KNOWLEDGE GRAPH AT NEO4J

How to represent interconnected information and concepts with the Neo4j platform for scientific research

IDC InfoBrief

May 2022

Table of Content

Knowledge Graphs: What Are They?

Knowledge graphs represent interlinked real-world descriptions of entities like people, objects, locations, events, and concepts.

Knowledge graphs provide a framework for data integration, management, and analytics.

Knowledge graphs also serve as one of the foundational pillars for artificial intelligence (AI), as well as human knowledge development.
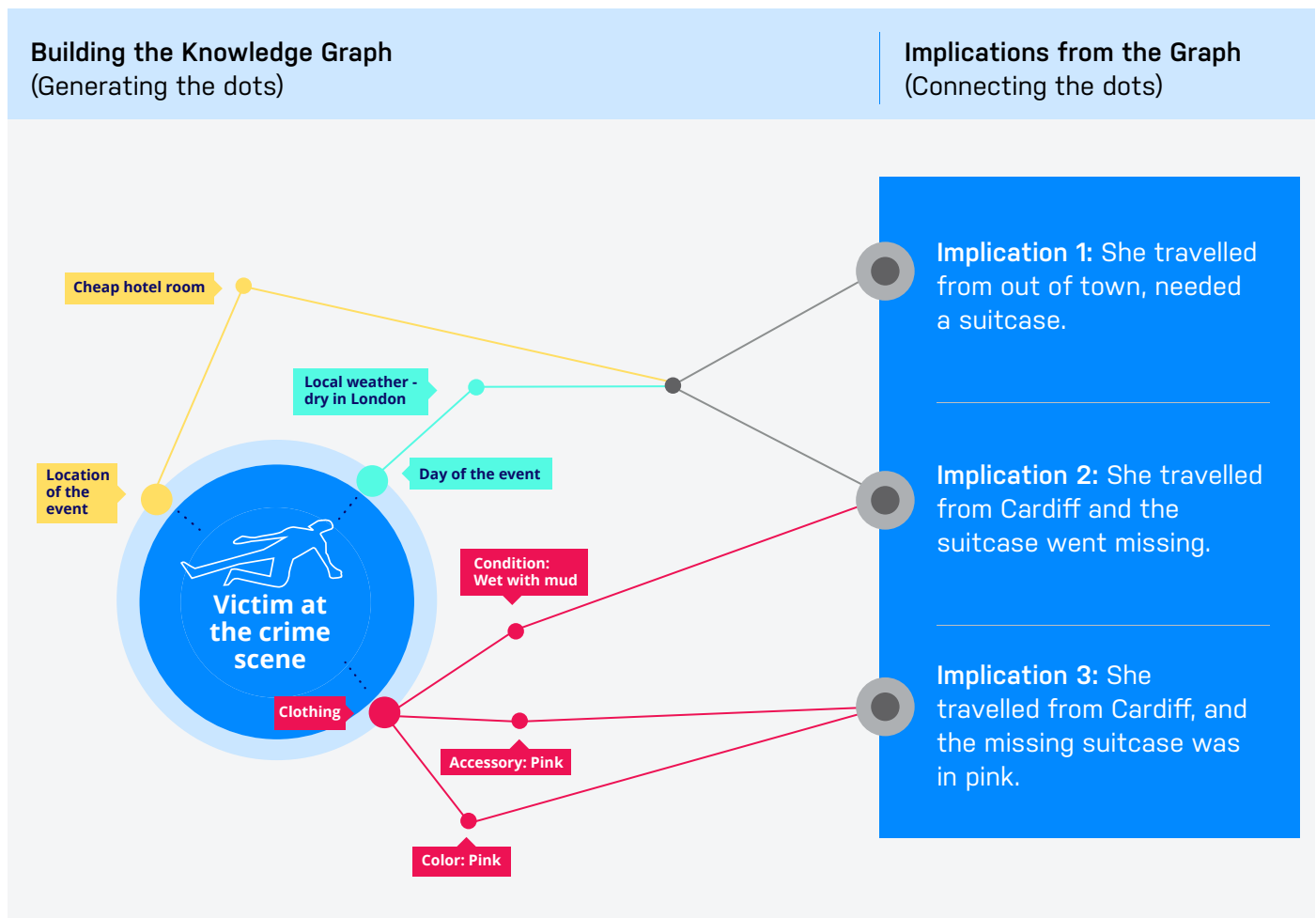
Everyone uses knowledge graphs, consciously or unconsciously; all of us in our daily lives link ideas and processes for problem solving. But technology can make them explicit. This IDC InfoBrief explores further the application of knowledge graphs in business, the use cases, and the related technologies, as well as provides guidance on scaling knowledge graphs for a digital-first economy.

