

Étude de cas



Lyft

Lyft accélère la découverte de données grâce à un outil basé sur Neo4j

SECTEUR

Transports

CAS D'USAGEDécouverte de données /
graphe de connaissances**OBJECTIF**

Permettre aux employés de Lyft de trouver rapidement les données à analyser

DÉFI

Il était difficile de trouver des données à cause de du volume et leur accumulation massive

SOLUTION

Création d'un outil qui accélère la découverte de données en utilisant Neo4j.

RÉSULTATS

- Productivité accrue de 30 % pour les data scientists de Lyft
- Note de satisfaction client de 8,5 sur 10 par les utilisateurs d'Amundsen

Avec un volume de données croissant et des collaborateurs de plus en plus nombreux, Lyft avait du mal à trouver les bonnes données à analyser. Lyft a alors mis au point l'outil Amundsen pour rationaliser la découverte de données.

L'entreprise

Implantée à San Francisco, [Lyft](#) est une entreprise de la « gig economy » (économie des travailleurs indépendants) qui passe par une application pour fournir plus de 50 millions de courses par mois à ses clients par le biais de son réseau de chauffeurs contractuels. Ceux-ci utilisent l'application Lyft qui génère des données sur la base de toutes les courses proposées. Lyft fait figure de principal concurrent d'Uber, opère dans 640 villes aux États-Unis et dans neuf villes au Canada.

La donnée est au cœur de toutes les décisions chez Lyft. Une fois les décisions prises, leur impact est évalué en utilisant les données.

Compte-tenu de ce rôle capital des données et de leur analyse pour l'entreprise, la vitesse à laquelle les utilisateurs sont en mesure d'identifier ces données, de les analyser et d'en extraire une visibilité est stratégique.

La data discovery ou découverte de données – qui consiste à trouver les bonnes données et à les comprendre – s'avérait jusque-là lente et inefficace. Certains tableurs avaient des noms similaires, par exemple `driver_rides_completed` et `rides_driver_total.lifetime_completed`. Et les collaborateurs se demandaient mutuellement de l'aide en se contactant sur Slack, ou consultaient Github pour voir comment était généré un tableur. Souvent, ils en extrayaient les cent premières lignes pour se faire une idée des contenus.

Avec la croissance de Lyft, le défi de la découverte des données s'est amplifiée. Et Lyft comptait déjà 10 pétaoctets répartis entre plusieurs milliers de tableaux et différents stockages, selon Tamika Tannis, ingénieure logiciel chez Lyft. Avec cette croissance, l'application mobile et d'autres services continuaient de générer toujours plus de données. De nouveaux talents ont alors été embauchés si bien que le nombre d'utilisateurs internes procédant à la découverte des données augmentait également.

Lyft a donc dû trouver un meilleur moyen de prendre en charge la découverte de données pour chaque personne dans l'entreprise. Afin d'évaluer le problème, l'équipe de Tamika Tannis s'est d'abord penchée sur les data scientists et a constaté que la découverte des données occupait près d'un tiers de leur temps.

La solution

Les ingénieurs de Lyft ont alors décidé de développer un outil pour simplifier la découverte de données. Ils ont d'abord ciblé le public des utilisateurs les plus fréquents : les analystes et les data scientists.

Nommé [Amundsen](#), cet outil allait offrir trois moyens complémentaires d'effectuer de la découverte de données : par recherche, par lignage et en réseau.

Étude de cas



« Amundsen est une vraie réussite chez Lyft. Actuellement, nous avons des ingénieurs, des responsables produits et même des collaborateurs du service client qui l'utilisent pour trouver ce dont ils ont besoin »

– Tamika Tannis,
ingénieure logiciel, Lyft

La priorité était d'aboutir à une recherche efficace qui classe les résultats par popularité et par pertinence. La découverte par lignage suit les relations entre jeux de données. La découverte de données en réseau connecte les données et les personnes, un atout pour les nouveaux collaborateurs en particulier.

