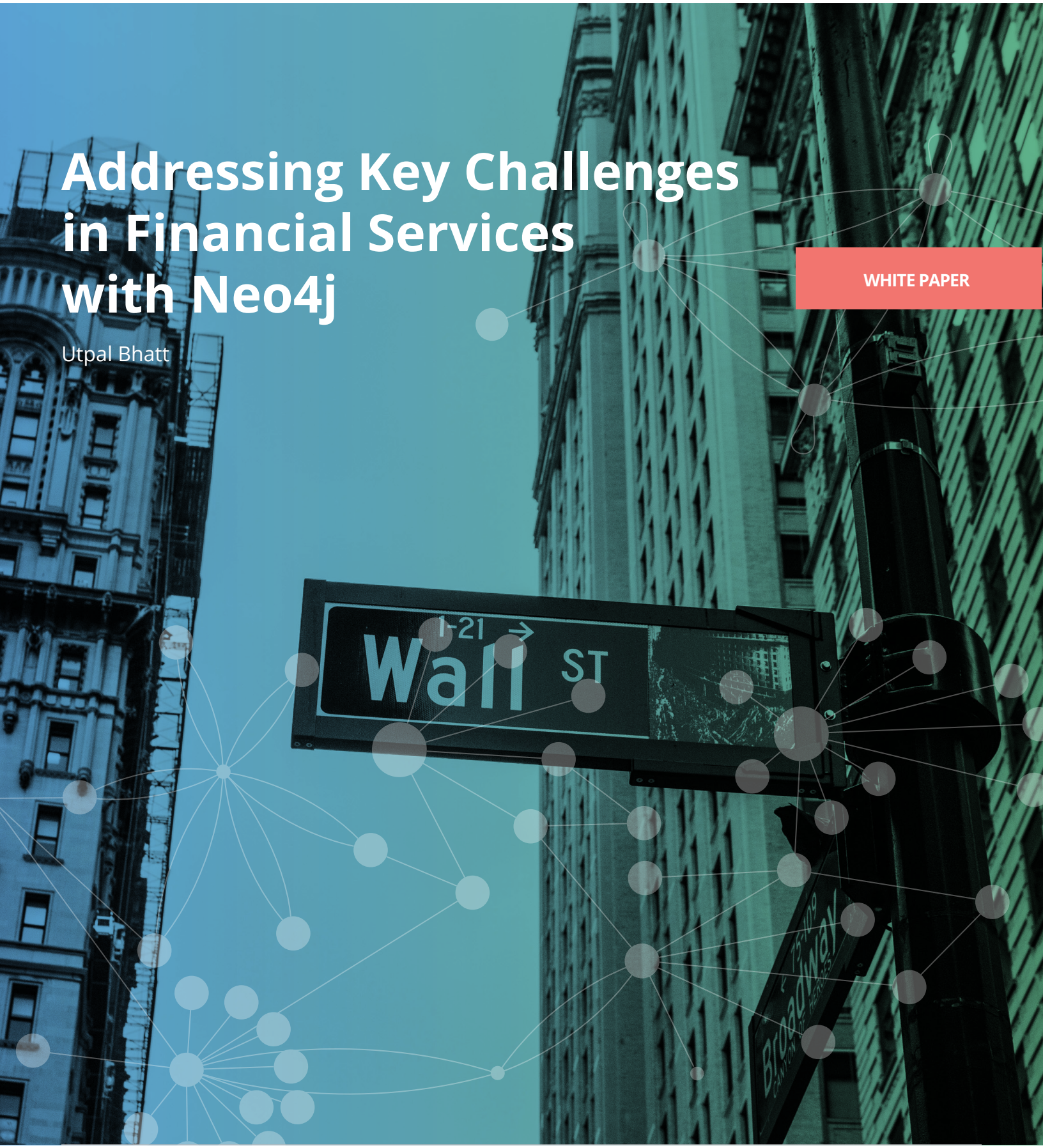
Addressing Key Challenges in Financial Services with Neo4j

WHITE PAPER

Utpal Bhatt



White Paper

TABLE OF CONTENTS

Addressing Key Challenges in Financial Services with Neo4j

UTPAL BHATT

Introduction

In today's regulatory environment, financial services firms are beginning to experience the impact of graph databases across a number of functions ranging from fighting financial crimes, preventing and responding to cyber threats and ensuring compliance.

Meanwhile, as the continuous digitization of processes requires financial services firms to evolve their customer engagement strategies to meet rising customer expectations, graph databases are helping financial services firms gain competitive advantage from digitization to drive new sales, reduce costs and build closer relationships with customers.