Knowledge Graphs – How Are They Used?

Finding Answers/Patterns Hidden in Complex Interlinked Events

A Crime Scene

[1]It's a grey but dry October morning in London. A middle-aged woman has been found dead on the floor of a cheap hotel room. She is expensively dressed all in pink except for her fashionable but very muddy shoes. A pink handbag also lies untouched next to her. What had happened?

Building the Knowledge Graph
(Generating the dots)

Implications from the Graph
(Connecting the dots)

Cheap hotel room

Local weather - dry in London

Location of the event

Day of the event

Victim at the crime scene

Condition: Wet with mud

Clothing

Accessory: Pink

Color: Pink

Implication 1: She travelled from out of town, needed a suitcase.

Implication 2: She travelled from Cardiff and the suitcase went missing.

Implication 3: She travelled from Cardiff, and the missing suitcase was in pink.

[1] Based on "A Study in Pink", Season 1, Episode 1 of the BBC TV series, Sherlock, 2010

Building a Knowledge Graph

And the Need for Technologies to Scale.

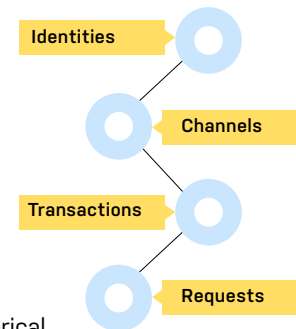|	The **data** that can be structurally explored and inquired	The business or domain knowledge that puts the data in **context**	The **graph** where entities and their relationships can be made explicit and examined
Why we need knowledge graph technology	**Step 1** Collecting The Data	**Step 2** Adding the Context	**Step 3** Querying the graph
When we need to make sense out of a huge amount of data with complex contexts, in shorter time intervals, and in a scalable manner	**Too much data, too little knowledge** Humans faceinformation overload when dealing with a large volume of contextual information	**The increasing complexity of context** Humans faceinformation overload when dealing with a large volume of contextual information	**The need to decide quickly** With knowledge graph technology, deductions and decisions can be made in near real time

To Catch a Thief

Fraud Detection and Risk Profiling in Banking, Insurance, Retail, and eCommerce

For the financial sector the world over, fraud strikes across an increasingly convoluted network of transactions, channels, and identities. This makes timely detection of fraudulent events a huge and constant challenge.

Knowledge graphs adapt to changing fraud techniques

Fraud techniques evolve over time. It is not sufficient to detect fraud based on historical events. Knowledge graphs can be used to investigate connections across transactions, channels, identity characteristics, etc., to spot unusual behaviors and patterns.

Identities
Channels
Transactions
Requests

SINGAPORE	USA
The number of scams reported last year hit a record high, climbing 65.1% from 2019, as scammers took more than $201 million from their victims. (The Straits Times, Feb 2021)	Losses from identity theft cases cost $502.5 billion in 2019, increased 42% to $712.4 billion in 2020, and is projected to approximate $721.3 billion in 2022. (Insurance Information Institute)

Cybersecurity for IT Service Providers and Government Agencies

In APJ, the federal/central governments of China, Australia, and Singapore lead in spending on cybersecurity. By 2025, they will invest more than $724 million on intelligent solutions to enhance cybersecurity.[1]

The cyberworld is fast-evolving, and so is the world of malware. Viruses, trojans, bots, spyware, ransomware, worms, adware, phishing, distributed denial of service (DDoS), the list continues to grow. The task of cybersecurity is further complicated by the heterogeneity of IT and OT systems that can be potentially attacked – databases, web portals, edge devices, operational systems, and so forth.

