de frais généraux opérationnels et d'évolutivité liés aux technologies existantes afin de permettre une meilleure analyse du comportement des clients et améliorer les processus internes.

Solution

Confluent Cloud pour économiser du temps et de l'argent en réduisant les charges opérationnelles et permettant le traitement en temps réel et l'évolutivité facile des données d'événement.

Résultats

- Réduction des coûts d'infrastructure de 40 %
- Architecture de données simplifiée et évolutive
- Surveillance améliorée de l'infrastructure pour de meilleurs accords de niveau de service et la santé du système
- Élimination des pertes de données

Picnic est le supermarché seulement en ligne qui connaît la plus forte croissance en Europe. Il est actuellement présent en France, aux Pays-Bas et en Allemagne. L'entreprise s'appuie largement sur des décisions fondées sur les données pour offrir à ses clients la garantie du prix le plus bas. Picnic traite plus de 300 millions d'événements uniques par semaine provenant d'applications clients (c.-à-d., des applications d'achat) et de systèmes internes pour alimenter l'analyse prédictive dans son entrepôt de données.

La croissance ultra-rapide de Picnic, attribuable à la demande des consommateurs, à la reconnaissance de l'industrie et à un financement de série B réussi, a soutenu le besoin de trouver une plateforme de diffusion de données en continu plus performante et plus fiable, ainsi que des capacités d'analyse en continu améliorées.

C'est à ce moment-là que Picnic s'est tourné vers Confluent.

Le défi

Deux pipelines de données à croissance rapide

Picnic dispose de deux pipelines de données principaux. Le premier concerne l'application destinée aux clients, qui recueille des données sur le comportement des utilisateurs, pour ce qui concerne l'utilisation des produits et le passage à la caisse. Picnic analyse ces données pour déterminer comment améliorer les recommandations de produits et améliorer l'interface utilisateur (IU) de l'application.

Le deuxième pipeline de données traite les données destinées aux systèmes dorsaux internes, qui aident l'entrepôt de données, le traitement des paiements, l'état de disponibilité des produits et bien plus encore. Il permet à Picnic de planifier les achats, d'optimiser les opérations en entrepôt et de garantir aux clients une expérience positive en leur livrant les bons produits en temps voulu.

Au fur et à mesure que Picnic s'est développé, ils ont rencontré des problèmes en utilisant AWS Kinesis pour ces pipelines de données, en particulier :

- Une incapacité à stocker les données pendant plus d'une semaine. Les possibilités de relecture et de récupération sont essentielles pour que Picnic puisse corriger rapidement les événements mal formés ou les défaillances du pipeline, mais aussi pour réaliser des traitements de flux complexes.
- Picnic a commencé à avoir besoin d'un vaste outillage autour de ses pipelines de données pour permettre un flux de données en douceur vers d'autres systèmes sans avoir à créer des services clients à partir de zéro.
- Enfin et surtout, Picnic a eu besoin de la sémantique « une fois seulement » pour certains cas d'utilisation tels que l'évaluation dynamique des règles commerciales.

Solution technique

Mise à l'échelle facile et flexibilité des données

Picnic a estimé que le temps du changement était venu et a cherché une solution qui puisse répondre à tous les problèmes mentionnés ci-dessus en même temps. Leur recherche les a conduits à Confluent Cloud et à sa fondation Apache Kafka®.

Ils ont conçu à nouveau l'architecture des pipelines de données, ce qui a conduit à une simplification spectaculaire de leurs pipelines de services internes et à un traitement plus efficace des données de demande des clients.

Très rapidement, ils ont constaté les avantages suivants en passant à Confluent :

- Ils peuvent stocker les données dans Confluent aussi longtemps qu'ils le souhaitent, puisque le stockage infini de Confluent n'a pas de limite de volume ou de période de rétention.
- Confluent offre un riche écosystème de connecteurs préconstruits, entièrement gérés, prêts à être utilisés dès la sortie de la boîte et a éliminé la charge opérationnelle de leur maintenance
- La sémantique « une fois seulement » est enfin possible grâce à Kafka et aux connecteurs entièrement gérés de Confluent.