This white paper illustrates how financial services organizations are using graph databases, specifically Neo4j, to effectively solve these problems.

Graphs for Better Risk Management and Regulatory Compliance

Asset Graphs

The financial crises of 2008 showed that financial assets are incredibly complex and become even more so as they're bundled together and sliced into various sub-assets that are repackaged and resold. For example, a firm might offer an option, which is an instrument on top of an asset. It might also have a mutual fund that owns stocks and hedge funds, and it might have options positions on the same stock, and so on. This results in incredibly complex interdependent systems where risk both compounds and obfuscates simultaneously, creating a regulatory compliance nightmare. It can look like there's even exposure across ten assets. If you could drill down to the root asset, you might find that you're 90% exposed to one particular factor, but the exposure is just veiled by the layers above.

Financial services firms are building asset graphs with Neo4j to get a complete, clear and traceable understanding of relationships among different kinds of financial assets. Such an asset graph provides the firm with a complete understanding of risk.

As a fix, financial services firms are building asset graphs with Neo4j to get a complete, clear and traceable understanding of relationships among different kinds of financial assets. Such an asset graph provides the firm with a complete understanding of risk. In addition, firms are also using asset graphs to perform derivatives pricing in real time where the price calculation formula takes into account the many interdependencies between assets, and therefore accurately reflects the risk/reward ratio.

Regulatory compliance requires financial services firms to have visibility into data, information and process flows. This can be a challenge because the same data can be replicated across many different systems.

CASE STUDY: CERVED

Cerved is Italy's leader in credit risk analysis and one of Europe's biggest rating agencies. It offers the most complete range of products and services — employed by about 34,000 companies and financial institutions — to assess their business partners' solvency and credit-worthiness, as well as monitor and manage credit risk at all stages.

One of Cerved's applications identifies the "real owner" of a business who, in the financial analysis, owns and controls the company, either directly or indirectly holding at least 25%+1 of its share capital. Identification of real owners was introduced in Italian legislation in 2007 with Law 231 for the prevention of money laundering, and is of crucial importance for the world economy.

Cerved sought to improve the application's efficiency to generate a real-time response. Unsatisfied with a system based on a relational database, Cerved's specialists turned to Neo4j, which is capable of handling algorithms that can link accessible datasets. Development took about nine months, including production of the technology, which had been used in the past only in advanced research environments.

"The result is a robust, efficient, high-performance implementation for identifying the actual owners of businesses, jointly constructed with the participation of multiple teams. It's a solution that now allows us to promote graph database uses in other areas of the company as well. This expansion facilitated comprehension of the technology even by the people in the company most directly linked with the business, encouraging development of other uses currently underway," says Stefano Gatti, Innovation & Data Sources Manager, Cerved.

In addition, Neo4j greatly improved the application's efficiency, reducing calculation time from 12 seconds down to 67 milliseconds in cases that require tracking of up to 15 ownership links. "This allowed us to extend its use and improve the precision of the algorithm at the same time," Gatti says. In turn, customer confidence in Cerved's data has increased.

Information Management with a Metadata Graph

Regulatory compliance requires financial services firms to have visibility into data, information and process flows. The Sarbanes-Oxley Act (SOX), for example, requires public firms to understand who has access to what data, what data resides in which systems, and how data flows across the organization. This can be a challenge because the same data can be replicated across many different systems. For example, security master data, which might be sourced from Bloomberg, might spread to 150 systems, while the master set of the firm's products might be replicated across 200 systems. Knowing your data lineage and how data moves is not only necessary for regulatory compliance efforts, but it also helps speed up projects that depend on that data.

Financial services are using Neo4j to model their data lineage and data flows as a graph to get a complete understanding of data and systems across the organization.

It has become difficult for financial services organizations to identify and stop fraudulent activity. Discrete methods fall short when it comes to detecting fraud rings.

CASE STUDY: GLOBAL 500 FINANCIAL SERVICES COMPANY

A Global 500 financial services firm, residing near the median of that list, wanted to build an integrated data distribution platform. A key component of the platform is a knowledge base that would allow the firm to describe the lineage of the datasets and attributes from master systems to consumers. The firm chose to store the meta-model for the knowledge base in Neo4j.

The Global 500 financial services firm encountered several challenges while building its knowledge base, one of them being coverage. Not all datasets had been entered into the platform, so flexibility was essential in order to accommodate new data sources, datasets, consumers and rules. What's more, the knowledge base needed to be able to easily answer numerous questions such as:

- What datasets and attributes do we provide?
- How are the datasets related?
- Which consumers are using which attributes?
- How are users receiving our data?