Addresses
Requests
Incidents
Systems
Accounts
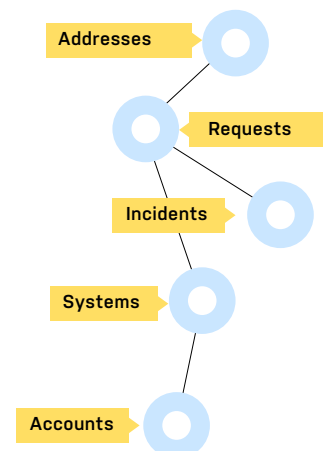
Cyberattacks are close to constant

In 2017, an experiment was conducted by a security company in England to investigate the extent of cyberattacks. Just 71 minutes after the company set up servers for the study, automated attack bots had already started scanning their system and looking for vulnerabilities that they can exploit. For the 170 hours this system was put online, the bot attacks continued unabated.

Source: [1] IDC Worldwide Artificial Intelligence Spending Guide

Patterns Hidden in Plain Sight

Biomarker Discovery for Life Science and Pharmaceutics

A clinical knowledge graph is an open-source platform containing about 20 million nodes and 220 million relationships between available experimental data, public data, and literature. It grows in size and allows the use of statistical analysis and machine learning to improve the efficiency of biomedical research, such as bio-marker studies. (Nature, 2022)

Biomarkers are traceable substances introduced into an organism to test organ function. For example, rubidium chloride is a radioactive isotope used to evaluate blood flow in the heart. But developing biomarkers has its challenges, as these chemicals interact with many aspects of the body's processes, even potentially leading to major health problems.
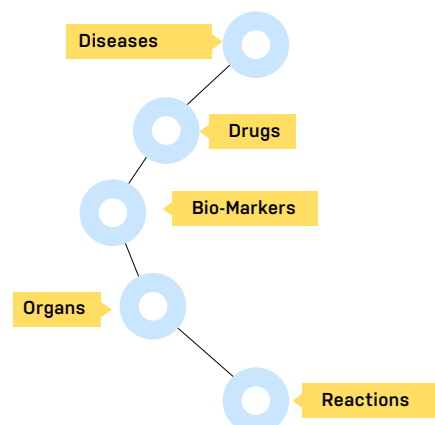
Knowledge graphs speed up the discovery of connections between biomarkers, individual genes, and metabolic processes to support the development of personalized medicine.

Integrating All of Biology into a Public Database

Dr. Daniel Himmelstein, Postdoctoral Fellow, University of Pennsylvania, started his PhD research with the question: How do you teach a computer biology? He found the answer in a heterogenous network (a.k.a., "HetNet"), which turned out to be another term for a labelled property graph.

After an attempt to create his own Python package for querying HetNets, Himmelstein turned to Neo4j. By importing open source drug and genetic information, he has developed a graph with more than 2 million relationships that can be mined for drug repurposing – in other words, finding new treatment uses for drugs that are already on the market – via a growing dataset of matching compound-disease pairs.

For each of the current 200,000 compound-disease pairs, his project computes the prevalence of many different types of paths and then uses a machine learning classifier to identify the patterns of the network, or the paths, that are predictive of treatment or efficacy. As an example, Himmelstein shows you how his HetNet project helped identify bupropion as a drug that not only treats depression but also nicotine dependence.

To know more, click here.

Smart Networking

APJ Telco companies expected to spend more than $584 million on advanced analytics for network optimization and preventative maintenance in 2022.

