neo4j

HELPING THE WORLD MAKE SENSE OF DATA

FROM GRAPH TO KNOWLEDGE GRAPH
A short journey to unlimited insights

Maya Natarajan

SENIOR DIRECTOR, PRODUCT MARKETING

Introduction

Unleashing the power of knowledge is imperative for enterprises looking for a competitive edge.

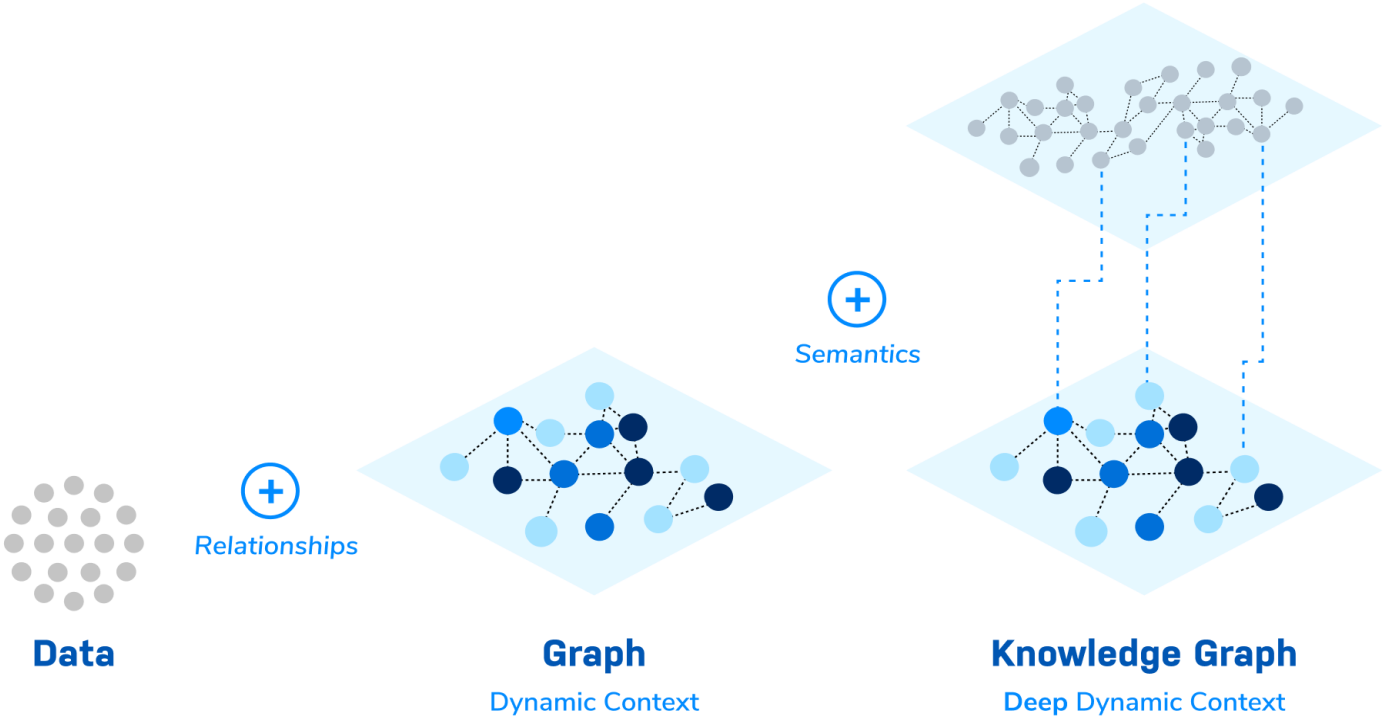
Everyone wants to capture knowledge, to connect everything that they know. However, turning data into knowledge is still very much an ongoing effort – though progress has been made, most data landscapes are far from mature.

We need a way to connect the data we have, across all systems, clouds, backups, and data lakes. Some of our data pipelines may be strong – serving real-time, clean data to dashboards and informing business processes – but that's not enough. That success must be replicated everywhere.

