

Étude de cas

 ICIJ**SECTEUR**

Médias & édition

CAS D'USAGE

Détection de fraude & recherche basée sur les graphes

OBJECTIF

Révéler une évasion fiscale en découvrant la dissimulation d'actifs personnels dans des sociétés offshore

DÉFI

Recherche et analyse de liens cachés au sein de 11,5 millions de documents sources

SOLUTION

Pratiques frauduleuses mises à jour à l'aide de la base de données de graphes Neo4j et de l'outil de visualisation des données Linkurious

RÉSULTATS

- Arrestations et démissions de personnalités politiques de premier plan et d'autres personnes
- Durcissement des règles financières à l'échelle mondiale, y compris aux États-Unis, dans l'UE et au Royaume-Uni

Le Consortium international des journalistes d'investigation (ICIJ)**Neo4j au cœur d'une investigation lauréate du Prix Pulitzer sur une évasion fiscale mondiale**

Lors de son investigation sur les Panama Papers, l'ICIJ a mené des recherches pour établir des liens dans un volume impressionnant de documents confidentiels sur 40 ans. Le consortium a utilisé Neo4j pour analyser pas moins de 11.5 millions de documents et mettre en lumière une évasion fiscale impliquant des présidents, des premiers ministres et jusqu'à des trafiquants de drogue.

L'organisation

L'ICIJ est un réseau international de près de 400 journalistes issus de 80 pays et de plus d'une centaine de médias tels que la BBC et le New York Times. Installée à Washington, l'équipe de douze personnes au cœur de l'ICIJ coordonne ce réseau pour enquêter sur des cas criminels, de corruption et d'autres affaires d'envergure mondiale. L'ICIJ, qui existe depuis 21 ans, a remporté de nombreux prix de journalisme, dont le plus honorifique de tous, le prix Pulitzer, pour son dossier sur les Panama Papers.

Le défi

L'enquête des Panama Papers représente la plus importante fuite de données de tous les temps. En 2015, une divulgation anonyme de données sur 11.5 millions de documents issus du cabinet d'avocats panaméen Mossack Fonseca a dévoilé l'usage illicite de comptes en banque offshore par des célébrités et des grandes fortunes du monde entier.

Ces ressources regroupaient 40 ans de documents confidentiels relatifs à plus de 200 000 entreprises dans 21 paradis fiscaux, allant de la Suisse à Hong-Kong et au Nevada ou aux États-Unis. Ces dissimulations étaient pratiquées par des particuliers pour camoufler leur véritable richesse aux autorités fiscales derrière un réseau de sociétés écrans et de comptes enregistrés aux noms de façade ou de proches parents.

Toutefois, au départ, les journalistes de l'ICIJ ont eu du mal à passer au crible cette foule de courriels, de tableurs financiers, de passeports et de dossiers d'entreprise, rédigés en anglais, français, espagnol, russe, mandarin et arabe.

« Ça a d'abord été un choc » relate Mar Cabra, responsable des données de l'ICIJ.

Elle a vite compris que l'ICIJ avait alors besoin d'une technologie accessible pour analyser cet ensemble de données interconnectées et dévoiler un maillage complexe de relations. De par sa précédente expérience, Mar Cabra en a déduit que la technologie des graphes pouvait apporter une réponse.

La solution

L'ICIJ avait déjà déployé un système de graphes en 2013 pour présenter publiquement le résultat de ses recherches dans le cadre d'une investigation sur des fuites de données offshore.

« Ce graphe était l'outil le plus efficace que l'ICIJ avait utilisé jusque là » indique Mar Cabra. « Il était possible de saisir un nom et en un double-clic, les réseaux apparaissaient. Des millions de personnes y étaient incluses. C'est pourquoi nous savions que dans cette enquête sur les Panama Papers, nous avions besoin des graphes pour mieux comprendre les données ».