Neo4j's flexible schema enabled the global firm to model all its data flows and rapidly answer questions about how and where its data is used. Given the success realized with Neo4j, the firm plans on widening its coverage of datasets and offering the solution to other parts of the bank.

Preventing and Detecting Financial Fraud

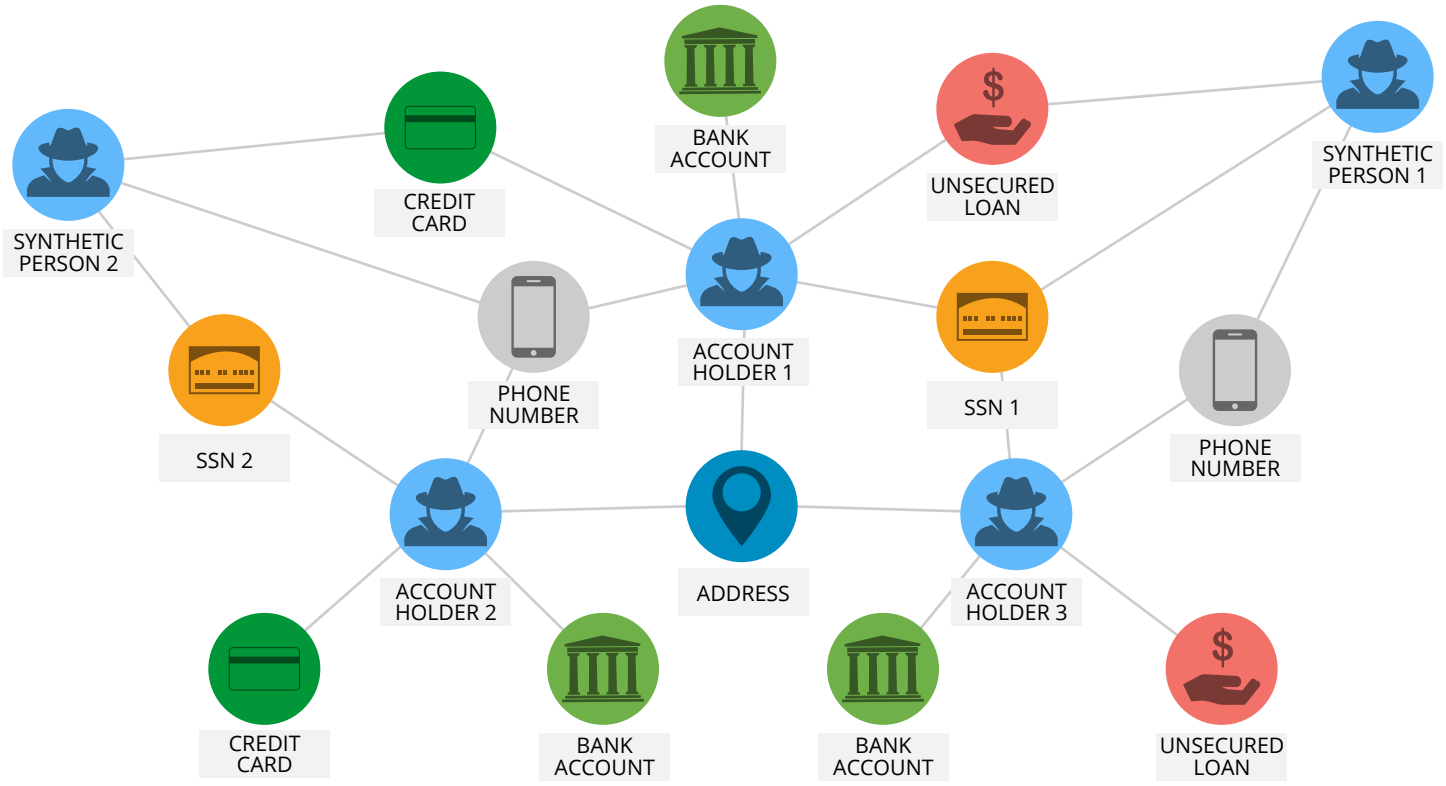
It has become difficult for financial services organizations to identify and stop fraudulent activity. Standard anti-fraud technologies — such as a deviation from normal purchasing patterns — use discrete data. This is useful for catching individual criminals acting alone, but discrete methods fall short when it comes to detecting fraud rings. Furthermore, many discrete methods are prone to false positives, which can impact customer satisfaction and result in lost revenue.

