

From Data to Decisions: AI + Knowledge Graphs

Philip Rathle

Chief Technology Officer, Neo4j



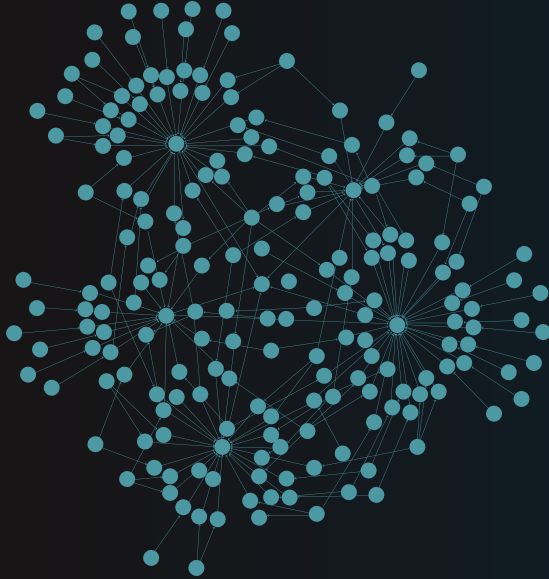
Part I: Graphs pre GenAI

Graph Database,
Graph Analytics & AI

Started with the Questions:

1. How do real & digital world systems show up & behave?
2. Can we contribute a new way of working with data that's 10x+ better for complex systems

Networks



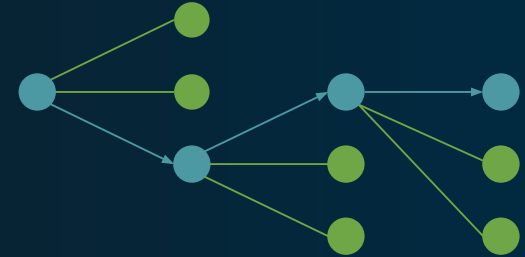
Computer, social,
communications, biological,
ecological, transportation,
payment, etc.

Hierarchies

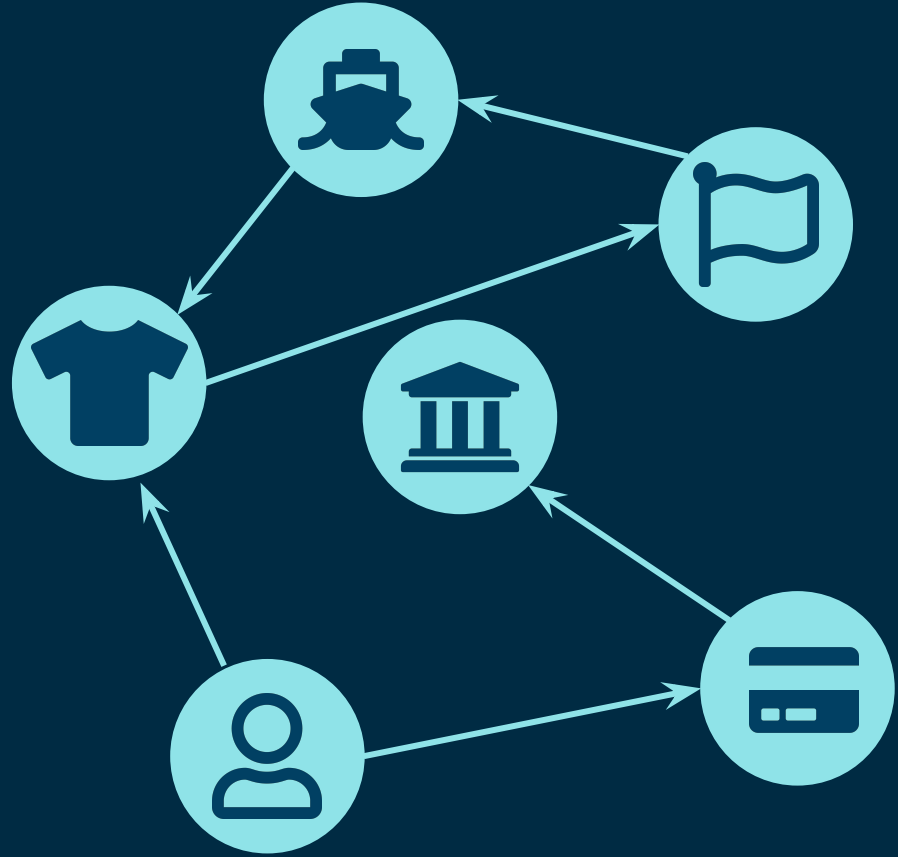


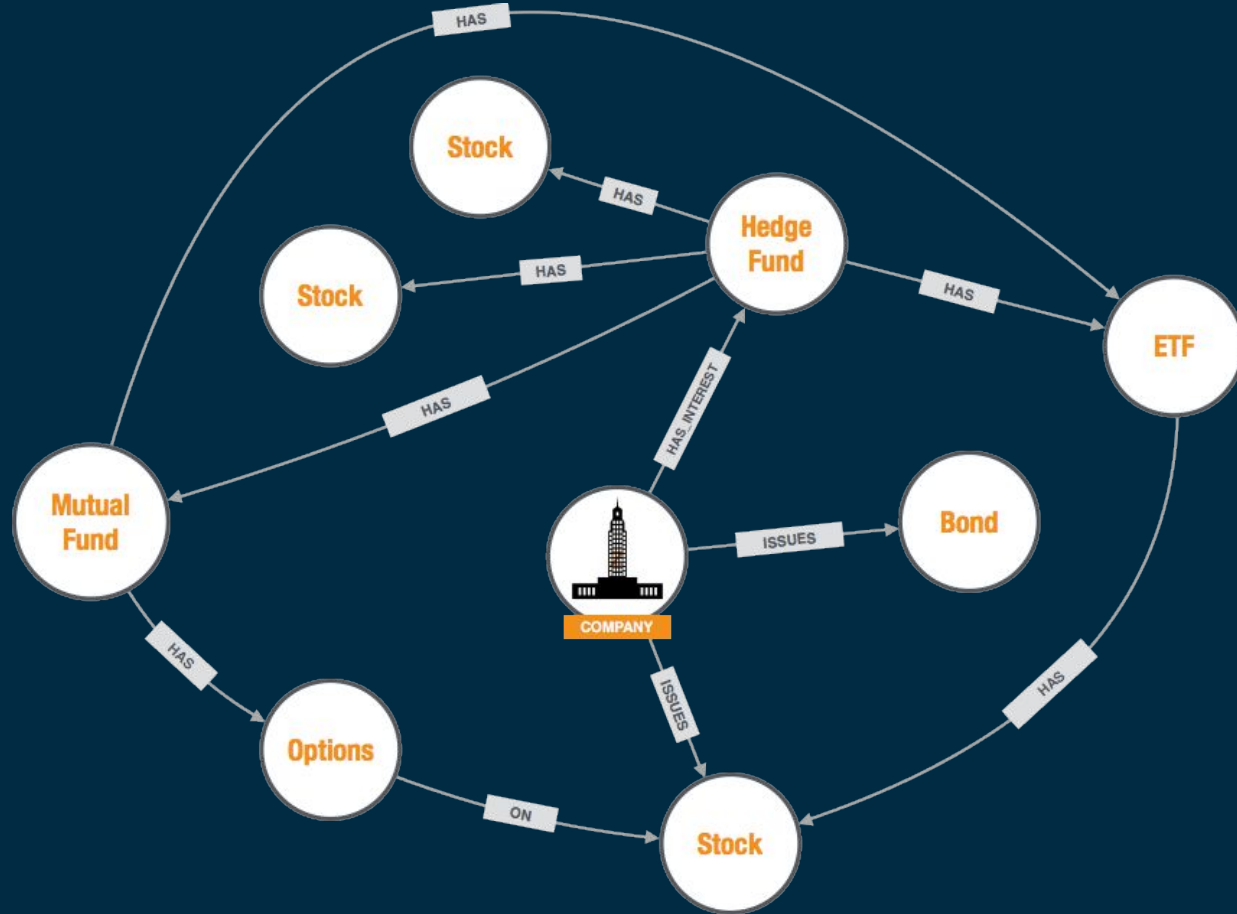
Org, supply chain,
bill of materials, asset
ownership, etc.

Journeys



Customer journey,
patient journey, career
journey...





Michael J. Hsu

Acting Comptroller of the Currency

Dates Served: May 10, 2021 – February 10, 2025

Michael J. Hsu served as Acting Comptroller of the Currency from May 10, 2021, to February 10, 2025. As Acting Comptroller of the Currency, Mr. Hsu was the administrator of the federal banking system and chief executive officer of the Office of the Comptroller of the Currency (OCC). The OCC ensures that the federal banking system operates in a safe and sound manner, provides fair access to financial services, treats customers fairly, and complies with applicable laws and regulations. It supervises nearly 1,100 national banks, federal savings associations, and federal branches and agencies of foreign banks that serve



Thinking Differently

From rows and columns to nodes and edges

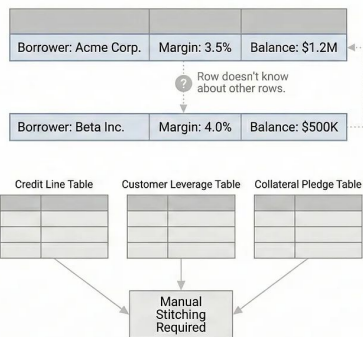
MIKE HSU
FEB 15, 2026

The Sapir-Whorf hypothesis postulates that the language a person speaks shapes the way they think and how they perceive the world. Language, in this sense, *limits* one's ability to understand things.

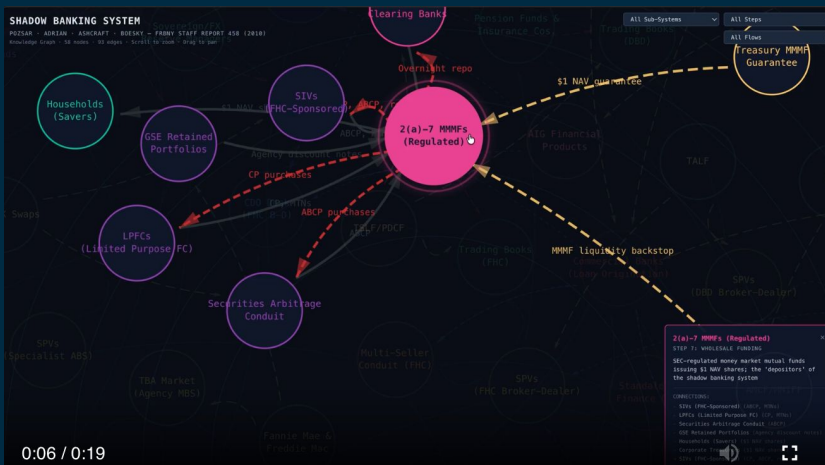
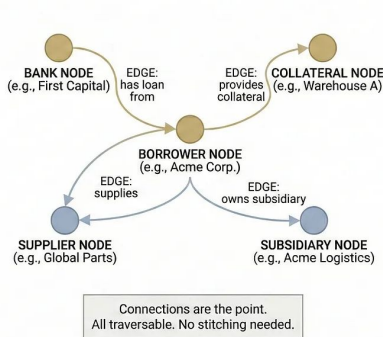
But therein lies an opportunity. If language is limiting, *learning new languages* can unlock new ways of seeing and engaging with the world.

This post argues that supervisors should learn the language of **knowledge graphs** — i.e., of **nodes and edges** — to see the world distinct from the rows and columns that have enabled and shaped our understanding of the financial system.

