ESTUDIO DE CASO

El Consorcio Internacional de Periodistas de Investigación (ICIJ)

Cómo el ICIJ utilizó la tecnología de base de datos de grafos para destapar el escándalo Swissleaks

INDUSTRIA
Medios de comunicación

RETO
Ayudar a los reporteros a penetrar en las complejas bases de datos de Swissleaks para un periodismo de investigación mejor.

ESTRATEGIA
- Crear una nueva herramienta de análisis de bases de datos de grafos para que los periodistas puedan investigar y descubrir nuevos hechos para sus historias, de una forma rápida y fácil.
- Compartir la herramienta a escala mundial para recibir colaboración en el proyecto Swissleaks.

SOLUCIÓN
- Neo4j fue la herramienta idónea para gestionar búsquedas, consultas y datos complejos.
- La herramienta de visualización Linkurious puso el proceso de análisis a disposición de todos los reporteros del ICIJ.

RESULTADO
- Permite a periodistas internacionales buscar y descubrir datos complejos y poco estructurados de forma fácil y rápida.
- La base de datos de grafos desvela ángulos sorprendentes e historias que antes no se hubieran descubierto.
- Un punto de inflexión para el análisis de fraudes que da un giro al proceso de investigación periodística.

Demostración del poder de los grafos en el periodismo de investigación.

A principios de 2014, dos periodistas franceses de Le Monde, Gérard Davet y Fabrice Lhomme, consiguieron acceder a un conjunto de datos enormemente complejo, además de muy valioso. Los datos revelaban los nombres de más de 100 000 clientes y sus cuentas bancarias en la filial suiza del HSBC y, con ello, quedaban al descubierto prácticas poco éticas.

Davet y Lhomme se dieron cuenta del enorme valor de los datos para destapar una historia inaudita sobre fraude, evasión fiscal y crimen internacional. ¿El problema? Los datos eran sencillamente demasiado complejos de analizar utilizando los medios tradicionales.

El reto: acceder a complejas bases de datos para obtener información de gran valor

Normalmente los reporteros tienen que intentar encontrar relaciones entre datos y archivos Excel, realizar búsquedas manuales en Internet y, a veces, trazar físicamente conexiones entre personas y entidades para dar con los hechos correctos para sus historias.

No obstante, Davet y Lhomme reconocieron que el conjunto de datos Swissleaks era sencillamente demasiado complejo de analizar manualmente o en la empresa. De modo que acudieron al Consorcio Internacional de Periodistas de Investigación (ICIJ) e iniciaron una de las colaboraciones periodísticas más grandes de todos los tiempos.

Mar Cabra, editora de la unidad de datos e investigación en el ICIJ, sabía que necesitarían una herramienta que pudiera analizar más detalladamente las relaciones entre los datos, tanto para esta como para futuras investigaciones.

Los datos Swissleaks contenían información de titulares de cuentas del HSBC repartidos por más de 200 países y con un capital conjunto de más de 100 000 millones de dólares. Sin embargo, la información estaba desperdigada en miles de archivos sin una conexión directa entre sí. La complejidad de los datos, según Cabra y el ICIJ, requería una herramienta para analizar la gran cantidad de datos sin estructurar y encontrarles un sentido de una forma rápida y sencilla.

Cabra también tenía un requisito más: no quería que científicos ni desarrolladores de datos tuvieran que hacer de intermediarios. Cabra quería que el proceso de análisis y descubrimiento de datos fuera accesible para todos los reporteros del mundo, sin importar sus conocimientos técnicos.

La solución: El análisis de grafos es la clave para "seguir la pista del dinero"

El caso Swissleaks no era la primera experiencia de la organización con un conjunto de datos complejo. Por lo tanto, Cabra sabía que el ICIJ necesitaría una solución de base de datos de grafos.

ESTUDIO DE CASO

"Poder utilizar la tecnología Neo4j y Linkurious representó un punto de inflexión para nosotros".