Telcos operate in ever more complex vendor ecosystems, and their services must interact across all kinds of hardware and environments. This creates major problems for telco support teams, as multivendor support requires a significantly higher capacity to understand the interactions between technology layers.

- IoT and the proliferation of edge devices continue to create variations in network topology, making it increasingly hard to conduct fault diagnosis and implement rule-based policies.

- Software defined networks (SDN) and network function virtualization (NFV) initiatives decouple network services from physical units. As a result, multiple instances will be running on a single system, increasing both the cost of failure and difficulty to recover all services affected.
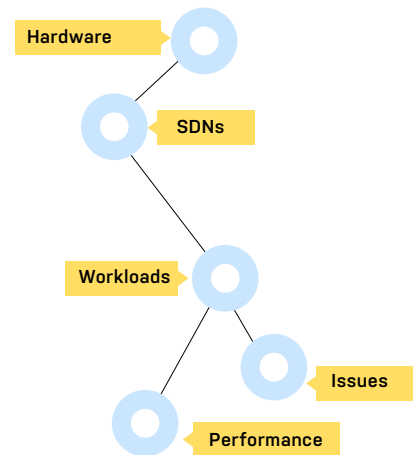
Knowledge graphs for telcos capture and model processes, sub-networks, equipment, and events to identify single points of failure and support SDN.

Context, Context, Context

Recommender Systems for eCommerce, Retail, Banking, Professional Services

By 2025, banks in APJ will spend an estimated $606 million on recommender systems to improve customer experience.

> *Knowledge graphs map the myriad behaviors of customers, thus allowing organizations to generate hyper-personalized product recommendations.*

Peers
Preferences
Promotions
**Itens**
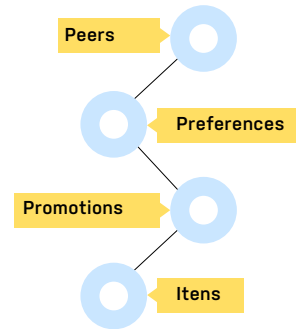
Secret of the world's most downloaded app

TikTok climbed to pole position as the world's most downloaded app since the last quarter of 2019. The app for short-form video sharing has 1 billion monthly active users - achieved by learning user behaviors and providing highly personalized recommendations in the format of content streams.

Digital Twins for Architecture, Construction, Manufacturing, Resource & Energy

A digital twin is a virtual model of a product or asset connected to, or aligned with, the related physical prototype – for instance, via data capturing means such as IoT, thus visualizing data flow, communication, and collaboration across operations from designing and engineering, to the supply chain and servicing.

> *Knowledge graphs enable the creation and analysis of a complex digital twin*

Principals
Procedures
Equipment
Flows
Exceptions

- An average manufacturing plant generates tens and thousands of heterogeneous measurement data points every second.

- Managing and making sense of the complex and dynamic models of a digital twin is challenging.

- Knowledge graphs overcome these challenges by enabling complex analytics to gain new insights.
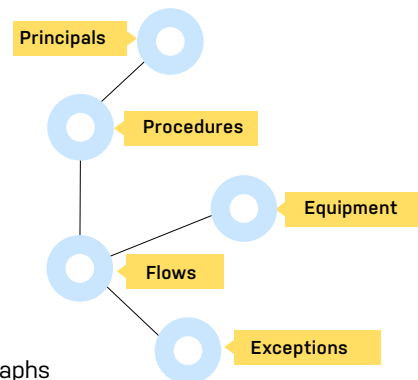
Resilient supply chain

Increasingly, companies are replacing linear supply chains with the notion of resilient ecosystems. Knowledge graphs are excellent in capturing and monitoring the dynamic relations between principals, agents, procedures, equipment, etc., in an ecosystem. For example, automobiles production is known for being highly automated and precise, and working with a myriad of suppliers of different parts and modules. Leading companies like BMW AG have invested in virtualized factory planning that translates customer preferences to detailed supply planning and optimized factory design, and allows real-time inter-organization collaboration.