It's time to connect data so it's manageable and useful. Anyone in your organization who needs data to do their job should know where to look for it, be confident in its accuracy, and be able to solve their own problems with ease.

You get there by starting small, gaining immediate value, and then expanding outward – from graph to knowledge graph – the start of a short journey to unlimited insights.

What contributes to the data challenges we see today?



Data Trends and Challenges

Silos

First, there's siloed data. Data stores and applications often serve individual groups or departments. The HR team has one platform, and the sales team may use another such as Salesforce. Systems of record are important because they maintain and control data and establish governance policies, but data silos also reduce speed to analytics, accuracy of reporting, and data quality.

Data Sprawl and Data Lakes

Most organizations have data lakes, data warehouses, relational databases that underlie systems of record, customer data, transactional data, product data, and order data, and the list goes on. This scattering of data contributes to data sprawl.

Data lakes are popular as a low-cost option for storing large amounts of structured, semi-structured, and unstructured data. Object storage such as Amazon S3 is often used to create data lakes.

From a cost standpoint, data lakes are attractive; they give you a place to store any type of data, including log files generated by applications and services. It's easy and convenient to put data into a data lake. However, governing that data and even knowing what is there becomes a challenge.

Cloud Storage

Cloud computing is revolutionary and yet remains a governance challenge, with more data stored in more systems. Consider your own personal cloud data, which may live in iCloud, Google Drive, Dropbox, and more, as well as in Evernote, Gmail, and Notes.

This leaves us with a massive amount of data that is not only in different data formats and partially redundant but also largely disconnected.

If GDPR compliance says you need to forget a person's data, do you know where all their data is? What about cookies that you may not even realize belong to that person? What about data in your logistics system and in your partner's systems?

Historical Data Goes Right Out the Window

Historical data fuels machine learning predictions. But the COVID-19 pandemic made historical data obsolete as disruption rippled across the economy. For example, historical data would normally be used to predict purchasing behavior.

The pandemic caused purchasing behavior to change almost overnight as online shopping came to dominate the market during the lockdowns. Due to these dramatic shifts in consumer behavior, historical data could no longer make accurate predictions about purchasing behavior.

With limited data, connections in the data increase in importance and value. Putting data in a graph database captures connections and relationships.

The data you have is valuable, but storing the relationships in and across that data – relationships that already exist – increases your ability to predict even in the absence of relevant historical data. That's because connections and relationships are the most predictive elements in the data.

All of these factors are driving enterprises like yours to move toward connecting data in a graph database to gain knowledge. The next section examines exactly how a graph becomes a knowledge graph.

> “ Murky, messy data is a compliance nightmare.

What Is a Knowledge Graph?

Here is a definition:

> *A knowledge graph is an interconnected dataset enriched with semantics so we can reason about the underlying data and use it confidently for complex decision-making.*

Now let's unpack that definition a little at a time.

The Value of Connected Data

> *A knowledge graph is an interconnected dataset enriched with semantics so we can reason about the underlying data and use it confidently for complex decision-making.*

Connecting data adds context and improves outcomes. The best doctors do their homework. If a doctor takes the time to review the reason for the appointment – your medical conditions, your medical history, any lab tests you've had, and your vital signs – that visit will be far more productive. The doctor has context: the context of you.

A more obvious example is LinkedIn, where the number of connections we have is considered a measure of success (500+ connections). Like all prominent social networks, LinkedIn is built on a graph data structure.

A satellite manufacturer wanted to integrate all of its processes and data across the entire life cycle of its highly complex products – Product 360.

"In the past we've had someone manually identify the root cause of a failure," said the chief data officer. "They'd look at everything that could have influenced a part's failure. Is it engineering? Procurement? Supplier? Is it a vendor issue or a manufacturing defect? The idea is to let the graph do those traversals and find variance and report it back instantaneously versus a human taking weeks to do it manually."

Enriched with Semantics

> *A knowledge graph is an interconnected dataset enriched with semantics so we can reason about the underlying data and use it confidently for complex decision-making.*

Connecting data makes that data inherently more valuable and provides dynamic context. Adding even more information – semantics – takes that value to a new level.