Auparavant, toutes les données provenant de RabbitMQ étaient envoyées dans un seul flux Kinesis, ce qui faisait peser la charge de la séparation des événements sur l'entrepôt de données et rendait l'évolutivité plus difficile.

Maintenant, Picnic utilise des connecteurs pour RabbitMQ afin de transférer les données de RabbitMQ vers les sujets Kafka de Confluent Cloud. Après quoi ils utilisent des connecteurs de réception entièrement gérés pour charger les données de manière transparente dans Snowflake et Amazon S3 pour une analyse plus approfondie par les équipes de science des données. Ils utilisent également la fonction Prévisualisation des données de Confluent sur des connecteurs entièrement gérés pour aider à réitérer et tester la sortie d'un connecteur avant de l'écrire en production.

Cette nouvelle configuration a permis une plus grande évolutivité et l'établissement de flux clairs de données homogènes. Elle a également bénéficié des connecteurs gérés de Confluent, qui ont permis à Picnic de surveiller plus facilement les API pour des solutions comme Prometheus.

Résultats commerciaux

Architecture simplifiée, économies d'argent

Réduction des coûts d'infrastructure : « Nos économies prévues ont diminué de 40 % et l'optimisation des coûts est l'un des plus grands impacts de Confluent chez Picnic », a déclaré Dima Kalashnikov, Responsable technique chez Picnic.

Une architecture informatique simplifiée : « La filière des services internes est passée de cinq services gérés à un seul et le nombre de services maintenus a diminué de 40 % pour l'ensemble des services gérés », a déclaré M. Kalashnikov.

Surveillance améliorée de l'infrastructure pour de meilleurs accords de niveau de service : Par Kalashnikov : « Grâce à la façon dont Confluent a simplifié notre infrastructure, cela nous a permis d'améliorer nos Accords de niveau de service sur la livraison des données et aussi d'éviter toute perte de données, ce qui est un Accord de niveau de service majeur pour nous. »

Quelle est la prochaine étape ?

À l'avenir, Picnic prévoit d'exploiter KSQLDB pour de nouveaux cas d'utilisation de l'analyse en continu, comme l'alimentation de tableaux de bord de rapports en temps réel, pour l'ETL en continu et pour permettre des choses comme les recommandations en temps réel basées sur le comportement de l'utilisateur.

Picnic souhaite également améliorer la livraison des données vers et depuis son centre de traitement automatisé et créer une plateforme en libre-service pour la livraison des données via des systèmes de déploiement automatisés pour les pipelines basés sur Confluent. Enfin, le but de ce projet est de livrer des données d'un centre de traitement automatisé à un entrepôt de données de façon que les analystes puissent les analyser en conjonction avec d'autres données disponibles.

« Nos développeurs d'apprentissage automatique et nos scientifiques de données peuvent désormais commencer à réfléchir et à travailler sur l'analyse des données en flux et les modèles en temps réel », a déclaré Kalashnikov. « Nous disposons désormais d'un écosystème vaste et florissant d'outils d'entreprise Confluent qui nous permet d'étendre notre plateforme d'analyse à des endroits où cela aurait été trop lourd par le passé. »

Learn More About Picnic

<https://picnic.app/de/>
<https://blog.picnic.nl/>

« Confluent offre exactement ce dont nous rêvions : un écosystème d'outils permettant de se procurer des données et de les réceptionner à partir du flux de données. Cela nous a apporté non seulement une grande agilité et une grande flexibilité du pipeline de données, mais aussi une infrastructure hautement simplifiée qui nous a permis de réduire les coûts. »

— DIMA KALASHNIKOV, RESPONSABLE TECHNIQUE