Sophisticated criminals continuously alter their strategies to circumvent detection by traditional solutions. For example, they use synthetic accounts to carry out what appear to be unrelated activities, when in fact they are well coordinated. A criminal might steal identity information from ten different people. The criminal then mixes and matches the social security numbers, addresses, phone numbers and email addresses to create "new" synthetic identities that are then used to open bank and credit card accounts, as well as personal lines of credit. Traditional fraud detection solutions won't flag these accounts. Firms need a way to follow the trail from one account to another to determine how activities that on the surface appear unrelated are in fact connected.

An increasing number of the world's leading financial institutions in the payments space are using Neo4j to model and monitor data about customers, accounts, devices, locations and other attributes to identify fraudulent activity from fraud rings using synthetic and stolen identities.

Uncovering rings with traditional relational database technologies requires modeling a graph as a set of tables and columns, and then carrying out a series of complex JOINs and self-JOINs. Such queries are very complex to build and expensive to run. Scaling them in a way that supports real-time access poses significant technical challenges, with performance deteriorating exponentially not only as the size of the ring increases, but also as the total dataset grows.

In contrast, graph databases are an ideal tool for overcoming these hurdles. Intuitive query languages like Cypher provide a simple semantic for detecting rings in the graph, navigating connections in memory and in real time in order to catch the activity as it happens. The graph data model below represents how the data actually looks to the graph database, and illustrates how one can find rings by simply walking the graph.



Modeling a fraud ring as a graph

Augmenting one's existing fraud detection infrastructure to support fraud ring detection can be done by running appropriate entity link analysis queries using a graph database, and running checks during key stages in the customer and account lifecycle, such as:

- At the time of account creation
- During an investigation
- When a credit balance threshold is hit
- When an overdraft occurs

Real-time graph traversals tied to the right kinds of events can help financial services firms identify probable fraud rings, during or even before a fraudulent transaction occurs.

Many traditional technologies aren't designed to connect the dots, which means that detecting money-laundering schemes requires a tremendous amount of manual effort. Teams of inspectors can spend months going through reams of data.

CASE STUDY: FORTUNE 500 FINANCIAL SERVICES COMPANY

A Fortune 500 financial services company conducts 2.2 million financial transactions per month. With the real-time graph analysis and simple data visualization provided by Neo4j, the company's analysts stop millions in fraudulent transactions annually.

The majority of transactions processed by the financial services company are approved or denied through an automated fraud detection system. However, a portion of transactions is submitted to an analyst for manual review. Prior to Neo4j, analysts queried a Microsoft SQL Server database to evaluate the customer information as well as enriched data from outside vendors and data on related customers, such as potential members of a fraud ring. Depending on the transaction request, queries could require four or more levels of traversals to extract and analyze the information. Queries were slow and returned complicated data that analysts had to review — all while the clock was ticking for customers waiting to transfer money.

The company needed a more efficient way to analyze data. This would require decreasing the amount of time needed to process fraud-detection queries against the database and delivering simple data visualizations to analysts. In order to achieve real-time results, the development team would need a database optimized for storing and traversing several levels of relationships.

Neo4j brought connected graph data into the company's architecture in real time and provided data visualizations that allow analysts to make fast, accurate decisions. The graph revealed emerging clusters and relationships that had previously gone unnoticed which improved the accuracy of real-time fraud detection. Each customer can be represented by up to 30 nodes, each with up to 60 properties totaling over 216 million nodes and 680 million relationships in the total dataset.

With Neo4j's simple data visualizations, it now takes an analyst about half the amount of time to manually review a requested financial transaction. This allows them to review nearly twice the number of transactions daily — stopping fraudulent transactions sooner and reducing wait times for non-fraudulent customers.

Anti-Money Laundering

Reducing the risk of money laundering presents a similar challenge. Firms need to know where funds come from and where they are headed, but criminals use indirection to make it difficult to follow money from one point to another. Instead of moving it directly, from point A to point B, they'll move it to a dozen other points in between with the hope that nobody notices. Unfortunately, many traditional technologies aren't designed to connect the dots across the many intermediate steps, which means that detecting money-laundering schemes requires a tremendous amount of manual effort. Teams of inspectors can spend months going through reams of data.

Anti-money laundering (AML) teams at financial services firms are using Neo4j to model companies, accounts and transactions as a graph to discover instances of money laundering. By graphing the relationships between all of these entities, AML teams can track how and where funds are moving through automated Cypher queries that map to traditional money laundering behaviors. Once a suspicious transfer of funds occurs, the system can automatically flag the transaction for review by an AML analyst.

Whether it's optimizing a network or an application infrastructure, managing change or providing more effective security-related access, more often than not these problems involve a complex set of physical and human interdependencies that can be quite challenging to manage.