RDF: The Semantic Elephant in the Room

In 1989, Dr. Tim Berners-Lee created what we now know as the web. He invented it because the scientists associated with CERN, a particle physics laboratory in Switzerland, had a hodge-podge of computers and if they could simply share their research, they could make progress much faster.

Not quite 10 years later, in 1998, Dr. Berners-Lee declared that the Semantic Web would be the next big thing. And work began on standards like XML and the Resource Description Framework, or RDF.

Since that time, much diligent work has been invested in ontologies such as FIBO and SnoMed using RDF-based vocabularies. Large public datasets like WikiData are also available as RDF.

How can those interested in a short trip to a knowledge graph make use of that kind of data?

Neo4j knowledge graphs incorporate RDF as well as any other type of data you may have. Neo4j is a property graph. That means it's simple, intuitive, and scalable. Simple because it can capture any information expressed in RDF or any other data model. Intuitive because it works on straightforward building blocks of nodes and relationships, which makes it easy to understand and explain. Scalable because it can accommodate billions of nodes and trillions of relationships.

The bottom line: Use the richness of RDF with the simplicity, intuitiveness, and scalability of Neo4j knowledge graphs.

In practical terms, if you create a graph, even a graph on a whiteboard, you will find yourself wanting to add more detailed information. That additional information is semantics: an overlay of meaning on top of the graph.

At NASA, David Meza sought to incorporate hundreds of millions of documents, reports, project data summaries, lessons learned, scientific research, medical analysis, and more into a knowledge graph. Meza ran the text of all these documents through a topic modeling algorithm that uncovered tags that could be applied to the content, enriching the graph.

According to Meza, NASA's Lessons Learned knowledge graph accelerated the mission to Mars, saving at least a year and over $2M in research and development.

So given his experience, what is Meza doing now? He's moved onto the next frontier: finding the right people to move NASA forward at a time when the private sector is diving in deep.

Talent is naturally a graph. NASA's new knowledge graph is used to find people with particular knowledge, skills, and abilities. But roles and job titles are not standardized, so Meza mapped them to existing ontologies from government databases, including the US Department of Labor's O*NET, OPM (Office of Personnel Management) classifications, and the EU's ESCO.

While NASA has particular needs when it comes to hiring talent, this is a broad challenge most organizations face – one where a knowledge graph could be of benefit no matter the mission.

Reason About the Underlying Data

> *A knowledge graph is an interconnected dataset enriched with semantics so we can reason about the underlying data and use it confidently for complex decision-making.*

Knowledge graphs provide deep, dynamic context. They enable people to find all related information in one place, with all of the relationships across that data. As you add more information, knowledge graphs become increasingly valuable.

For example, with NASA's Lessons Learned Knowledge Graph, the engineers could look at any particular part and gain insight into how it is used. Engineers could ask for a particular part and find out which subsystem it relates to. They might explore how all of the subsystems managed by different groups come together and try to get a feel for the system as a whole.

Fraud investigators find themselves drowning in thousands of alerts of potential fraudulent activity. Sifting through these alerts requires looking in several systems, which is time-consuming. A knowledge graph brings together many data sources, in this case customer attributes, credit scores, location, payment history, and more, so that investigators quickly hone in on fraudulent activities and see patterns they would not have noticed otherwise.

Use It Confidently

> *A knowledge graph is an interconnected dataset enriched with semantics so we can reason about the underlying data and use it confidently for complex decision-making.*

Data brawls. You may not have seen a bar room brawl, but most of us have seen a data brawl (or, more politely, a data disagreement). They arise from discrepancies where the numbers just don't match. "Where did you get your numbers?" someone asks. "That's not what my data says."

How can we use data confidently? By knowing where it came from, how much transformation it's been through, whether it's been cleansed, when it was last updated, and who updated it.

With a knowledge graph, it is easy to add this type of metadata to the graph. Knowledge graphs excel at capturing details about data, and this metadata becomes part of the knowledge graph. In the event of a data brawl (or data discussion), these types of details are invaluable in investigating any discrepancies.