TABULAR DATA: Isolated Islands



KNOWLEDGE GRAPHS: Relational Network



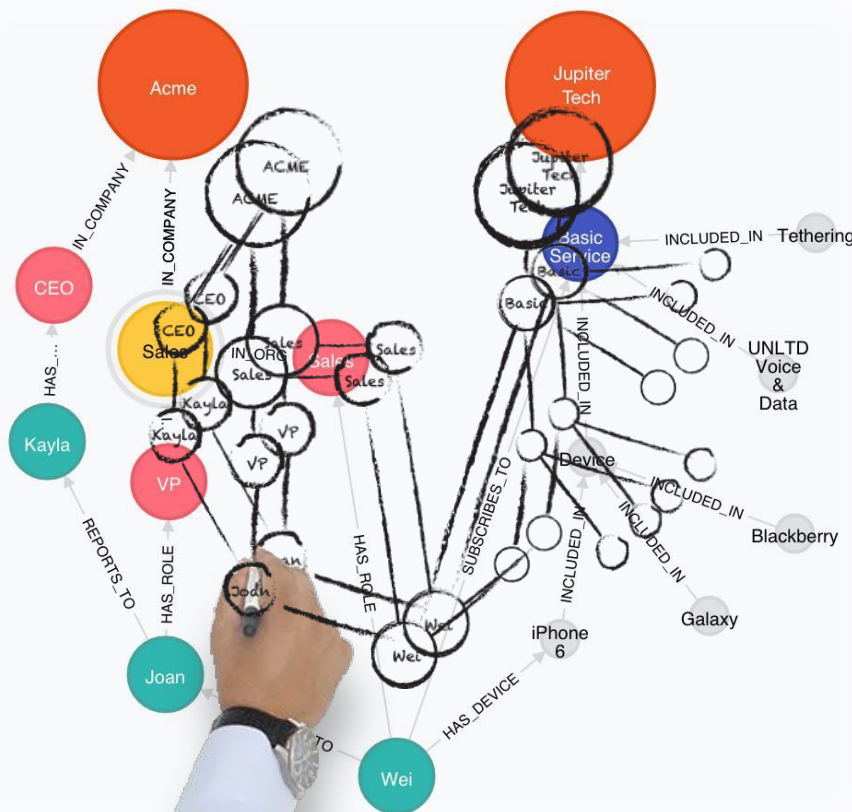
0:06 / 0:19



A R R I V A L

Graph DBMS

Data stored & queried as a graph vs tables



Stores & processes data the way we understand the world *and like the brain*

Highly efficient: 1000x performance with 10x less hardware

Flexible schema enables applications to be far more easily evolved

Agile, open, enabling data & use case network effects



ISO GQL standard 2024
(first since SQL '86)

ISO GQL: A defining moment in the history of database innovation



[Philip Rathle](#)

Chief Technology Officer, Neo4j

April 17, 2024 6 min read

ISO/IEC 39075:2024

Information technology — Database languages — G...

Published (Edition 1, 2024)

Table of Contents

- Foreword
- Introduction
- 1 Scope
- 2 Normative references
- > 3 Terms and definitions
- > 4 Concepts
- > 5 Notation and conventions
- 6 <GQL-program>
- > 7 Session management
- > 8 Transaction management
- > 9 Procedure specification
- > 10 Variable definitions
- > 11 Object expressions

Figures

Tables

① Rendering disclaimer

Languages: en

Search in document

Introduction

This document defines GQL, a database language for modeling structured data as a [graph](#), and for storing, querying, and modifying that data in a [graph](#) database or other [graph](#) store. There are two major [graph](#) data models in current use: the Resource Description Framework (RDF) model and the [Property Graph](#) model. The RDF model has been standardized by W3C in a number of specifications. GQL addresses the [Property Graph](#) model.

[Property graphs](#) organize data as entities called [nodes](#) (or, alternatively, vertices) and [edges](#) (or, alternatively, [relationships](#)). Each [graph element](#) (a [node](#) or an [edge](#)) can have associated [labels](#) and properties. The flexibility and intuitiveness of the data model and its emphasis on interconnections between [graph elements](#) make [property graphs](#) suitable for storing complex knowledge and for analytical tasks such as entity resolution, fraud detection, cyber security, and forecasting.

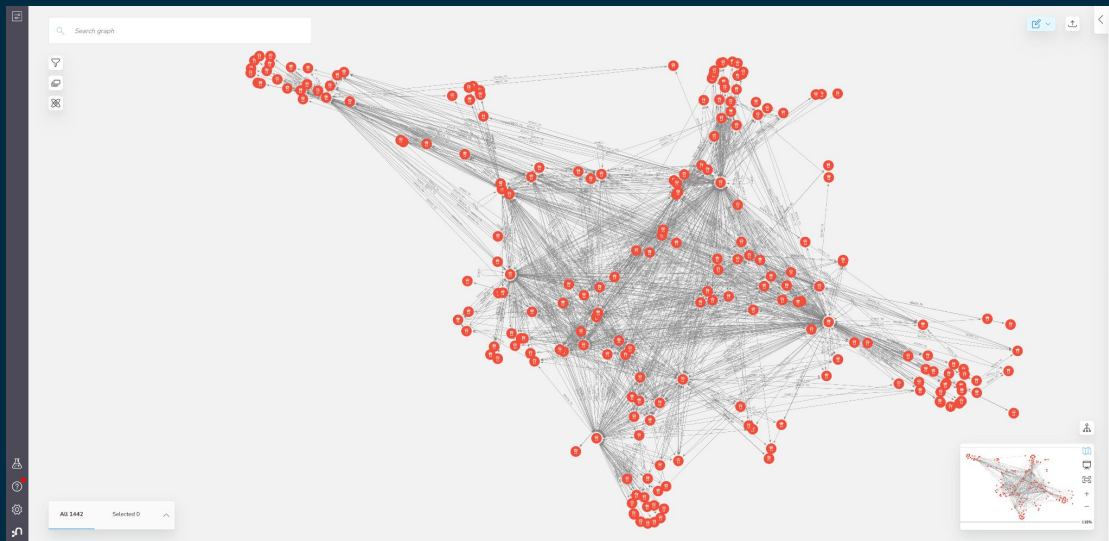
GQL is declarative and transactional, taking inspiration from SQL and from leading independently-developed [property graph](#) languages. [Property graphs](#) select data primarily through [path pattern](#) matching. Defining [path pattern](#) searches in a [graph](#) is often simpler or more flexible than defining the equivalent joins in SQL. The flexible data model, the availability of [path pattern](#) matching, and the efficiency of traversing [edges](#) compared to joining tables have led to increasing interest in [property graph](#) databases.

Various [graph](#) data models have been around for many decades, but it is only since the early 21st century that the demand has driven the rise of commercial [graph](#) database and [graph](#) analytical systems for [property graphs](#).

GQL provides a standard yet flexible common language for this growing market. GQL supports the



Graph Analytics: The Shape of the Data Tells a Unique Set of Stories



Visualizing & Identifying Supply Chain Bottlenecks

Graph Analytics = Graph Algorithms + Parallelized Graph Pattern Matching + Graph Embeddings + Graph ML + Visual Graph Exploration



25% reduction in fraud across 800k credit applications annually



5x faster time to spot incidents. 10% reduction in contestation. \$750M annual economic impact.



40% lower attrition. 13% increase in internal hiring across 8k employees



Near-instant response to zero-day attacks across 100M customer endpoints

Normal Behavior

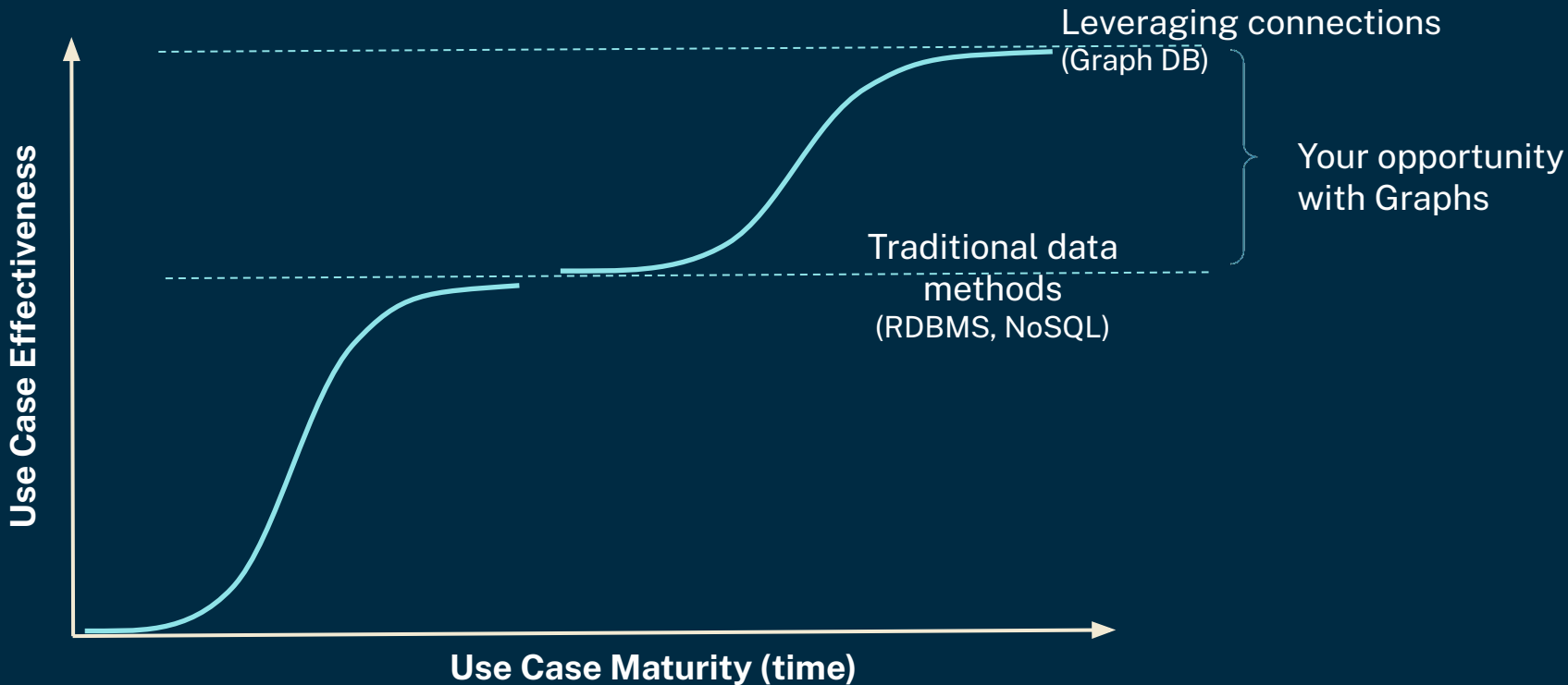


Previously
undetected
fraud.

*Detected 72hrs
into a PoC*

**Value: >\$10M+ in
fraud detected**

Connected Data Unlocks Step Function in Value

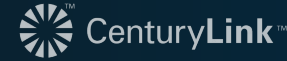




AIRBUS



Boston
Scientific



DAIMLER



Deloitte.



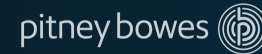
2,100 Customers
Worldwide

intuit.

LEVI STRAUSS & CO.