Étude de cas



« Le graphe permettait d'explorer ces réseaux de façon extrêmement aisée. Mes journalistes étaient impressionnés. Nous avons l'impression d'avoir des super pouvoirs. »

– Mar Cabra,
Responsable des données,
ICIJ

Pour prendre en main les 2,6 téraoctets de données des Panama Papers, l'ICIJ a extrait les métadonnées des documents en utilisant Apache Solr et Tika, puis a connecté toutes les informations entre elles dans la base de données de graphes Neo4j dont l'accès passait par l'outil de visualisation des données Linkurious. En parallèle, les journalistes du consortium ont utilisé la plateforme sociale open source OXWALL pour partager leurs conclusions, conseils et pistes - y compris des menaces - relatifs à l'investigation.

Les développeurs de l'ICIJ ont élaboré le graphe Neo4j autour des principales entités de données divulguées telles que les entreprises, leurs clients et leurs responsables. Ainsi, les journalistes ont pu mettre à jour les relations entre ces nœuds centraux - par exemple en faisant correspondre des comptes en banque et des personnes partageant la même adresse, des liens familiaux ou d'affaires, ou en contact régulier par e-mail.

Ce graphe comprenait 840 000 nœuds et 1,3 million de relations, mais les journalistes n'avaient qu'à saisir le nom d'une personne pour que son réseau de connexions apparaisse instantanément. Ils ont également pu explorer les données de façon approfondie grâce à des requêtes Cypher avancées.

Cette solution a permis de faire un grand bond en avant par rapport aux technologies précédemment utilisées.

« Le graphe permettait d'explorer ces réseaux de façon extrêmement aisée et compréhensible par tout un chacun » apprécie Mar Cabra. « Mes journalistes étaient impressionnés. Nous avons l'impression d'avoir des super pouvoirs et nous nous disions « Mince ! je n'avais pas vu ces relations plus tôt en consultant les documents et maintenant je découvre de nouveaux scoops ». Pour eux, c'était magique. Avec les bases de données de graphes, vous êtes fondamentalement capable de trouver des relations que vous ne pouviez pas voir en travaillant dans une base de données SQL. »

Les résultats

Depuis avril 2016, l'ICIJ a publié une série de dossiers sur les Panama Papers qui ont déclenché des centaines d'investigations et d'audits dans plus de 82 pays et donné lieu à de nouvelles règles financières aux États-Unis, dans l'UE, au Royaume-Uni et en Suisse.

En utilisant Neo4j, l'ICIJ a révélé que 500 banques avaient enregistré près de 16 000 entreprises écrans pour le compte des clients de Mossack Fonseca, parmi lesquels 40 dirigeants mondiaux actuels ou anciens, des milliardaires, des trafiquants de drogue, des célébrités, des vedettes du sport et un réseau de proches du président russe Vladimir Poutine qui ont brassé quelques 2 milliards de dollars dans le monde.

Ces révélations ont conduit à l'arrestation et à la démission de personnalités de premier plan dans le monde entier. Parmi eux, le président islandais Ólafur Grímsson s'est retiré après que le graphe ait fait le lien entre son épouse et une société offshore non déclarée détenant des millions de dollars en obligations bancaires. L'ancien Premier ministre pakistanais Nawaz Sharif a également été emprisonné pour corruption. En 2017, les fondateurs de Mossack Fonseca eux-mêmes ont été arrêtés pour blanchiment d'argent et en mars 2018, la société a fermé ses portes.

Pour l'ensemble de ces raisons, l'ICIJ a ensuite reçu le Pulitzer 2017 dans la catégorie Explanatory Reporting (journalisme explicatif).

« Il s'agit de la plus grande enquête collaborative de l'histoire du journalisme et Neo4j et Linkurious y ont joué un grand rôle » conclut Mar Cabra.

Neo4j est la plus importante plateforme de bases de données de graphes qui permet à Airbus, Comcast, eBay, la NASA, UBS, Walmart et d'autres d'innover et de rester compétitifs. Des milliers de déploiements par la communauté et plus de 300 clients mettent à profit les données connectées avec Neo4j pour 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, dont l'intelligence artificielle, la détection de fraude, les recommandations en temps réel et les données de référence. Pour en savoir plus, consulter neo4j.com.

Des questions sur Neo4j ?

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