

CASE STUDY



ICIJ

ICIJ ha utilizzato la tecnologia di database a grafo per indagare sul caso Swiss Leaks

SETTORE

Media

LA SFIDA

Aiutare i reporter a comprendere i complessi dati di Swiss Leaks per un migliore giornalismo investigativo.

LA STRATEGIA

- Impostare un nuovo strumento di analisi a grafo per consentire ai giornalisti di analizzare e scoprire in modo rapido e semplice nuovi elementi per le inchieste.
- Condividere lo strumento a livello mondiale per consentire la collaborazione sul progetto Swiss Leaks.

LA SOLUZIONE

- Neo4j è stato lo strumento perfetto per la gestione di ricerche, interrogazioni e dati complessi.
- Lo strumento di visualizzazione Linkurious ha consentito l'accesso al processo di analisi da parte di tutti i reporter ICIJ.

IL RISULTATO

- Consente ai giornalisti in tutto il mondo di cercare e scoprire in modo semplice e rapido dati complessi e non strutturati.
- Il database a grafo consente di individuare aspetti e notizie rivoluzionarie impossibili da scoprire prima.
- Uno strumento rivoluzionario per l'analisi di truffe, che trasforma i processi di giornalismo d'inchiesta.

L'efficacia dei grafi nel giornalismo d'inchiesta.

All'inizio del 2014, due giornalisti francesi della testata Le Monde, Gérard Davet e Fabrice Lhomme, ebbero accesso a un set di dati estremamente complesso, ma assai utile, che conteneva un elenco dettagliato di oltre 100.000 clienti e dei relativi conti bancari presso la filiale svizzera della banca HSBC, mettendo così in luce diverse pratiche non etiche.

Davet e Lhomme notarono il potenziale di questi dati, che avrebbe potuto dare il via a un'esclusiva indagine su frodi, evasione fiscale e criminalità internazionale. Tuttavia si poneva un problema. I dati erano troppo complessi per essere analizzati con strumenti convenzionali.

La sfida: Dipanare la complessità dei dati per giungere a una comprensione corretta

Per giungere ad elementi corretti su cui basare le proprie indagini, i reporter normalmente devono cercare di trovare le relazioni tra i dati presenti in file Excel, effettuare manualmente ricerche su Internet e talvolta agire di persona per desumere i legami tra persone ed entità. Tuttavia, nel caso di Swiss Leaks, Davet e Lhomme riconobbero che il set di dati era troppo complesso per essere analizzato manualmente o autonomamente. Perciò si rivolsero all'International Consortium of Investigative Journalists (ICIJ), che diede vita a una delle collaborazioni giornalistiche di maggiore portata mai realizzate.

Mar Cabra, direttrice del dipartimento Data and Research presso ICIJ, sapeva che per questa inchiesta e quelle future avrebbero avuto bisogno di uno strumento in grado di analizzare correttamente le relazioni tra i dati.

I dati di Swiss Leaks includevano informazioni relative a titolari di conti HSBC in oltre 200 Paesi per un ammontare pari a oltre 100 miliardi di dollari. Le informazioni tuttavia erano disseminate in migliaia di file e non presentavano connessioni dirette tra loro. La complessità dei dati richiedeva quindi uno strumento in grado di analizzare l'enorme quantità di dati non strutturati e che consentisse di chiarirli in modo semplice e rapido.

Cabra poneva inoltre un altro requisito: non voleva che venissero utilizzati come intermediari data scientist o sviluppatori. Il processo di scoperta e analisi dei dati doveva infatti essere accessibile ai reporter di tutto il mondo, indipendentemente dalle loro competenze tecniche.

La soluzione: L'analisi a grafo è la chiave per "seguire la scia del denaro"

L'inchiesta Swiss Leaks non è stata la prima esperienza dell'organizzazione con un set di dati complesso. Di conseguenza, Cabra sapeva che l'ICIJ avrebbe avuto bisogno di una soluzione con un database a grafo.

