

## GRAPH DATABASES EMERGE AS MAJOR DRIVER OF INNOVATION

With the huge growth and ready availability in compute horsepower, the value of interconnected data sets – data sets that were once unconnected – becomes both visible and accessible to data and executive teams alike.

But it's that step of connecting these data sets, to show correlations that lead to potential decisions using that data, that's critical.

With that simple but powerful idea in mind, many organizations are now rushing to invest in graph databases to add a critical missing relationship perspective to their data. Without that key context, most organizations have little more than a relational database listing of who might have done what and when. A graph database adds a critical perspective that turns all the data an organization collects into truly actionable intelligence.

Graph databases enable organizations of all sizes to more easily and effectively visualize massive amounts of data. There are already hundreds of organizations taking advantage of a graph database from Neo4j to address everything from visualizing complex data sets to driving applications processing data in real time.

While many organizations are only just now seeing the tip of the iceberg when it comes to developing a wide range of next-generation applications, early adopters of graph databases are already reaping the benefits of a technology platform that turns data into a real asset.

### GRAPH DATABASES: A BRIEF HISTORY

At its most basic level, a graph database differs from traditional relational databases in that it produces graph structures to create semantic queries against data sets that have been previously linked together. That approach makes it simpler to launch queries against those data sets versus trying to join various tables in a relational database to replicate the same capability.

Graph databases are generally made up of three elements:

- **Nodes:** These represent entities such as people, businesses, accounts, or any other item to be tracked.
- **Edges:** Also called graphs or relationships, these are the lines that connect nodes to other nodes; they represent the relationship between them.
- **Properties:** These are other data sets that are especially germane to a specific node.

While other classes of databases have been used to model these relationships over the years, graph databases came into their own in the late 1990s when they were first employed mainly to index web pages. Since then, the ability of graph databases to support transactional and real-time applications has made them an attractive alternative to relational databases for supporting complex queries. That's because relational databases are unable to establish a relationship between data sets until a schema is built to join tables using complex SQL queries that are fundamentally inefficient.

Early adopters of graph databases are already reaping the benefits of a technology platform that turns data into a real asset.



Walmart relies on the Neo4j graph database to optimize upselling and cross-selling online across all its major product lines in core markets. The result is a superior digital business experience for end customers.

## MAKING THE CASE FOR A DEDICATED GRAPH DATABASE

Graph databases are now being broadly adopted, even to the point where some providers of legacy database technologies are trying to graft graph database schemas on top of existing relational database architectures. While that approach may reduce costs in theory, in reality, it can limit and compromise the performance of the queries launched against the database.

Because of that issue, almost every instance of a graph database being used to transform how an organization operates relies on a standalone instance of a graph database. Imitation may be the sincerest form of flattery, but when it comes application performance, there can never be a substitute for the real thing.

With that issue in mind, here are several examples of how standard graph databases are not only reinventing how tasks are completed, but are also enabling new application experiences that were once impossible.

## HOW GRAPH DATABASES DRIVE DIGITAL BUSINESS: WALMART

When it comes to digital business initiatives, Walmart's efforts are seen as an inspiration for transforming a traditional brick-and-mortar company into a digital business powerhouse capable of competing with the likes of Amazon.

Walmart relies on the Neo4j graph database to optimize upselling and cross-selling online across all its major product lines in core markets. Instead of employing a complex batch process on top of a legacy relational database, Walmart makes extensive use of a real-time recommendation system built on top of a graph database capable of processing low-latency queries. By design, graph databases can be used to uniquely query customers' past purchases during an online visit to match historical and session data in a way that dramatically outperforms traditional relational databases.

The result is a superior digital business experience for end customers navigating one of the largest online product portfolios in existence today.

## HOW GRAPH DATABASES TRANSFORM MEDICAL RESEARCH: CANDIOLLO CANCER INSTITUTE

Advances in modern medicine require researchers to be able to easily perform tests where they can manipulate massive amounts of data in real time. One research organization that relies on the Neo4j database to achieve that goal is the Candiolo Cancer Institute (IRCC) based in Candiolo (Torino), Italy.

IRCC researchers needed to develop a laboratory information management system to track data such as the biological and molecular properties of cancer samples along with what scientific procedures were performed on those samples. The challenge IRCC faced was that much of this data is structurally complex; it tends to be hierarchical in nature because of intricate and frequently changing relationships.

Initially, IRCC tried to model the data using a relational database. But all the SQL joins involved resulted in sluggish queries, as well as challenges with data integration and coherency. To solve that problem, IRCC developed a production version of its database that relies on MySQL to store legacy data and track entities, characteristics and laboratory procedures.

This data is then sent to a Neo4j graph database that continually imports data from publicly available data resources. A MongoDB document database is used to store raw, complex data, while the rest is natively stored on Neo4j to identify complex relationships, analyze experimental procedures, and model the genomic domain and complex semantics for genomic knowledge.

In the future, IRCC plans to remodel its database. It will use Neo4j as a more abstract layer to generate data models for each instance to create an abstract ontology to determine relationships. This approach enables IRCC to model relationships between concepts that evolve and accommodate continually changing biological research. The result is not only the ability to process queries faster, but the IRCC also gains a means to track its workflow in a way that can be shared with researchers around the world.