NORDSTROM

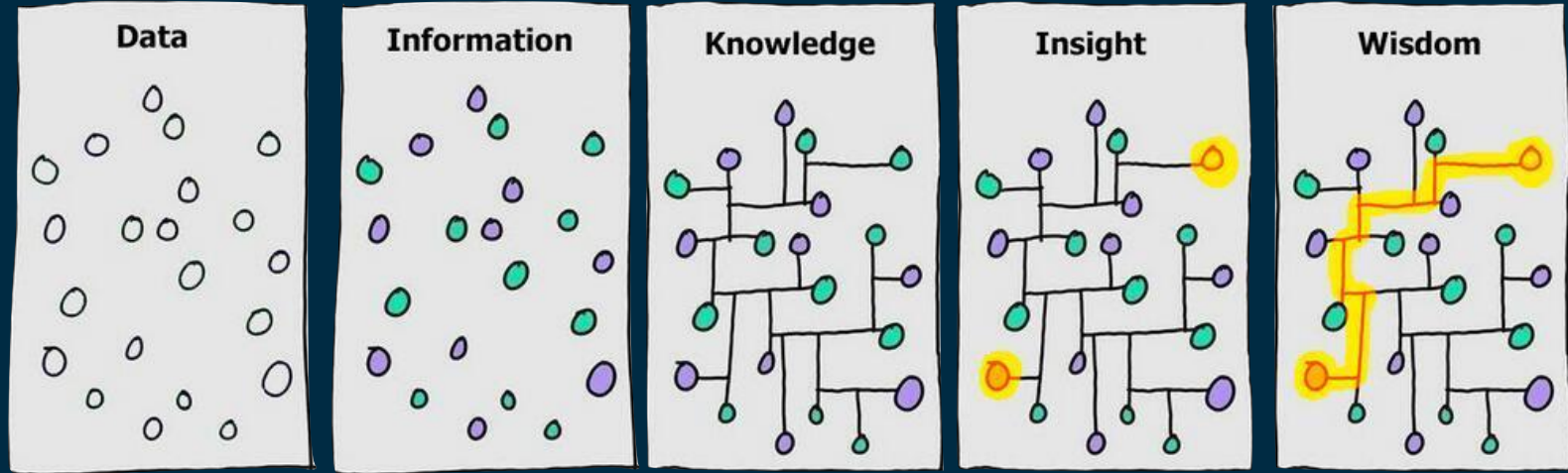


SAMSUNG



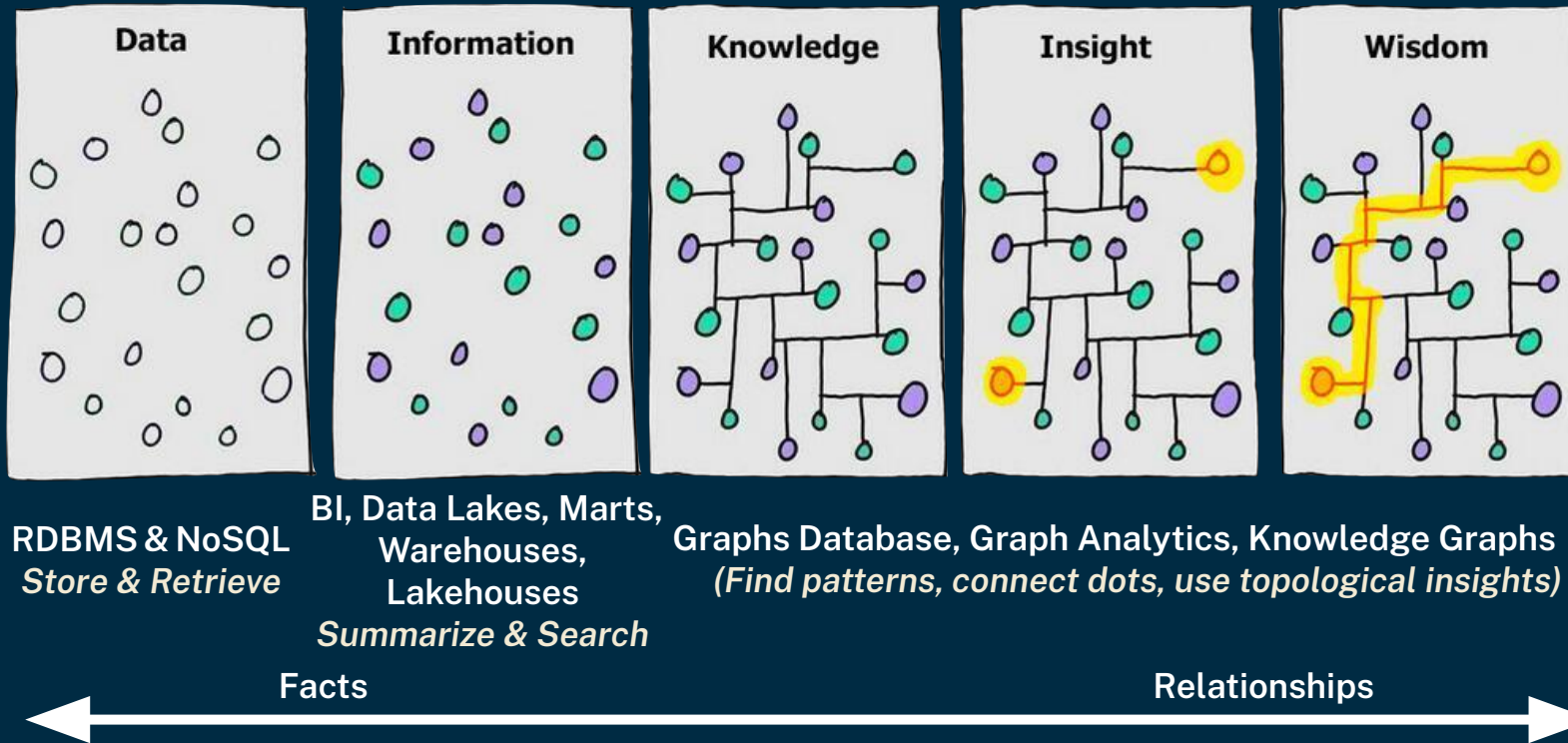


Progression of Human Understanding



Progression of Organizational Understanding

The Database Landscape





Part II: Graphs in the GenAI Era

GraphRAG, Knowledge Graphs,
Context Graphs, Ontology, Semantics,
Memory



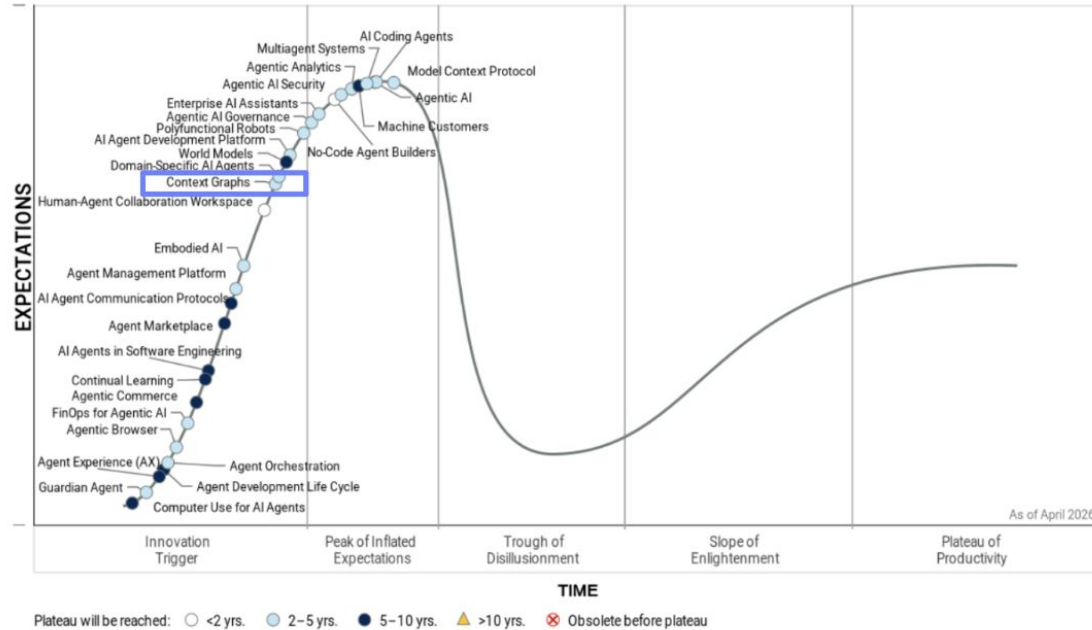
“D&A leaders must redesign the D&A architecture to make the context layer the central brain for AI agents to deliver trusted intelligence.”

- Gartner®

Figure 1: Hype Cycle for Agentic AI

Download Image

Hype Cycle for Agentic AI



AI's Trillion-Dollar Opportunity: Context Graphs

Dec 22, 2025

Forbes

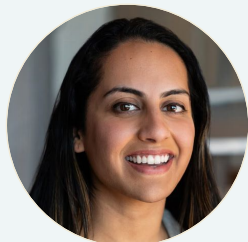
INNOVATION > VENTURE CAPITAL

VCs Say Context Graphs Might Be The Next Big Thing In AI

By **Josipa Majic Predin**, Contributor. G I'm a founder, writer and lecturer focus...

[Follow Author](#)

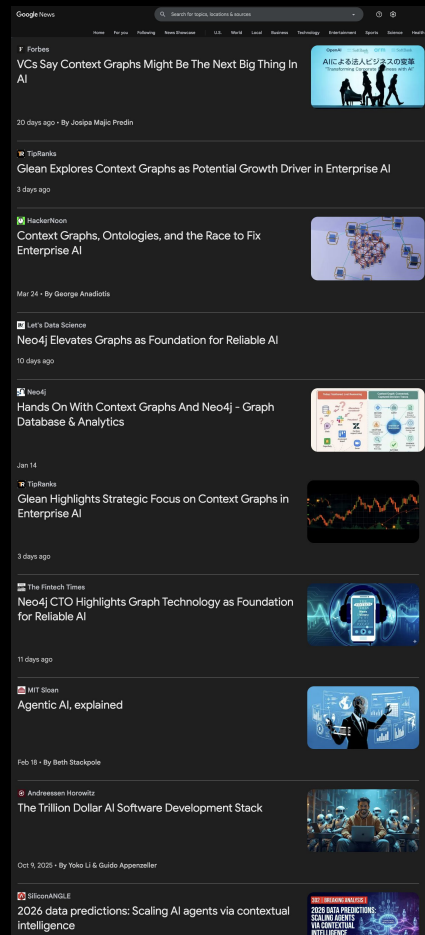
Published Apr 03, 2026, 08:00am EDT



Jaya Gupta



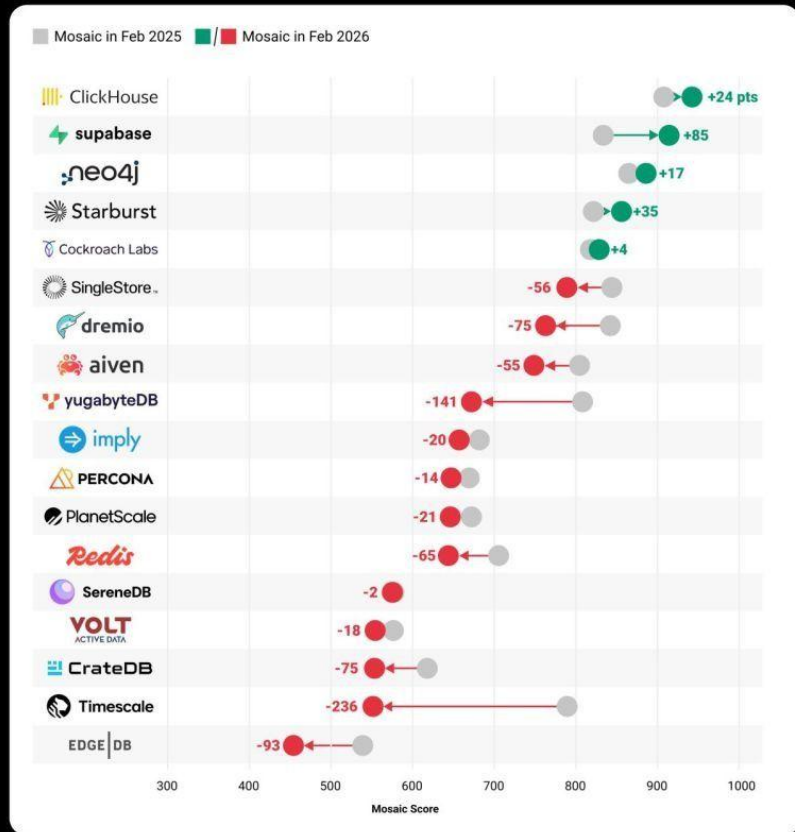
Ashu Garg



Emerging database leaders

One-year mosaic score changes reveal market leaders

CBINSIGHTS



© 2026 CB Insights | Data on 1-year Mosaic score (private company success probability) changes for companies in CB Insights' Relational database management systems and Open-source database markets.

CBINSIGHTS

Neo4j: "Momentum reflects scale and defensibility: \$200M ARR, 84% Fortune 100 penetration, and accelerating GraphRAG adoption, contributing to one of the highest IPO probabilities we track."
- CBInsights



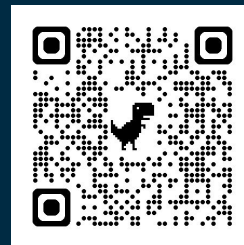
What's the big deal?



95%
AI pilots
fail



Massachusetts
Institute of
Technology





“For high-stakes work,
I need a system that
accumulates
knowledge and
improves over time.”

“Model quality fails
without context.”



