

Étude de cas



Novartis

Novartis, à la pointe des connaissances en biologie pour mettre au point de nouveaux médicaments

SECTEUR

Santé / Sciences de la vie

CAS D'USAGE

Graphe de connaissances

OBJECTIF

Accélérer le processus de mise au point de médicaments

DÉFI

Trouver des associations à travers de vastes ensembles de données expérimentales issus de la recherche médicale

SOLUTION

Création d'un graphe de connaissances biologiques liées à la recherche

RÉSULTATS

- Permettre aux chercheurs de corréler les données expérimentales avec la recherche médicale
- Identifier des composés prometteurs pour accélérer le développement de médicaments

Novartis cherchait un moyen pour ses chercheurs de se poser des questions clés et d'obtenir des réponses rapides étayées par la recherche médicale. L'équipe de recherche biomédicale de Novartis utilise depuis Neo4j pour explorer et organiser des volumes massifs de données biologiques sur lesquelles s'appuie la mise au point de médicaments de nouvelle génération.

L'entreprise

Novartis est une entreprise pharmaceutique mondiale basée à Bâle, en Suisse, qui fournit des solutions pour répondre aux évolutions des besoins des patients. Il s'agit de l'une des plus grandes sociétés du secteur de la santé tant par sa capitalisation boursière que par ses ventes. Les Instituts Novartis pour la recherche biomédicale constituent la branche innovation de Novartis, avec 6 000 chercheurs répartis sur six sites dans le monde.

Le défi

Depuis des décennies, Novartis accumule des données sur l'effet de divers composés sur des cibles biologiques, notamment des protéines telles que les enzymes, ce qui représente près d'un milliard de données en tout. Ces données historiques sont stratégiques, mais minimes par rapport à l'incroyable granularité des données actuellement collectées.

Aujourd'hui, Novartis utilise aussi un processus automatisé pour saisir des données d'images au contenu dense qui montrent comment un composé en particulier affecte une culture entière de cellules. Il en résulte des données phénotypiques de plusieurs téraoctets.

Novartis devait relever le défi de combiner ses données stockées historiquement avec ces données phénotypiques émergentes. Il lui fallait également trouver le moyen de mettre ces données dans le contexte plus large de la recherche médicale en cours dans le monde entier.

L'équipe de Novartis souhaitait croiser ses données avec les informations médicales de la base PubMed du National Institute of Health (Institut Américain de la Santé). PubMed contient plus de 25 millions d'extraits issus d'environ 5 600 journaux scientifiques.

L'équipe recherchait également comment permettre aux chercheurs de poser des questions reliant les points entre toutes ces données dans le contexte de la recherche médicale la plus récente.

Comme l'explique Stephan Reiling, scientifique principal chez Novartis, « lorsque nous essayons d'analyser ces données, il apparaît clairement qu'il nous faut un moyen de stocker les connaissances biologiques puis de les soumettre à des requêtes. »

Étude de cas



« En fusionnant ces données, nous créons un graphe géant pour mieux comprendre la biologie et la façon dont nous pouvons utiliser ces connaissances scientifiques pour développer la nouvelle génération de médicaments »

– *Stephan Reiling,*
scientifique principal
chez Novartis

La solution

Collecter et connecter des données sur les maladies, les gènes et les composés – en identifiant en parallèle la nature des relations entre ces éléments – promettait d'accélérer les découvertes en matière de médicaments.

L'équipe de Novartis voulait relier les gènes, les maladies et les composés selon un modèle triangulaire. « Pour réussir à découvrir des médicaments, il faut naviguer dans ce triangle » explique Stephan Reiling. L'équipe a décidé de créer un graphe de connaissances stocké dans Neo4j et a conçu un procédé pour intégrer les données des dernières recherches médicales.

Au début de ce procédé, les textes sont explorés pour extraire les données pertinentes de PubMed. Ces données sont ensuite envoyées dans Neo4j avec les données historiques et les données d'images de Novartis. Le circuit suivi par les données alimente les 15 types de nœuds qui ont été définis pour coder les données. À l'étape suivante, les informations sur les relations entre ces nœuds sont entrées. L'équipe a ainsi identifié plus de 90 relations différentes.

Novartis utilise les algorithmes de graphes de Neo4j pour traverser le graphe et identifier un modèle triangulaire de nœuds escompté, reliant les trois classes de données entre elles. L'analyse du graphe repère non seulement les nœuds pertinents pour la relation triangulaire recherchée, mais utilise également un indicateur conçu par l'équipe pour évaluer le poids associé entre chaque nœud de chaque triangle. Grâce à cette fonction, l'équipe élabore des requêtes pour trouver les données liées au modèle de nœuds recherché avec un poids associé donné, puis classe les triangles en fonction de cet indicateur.

Lorsque les chercheurs interrogent le graphe de connaissances, les résultats indiquent le niveau de corrélation entre les éléments. Si un chercheur connaît déjà une corrélation spécifique, il peut choisir d'en investiguer d'autres pouvant orienter son travail dans de nouvelles directions.

Le résultats

Neo4j permet à Novartis de naviguer en toute souplesse dans toutes ses sources de données, ce qui est un atout pour la recherche.

« En fusionnant ces données, nous créons un graphe géant pour mieux comprendre la biologie et la façon dont nous pouvons utiliser ces connaissances scientifiques pour développer la nouvelle génération de médicaments » indique Stephan Reiling.

Le graphe de connaissances de Neo4j capture les éléments nécessaires à un système opérationnel de compréhension biologique qui continue de croître avec l'avancée des sciences médicales. Actuellement, ce graphe de connaissances compte un demi-milliard de relations et l'équipe s'attend à tripler facilement ce chiffre avec l'ajout de nouvelles données.

En utilisant Neo4j, les chercheurs voient quels composés et quels gènes sont le plus étroitement associés aux maladies. Ils explorent aussi la littérature médicale pour examiner les preuves de cette association. « Un vaste volume de données biologiques est à disposition, avec des sources de données remarquables. En regroupant réellement toutes ces données, je peux pour la première fois décider de trouver des composés similaires à tel autre avec des indications sur telle maladie. La flexibilité avec laquelle naviguer dans toutes ces sources de données est vraiment très puissante » ajoute Stephan Reiling.

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