For Complex Decision-Making

> *A knowledge graph is an interconnected dataset enriched with semantics so we can reason about the underlying data and use it confidently for complex decision-making.*

"When you start looking at what kind of documents you have and how you're able to transform those into actionable knowledge for your end users, you improve your decision-making," said David Meza, senior data scientist at NASA.

Dooloo, a startup based in France, has developed a knowledge graph-based platform to assist people who experience chronic pain. By combining a person's unique history with the latest research and treatments, patients and their caregivers quickly discover next steps to improve their quality of life and better manage their symptoms.

Categories of Knowledge Graphs

The world of knowledge graphs, like knowledge itself, is multifaceted and broad, almost endless. In general, knowledge graphs fall into two categories: **actioning** knowledge graphs and **decisioning** knowledge graphs.

Actioning Knowledge Graphs for Data Management

Data management is an important use case for knowledge graphs. Born-digital companies like Lyft and Airbnb thrive on data, and enabling their data scientists to find the latest, freshest data is key to their success.

These companies and others have used knowledge graphs to create metadata hubs to capture data lineage: where their data came from, how it was transformed, and how it was cleansed. Knowledge graphs model complex data pipelines so that consumers and producers of data can be easily identified and new data sources can be integrated.

With a strong foundation regarding the provenance of data, you can take action on that data, confident of where it came from and who its producers and consumers are.

In addition to data management use cases, actioning knowledge graphs are used for personalization and recommendations. Actioning knowledge graphs bring together all data about customers, products, and more into a 360-degree view that drives a wide variety of actions such as identification of customers in danger of churning, as well as recommendations about the kind of offer that might persuade them to stay.

Decisioning Knowledge Graphs for Data Analytics

Knowledge graphs form the foundation of modern data and analytics. With data captured in a knowledge graph, you no longer need to guess at correlations: all the relationships inherent in your data are captured and stored. In this way, knowledge graphs represent a more faithful representation of data and enable you to unlock its predictive power.

With a decisioning knowledge graph, the ultimate goal is to make a better decision, whether that's a human decision or an algorithmic decision. Those decisions can be supported in several ways.