Knowledge Graph Based Analytics for ML and AI

Main types of Graph Analytics
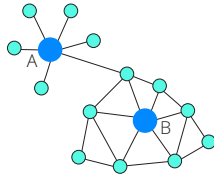
PATH FINDING & ANALYSIS

Estimate the relationships between nodes, often used to find the shortest path (connected edges) in logistics and transportations.
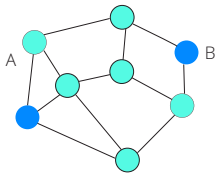
How to travel from A to B in the shortest time?
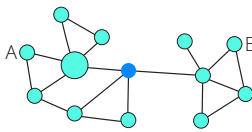
CENTRALITY FINDING & ANALYSIS

Determine how important a node in the network is. There are different types of centrality – degree centrality, betweenness centrality, etc.

Which social network influencer to choose for the next campaign, A or B?
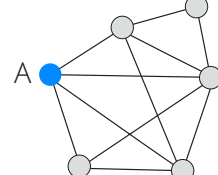
CONNECTIVITY FINDING & ANALYSIS

Estimate how strongly or weakly connected two nodes are. Find connections that everything goes through.

Which node is the most critical for information to flow from A to B?

COMMUNITY FINDING & ANALYSIS

Find groups of people interacting frequently with each other. Evaluate the tendency for such interactions to strengthen or break apart.

Knowing A is a fraud agent, what about the nodes interacting with A in the event of fraud?

SIMILARITY FINDING & ANALYSIS

Understand the similarity of nodes based on their neighborhoods or their properties.
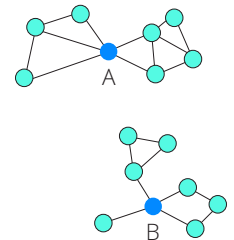
How likely will A make similar buying decisions of B?

Graph analytics play a huge role in machine learning and artificial intelligence

1 Enriched Features

2 Better Predictability

3 Many Use Cases

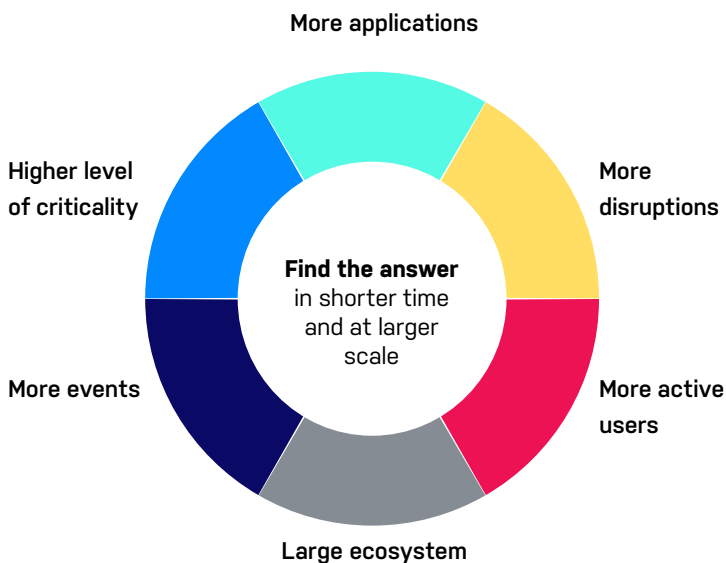# Scaling Knowledge Graphs for the Digital-first Economy

A higher proportion of gross domestic product (GDP) will be contributed by digital products and services. Increasingly, companies will face new challenges in the digital-first economy, and they will be required to scale the solutioning process with greater agility.