CASE STUDY: GLOBAL MONEY TRANSFER COMPANY

One of the world's largest money transfer companies moves approximately $600 billion each year. The company not only complies with stringent anti-money laundering (AML) requirements in every country where it operates, but it sets the benchmark for investigatory compliance across its industry. The secret: Neo4j.

"Smurfing" — a process by which illicit funds move into the financial system — is a key threat faced by banks and money transfer companies. Smurfing involves splitting illicit funds into smaller amounts in order to bypass the regulatory limits on the amount of money a customer can send or receive, thereby avoiding suspicion by sending smaller sums to a network of beneficiaries.

A typical AML investigation involves checking about 10,000 transactions — a task that was impossible to do efficiently using the static visualization presented by the company's SQL-based system. The global money transfer company surveyed the software market and found that Neo4j enables investigators to visualize the relationships between transfer accounts as well as dig deeper by querying the data dynamically and in real time. This provides greater visibility and enables the company to make connections more rapidly.

Given the success of Neo4j in the company's European Financial Crimes Department, the graph database was rolled out to investigatory teams in the US, Canada, India, Malaysia and Australia. It is fully integrated with the company's daily operations and its existing internally developed investigative software.

"At the beginning, compliance was perceived as a constraint," said the company's Compliance Manager for Europe,. "Now — and this is really the case — we consider it a real competitive edge. It is an advantage to have the best practice inside your company and to be one step ahead of the rest of the market. If we deploy Neo4j and impress all the regulators across the world with this solution, then we define the benchmark."

Graphs for Reliable, Secure IT and Improved Customer Experience

Network and IT Infrastructure Monitoring

Discovering, capturing and making sense of complex interdependencies is central to running IT organizations more effectively, and it is also a critical part of running the businesses IT serves. Whether it's optimizing a network or an application infrastructure, managing change or providing more effective security-related access, more often than not these problems involve a complex set of physical and human interdependencies that can be quite challenging to manage.

Moreover, once any two or more of these areas are brought together, the relationships are rarely linear or purely hierarchical. They form, in the computer science sense of the term, a graph. These domains are rarely static. In fact, they tend to change with reasonable frequency as a result of factors such as reorganization and personnel changes, mergers and acquisitions, new applications being developed and old ones retired, and ongoing data center improvements.

Because of the evolving complexity and interrelated nature of today's IT infrastructure, only a graph database can properly map and query all of these connected data points.

Financial services firms are augmenting their IT and network infrastructure monitoring capabilities by using Neo4j to graph their data centers, networks and other IT architecture. Broad IT use cases include everything from dependency management, impact analysis, network planning, downtime reduction, root-cause analysis and routing and quality-of-service mapping. Because of the evolving complexity and interrelated nature of today's IT infrastructure, only a graph database can properly map and query all of these connected data points.

RBS
The Royal Bank of Scotland

CASE STUDY: ROYAL BANK OF SCOTLAND

As a large enterprise, Royal Bank of Scotland has a number of legacy integrations and dependencies to contend with each time an application is deployed. In an attempt to improve developer productivity, the financial services firm sought to automate deployment.

RBS had reached a point where their legacy deployment process was causing butterfly effects. IT would deploy a library to user acceptance testing (UAT), development, production and then make a few minor changes that triggered significant and unintended ramifications. To solve these problems, RBS created hooks into a SaaS-based continuous integration platform, but it soon became evident that any pre-existing application would require a complete rewrite.

RBS needed to integrate with existing GoldenSource services that included every core operation inside the bank such as teams, users and hosts. RBS also wanted to expose the information and interact with its release system through a REST API and avoid being restricted in future user experience development, including mobile. RBS needed full flexibility and control, which could only be had by building its own release tool.

To solve these problems, RBS created Dart ("Damn, Another Release Tool") based on Neo4j. Unlike a relational database, Neo4j gives RBS the flexibility to change its data model easily and on a continual basis. Most importantly, Neo4j makes integrations easy. RBS now aspires to have a true microservices architecture and deploy Dart as another microservice.

"It's very easy to embed Neo4j in your service and spin up a cluster without any external housekeeping service required. We knew this would be incredibly important when deploying our conceptual services to the cloud because they would deploy as one package," says Stelios Gerogiannakis, senior engineer, Royal Bank of Scotland.

Dart has been a huge success for RBS. Over the course of seven months, usage grew by 50% each month. Users find it easier to automate their deployments and push software into development and UAT, which translates to increased productivity. As a result, Dart will be rolled out to roughly 10,000 developers across all of the organizations in the RBS group.