Massachusetts
Institute of
Technology



Top Causes of AI Failure



1

AI Hallucinates

LLMs make mistakes – that's unacceptable for mission critical applications

2

AI is a Black Box

Very hard to build trust with stakeholders, regulators, and customers if you can't explain anything about AI decisions

3

LLMs Have No Judgment

Any data is fair game to AI models. No ability to apply privacy, security, or other restrictions on appropriate, legal, and ethical data use

4

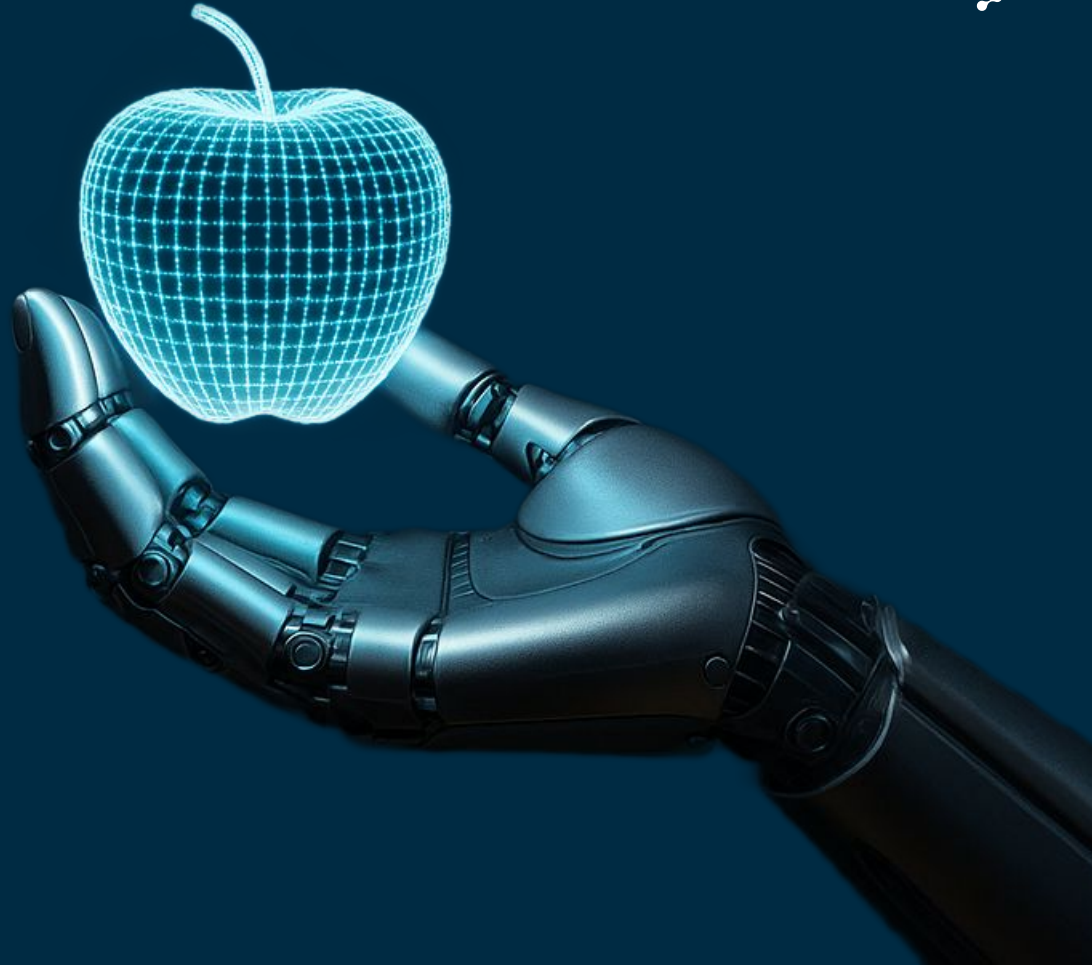
AI Lacks Context

Models don't have access to relevant enterprise context for each decision



Tracing the root cause

What do
you feed
your
agents?





Data $\neq$ Knowledge

I Studied 1,500 Academic Papers on Prompt Engineering. Here's Why Everything You Know Is Wrong.



Aakash Gupta

Follow

7 min read · Aug 17, 2025



Myth 1: Longer, Detailed Prompts Equal Better Results

The most pervasive myth in prompt engineering is that more detailed, longer prompts automatically produce better results. This intuition makes sense — if you're asking a human for help, providing more context and specific instructions typically leads to better outcomes.

But AI models don't work like humans. Research consistently shows that well-structured short prompts often outperform verbose alternatives while reducing costs significantly.



Agents Need Knowledge. Specifically:



1

Real-Time Updates &
Low-Latency Retrievals

2

Efficient Storage of
Structured, Semi-structured
and Unstructured Data

5

Data Access Controls
and Explainability
for Trust,
Transparency
and Compliance

3

Connects Enterprise Data
Within and Across Silos to
Provide Context for AI

4

Stores RAG Corpus and
Agentic Memory in a
Unified Platform

Knowledge layer for AI: the need and the challenge

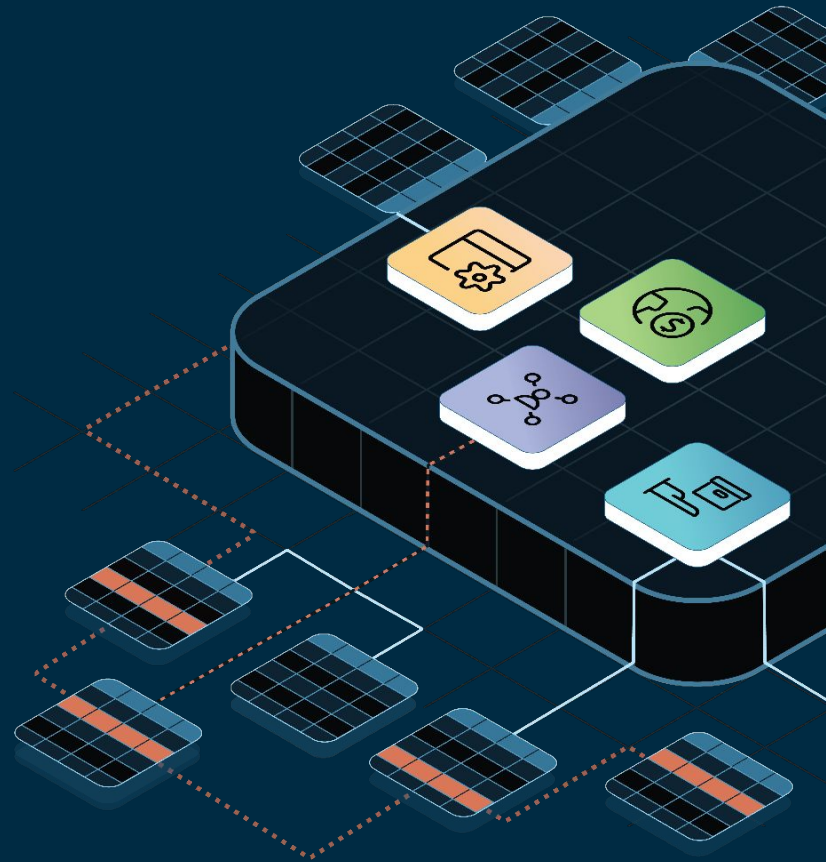
Foundation models trained on the world's data lack specific domain knowledge

Small models trained on specific data lack the latest data, are opaque, lack discernment, and can still hallucinate the results

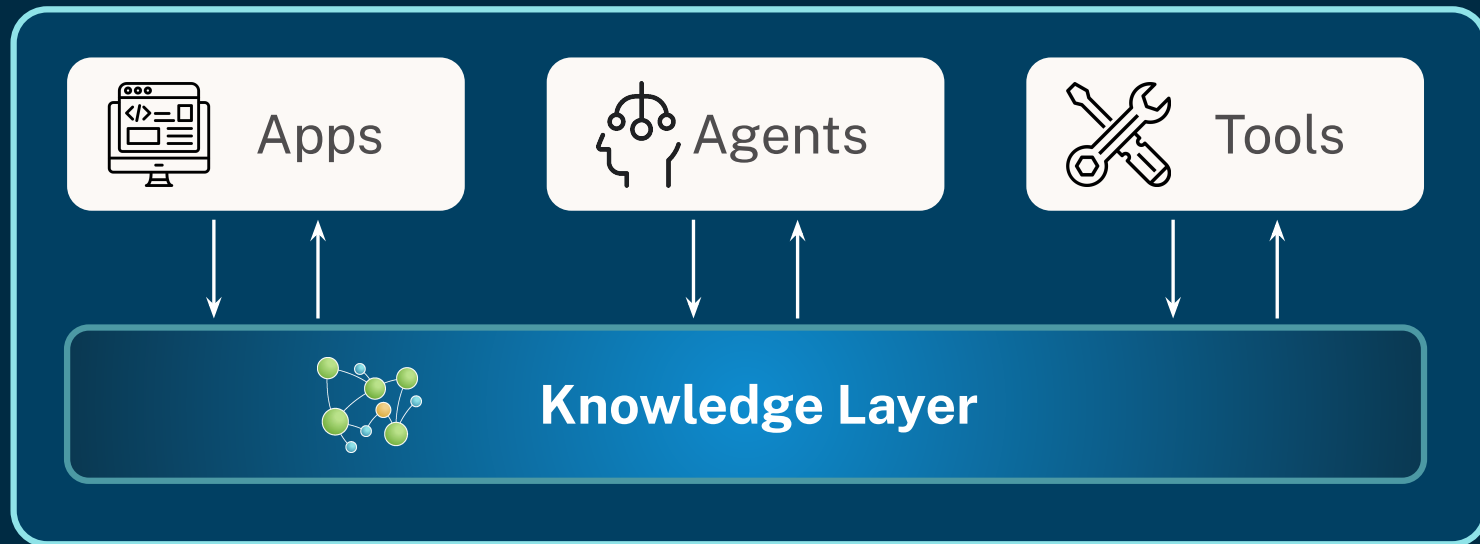
Data structured as static rows and columns is organized to support fixed & simple questions

Data in unstructured form is difficult to mine and make sense of

Organizational data is spread across silos and includes lots of noise in addition to the signal



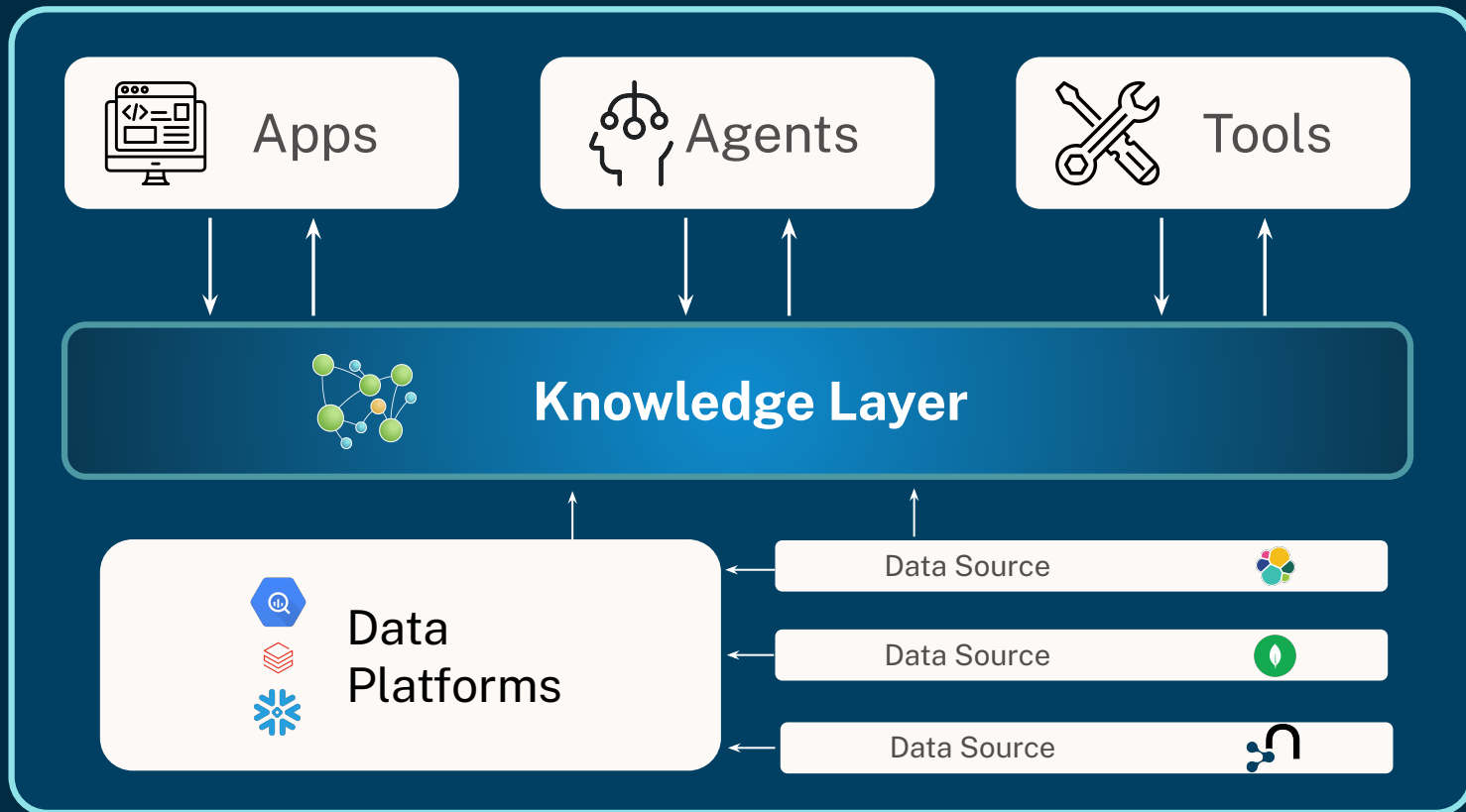
The AI Knowledge Layer



- Holds structure, context and meaning
- Bridges human and machine understanding
- Uniform memory & retrieval layer for agents
- Links and unifies data sources and objects

