Rethink Your Master Data

How Connections Will Define the Future of MDM

Nav Mathur, Senior Director of Global Solutions, Neo4j

White Paper

Rethink Your Master Data

How Connections Will Define the Future of MDM

Nav Mathur, Senior Director of Global Solutions, Neo4j

Introduction

Data is both our most valuable asset and our biggest ongoing challenge. As data grows in volume, variety and complexity, across applications, clouds and siloed systems, traditional ways of working with data no longer work.

Increasingly, businesses are recognizing a need to harness all of their data, particularly their data around customers, products, partners and more – often called master data. Pressing business priorities such as compliance and digital transformation require a holistic view of this **master data.**

Achieving that holistic view requires connecting data across a myriad of sources and silos. Connecting data using flexible graph technology offers a proven approach to solving these data challenges, capturing not only data but an unlimited number of connections and relationships between data.

This paper describes the power of connecting your most important data about customers, products, employees, business partners and more using graph technology. Along the way, real-world use cases from global enterprises to disruptive startups illustrate the power of connected data.

A *Big* Data Problem

Organizations have lots of data, but it's siloed and disconnected. Data is spread across:

- Different departments (CRM, Sales, Marketing, Product Development, Finance, HR)
- Different divisions (product lines)
- Different platforms (web team, data warehouses, NoSQL systems, data lakes)
- Different locations (cloud, on-premise, edge, IoT, mobile)
- Different formats (XLS, database schema, unstructured, object storage)
- Different systems (CRM, ServiceNow, Salesforce, Slack, Office 365)

Disconnected data creates problems.

Rethink Your Master Data

Without a holistic view of your data, fragmentation, misunderstandings, inaccuracies and mistakes abound. Worse, disconnected data creates friction that makes compliance more difficult, customer 360 impossible and new business opportunities hard to see, let alone execute.

Modern Master Data Management

Master data is the authoritative record of everything vital to your organization's operations including information on users, customers, products, accounts, partners, locations, business units and more. Typically, this data is stored in many different places, with lots of redundancy, variable formats, uneven quality and inconsistent access. Master data management – at its essence – involves connecting and organizing all of your most important data.

If data is the lifeblood of your enterprise, then MDM is hematology – the discipline of the entire system. Simply put, MDM is a set of methods, systems and technologies that ensure the quality, accuracy, completeness, timeliness and consistency of all reference data in the organization. It encompasses virtually every element of the enterprise including databases, applications, business processes, organizational units and geographies. MDM provides the authoritative foundation for all information across the enterprise and a single source of truth, with the aim of building a "golden record" that has the approved version of the latest and most important data about customers, suppliers, products and the like.

In the past, MDM systems required a centralized approach. Such systems were implemented as major corporate initiatives, complex and expensive long-term projects that required executive buy-in and alignment across numerous stakeholders. Further, such systems used a rigid schema that made changes and additions time-consuming.

Modern MDM requires the capability to work across silos, absorb new technologies and sources of information, find hidden relationships, quickly generate insights and deliver results in real-time at scale. It offers agility to answer any questions that arise, not just those anticipated in advance.

Why Graph Technology?

Graph database technology offers a proven way to connect master data. It enables you to start right where you are with a use case that solves pressing problems and creates immediate business value. It gives you the flexibility to connect data across existing MDM systems, or use the graph data store itself as your MDM system.

There is good reason that graphs lie at the core of the most disruptive companies of our modern era, including Google, Facebook, LinkedIn and Amazon. These companies continue to demonstrate the competitive advantage of understanding networks and mastering connected data.

The Limits of Relational Databases

Most MDM systems rely on relational databases with grid-like structures that are not optimized for traversing relationships. Despite the name, relational databases are not designed to capture relationships between data points. Even with all the recent advances in computer processing and high-speed networks, the performance of relational database applications continues to lag when it comes to ad hoc, multi-hop queries.

The root cause usually boils down to one factor: queries about **data relationships**.

If data is the lifeblood of your enterprise, then MDM is hematology – the discipline of the entire system.