« On peut avoir besoin de voir certaines données utilisées par son manager ou ses collègues, cela permet d'accéder à des données fiables que tout le monde emploie déjà dans des buts similaires » indique Tamika Tannis.

Amundsen a recours à une architecture de micro-services. Son service Databuilder prend en charge les données dans le service de recherche, qui s'appuie sur Elasticsearch, et dans le service de métadonnées exécuté par la [base de données de graphes Neo4j](#). Elasticsearch alimente la recherche en fournissant une pertinence en fonction des termes de recherche, de la fonction de l'utilisateur dans l'entreprise et de la popularité des tableaux ; mais toutes ces relations sont effectuées en premier lieu dans Neo4j.

Lyft a choisi Neo4j parce que la solution saisit la configuration de son écosystème de données, lequel est naturellement représenté par un graphe. La flexibilité de Neo4j est très avantageuse pour faire des itérations rapides sur de nouvelles fonctions.

« Lorsque nous avons un nouveau cas d'usage et un nouvel élément de métadonnées à représenter, il nous suffit de créer cette relation » explique Tamika Tannis.

Chez Lyft, Neo4j est un composant essentiel de l'architecture d'Amundsen et sert de source de référence pour les métadonnées modifiables. Neo4j apporte également un socle aux nouveaux projets, notamment dans la conformité et la qualité des données. « L'avenir, à mon avis, consiste à disposer d'un référentiel de métadonnées complet sur lequel construire de nombreuses applications » déclare Mark Grover, responsable produit chez Lyft.

Les résultats

Amundsen connaît une forte adoption. « L'adoption monte à 90 % chez nos spécialistes en science de la donnée qui font partie de nos publics cibles » précise Tamika Tannis. « Cela veut dire que 90 % de nos data scientists utilisent Amundsen dans leur travail hebdomadaire. Nous constatons également que cet outil a renforcé la productivité de près de 30 % dans tout notre service de science de la donnée » poursuit-elle.

Avant même que l'équipe n'ouvre Amundsen à de nouveaux utilisateurs, certains se sont mis à l'utiliser. « Amundsen est une vraie réussite chez Lyft. Actuellement, nous avons des ingénieurs, des responsables produits et même des collaborateurs du service client qui l'utilisent pour trouver ce dont ils ont besoin » relate Tamika Tannis.

Les utilisateurs partagent leurs commentaires directement dans l'application. Ils peuvent ainsi se faire entendre pour évaluer l'outil ou demander de nouvelles fonctionnalités. « Si quelque chose semble incorrect, ils peuvent tout de suite remonter le bug » explique Tamika Tannis. Amundsen reçoit des notes élevées des employés de Lyft. Sa note de satisfaction client (CSAT) est de 8,5 sur 10.

Pour donner davantage d'impact à Amundsen, Lyft a décidé de proposer l'outil en open source. Une communauté active grandit autour de lui et plus de dix entreprises contribuent maintenant au projet.

Neo4j est le leader de la technologie des bases de données de graphes. Avec le plus grand nombre de déploiements au monde, Neo4j aide des entreprises mondiales - comme Airbus, [Michelin](#), [NASA](#), [Crédit Agricole](#) et [Volvo Cars](#) - à prédire et identifier la façon dont les personnes, les processus, les lieux et les systèmes sont interconnectés. Grâce à cette approche par les relations, les [applications](#) mises au point en utilisant Neo4j relèvent les défis associés aux données connectées, tels que l'[analytique et l'intelligence artificielle](#), la [détection de fraude](#), les [recommandations](#) en temps réel et les [graphes de connaissance](#). Pour en savoir plus, merci de consulter [Neo4j.com](#) et [@Neo4jFr](#).

Des questions sur Neo4j ?

Contactez-nous :
info@neo4j.com
neo4j.com/contact-us