STORES DATA AS A GRAPH

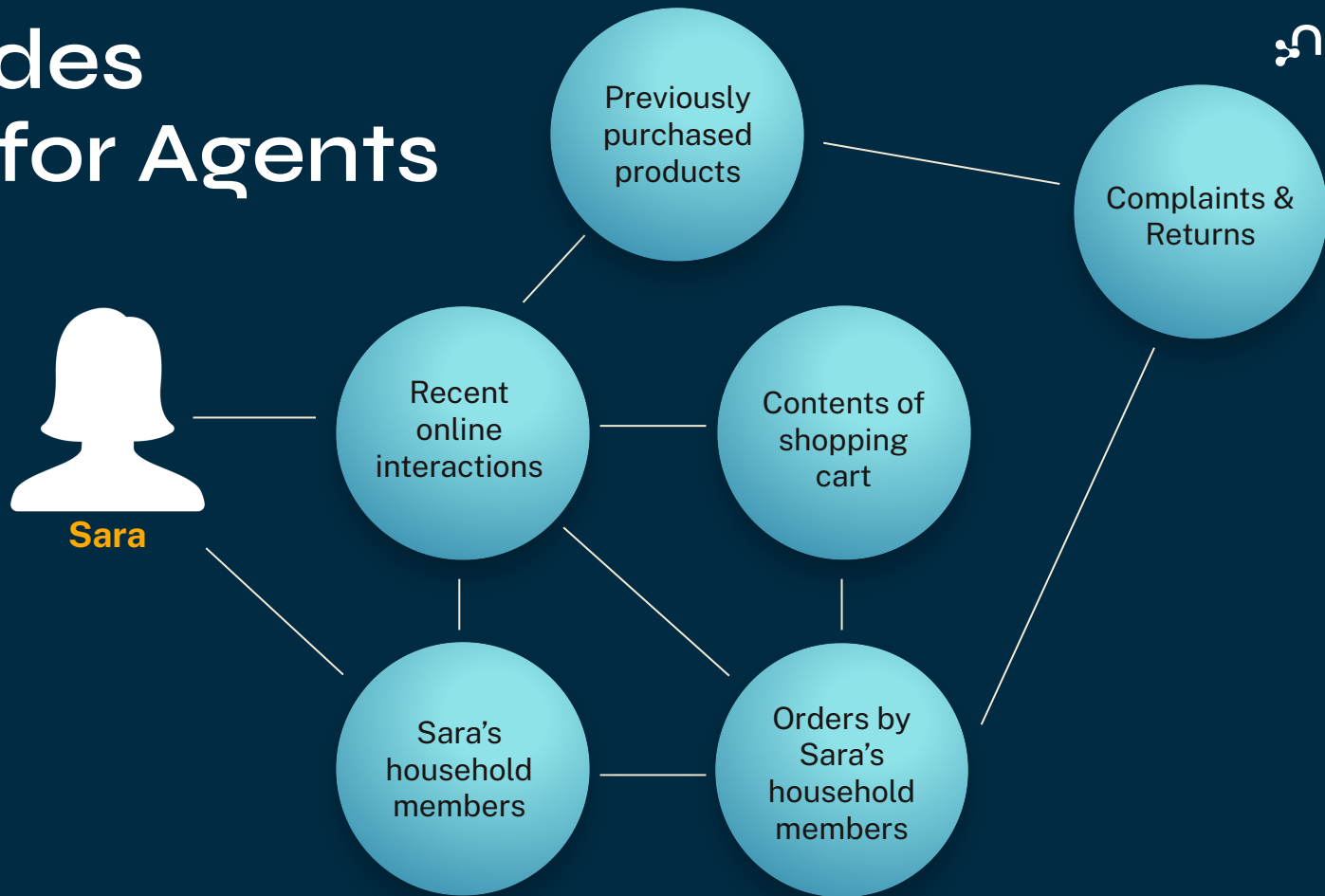
The AI Knowledge Layer



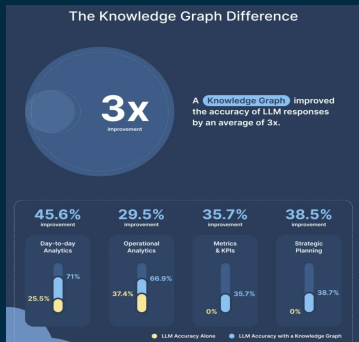
#1: Provides Context for Agents

Agents need to know what real-world topic or actor is being treated

And how it interacts with the world



And therefore Better Answers



GraphRAG with GNN+LLM

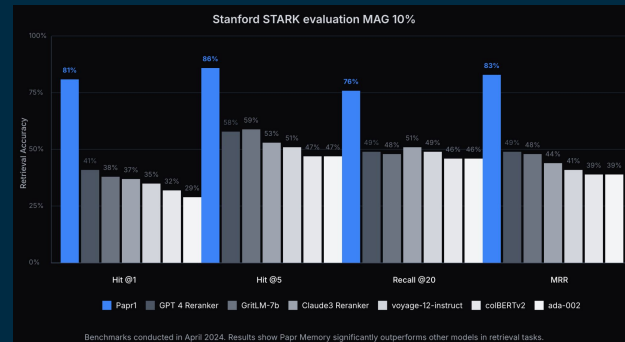
Brian Shi | Alfred Clemedtson | Zach Blumenfeld

Experiments: G-retriever with STaRK-Prime

Method	Hits@1	Hits@5	Recall@20	MRR
Pipeline*	32.09	48.34	47.85	38.48
G-Retriever	32.27 ± 0.3	37.92 ± 0.2	27.16 ± 0.1	34.73 ± 0.3
Subgraph Pruning**	12.60	31.60	35.93	20.96
Baseline	15.57	33.42	39.09	24.11

*Pipeline adds nodes in subgraph context to G-retriever output.

**Frozen LLM. No fine-tuning.



AdaptJobRec: Enhancing Conversational Career Recommendation through an LLM-Powered Agentic System

Qixin Wang^{1,2}, Dawei Wang¹, Kun Chen¹, Yaowei Hu¹
 Puneet Girdhar¹, Ruoteng Wang¹, Aadesh Gupta¹, Chaitanya Devella¹
 Wenlai Guo¹, Shangwen Huang¹, Bachir Aoun¹, Greg Hayworth¹
 Han Li^{1*}, Xintao Wu^{2*}

¹Walmart Global Tech
 han.li@walmart.com

²University of Arkansas
 xintaowu@uark.edu

Methods	Hit@10 ↑	NDCG@10 ↑	MAP@10 ↑	Methods	Conv Round ↓	Resp Latency (ms) ↓
RAG	0.2000	0.0391	0.0136	RAG	7.10	1065
ReAct	0.2789	0.0747	0.0342	Plan & Execute	6.38	957
Plan & Execute	0.3127	0.0799	0.0364	ReAct	3.86	579
MACRS	0.3120	0.0775	0.0350	MACRS	3.68	552
AdaptJobRec	0.3176	0.0810	0.0371	AdaptJobRec	3.32	498

Table 1: Comparison on Job Recommendation Task

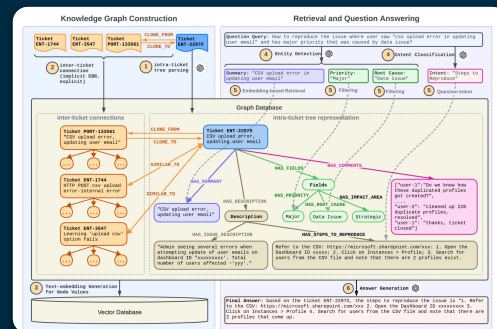
Table 5: Comparison of Average Conversation Rounds and Response Latency for the Pilot Group

Linked in

GraphRAG reduces median per-issue resolution time by

28.6%

for LinkedIn's customer service team



GenAI

Knowledge Graph

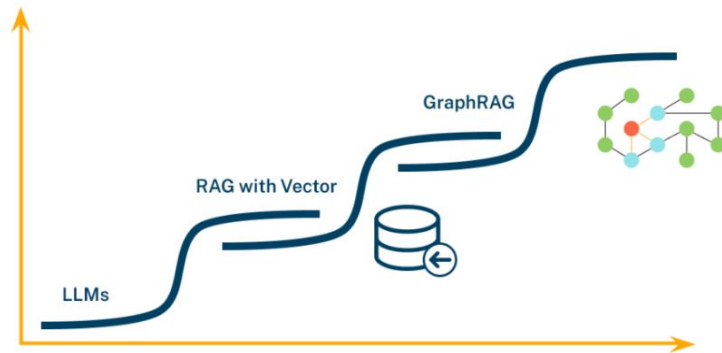
The GraphRAG Manifesto: Adding Knowledge to GenAI



[Philip Rathle](#)

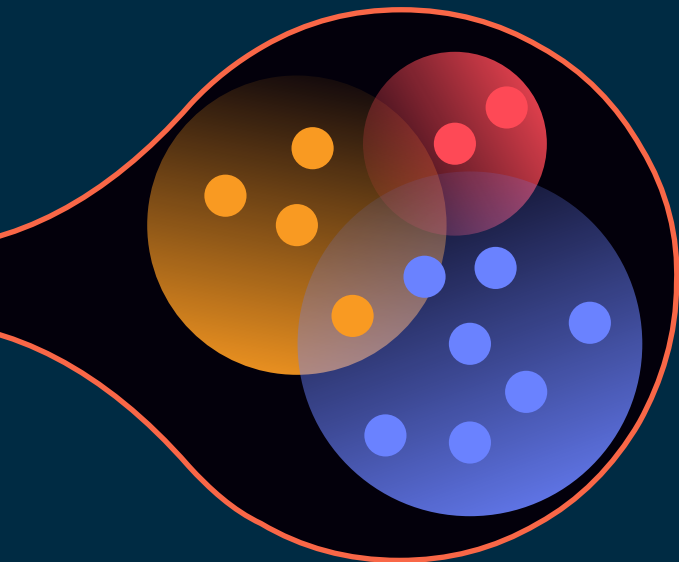
Chief Technology Officer, Neo4j

July 11, 2024 21 min read



RAG

Vector only approach



**Better accuracy though
concept matching**

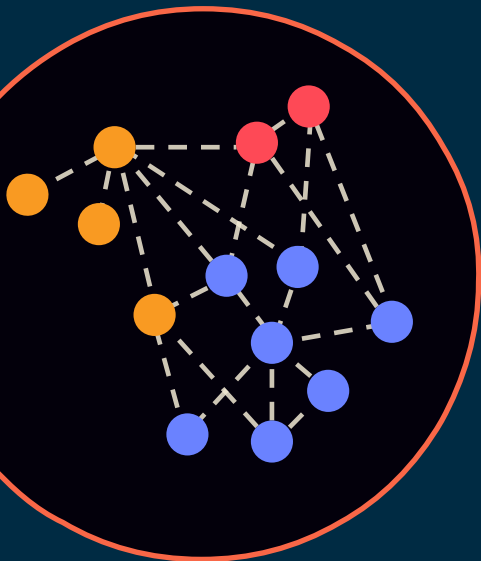
Which asset managers are vulnerable to lithium shortage?