– Mar Cabra

"Trabajando en casos como el Offshore leaks, aprendí la importancia de los análisis de grafos en la investigación de corrupción fiscal", dijo Cabra. "Las conexiones son cruciales para ver la realidad de una historia: te muestran quién hace negocios con quién. En seguida supimos que necesitábamos un medio basado en grafos para el caso del HSBC".

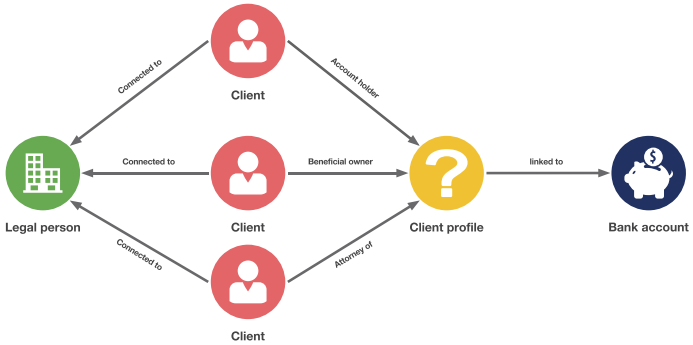
El primer paso de la Unidad de datos e investigación fue recrear la base de datos de clientes del HSBC a partir de los archivos planos de Excel proporcionados. A continuación, enlazaron cada nombre con uno o más países (ambos referidos como los "nodos" en la base de datos de grafos). Finalmente, convirtieron los datos en un formato de grafos para explorar las conexiones entre nodos.

En total, la filtración contenía unos 60 000 archivos con información sobre más de 100 000 clientes en 203 países. La base de datos de grafos resultante tenía más de 275 000 nodos con 400 000 relaciones entre ellos.

El ICIJ trabajó con Talend, un especialista en software de integración de código abierto, para transferir la base de datos original a la base de datos de grafos Neo4j de Neo Technology. Otro colaborador de Neo, Linkurious, proporcionó una aplicación web como interfaz de usuario para que los reporteros pudieran ver y entrar en la base de datos de grafos fácilmente.

La visualización de grafos permitió a los periodistas del ICIJ identificar las conexiones entre personas y cuentas bancarias, lo que les ayudó a "seguir la pista del dinero" para detectar numerosos casos de fraude, corrupción y evasión fiscal.

A continuación, aparece un ejemplo que ilustra estos tipos de relaciones gráficas.



Esquema que representa las entidades y las conexiones encontradas en el conjunto de datos de Swissleaks.

El resultado: El desbloqueo de datos complejos para la investigación del fraude

Después de importar los datos a Neo4j y Linkurious, Cabra y su equipo notaron una gran diferencia en cómo los periodistas podían utilizar los datos de Swissleaks.

"Poder utilizar la tecnología Neo4j y Linkurious supuso un punto de inflexión para nosotros", comentó Cabra. "Permitieron a nuestros periodistas investigar datos financieros complejos de formas muy diferentes de cómo lo hacíamos antes".

La solución Neo4j-Linkurious también satisfizo el requisito de Cabra de facilidad de uso.

"Era tan intuitivo que con muy poca formación los reporteros ya podían utilizarlo", dijo Cabra. "Solo realizamos una breve demostración en línea y un seminario web. Nuestros reporteros quedaron sorprendidos de lo fácil que era consultar los datos; no necesitaban conocimientos avanzados de informática para ello".

Después de que el equipo de Cabra mostrará el funcionamiento de la herramienta en la sala de redacción virtual del ICIJ, periodistas de todo el mundo accedieron al conjunto de datos y la herramienta de análisis de grafos desde sus regiones y se pusieron a consultar datos a escala mundial. Gracias a la fácil visualización de redes alrededor de clientes y cuentas, encontraron muchas más conexiones de las que tenían antes, lo que les llevó a nuevas historias que más tarde se convertirían en portadas en todo el mundo. Antes del uso de esta tecnología, los reporteros tenían que establecer conexiones a mano con

ESTUDIO DE CASO

"Es una herramienta innovadora y revolucionaria que ha transformado nuestro proceso de investigación periodística."