Rethink Your Master Data

Relational databases were not built to handle connected information, so queries about data relationships require numerous JOIN tables. These operations are costly in terms of computing and memory – and the burden rises exponentially with the size and complexity of queries. Lengthy SQL statements are required to accomplish simple operations. Performance degrades sharply with the number and levels of data relationships (hops) and the size of the database.

While relational databases continue to serve many purposes, **they do not serve connected data use cases effectively**. Because JOINs are expensive, they can't analyze relationships beyond three hops. These multi-hop queries are time-consuming and may even hang, never returning an answer.

The Power of Graph Technology

Graph databases connect all types of data stores – both flexibly and at scale – providing a sweet spot that complements existing databases. Graphs enable next-generation approaches that connect master data wherever it is by building a metadata fabric that weaves connections in the underlying data.

Graph queries are fast, nimble and able to identify and exploit the natural connections hidden in data – and this advantage increases with scale and complexity. With graph databases, queries are much faster – ten times faster is normal but in some cases performance is a thousand or even a million times faster than a relational database.

The advantages of graph technology include:

- Support for any query

- Lightning fast, no matter how many connections (hops)

- Simple query language

- Complements existing systems; no need to rip and replace

- AI/ML on connected data using graph algorithms

- Visualization and communication (whiteboard style structure)

Introducing Neo4j

Neo4j is the leading graph database platform. Hundreds of organizations have turned to Neo4j from industries such as financial services, government, energy, software, retail, media, manufacturing and more.

Neo4j stores and queries data as nodes (entities) and relationships (connections). Nodes linked by relationships form a network. Think of nodes as nouns and relationships as verbs. Properties can be attached to both nodes and relationships, akin to adjectives and adverbs, respectively.

Relational databases force data into a pre-defined model; in contrast, graphs capture the natural structure of a given dataset. Information is stored according to how it is retrieved – thus revealing how individual entities are naturally connected.
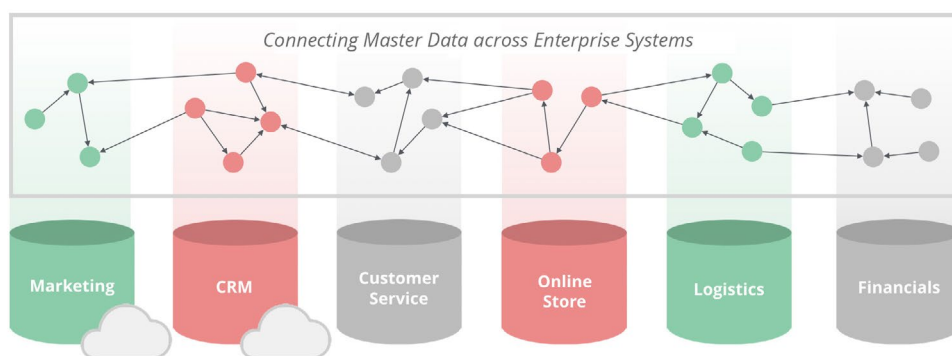
The relationships between data are as important as the data points themselves. By contrast, relational databases compute relationships at query time through expensive JOIN operations. Graph databases excel at managing highly connected data and complex queries. Neo4j uses the Cypher query language (similar to SQL but designed for graphs). With a native graph database, you can traverse millions of connections per second.

Rethink Your Master Data

How Graphs and MDM Intersect

Relational databases are not going away, so there's no need to rip and replace. A more practical approach is to add a solution that complements your existing system and allows you to continue to reap the benefits of past investments. Neo4j provides the missing connections and insights.

Many organizations maintain multiple data sources – CRM systems, work management systems, accounts payable and receivable and so on. In most cases, it simply is not feasible to move all data into a single location. A Neo4j-powered MDM solution can be layered atop legacy systems, work across silos, provide a consistent source of information and reveal the relationships hidden in your master data.

Connecting Master Data across Enterprise Systems

Marketing   CRM   Customer Service   Online Store   Logistics   Financials

Just ask the global booking platform Airbnb.