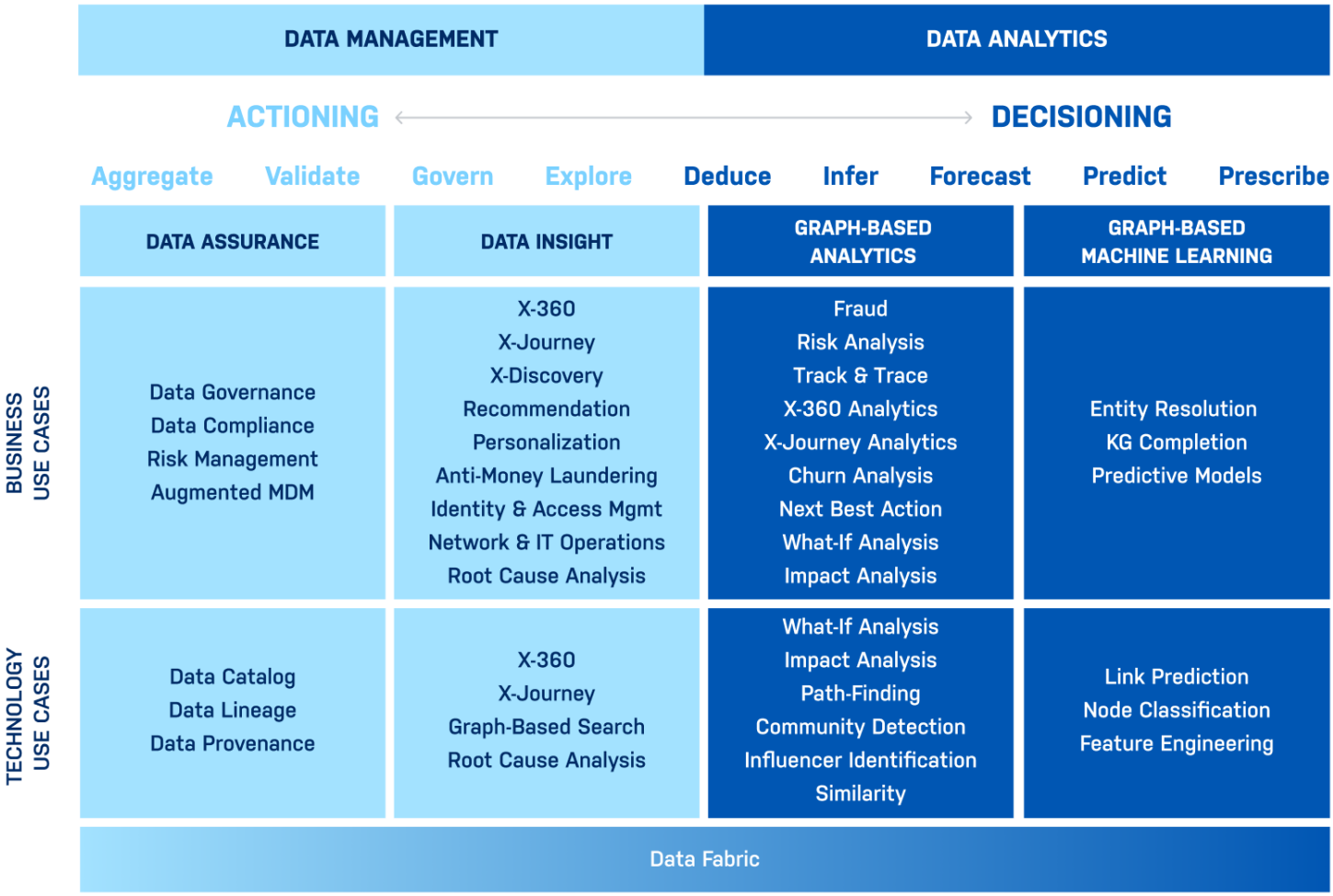
Graph queries enable you to answer any question of your knowledge graph, at scale. Boston Scientific uses advanced queries to do root cause analysis and identify combinations of at-fault components that result in defects (an anti-recommendation of sorts).

Graph algorithms identify patterns in your data, such as the shortest path between two points or the most influential customers.

OrbitMI uses a decisioning knowledge graph to perform complex route planning for container ships. Using pathfinding algorithms, they plan maritime routes in less than a second. Furthermore, their knowledge graph backs their SaaS analytics offering. The knowledge graph's impact is not only economic: greater efficiency in complex route planning reduced carbon emissions by 60,000 tons.

Graph queries and algorithms can also uncover predictive features for machine learning.

AstraZeneca uses graph algorithms and machine learning on its knowledge graph to identify patient journey archetypes and patterns. This research enables the company to identify influential touchpoints for early interventions to improve patient outcomes for illnesses like kidney disease.



A Short Journey to Unlimited Insights

A graph becomes a knowledge graph when we overlay an organizing principle. In this way, knowledge graphs make data smarter.

The transition from graph to knowledge graph is subtle and organic. Create a graph of your supply chain, for example. Add in information about how long you have worked with partners in that supply chain, their locations, modes of transportation, discounts, and any delays in shipping that must be factored in – and you've got yourself a knowledge graph.

This is the beauty of knowledge graphs in Neo4j. You can start small, solving a practical challenge of any kind, and enrich your knowledge graph incrementally to solve additional use cases and serve additional stakeholders.

If you decide to tweak the structure of your knowledge graph or add data sources, Neo4j makes it simple to modify the structure of your graph and reflow the data into that structure. That agility is part of the underlying graph database.

This deceptively simple structure supports advanced AI and machine learning. It is no wonder that every graph data science project starts with a knowledge graph.

Neo4j offers the most comprehensive knowledge graph on the market for data management, data analytics, and all the way to machine learning.

Learn more about the fastest path to knowledge graphs at neo4j.com/knowledge-graph.

Questions about Neo4j?

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