Over time, the centralized entitlements structure, whether represented in Active Directory or an LDAP directory, grows in such a way that people end up having more permissions than they need. And because of the complexity of the infrastructure, administrators can't easily determine who has what permissions.

Identity and Access Management

Within the IT organization, identity and entitlements management must also be managed to minimize risk. Over time, the centralized entitlements structure, whether represented in Active Directory or an LDAP directory, grows in such a way that people end up having more permissions than they need. And because of the complexity of the infrastructure, administrators can't easily determine who has what permissions. Users in one group may be granted permissions to assets they don't necessarily need simply because that group belongs to another group that does need those permissions. This "permissions creep" results in an increased risk of data breaches by both the insiders who are authorized to access it and by malicious outsiders who target privileged credentials to gain unauthorized access.

Corporate information security groups at financial services firms are using Neo4j to build entitlement graphs to audit the entitlement structure and issue queries in real-time to determine entitlements regardless of the complexity. With Neo4j, financial services firms are able to seamlessly track all identity and access authorizations and inheritances with substantial depth and real-time results.

The interconnected view of graph data produces better real-time insights and controls into entitlement access than custom-built directories. As a result, identity and access management happens quickly and effectively without complex, hours-long queries into the database.

CASE STUDY: GLOBAL 500 FINANCIAL SERVICES COMPANY

When building a cross-asset investment banking platform, one Global 500 financial services company ran into significant challenges with user entitlement and authorization access. Both externally and internally, various levels of users required access to the platform for activities as diverse as portfolio management, equities trading and foreign currency exchange. For such a high-profile platform, the firm required a robust permissions management solution.

This Global 500 financial services firm uses Neo4j to manage user identity and access management for its investment banking platform for both firm employees and their clients. Using Neo4j, the firm can easily track which users have access to which assets, activities, decision-making tools and more, and queries can be run in real time as users are added, removed or audited for fraud prevention.

The flexibility of the graph data model allows the firm to add, modify or delete new roles, assets and access rights without disrupting the day-to-day operations of the database or the investment platform. The current database has over one million nodes and six million relationships, and the overall database size is growing by 20% per year with no performance challenges.

Cybersecurity

Cybersecurity is of critical importance to financial services firms. IT organizations must mitigate cybersecurity risk, both in terms of external access to the infrastructure and in understanding how a malware infection or attack can move across a particular network. These efforts are made challenging by the complexity of the data center. A typical bank will have tens of thousands of Linux, Windows and other servers as well as hundreds of thousands of desktops. Everything is connected to everything else across many physical and virtual networks and subnets.

Using a graph data model, cybersecurity experts no longer have to think of cyberattacks in terms of discrete tables, lists and logs, but instead can track attacks (or potential attacks) more intuitively across their entire IT infrastructure.

In order to implement the proper controls and shore up exposed assets, IT must have visibility across the entire infrastructure. IT must know which customers and which employees are accessing what pieces of the infrastructure and what resources are made available to them. IT must also track which systems are connected.

Using a graph data model, cybersecurity experts no longer have to think of cyberattacks in terms of discrete tables, lists and logs, but instead can track attacks (or potential attacks) more intuitively across their entire IT infrastructure. Neo4j can be used to identify single points of failure within a network before the vulnerability is known to attackers, and real-time, connected data queries can quickly alert cybersecurity experts if typical attack patterns are occurring within a network, even if individual data points don't appear to be malicious. By strengthening cybersecurity across a financial services firm, Neo4j keeps business running as usual while protecting mission-critical systems from would-be criminals.

MITRE

CASE STUDY: MITRE CORPORATION

In their efforts to stop cyberattacks, analysts track large amounts of detailed information, such as network and endpoint vulnerabilities, firewall configurations and intrusion detection events. The solutions they use to analyze this data typically track data points. But to be successful, analysts need to understand how data points are related. To address these challenges, researchers at the MITRE Corporation, a US federally funded, not-for-profit company, are developing a tool for cyberwarfare analytics, visualization and knowledge management.

CyGraph brings together isolated data and events into an ongoing overall picture for decision support and situational awareness. It prioritizes exposed vulnerabilities, mapped to potential threats, in the context of mission-critical assets. It also correlates intrusion alerts to known vulnerability paths and suggests the best course of action for responding to attacks. For post-attack forensics, CyGraph shows vulnerable paths that warrant deeper inspection.

Rather than being fixed, the model schema in the CyGraph architecture is free to evolve with the available data sources and desired analytics. The data model is based on a flexible property-graph formulation implemented in Neo4j. REST web services provide interfaces in CyGraph for data ingestion, analytics and graph visualization. Data in the wild is mapped to the common CyGraph data model in a two-step process. Data is normalized and stored in a document-oriented database and then represented as a graph in Neo4j.