Internal MDM for Increased Productivity: Airbnb

Airbnb had more than 3,500 employees across the globe and faced the huge challenge of managing the volume and complexity of its data. The company managed more than 200,000 tables in its data warehouse, all spread across multiple clusters.

"It wasn't evident how you even found the right table," recalled Airbnb software engineer John Bodley. In surveys, employees gave the company poor reviews when asked whether they had the right information to do their jobs.

Using Neo4j, the company created an internal MDM tool called the Dataportal to connect all these silos, enabling employees to find the data they need with ease. Neo4j served as the perfect fit for the company's operations. As Bodley explained, "Our company is a graph."

> *Ask yourself:* Is your company – and your company's data – a graph?

A Neo4j-powered MDM solution can be layered atop legacy systems, work across silos, provide a consistent source of information and reveal the relationships hidden in your master data.

Rethink Your Master Data

Connecting Silos to Fight Diabetes: DZD

The German Centre for Diabetes Research (DZD) sought a way to bring together all the information spread across the organization and its various research activities. The DZD wanted a centralized data and knowledge management system for technical reasons and human ones too – especially to promote cross-disciplinary collaboration.

DZD's research network accumulates a huge amount of data distributed across various locations and consolidates it into a single, master database. This central database provides DZD's 400-strong team of scientists with a holistic view of available information, enabling them to gain valuable insights into the causes and progression of diabetes.

With Neo4j, DZD runs queries across many locations – and already has discovered intriguing connections and patterns for future research.

"Creating the first data models with Neo4j was very fast," said Dr. Alexander Jarasch, head of bioinformatics and data management at DZD.

"In the first week, I was able to connect metadata from our scientists into a data model, test the model and show the added value of the graph database," said Jarasch. "Thanks to the high scalability and performance of Neo4j, the data integration possibilities are limitless. We're employing AI and graph analytics to find connections with other diseases, including cancers."

Ask yourself: **What could you do in a week by connecting your data silos?**

Product 360: Lockheed-Martin

You don't have to be a rocket scientist to see the strategic value of graph databases. But it doesn't hurt to be one either.

Lockheed Martin Space (LMS) builds satellites, space vehicles and other astronautical equipment. As the premier government contractor for NASA, it has built more interplanetary spacecraft than all U.S. companies combined. The company had many silos – all filled with data.

Ann Grubbs, LMS chief data engineer, described the environment as "hundreds, maybe thousands of data systems and tens of thousands of datasets."

Lockheed Martin Space connects all of its data silos by storing the relationships between the data and those systems in a graph database. This lightweight manner of connecting data silos by storing the pointers between them made it possible to quickly answer questions that formerly required weeks of querying different systems.

Graph technology now reveals connections never visible before. In one case, LMS analyzed which spacecraft parts were most important.

"To our surprise," chuckled Grubbs, "it turned out to be a tube of adhesive that had the most influence."

Ask yourself: **What questions could you answer if your data sources were connected?**

Rethink Your Master Data

Transformative MDM: Pitney Bowes

For some organizations, graph-based MDM has been transformational.

Although best known for postage meters and mailing services, Pitney Bowes actually is one of the top software companies in the world. Having built a slew of back-end processes to run its global business (routing mail around the world requires a lot of coordination), it is effectively a tech company.

"The main go-to-market focus we have is around the single view of customers, which is the Master Data Management (MDM) use case at its heart," said Aaron Wallace, principal product manager, Pitney Bowes.

Pitney Bowes had more than 150 different systems spread across the globe. The number grew constantly as the company made up to a dozen or more acquisitions every year. The company needed a centralized hub that all of these systems could plug into. At first, the company took a typical silo approach with an MDM stack that was highly centralized, controlled and governed.

They then realized that a single-system solution to MDM wasn't conducive to making their systems efficient. Seeking an enterprise-wide solution, Pitney Bowes became an early adopter of graph databases and a Neo4j partner.

