

## Estudio de caso



## Novartis

## Novartis recopila los datos biológicos más recientes para desarrollar fármacos

**INDUSTRIA**

Asistencia sanitaria  
Sanidad/ Ciencias biológicas

**CASO DE USO**

Grafo de conocimientos

**OBJETIVO**

Acelerar el proceso de desarrollo de fármacos que salvan vidas

**RETO**

Encontrar asociaciones entre multitud de datos experimentales e investigaciones médicas

**SOLUCIÓN**

Crear un grafo que capture conocimientos biológicos vinculados a investigaciones.

**RESULTADOS**

- Capacitar a los investigadores para correlacionar los datos experimentales con la investigación médica.
- Identificar compuestos con potencial para acelerar el desarrollo de fármacos

*Novartis quería ayudar a sus investigadores a plantear preguntas importantes y obtener respuestas rápidas, respaldadas por investigaciones médicas. El equipo de investigación biomédica de Novartis utiliza Neo4j para extraer y organizar grandes volúmenes de datos biológicos que sustentan el desarrollo de la próxima generación de medicamentos.*

**La compañía**

Novartis, una compañía internacional de productos médicos con sede en la ciudad suiza de Basilea, ofrece soluciones para las necesidades cambiantes de los pacientes. Su capital de mercado y sus ventas la sitúan entre las principales farmacéuticas del mundo. Los Institutos Novartis para la Investigación Biomédica son el brazo innovador de Novartis, con 6 000 investigadores a su servicio en seis ubicaciones en distintos países.

**El reto**

Novartis lleva décadas recopilando información sobre el efecto de diversos compuestos en ciertas proteínas (por ejemplo, enzimas) y dispone de aproximadamente mil millones de datos. A pesar de la importancia de esa información, su envergadura es ínfima comparada con la asombrosa granularidad de los datos que se recopilan actualmente.

Hoy en día, Novartis utiliza un proceso automatizado que captura datos de imágenes de alto contenido que muestran cómo un compuesto concreto ha afectado a todo un cultivo celular. Esto genera terabytes de datos fenotípicos.

Novartis afrontaba el reto de combinar sus almacenes de datos históricos con estos datos fenotípicos en expansión. También tenían que ubicar todos esos datos dentro del contexto más amplio de investigaciones médicas activas en todo el mundo.

El equipo de Novartis quería combinar sus datos con información de las bibliotecas médicas PubMed de los Institutos Nacionales de la Salud estadounidense. En PubMed hay alrededor de 25 millones de sinopsis de unas 5 600 publicaciones científicas.

El equipo de Novartis buscaba una forma de capacitar a los investigadores para hacer preguntas destinadas a conectar todos esos datos en el contexto de las investigaciones médicas más recientes.

En palabras de Stephan Reiling, científico sénior en Novartis, «cuando tratamos de analizar estos datos, se hace mucho más evidente que necesitamos una solución para almacenar los conocimientos biológicos y ejecutar consultas con ellos».

**La solución**

La captura y conexión de datos sobre enfermedades, genes y compuestos, así como la identificación de la naturaleza de las relaciones entre esos elementos, prometía acelerar el descubrimiento de fármacos.

## Estudio de caso



« Al fusionar estos datos, estamos creando un grafo gigante para comprender mejor la biología y cómo podemos usar estos conocimientos científicos para desarrollar la próxima generación de fármacos. »

– *Stephan Reiling,*  
*científico sénior*  
*en Novartis*

El equipo de Novartis quería vincular genes, enfermedades y compuestos en un patrón triangular. «Para descubrir fármacos eficaces, hay que saber usar ese triángulo», explicó Reiling.

El equipo de Novartis decidió crear un grafo de conocimientos almacenado en Neo4j e ideó un flujo de procesamiento para capturar las investigaciones médicas más recientes.

Al principio del procesamiento también se utiliza la extracción de texto, para obtener datos relevantes de PubMed. Esos datos se proporcionan a Neo4j, junto con los datos históricos de Novartis y sus datos de imágenes.

En el flujo de procesamiento se rellenan los 15 tipos de nodos establecidos para codificar los datos. En la fase siguiente se rellena la información de relaciones entre los nodos. El equipo identificó más de 90 relaciones diferentes.

Novartis utiliza algoritmos de grafos de Neo4j para examinar el grafo e identificar un patrón de nodo triangular deseado entre las tres clases de datos. Los análisis de grafos encuentran nodos relevantes en una relación triangular deseada, y además utilizan una métrica diseñada por el equipo para calcular lo estrecha que es la relación entre los nodos de un triángulo. Usando esta función, el equipo ideó consultas para encontrar datos vinculados por el patrón de nodos deseado, con una magnitud de asociación determinada, y luego ordenar los triángulos de acuerdo con esta métrica.

Cuando los investigadores consultan el grafo de conocimiento, los resultados muestran la magnitud de la correlación entre elementos. Si un investigador ya está al tanto de una correlación con una asociación estrecha, puede optar por investigar otras. Esto es algo que podría abrir nuevas vías en su trabajo.

### El resultado

Neo4j permite a Novartis navegar de manera flexible por todas sus fuentes de datos, lo que constituye una ayuda valiosa para la investigación.

«Al fusionar estos datos, estamos creando un grafo gigante para comprender mejor la biología y cómo podemos usar estos conocimientos científicos para desarrollar la próxima generación de fármacos», dijo Reiling.

El grafo de conocimientos de Neo4j captura elementos necesarios para un sistema operativo de comprensión biológica que crece a la par de la expansión de la ciencia médica. Actualmente hay 500 millones de relaciones en el grafo de conocimientos y el equipo cree que es fácil triplicar ese número al ir agregando datos.

Con Neo4j, los investigadores ven qué compuestos y genes están más estrechamente asociados con ciertas enfermedades, y profundizan en la bibliografía médica para examinar la evidencia de la asociación.

«Hay una gran cantidad de datos biológicos disponibles y magníficas fuentes de datos. Aunando todo eso podemos decir: 'Quiero encontrar compuestos similares a este otro que tengan anotaciones sobre esta enfermedad'. La flexibilidad para navegar por todas estas fuentes de datos es realmente eficaz», afirmó Reiling.