Based on the context provided, here is a summary of the asset managers that may be vulnerable to lithium storage:

- The company that produces batteries and other products incorporating battery technology is vulnerable, as increased demand for batteries could place capacity constraints on their lithium supply chain.
- The company that sources steel and copper from Chile and Australia is vulnerable, as those countries are expected to experience severe weather due to climate change.
- The company that relies on suppliers for components and systems necessary for their products has some vulnerability. If there are lithium shortages it could result in supply constraints, cost pressures, or component shortages.
- The company uses lithium-ion batteries or other battery technology in their products is vulnerable to lithium shortages which could reduce battery supply, increase costs, or otherwise negatively impact their business.

GraphRAG

Vector + Graph



**Knowledge graph yields
more complete &
actionable answers**

Which asset managers are vulnerable to lithium shortage?

Based on the context provided, the asset managers that may be vulnerable to lithium shortage are:

- Quantamental Technologies LLC
- DEPRINCE RACE & ZOLLO INC
- ATRIA Wealth Solutions, Inc.
- CITIGROUP INC
- CHARLES SCHWAB INVESTMENT MANAGEMENT INC
- ROYAL LONDON ASSET MANAGEMENT LTD
- STIFEL FINANCIAL CORP
- TUCKER ASSET MANAGEMENT, LLC
- Patriot Financial Group Insurance Agency, LLC
- SeaCrest Wealth Management, LLC

The context indicates that Stanley Black & Decker uses lithium batteries in many of its products. With increasing demand for lithium, there could be supply constraints and increased costs for lithium. This could negatively impact the profitability of Stanley Black & Decker. The asset managers listed above have holdings in Stanley Black & Decker, and therefore may be vulnerable to the risks of lithium shortage

Reducing Hallucinations in Complex Question Answering using Simple Graph-based Retrieval-Augmented Generation (long version)

Christopher J. Wedge, Joshua Stutter, Danny Dixon, Jacek Cala
National Innovation Centre for Data, Newcastle University
Newcastle upon Tyne, UK
{Chris.Wedge, Joshua.Stutter, D.J.Dixon2, Jacek.Cala}@ncl.ac.uk

ABSTRACT

Large language models (LLMs) have fundamentally transformed the landscape of Natural Language Processing (NLP), although they remain susceptible to errors. Retrieval-augmented generation (RAG) systems have emerged as a common deployment scenario seeking to both avoid the well known risk of the LLM “hallucinating” information, and to enable reasoning and question answering over proprietary information that the LLM did not have access to during training without resorting to expensive model fine-tuning.

In this work, we explore the idea of using a lightweight graph structure with a relatively simple graph schema, to support the RAG subsystem via a dedicated toolset. We design an agentic system with a variety of vector search and graph query tools operating over a structured dataset based on a curated subset of English Wikipedia articles, and evaluate its performance on questions from MoNaCo, a challenging Wikipedia based benchmark of complex question answering (QA) tasks.

Our results show that the introduction of graph-based tools can significantly increase the precision and recall of factual correctness, can halve the number of hallucinated answers, and achieves the highest fine-grained truthfulness score among the three evaluated scenarios. All this with a modest increase in token usage.

1 INTRODUCTION

Since the advent of the transformer architecture many competing Large language models (LLMs) have been developed with strong capabilities in understanding, reasoning, summarisation, and ques-

omissions, missed relevant context, or failure to invoke the correct tools. In practice, such errors often occur simultaneously in multiple stages of system operation [12]. Therefore, any method that improves the quality of a model or a model-based system is critical, insofar as it reduces both types of error simultaneously rather than merely shifting performance along the precision-recall tradeoff. This consideration becomes even more vital for complex question answering problems, such as multi-hop (MHQA), cross-document (CDQA) and multi-entity question answering (MEQA) [5, 38, 57], where knowledge from multiple sources must be retrieved, aggregated, and reasoned over in order to provide an answer.

As an example, consider the following question:

Can you name all the battles between the Dutch and English in the First, Second and Third Anglo-Dutch Wars, and list the victor of each battle?

To answer this question, one needs to perform a sophisticated retrieval and reasoning process. In fact, it requires multi-entity and multi-hop reasoning, and cross-document access, all at once [50].

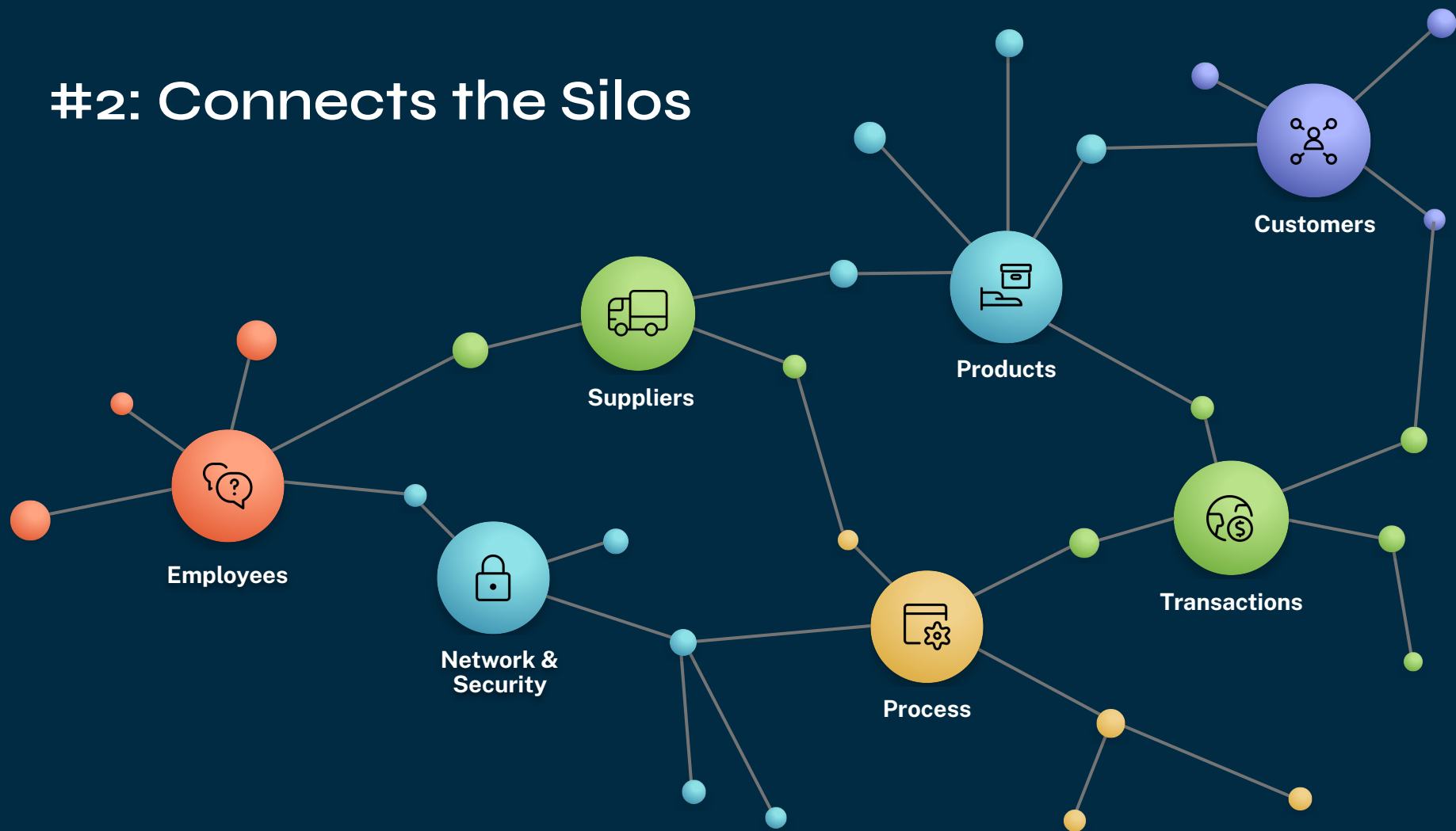
First, as with MEQA, the process must operate on a set of entities (in this case nations, wars, battles and victors) and needs to perform per-entity analysis and filtering. Second, as with MHQA, it requires hierarchical traversal, repeated relational expansion and nested decomposition. Finally, as with CDQA, it needs to fetch information from multiple sources: war documents, battle descriptions, and historical records. A single document is unlikely to answer this question; the information required is typically dispersed across dozens of documents.

These types of questions pose a significant challenge to current state-of-the-art LLM-based systems. There are multiple reasons for

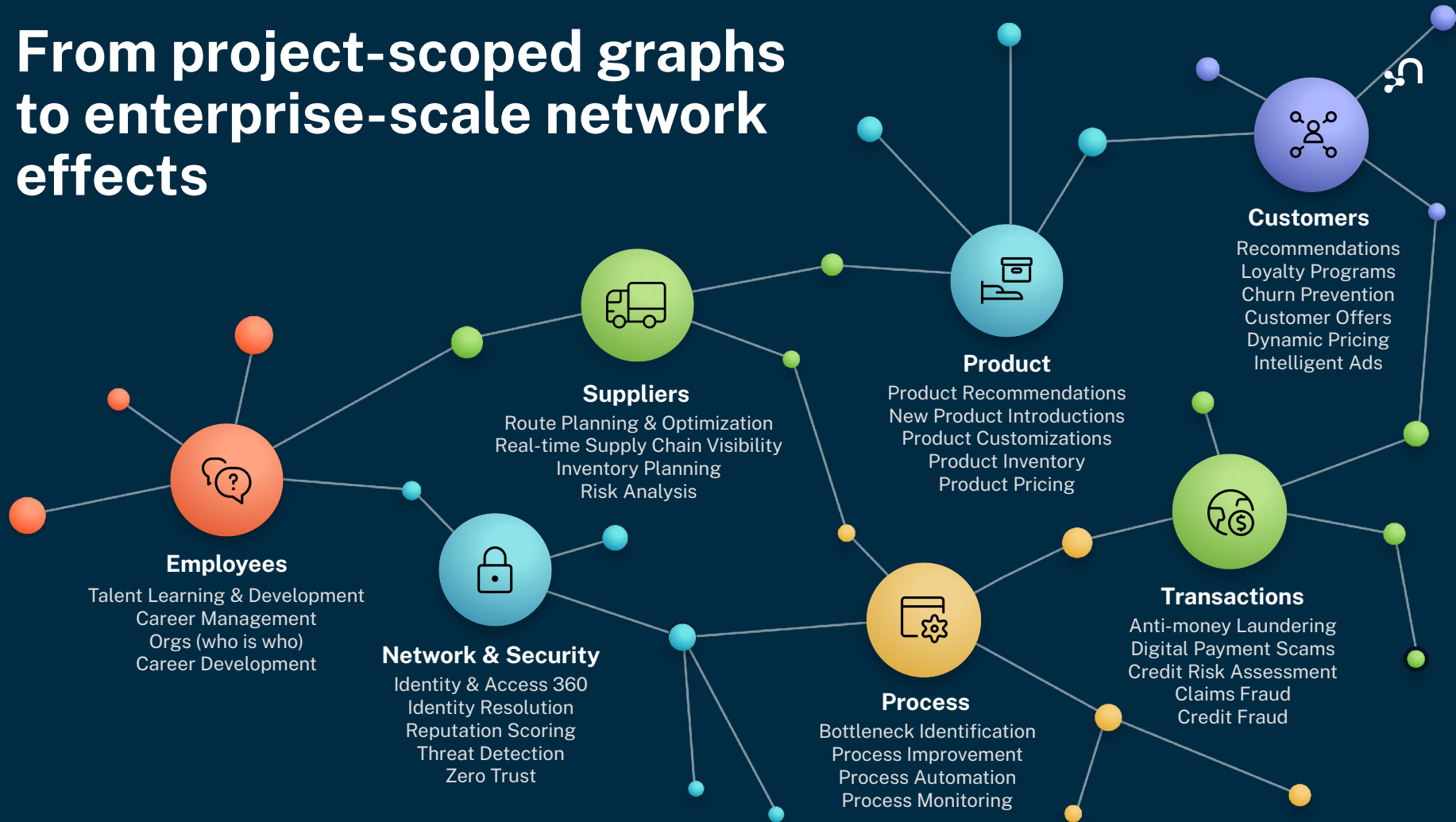
Agents using GraphRAG vs. vector-only RAG