Built on Neo4j, the solution provides a visualization of data moving through the organization. For example, the Pitney Bowes data-matching engine generates a record from multiple data repositories and matching algorithms resolve discrepancies. One individual may appear as "Charles Kane" in one data record, "Chuck Kane" in another and "Citizen Kane" elsewhere. Similarly, an individual's address may reside in one database, the email in a second database and mobile phone and social media information in a third. The system merges all those records into a single graph.

The efforts proved so successful that Pitney Bowes began offering an MDM solution to its own customers called the Spectrum Data Hub Module – powered, naturally, by the Neo4j graph database.

> *Ask yourself:* Could you productize your company's secret sauce?

Governance, Risk and Compliance

Just as data is growing more voluminous and complex, so are regulatory requirements. Compliance is a major driver of MDM initiatives. Graph technology helps enterprises navigate this new regulatory environment.

In 2016, the European Commission ratified the General Data Protection Regulation (GDPR). Under the law, companies must allow customers to transfer their personal information to competitors and allow people to exercise their "right to be forgotten," which requires the organization to erase all their personal information.

The GDPR comes amid a broader regulatory movement. The California Consumer Privacy Act of 2018 (CCPA) imposes stiff penalties on those that misuse and resell consumers' private information. Nevada and New York have followed suit, and many other states and nations are considering similar legislation.

Rethink Your Master Data

Companies must not only safeguard customer data, but also track how it is collected, used, stored, shared, accessed by third parties and protected. The "right to be forgotten" poses a disruptive requirement because organizations historically have focused on protecting and preserving information. Purging data case-by-case requires new capabilities. Compliance demands traceability, time stamps and mapping all the activity around personal data. As a consequence, organizations must adopt a new approach to data governance – and a platform to match.

With graph technology, companies track the data lifecycle, build "reverse lineage" maps of data flow and provide a full accounting to regulatory authorities.

Compliance Leads to Innovation: Convergys

There is a silver lining when you solve compliance with graph technology: new opportunities.

Convergys is a global customer care outsourcing firm whose clients include about half of the Fortune 500. It employs more than 115,000 employees worldwide and handles about 8 billion contacts per year.

With extensive operations in the EU, the company was alarmed by the requirements of GDPR. The company managed about 120 applications, internal storage and collaboration systems, plus more than 100 customers and 43 sites affected by the EU regulations. Initially, the company tried to build its own compliance solution.

"There's one problem with all of these apps – they were all built to put data in," said Lloyd Byrd, Convergys vice president of application development and technical solutions. "None of these apps were built to take data out, records at a time, or manipulate individual records. We didn't have a good way to address this problem."

The company partnered with Neo4j and within a couple of months built a graph-based GDPR solution running on the cloud. The solution was designed to extend well beyond the EU because data accountability is becoming a global reality.

The graph database solution opened the door to other benefits – better operational analytics, a tenfold improvement in large data loading, employee knowledge graphs and statistical insights.

"There are probably more use cases than people imagine where graph technology can enable better results," said Byrd. "For us, we think it's around operational results and being able to connect data better."

> *Ask yourself: What compliance challenge could you overcome using connected data?*

Data Lineage: UBS

Another powerful MDM capability is tracking data lineage. The UBS Group – one of the largest financial institutions in the world – built a data lineage and data governance platform using Neo4j.

The project began as a compliance mission but turned into something with broader benefits. The 2007 global financial crisis showed that banks lacked capabilities for risk data aggregation and practices for risk reporting. In response, the Basel Committee on Banking Supervision issued standard 239 (BCBS 239) to strengthen systems for risk data aggregation and internal risk reporting.

Data lineage is an essential component of risk management. It involves tracking the entire lifecycle of information – its origin, evolution and movement through the organization. With these tools, organizations trace information as it flows through the enterprise, monitor quality, discover errors, fix mistakes and reduce duplication.

Rethink Your Master Data

After attempting a solution with a relational database, UBS switched its data lineage system to Neo4j. UBS used Neo4j to evaluate data lineages and depict the results in a lineage diagram.