## HOW GRAPH DATABASES TRANSFORM AGRICULTURE: MONSANTO

Agriculture giant Monsanto is constantly trying to find ways to increase crop yields. That means analyzing various strains of seeds over decades of time to determine which one will lead to better crops being harvested. Relying on legacy relational databases to analyze all that genetic data has become untenable. Almost every query now being launched employs modern data analysis techniques that need to run in real time. Traditional relational databases would take seconds to minutes to hours to perform one round of analysis. Obviously, that approach will not scale across thousands of queries.

Using a Neo4j database makes it simple to model those same queries into a graph, which in turn allows analysis that used to take minutes or hours to be processed in seconds. In fact, the query can now be launched against several million objects at once.

The Neo4j cluster holds the Monsanto genetic history data, which is accessed via a rich application programming interface (API) that enables both geneticists and application developers to speak and execute complex algorithms in simple terms. Thus far, over 700 million REST API requests have been made to the Neo4j cluster.

## HOW GRAPH DATABASES TRANSFORM E-COMMERCE: EBAY SHUTL

Same-day delivery of online orders is the holy grail of ecommerce. In the search for this holy grail, eBay is increasingly making use of a platform for managing the process – based on Neo4j – that it acquired via its purchase of the delivery service provider, Shutl.

Originally developed on a legacy relational database, it quickly became apparent that the queries used to select the best courier to deliver a package in the same day were simply taking too long. The Neo4j database takes advantage of queries that can be launched as a graph to eliminate one of the biggest roadblocks there is between retailers and instant gratification for online shoppers.

Specifically, eBay Shutl can process queries thousands of times faster than the prior relational database solution, using anywhere from tens to hundreds of times less code. That reduction in code turns out to be critical. Now, there's room for additional custom code within the application, which has allowed the company to add functionality that wasn't possible before.

## HOW GRAPH DATABASES TRANSFORM SCIENCE: NASA

Often the biggest data goal many organizations have is simply to determine what they already know. At NASA, the agency has deployed an instance of Neo4j alongside a MongoDB document database to help employees see what data they have. The database serves as a repository for the data that is connected to a Neo4j database that makes it easier to visualize who in NASA owns any data related to a key term, using what NASA describes as a Knowledge Architecture to break down organizational silos.

Previously, a NASA engineer had to search for a key term across 20 million documents. Now NASA engineers can link multiple key terms together to more efficiently search the repository to identify not only where those documents reside, but also who authored them inside the agency.

These capabilities enabled a team member from NASA's Orion project to find information from the Apollo project that prevented an issue, saving well over two years of work and \$1 million of taxpayer funds. It has also fostered greater levels of collaboration between engineers and scientists around the globe, which for a scientific agency such as NASA, is priceless.

Using Neo4j, a team member from NASA's Orion project found information from the Apollo project that prevented an issue, saving well over two years of work and \$1 million of taxpayer funds.



## HOW GRAPH DATABASES TRANSFORM JOURNALISM: ICIJ

The International Consortium of Investigative Journalists (ICIJ) is a small team of data journalists specializing in investigating cross-border crime and corruption. Their most well-known work involves the disclosure of the existence of documents showing assets belonging to government officials and private individuals that were being hidden from tax collectors. Known as “The Panama Papers,” analyzing that trove of documents required identifying relationships across 11.5 million documents spanning 2.6TB of data.

ICIJ relied on a Neo4j database to reveal connections in both text-based and account-based data that had been leaked. Journalists with limited technical skills were able to identify relationships between people, corporations, accounts, shell companies and offshore accounts.

## HOW GRAPH DATABASES TRANSFORM IT MANAGEMENT: ORANGE BUSINESS SERVICES

Orange Business Services is employing the Neo4j graph database to unify IT management by being able to connect server or network elements to provide a comprehensive view of the IT environment. Historically, each class of IT infrastructure needed to be managed in isolation. Orange Business Services is now using Neo4j to break down IT silos by identifying the relationship between all the components that make up the IT environment.

Internally, Orange Business Services leveraged that capability to identify potential security issues in its information systems. But that effort has been expanded to include a managed service that the IT services provider now makes available to Fortune 500 companies.

Orange Business Services is looking at additional use cases involving employee churn reduction and training recommendations for human resources.

## SUMMARY

These use cases are just the start in terms of new application experiences that can be created when the relationship between multiple data sets is already baked into the core database. Graph databases are one of the hottest segments of the so-called NoSQL database market for good reasons.

While relational databases have their limitations as well as their uses, there's no better reason to embrace a new technology platform than to provide a capability that adds immediate business value. After all, the only limitation to graph databases isn't the technology itself; rather, it's the collective imagination of the organizations that have yet to embrace them.

The only limitation to graph databases isn't the technology itself; rather, it's the collective imagination of the organizations that have yet to embrace them.



RTInsights is an independent, expert-driven web resource for senior business and IT enterprise professionals in vertical industries. We help our readers understand how they can transform their businesses to higher-value outcomes and new business models with IoT real-time analytics. We provide clarity and direction amid the often confusing array of approaches and vendor solutions. We provide our partners with a unique combination of services and deep domain expertise to improve their product marketing, lead generation, and thought leadership activity.

Neo4j, Inc. is the graph company behind the leading platform for connected data. The Neo4j graph platform helps organizations make sense of their data by revealing how people, processes and digital systems are interrelated. This connections-first approach powers intelligent applications tackling challenges such as artificial intelligence, fraud detection, real-time recommendations and master data.

More than 250 commercial customers, including global enterprises like Walmart, Comcast, Cisco, eBay and UBS use Neo4j to create a competitive advantage from connections in their data.