The Neo4j native graph pattern-matching language supports a library of domain-specific queries as well as flexible ad hoc queries. CyGraph then provides a variety of clients for specialized analytic and visual capabilities including graph dynamics, layering, grouping, filtering and hierarchical views.

The use of Neo4j at the MITRE Corporation provides insight into the mission impact of cyber activities. Graph layers (network infrastructure, cyber defensive posture and threats, mission dependencies and so on) define subsets of the overall model space with relationships within and across each layer. Analysts can also gain visibility into operations for global situational awareness.

Data tends to be locked away in silos across the organization without any way to leverage the connections between data and innovate based on those connections.

Improving Customer Experience with a 360-Degree View

Customer expectations are rising at a time when customer service is a significant differentiator. Customers expect companies to deliver personalized service that reflects an understanding of who they are, their communication preferences, the products and services they've purchased in the past and what they might be interested in in the future.

Even as financial services firms work to meet these rising customer service expectations, they must also determine how to use and integrate online channels to drive more revenue.

All of these efforts require a 360-degree view of customers, and that's something most financial services firms simply don't have. Data tends to be locked away in silos across the organization without any way to leverage the connections between data and innovate based on those connections. Harnessing the power of connected data (i.e., data relationships) is essential to sustainable competitive advantage in today's ever-more-connected, ever-more-competitive world.

When a financial services firm uses Neo4j for Master Data Management (MDM) and a 360-degree view of its data, new applications tap into the power of a graph database while co-existing with all of the firm's current data infrastructure for legacy applications. In addition, firms gain a deeply connected view of their data across both old and new applications, enabling them to become more competitive without business disruption.

CASE STUDY: FORTUNE 500 CREDIT CARD PROVIDER

This Fortune 500 technology company in the global credit card payments industry had a number of data silos across its organization, including data around merchant accounts, end customer accounts, transaction logs, authorizations, procurement data, merchant locations and many others. Before using Neo4j, the firm lacked a 360-degree view of partners, merchants and customers.

By integrating all of its data silos into a single Master Data Management (MDM) solution powered by Neo4j, the Fortune 500 credit card provider can now gain a fuller and more complete view of how its merchant, partner and customer ecosystems all interact with one another.

The integrated MDM database contains over 1 million merchant nodes, 100 million end user accounts, billions of data relationships and approximately 1 terabyte of storage.

Neo4j can be used either as a central database platform or alongside an existing data infrastructure in order to add graph capabilities to new or existing applications.

Neo4j – Benefits and Key Features

Despite the variety of challenges faced by the financial services industry and the fact that they touch different parts of the organization, these challenges share an underlying commonality: the need to understand connected data relationships. A graph database such as Neo4j does exactly that.

Neo4j is the essential database for connected data, thanks to its emphasis on ***native graph storage*** and ***real-time query processing***. Its real-time performance when traversing data, ACID (Atomic, Consistent, Isolated, Durable)-compliant transactions for graph data, flexibility of data modelling and wide array of developer productivity tools allow financial services organizations to deliver powerful new insights from data connections with greater agility and at lower cost than has previously been possible.

Specific benefits of Neo4j include:

- **Minutes-to-milliseconds query performance of traversals.** Neo4j can traverse any level of data in real-time due to its native graph architecture. RDBMS and other NoSQL databases typically see a significant performance degradation when traversing data beyond three levels of depth.
- **Data integrity.** Neo4j is a fully ACID transactional database for storing critical data; this means that transactions either fully succeed or fail gracefully without corrupting the data within. Neo4j's fully ACID transactions and Causal Clustering architecture ensures data integrity of the graph at all times, making it the ideal database for all types of mission-critical applications.
- **Flexibility.** With Neo4j, development teams don't need to struggle to fit data into predefined tables, and when business requirements change, the data model can flexibly adapt in minutes – not months.
- **Developer productivity.** Neo4j delivers unparalleled developer productivity tools to build graph applications. With tools including Cypher (the industry's most intuitive and expressive graph query language), official language drivers, stored procedures and native Java APIs, developers can build and maintain graph applications with greater efficiency.

Specific technological innovations that make Neo4j powerful include:

- Enterprise-grade clustering
- Index-free adjacency enabling real-time queries evaluating millions of relationships
- Advanced security architecture

Neo4j can be used either as a central database platform or alongside an existing data infrastructure in order to add graph capabilities to new or existing applications.

Using a graph database such as Neo4j in conjunction with or in place of a relational database management system or NoSQL database can help financial services institutions create a sustainable competitive advantage.