"Mentre lavoravo a inchieste come Offshore Leaks, ho compreso quanto fosse importante l'analisi a grafo durante indagini sulla corruzione finanziaria", ha affermato

CASE STUDY



“Le tecnologie Neo4j e Linkurious sono state davvero rivoluzionarie per noi.”

- Mar Cabra

Cabra. "I collegamenti sono la chiave per capire qual è la vera notizia: mostrano chi sta facendo affari con chi. Perciò per HSBC Leaks abbiamo deciso immediatamente che era necessario adottare un approccio basato su grafi".

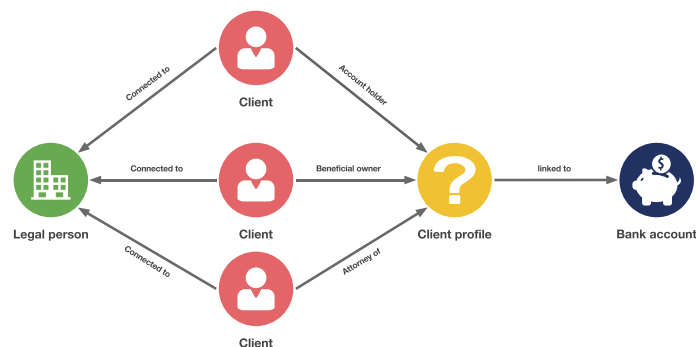
La prima mossa del dipartimento Data and Research è stata quella di ricreare il database di clienti HSBC dai semplici file Excel a disposizione. Successivamente, ogni nome è stato collegato a uno o più paesi (i "nodi" nel database a grafo). Infine i dati sono stati trasformati in un grafico per comprendere i collegamenti tra i nodi.

Il leak presentava circa 60.000 file con informazioni su oltre 100.000 clienti in 203 Paesi. Il database a grafo presentava oltre 275.000 nodi con 400.000 relazioni.

L'ICIJ per trasferire il set di dati originali nel database a grafo Neo4j di Neo Technology ha collaborato con Talend, specialista di software di integrazione open source. Un altro partner Neo, Linkurious, ha invece fornito un'app Web come interfaccia utente in modo che il database a grafo potesse essere visualizzato e accessibile facilmente dai reporter.

La visualizzazione a grafo ha consentito ai giornalisti del consorzio ICIJ di identificare i legami tra le persone e i conti bancari, permettendo loro di seguire la scia del denaro per identificare decine di esempi di frode, corruzione ed evasione fiscale.

Di seguito è riportato un esempio per illustrare il collegamento delle relazioni mediante grafo.



Schema delle entità e connessioni nel set di dati Swiss Leaks.

Il risultato: La chiave per districare una grande mole di dati complessi per inchieste su frodi fiscali

Dopo aver importato i dati in Neo4j e Linkurious, Cabra e il suo team hanno notato una netta differenza nel modo in cui i giornalisti erano in grado di utilizzare i dati Swiss Leaks.

"Le tecnologie Neo4j e Linkurious sono state davvero rivoluzionarie per noi", ha commentato Cabra. "Hanno consentito ai nostri giornalisti di analizzare complessi dati finanziari in modi molto diversi da quelli utilizzati in precedenza".

La soluzione Neo4j-Linkurious rispondeva inoltre perfettamente alla richiesta di Cabra in merito alla semplicità d'uso.

"Era talmente intuitiva che i reporter potevano iniziare a utilizzarla con un training minimo", ha dichiarato Cabra. "Abbiamo predisposto una breve demo e un webinar online. I nostri reporter erano davvero sorpresi da quanto fosse semplice effettuare un'interrogazione dei dati, anche senza competenze IT avanzate".