UBS attained better transparency into its own data. When generating a lineage, the company no longer suffers the headaches of JOINing multiple tables of a relational database. With Neo4j, the results are obtained easily and displayed in an intuitive graph visualization.

"Neo4j helps us understand the flow of data through the organization," explained Sidharth Goyal, a senior software engineer and technical lead at UBS. "It helps us understand how changes in one application are going to impact the entire organization. It helps us understand how errors can propagate through the system."

Ask yourself: How could graphs illuminate your data flows?

From MDM to Innovation

Once data is connected, the use cases are endless.

Let's say you have connected customer and product data. What if you add another node with another data source, such as partners or transactional data? In the language of innovation, the small step from your first use case to your next one is called the adjacent possible.

Graph technology also triggers healthy cultural shifts. If an organization has a data silo problem, it probably has knowledge silos too. Graph thinking catalyzes better insights and captures organizational know-how.

Knowledge Graphs

Talent Management

Research and Development

Customer 360

Fraud Detection

Churn Prediction

Compliance

With master data connected in Neo4j, a world of use cases opens up

Asset Management

Recommendations

> If an organization has a data silo problem, it probably has knowledge silos too. Graph thinking catalyzes better insights and captures organizational know-how.

NASA: Space-Age Knowledge Graphs

We are living in the century of the knowledge worker.

Few organizations have amassed as much institutional brainpower as NASA, the organization that put astronauts on the moon half a century ago and constantly expands the frontiers of science and engineering. To build its knowledge architecture, the space agency turned to Neo4j.

The challenge: a super abundance of brainpower and a shortage of connections. NASA operates 20 locations with 80,000 employees and has been amassing data since the 1950s – hundreds of millions of documents and reports stored in a database growing by the day. NASA keeps a database of lessons learned and encourages engineers to read about past projects – but finding the right information was like looking for a needle in a vast haystack.

"We have to try to break down those silos, which is exactly the capability that graph databases provide," said David Meza, the chief knowledge architect at NASA.

In the past, searches were time-consuming, inefficient, yielded unsorted results and only scratched the surface of millions of documents. NASA turned to a graph approach and began to convert its document-oriented database into a graph-oriented one using Neo4j.

The results are impressive.

"Using Neo4j, someone from our Orion project found information from the Apollo project that prevented an issue, saving well over two years of work and one million dollars of taxpayer funds," said Meza.

> *Ask yourself:* How could you capture the know-how in your company using graphs?

Graph technology is designed to reveal and handle connected data – the revolutionary approach that is already defining the future of the digital era.

Conclusion

Neo4j closes the innovation gap.

It brings together disparate sources of master data, connects silos and enriches it all with metadata. It keeps pace with the millisecond pace of modern business and scales without limits. Most importantly, graph technology is designed to reveal and handle connected data – the revolutionary approach that is already defining the future of the digital era.

Graph technology allows you to start where you are: with a single area, connecting data across disparate systems, in a nondisruptive fashion to solve a pressing problem. Connecting your data in this way creates a snowball effect, supporting new opportunities and use cases with minimal additional effort.

The Neo4j Graph Platform connects data at scale, powering millisecond queries on vast amounts of connected data. Furthermore, with its large library of graph algorithms, it paves the way for AI and machine learning on all of that data.

Graphs hold immense strategic value for master data management and beyond. Neo4j transforms your data managers and data scientists into data strategists. Armed with the power of graph technology, these strategists discover relationships, generate insights, drive innovation and capture competitive advantage.

Neo4j is the leader in graph database technology. As the world's most widely deployed graph database, we help global brands – including Comcast, NASA, UBS, and Volvo Cars – to reveal and predict how people, processes and systems are interrelated.

Using this relationships-first approach, applications built with Neo4j tackle connected data challenges such as analytics and artificial intelligence, fraud detection, real-time recommendations, and knowledge graphs. Find out more at neo4j.com.

Questions about Neo4j?

Contact us around the globe:
info@neo4j.com
neo4j.com/contact-us