The Neo4j Ecosystem: Accelerating Speed to Market

When companies choose Neo4j, they become part of a passionate, thriving community of hundreds of thousands of users and developers who inspire and support each other, sharing skills and accelerating delivery of graph technology. Neo4j has over 200 customers that include Walmart, UBS, Cisco, HP, adidas and Lufthansa, in addition to a whole range of exciting startups.

Neo4j, the world's leading graph database, enables financial services firms to address the broad range of challenges described in this white paper. Existing systems represent data in tables and columns, making it hard to trace connections across datasets. Queries that take hundreds of lines of SQL can be reduced to less than 10 lines of Neo4j's intuitive Cypher query language. Better still, answers to those queries are returned in seconds, not hours.

Using a graph database such as Neo4j in conjunction with or in place of a relational database management system or NoSQL database can help financial services institutions create a sustainable competitive advantage.

CASE STUDY: THE ICIJ

Reporters traditionally draw out connections between people and entities to get the facts for their stories. But when it came to the Swiss Leaks data, a new approach was in order. The data was simply too complex to analyze manually, driving two French journalists to consult the International Consortium of Investigative Journalists (ICIJ).

The Swiss Leaks data included information from account holders located in more than 200 countries collectively holding sums in excess of $100 billion. But the information was scattered in 60,000 files with no straightforward connection between them. Reporters had to establish connections by hand with the information of dozens of files — a time-consuming task that could yield inaccurate results. The ICIJ needed a more efficient way to analyze the vast amounts of unstructured data. What's more, the data discovery and analysis process needed to be accessible to reporters worldwide — regardless of their technical expertise. ICIJ knew it needed a graph database.

"While working on stories like Offshore Leaks, I learned how important graph analysis is when investigating financial corruption," said the ICIJ's editor of the Data and Research Unit, Mar Cabra. "Connections are key to understanding what the real story is: They show you who's doing business with whom. We decided early on that we needed to use a graph-based approach."

ICIJ used Neo4j to build a graph database with more than 275,000 nodes with 400,000 relationships among them. Using Neo4j, journalists easily identified major players, intermediaries and beneficiaries in the scandal, and saw how they were connected within milliseconds. By being able to easily visualize the networks around clients and accounts, they found many more connections than they had before, which lead to stories that became front page news all around the globe.

The six-month, 150-journalist project was awarded the prestigious Data Journalism Award (Investigation of the Year category) by the Global Editors Network. In addition, the ICIJ used Neo4j to power the even-bigger Panama Papers investigation (currently the largest known data leak in history at 2.6TB). For their work analyzing over 11.5-million documents using Neo4j, the ICIJ received the 2017 Pulitzer Prize for explanatory journalism.

Financial services firms can no longer conduct business as usual, with data locked away in heterogeneous silos throughout the organization. Firms need an easy way to tie together data from these disparate systems without disrupting their day-to-day business.

Conclusion

Financial services firms can no longer conduct business as usual, with data locked away in heterogeneous silos throughout the organization. Firms need an easy way to tie together data from these disparate systems without disrupting their day-to-day business. They need a way to gain a connected view and the ability to derive value from data relationships.

Neo4j meets these requirements, thus enabling financial services organizations to solve key challenges and uncover new business opportunities. In addition to real-time processing of data relationships at scale, Neo4j is ACID-compliant and therefore suitable for master data, transactions and data reliability. It scales with blazing fast queries and delivers consistent query response times — even as datasets grow over time. Also, Neo4j's powerful and expressive query language, Cypher, takes less time to learn and requires 10 to 100 times less code than SQL.

With Neo4j, financial services firms detect fraud rings more efficiently, model complex assets in a changing marketplace, connect disparate systems and data sources and oversee identity and access management — all while increasing customer engagement and driving new forms of revenue.

To learn more about adding Neo4j to your environment, contact a Neo4j expert or download a 30-day free trial of Neo4j Enterprise Edition.

Neo4j is an internet-scale, native graph database that leverages connected data to help companies build intelligent applications that meet today's evolving challenges including machine learning and artificial intelligence, fraud detection, real-time recommendations and master data. As the #1 platform for connected data, Neo4j has over three million downloads, the world's largest graph developer community, and over thousands of graph-powered applications in production.

The world's most sophisticated organizations worldwide, from enterprises like Walmart, eBay, UBS, Cisco, HP, adidas and Lufthansa to hot startups like Medium, Musimap and Glowbl, use Neo4j to harness the connections in their data.

Questions about Neo4j?

Contact us:
1-855-636-4532
info@neo4j.com