- 80% more truthful
- 2x more questions answered
- Fewer hallucinations
- More efficient token use

#2: Connects the Silos

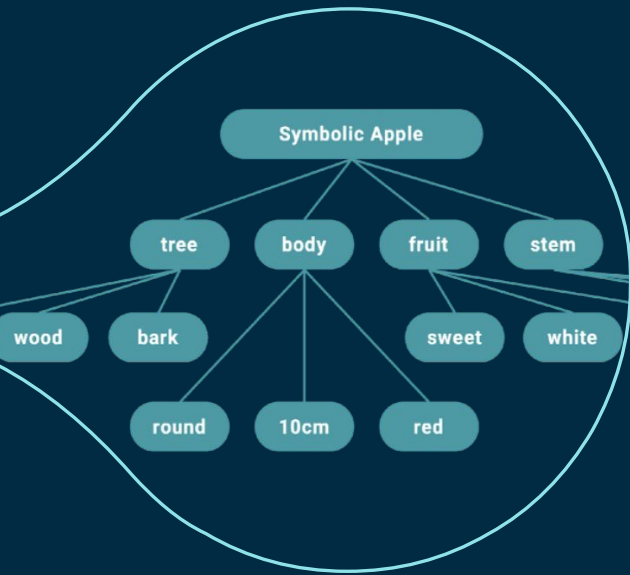


From project-scoped graphs to enterprise-scale network effects

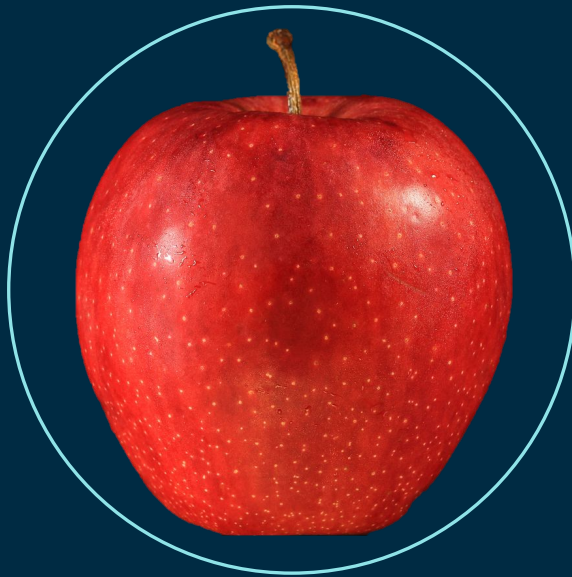


#3: Provides Knowledge Transparency and (if you want it) Deterministic Reasoning

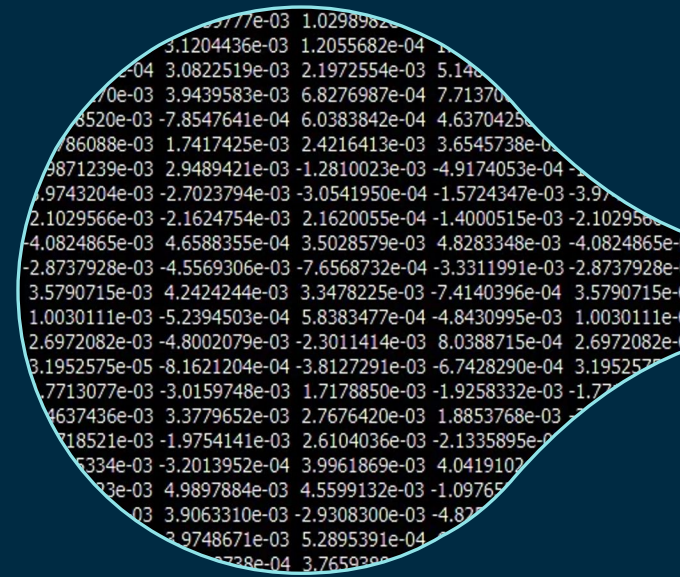
Knowledge Graph View
of an Apple



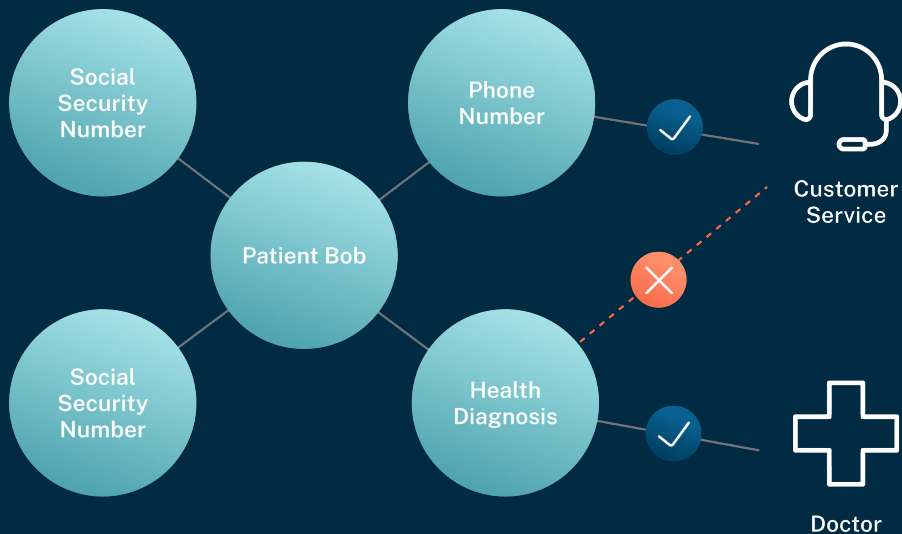
Human View
of an Apple



Vector View
of an Apple



#4: Data Protection & Discernment



"Neo4j enables us to enforce robust information security by applying access controls at the subgraph level."

-Eric Costantini, Chief Business Officer at Data^2.

#5: Handling of Complex Questions



Plan: To determine the number of unique devices of the top 5 most active users with high login records in the top country code with the highest number of IP usage, we need to follow these steps:

1. Identify the country code with the highest number of IP usage.
2. Identify the top 5 users with the highest number of login records in that country.
3. Retrieve the unique devices associated with these users.

Let's start by identifying the country code with the highest number of IP usage.

Question: What is the number of unique devices used by the top 5 most active users with the highest login records in the country with the highest number of IP usages?

Agentic Query Generation:

“Cypher >> SQL for Complex Queries”



Problem	Correctness	Comprehension	Open Challenges & Plans
SQL vs Cypher			
<i>“Find missing documents that are required by the trigger but are not currently connected to the onboarding graph”</i>			
With Neo4J		With MySQL	
<pre>MATCH (og @DatabaseName)-[:REQUIRES_TRIGGER]-(tr @DatabaseName)-[:EVALUATES]->(trigger @DatabaseName) WHERE trigger.trigger_required_expression CONTAINS '({{trigger_required_expression}})' WITH og, trigger [doc ON apoc.text.contains(trigger.trigger_required_expression, '{{trigger_required_expression}}') doc[1]] AS required_docs MATCH (og)-[:DOCUMENTS]->(doc_type @DatabaseName) WHERE doc_type.document_type_name = 'required_docs' WITH og, trigger, required_docs, COLLECT(doc_type.document_type_name) AS connected_docs RETURN og AS onboard_graph_id, trigger AS trigger_id, required_docs AS required_docs, connected_docs AS connected_docs, apoc.coll.subtract(required_docs, connected_docs) AS MissingDocuments</pre>		<pre># -- Step 1: Identify relevant triggers and extract required document IDs WITH (SELECT og_id AS onboard_graph_id, tr_id AS trg_trigger_id, trd_document_type AS required_document_type FROM onboard_graph_id INNER JOIN requires_trigger tr ON og_id = tr.onboard_graph_id INNER JOIN trg_trigger trd ON tr.trg_trigger_id = trd.trg_trigger_id INNER JOIN evaluates ev ON trd.trg_trigger_id = ev.trg_trigger_id INNER JOIN trg_trigger trd ON ev.trg_trigger_id = trd.trg_trigger_id INNER JOIN trigger_required_documents trd ON trd.trg_trigger_id = trd.trg_trigger_id WHERE tr.trigger_required_expression LIKE '{{trigger_required_expression}}') AS tr # -- Step 2: Find connected document types SELECT og_id AS onboard_graph_id, tr.onboard_graph_id, tr.trg_trigger_id, trd_document_type AS required_document_type FROM onboard_graph_id INNER JOIN requires_document_type rdt ON rdt.onboard_graph_id = og_id INNER JOIN document_type dt ON rdt.document_type_id = dt.id WHERE dt.document_type_name = tr.required_document_type # -- Step 3: Aggregate results and identify missing documents SELECT og_id AS onboard_graph_id, tr.onboard_graph_id, tr.trg_trigger_id, trd_document_type AS required_document_type, GROUP_CONCAT(DISTINCT trd_document_type) AS MissingDocuments FROM onboard_graph_id INNER JOIN requires_document_type rdt ON rdt.onboard_graph_id = og_id INNER JOIN trg_trigger trd ON tr.trg_trigger_id = trd.trg_trigger_id INNER JOIN document_type dt ON rdt.document_type_id = dt.id INNER JOIN trigger_required_documents trd ON trd.trg_trigger_id = trd.trg_trigger_id GROUP BY og_id, tr.onboard_graph_id, tr.trg_trigger_id, trd_document_type</pre>	

Role of the AI Knowledge Layer

- 1 Provides Context for Agents, for Better Answers
- 2 Connects the Silos
- 3 Knowledge Transparency & Option for Determinism
- 4 Data Protection for AI Discernment
- 5 Handling of Complex Questions

Top Enterprise AI Challenges



1 AI Hallucinates

LLMs make mistakes. That's unacceptable for many mission-critical applications.



2 AI is a Black Box

Very hard to build trust with stakeholders, regulators, and customers if you can't explain anything about AI decisions



3 LLMs Have No Discernment

Any data is fair game to AI models. No ability to apply privacy, security, or other restrictions on appropriate, legal, and ethical data use