– Mar Cabra

la información encontrada en decenas de archivos: una tarea tediosa que podía generar resultados inexactos.

Con Neo4j a su disposición para investigar los datos filtrados del HSBC, los periodistas identificaron fácilmente a los principales implicados, intermediarios y beneficiarios en el escándalo (independientemente de la ubicación), y vieron cómo estaban conectados, en la milésima de segundo de un clic.

Puesto que el funcionamiento de la herramienta de visualización Linkurious era también era entendible más allá de barreras culturales y geográficas, era el método perfecto para conseguir colaboración internacional en el proyecto, y demostró ser una herramienta inestimable para la realización de informes y la comprobación de hechos a medida que los periodistas preparaban sus textos.

"Es una herramienta innovadora y revolucionaria que ha transformado nuestro proceso de investigación periodística", destacó Cabra. "Tenemos un potencial enorme ante nosotros que nos permite utilizar grafos para investigar otros escándalos financieros o corporativos. Sencillamente, todo esto antes no hubiera sido posible a esta escala; ¡es mágico!"

Ahora Neo4j se ha convertido en una parte integral del ICIJ para sus proyectos de datos masivos o Big data. Tras las revelaciones de Swissleaks, algunos periodistas externos también han llamado a las puertas del ICIJ para utilizar las herramientas de análisis de grafos para sus proyectos.

La repercusión: El desbloqueo de datos complejos para la investigación del fraude

En febrero de 2015, más de 50 organizaciones de prensa de todo el mundo (incluida Le Monde) revelaron que el HSBC había ayudado a criminales, traficantes y evasores fiscales y había sacado provecho de dichos negocios, puesto que ayudó a ocultar dinero a más de 100 000 clientes con cuentas por valor de 100 000 millones de dólares en Suiza.

El proyecto, que duró 6 meses y requirió el trabajo de 150 periodistas, recibió el prestigioso premio Data Journalism Awards (en la categoría Investigación del año), los premios de periodismo de datos organizados por la Global Editors Network. Y todavía continúan publicándose casos.

Acerca del ICIJ: Investigadores expertos del cuarto poder

El Consorcio Internacional de Periodistas de Investigación (ICIJ) es una red global de más de 190 periodistas de investigación repartidos en más de 65 países que trabajan juntos en artículos de investigación exhaustiva.

Fundado en 1997 por el periodista estadounidense Chuck Lewis, el ICIJ comenzó como un proyecto del Centro para la Integridad Pública con el fin de ampliar el estilo de periodismo de investigación del centro, centrándose en temas como la delincuencia transfronteriza, la corrupción y la responsabilidad pública del poder. Respaldado por el centro y sus especialistas en informes computarizados, expertos en registros públicos, comprobadores de datos y abogados, los reporteros y editores del ICIJ proporcionan recursos en tiempo real y las herramientas y técnicas más recientes a periodistas de todo el mundo.

Neo4j es el líder en tecnología de base de datos de grafos. Siendo la base de datos de grafos más utilizada en el mundo, ayudamos a marcas globales - que incluyen Comcast, NASA, UBS y Volvo Cars – a revelar y predecir como las personas, procesos y sistemas se interrelacionan. Usando el enfoque en las relaciones, aplicaciones desarrolladas usando Neo4j afrontan problemas de conexión de datos tales como análisis e inteligencia de datos, detección de fraude, recomendaciones en tiempo real y grafos de conocimiento. Más información en neo4j.com.

UK uk@neotechnology.com
France ventes@neotechnology.com
Nordics nordics@neotechnology.com
DACH vertrieb@neotechnology.com
Southern Europe southern-europe@neotechnology.com

© 2021 Neo4j, Inc. Todos los derechos reservados.