Successivamente alla condivisione dello strumento nello studio virtuale ICIJ da parte del team di Cabra, i giornalisti di tutto il mondo hanno effettuato l'accesso al set di dati e allo strumento di analisi a grafo dai loro rispettivi Paesi, procedendo con interrogazioni di dati su scala mondiale. La visualizzazione semplice delle reti attorno a clienti e conti ha consentito di trovare ancora più connessioni, che hanno portato a notizie esclusive pubblicate in seguito sulle prime pagine dei giornali di tutto il mondo. Prima dell'introduzione di questo processo, i reporter dovevano stabilire manualmente da soli le connessioni servendosi di

CASE STUDY

“È uno strumento di scoperta rivoluzionario che ha trasformato il processo di giornalismo d’inchiesta.”

– Mar Cabra

decine di file: un’attività dispendiosa in termini di tempo, che inoltre poteva condurre a risultati imprecisi.

L’utilizzo di Neo4j per esaminare i dati relativi ai leak HSBC ha consentito ai giornalisti di identificare in modo semplice soggetti principali, intermediari e beneficiari coinvolti nello scandalo (indipendentemente dal Paese) e comprendere i collegamenti, il tutto con pochi clic.

Lo strumento di visualizzazione Linkurious era inoltre adatto alle varie culture e Paesi in cui era utilizzato ed era l’ideale per ottenere una collaborazione internazionale sul progetto. Ciò lo ha reso uno strumento inestimabile per presentare la notizia e verificare i fatti mentre i giornalisti approntavano le loro copie.

“È uno strumento di scoperta rivoluzionario che ha trasformato il processo di giornalismo d’inchiesta”, ha affermato Cabra. “Il potenziale dei grafi per analizzare scandali aziendali o finanziari è enorme. Prima non sarebbe stato possibile su così larga scala: è davvero fantastico”.

Neo4j è ora parte integrante dei progetti con big data di ICJ. A seguito delle rivelazioni Swiss Leaks alcuni giornalisti esterni hanno chiesto a ICJ di poter usare gli strumenti di analisi a grafi.

Le conseguenze: La chiave per districare una grande mole di dati complessi per inchieste su frodi fiscali

A febbraio 2015, oltre 50 organizzazioni giornalistiche in tutto il mondo (incluso Le Monde) hanno rivelato come HSBC abbia aiutato criminali, trafficanti ed evasori fiscali e abbia tratto profitto dagli affari fatti coprendo oltre 100.000 clienti con conti da oltre 100 miliardi di dollari in Svizzera.

Il progetto, durato sei mesi e che ha visto coinvolti 150 giornalisti, ha ricevuto il prestigioso Data Journalism Award (categoria Investigation of the Year), insignito dal Global Editors Network. La pubblicazione delle notizie è ancora in corso.

Informazioni su ICJ: Gli investigatori del quarto potere

L’International Consortium of Investigative Journalists (ICJ) è una rete globale composta da oltre 190 giornalisti in oltre 65 Paesi, che collaborano su inchieste approfondite.

Fondato nel 1997 dal giornalista americano Chuck Lewis, il consorzio ICJ è stato lanciato come progetto del Center for Public Integrity, per ampliare il giornalismo di controllo del Centro, focalizzandosi su problemi come il crimine transfrontaliero, la corruzione e la responsabilità del potere. Sostenuti dal centro e da specialisti del giornalismo di precisione, esperti di archivi dati, esperti di verifica dei fatti e avvocati, i reporter ICJ forniscono risorse in tempo reale e gli strumenti e le tecniche più all’avanguardia ai giornalisti di tutto il mondo.

Informazioni su Neo Technology

Neo Technology ha creato Neo4j, il database a grafo leader del settore in grado di portare le relazioni fra i dati a un livello superiore. Aziende che offrono consigli personalizzati su prodotti e servizi, siti Web che aggiungono funzionalità social, operatori di telecomunicazioni che devono diagnosticare i problemi di rete, aziende che ridefiniscono i modelli per dati master, identità e accessi: le organizzazioni adottano i database a grafo per modellare, archiviare ed eseguire query nei dati e nelle relazioni fra di essi. Grandi aziende come Walmart, eBay, UBS, Nomura, Cisco, HP e Telenor, oltre che startup come CrunchBase, Medium, Polyvore e Zephyr Health utilizzano Neo4j per sfruttare le relazioni fra i dati.

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