IDC *PREDICTION*

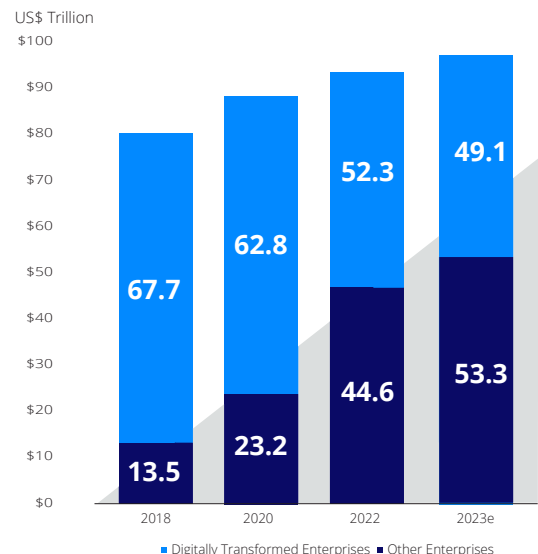*IDC predicts that 65% of global GDP will be digitalized, driving over $ 6.8 trillion in spending from 2022 to 2023.*

IDC FutureScape: Worldwide Digital Transformation 2021 Predictions

Are You Ready for the Challenge of a Digital-first Economy?

More applications

Higher level of criticality

Find the answer in shorter time and at larger scale

More events

Large ecosystem

More disruptions

More active users

Source: IDC TechBrief: Graph Databases (Carl W Olofson), IDC's IT Industry Vision for the Digital-First Enterprise (Rick Villars)

Worldwide Nominal GDP Driven by Digitally Transformed vs. Other Enterprises

US$ Trillion

Year	Digitally Transformed Enterprises	Other Enterprises
2018	67.7	13.5
2020	62.8	23.2
2022	52.3	44.6
2023e	49.1	53.3

■ Digitally Transformed Enterprises ■ Other Enterprises

IDC Recommendations:

1 Know the concept
Understand the characteristics and steps to build knowledge graphs, and their role as a fundamental pillar of problem solving – in our daily lives and in the business world.

2 List out business requirements
Knowledge graphs can be used in catching anomalies, discovering patterns, and laying out a solid foundation for intelligent solutions such as recommender systems and digital twins. All companies can find relevant use cases.

3 Go for native technology
IDC recommends avoiding any database system that says, "Oh, we support graphs too" with reference to technology that is really aimed at some other workload. Go for native graph technology.

Carl W. Olofson, Vice President, IDC Worldwide

Message from the Sponsor

Neo4j Knowledge Graph: Making Sense of Complex Data for Intelligent Decision-making

A Neo4j knowledge graph is defined as an interconnected dataset enriched with meaning so we can reason about the underlying data and use it confidently for complex decision-making.
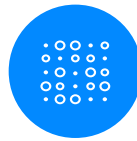
Neo4j knowledge graphs:

Drive intelligence into data to significantly enhance its overall value

Turbocharge artificial intelligence/machine learning for better predictions

Enable use cases all across the data spectrum, from management to analytics and machine learning

Bridge data silos and form the foundation for innovative applications like data fabric and digital twins

Neo4j knowledge graphs are deployed across industries such as financial services, healthcare, life sciences, supply chain and logistics, retail, and manufacturing for various use cases from fraud detection and real-time recommendations to patient journeys, digital twins, bill of materials and more.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

IDC Asia/Pacific
83 Clemenceau Avenue
#17-01 UE Square, West Wing
Singapore 239920
T 65.6226.0330

idc.com

Neo4j is the world's leading graph data platform. We help organizations – including **Comcast**, ICIJ, **NASA**, **UBS**, and **Volvo Cars** – capture the rich context of the real world that exists in their data to solve challenges of any size and scale. Our customers transform their industries by curbing financial fraud and cybercrime, optimizing global networks, accelerating breakthrough research, and providing better recommendations. Neo4j delivers real-time transaction processing, advanced AI/ML, intuitive data visualization, and more. Find us at **neo4j.com** and follow us at **@Neo4j**.

Questions about
Neo4j? Contact us
around the globe:

info@neo4j.com
neo4j.com/contact-us