4 AI Lacks Context

Models don't have access to relevant enterprise context for each decision







Deeper dive into the knowledge layer

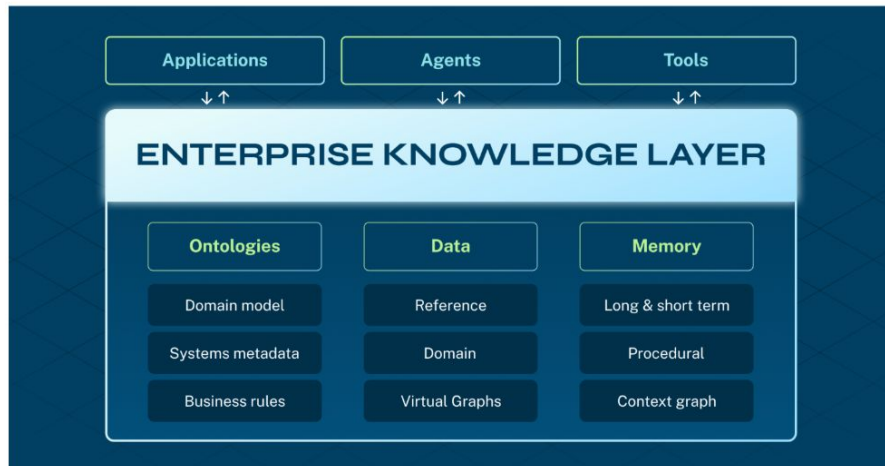
Agentic AI

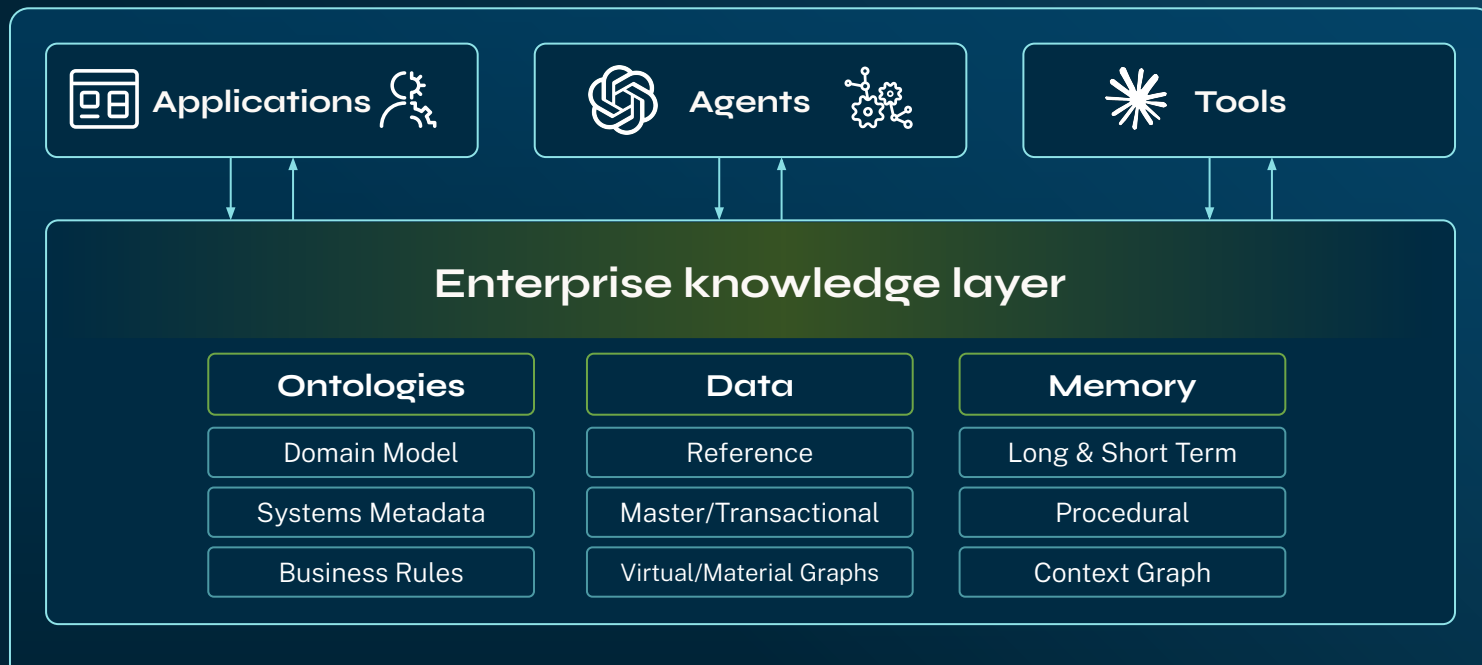
The Enterprise Knowledge Layer

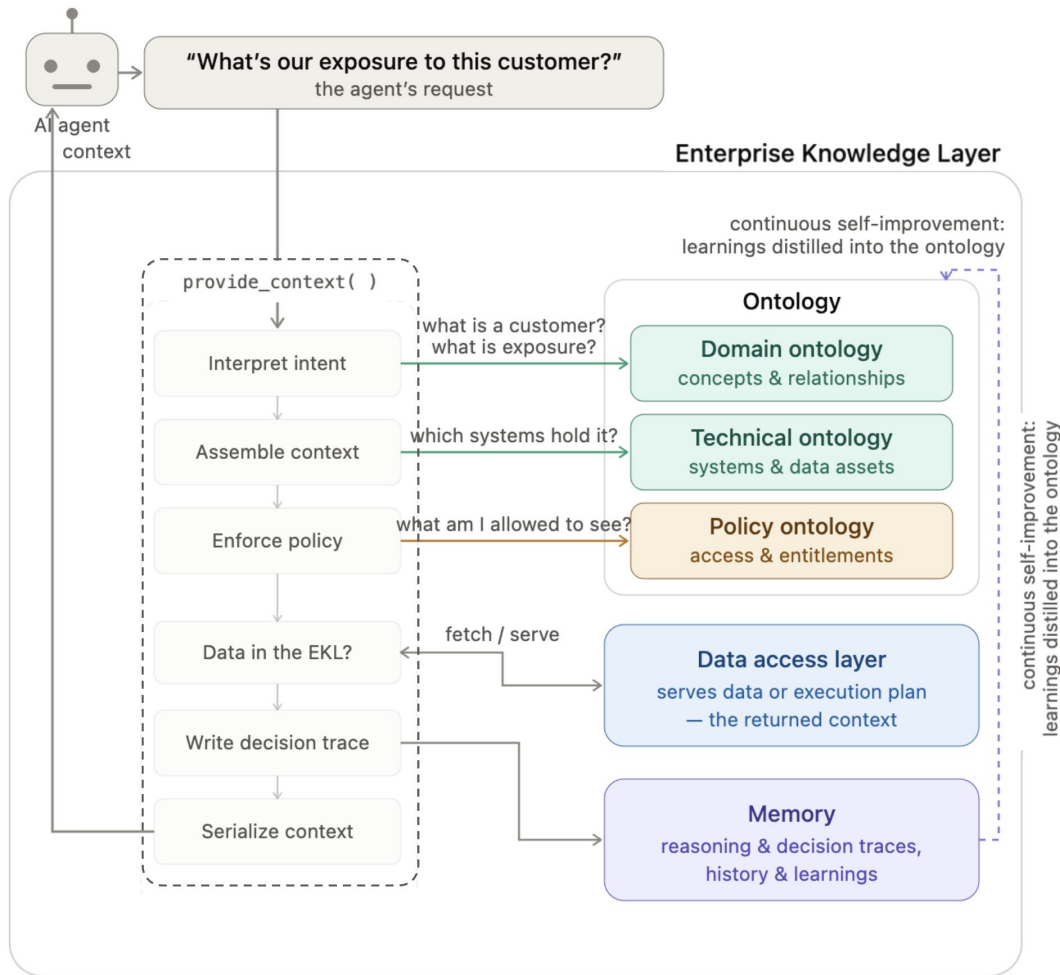
[Jesús Barrasa](#)

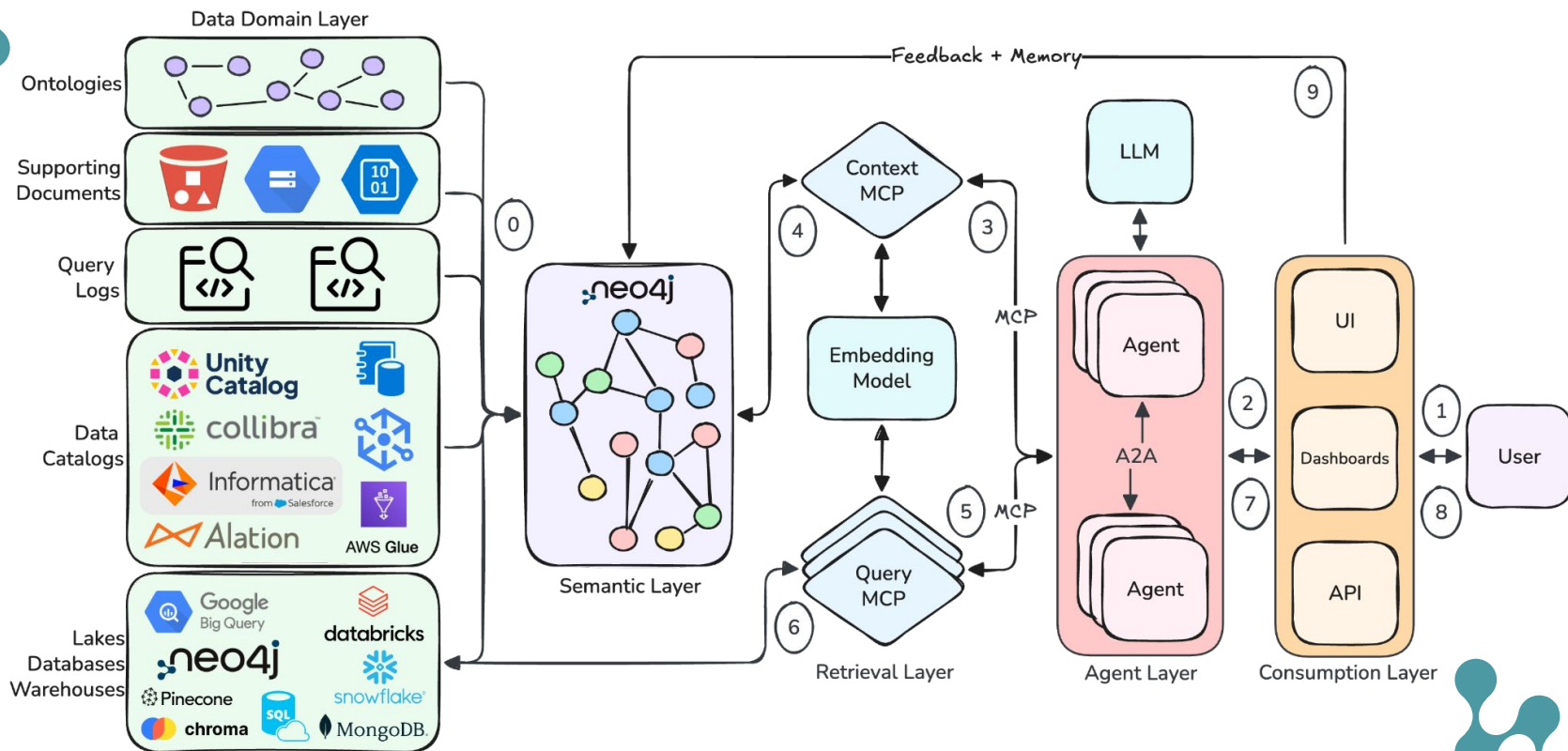
AI Field CTO, Neo4j

July 20, 2026 18 min read









Graph Queries & Computation: Fundamentals

Graph Algorithms

Knowledge enhancements
computed from graph topology
*Centrality, Clustering,
Link prediction, et al.*



Graph Querying

- **Transactional** - “spider” through the graph from a known starting point to return context or a “connect the dots” answer
- **Analytic** - match for patterns across a large graph corpus (e.g. “*find money laundering patterns*”)

GNNs

Topological vectors:
multidimensional vector
array representing the
shape of the subgraph
surrounding a node



Similarity & Search

String/number, fulltext,
semantic search (aka vector
similarity), geo coordinates,
etc.

Graph Visualization

(For the humans)
Entity-link navigation, AI-generated
dashboards, etc.

Graph Intelligence Platform

Graph AI

Aura Agents

Build multi hop agents in seconds.



Agentic Ecosystem

Integrate: AWS, GCP, Azure, LangChain, ServiceNow, Salesforce, Databricks & Snowflake



AI FRAMEWORKS



AGENT FRAMEWORKS

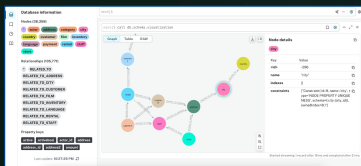
Agentic Brain

Provides foundational services like memory, context-graph to power all AI platforms.

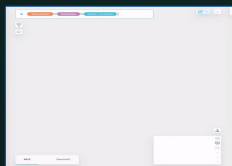


AI Powered Graph Tools

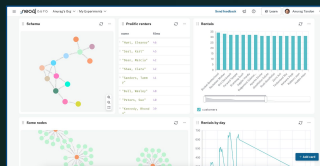
Browser (Query)



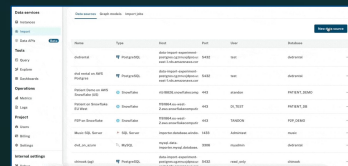
Bloom (Explore)



NeoDash(Dashboards)



Data & Document Import



GraphQL



Database & Graph Algorithms

Database

Intelligent Operational Database

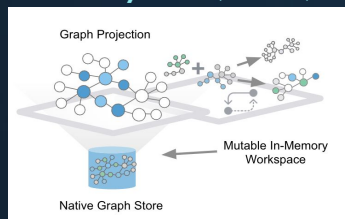
Better Predictions for Analytics



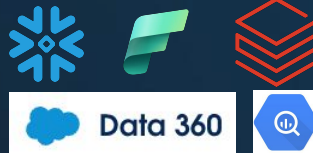
Graph Transactions, Storage & Querying

Graph Analytics, ML & Data Science

Analytics (GDS)



Graph Engine (Virtual Graph)



neo4j
graphtalk
San Francisco