neo4j

Hands on with GraphRAG

Hi, ich bin Martin 👋

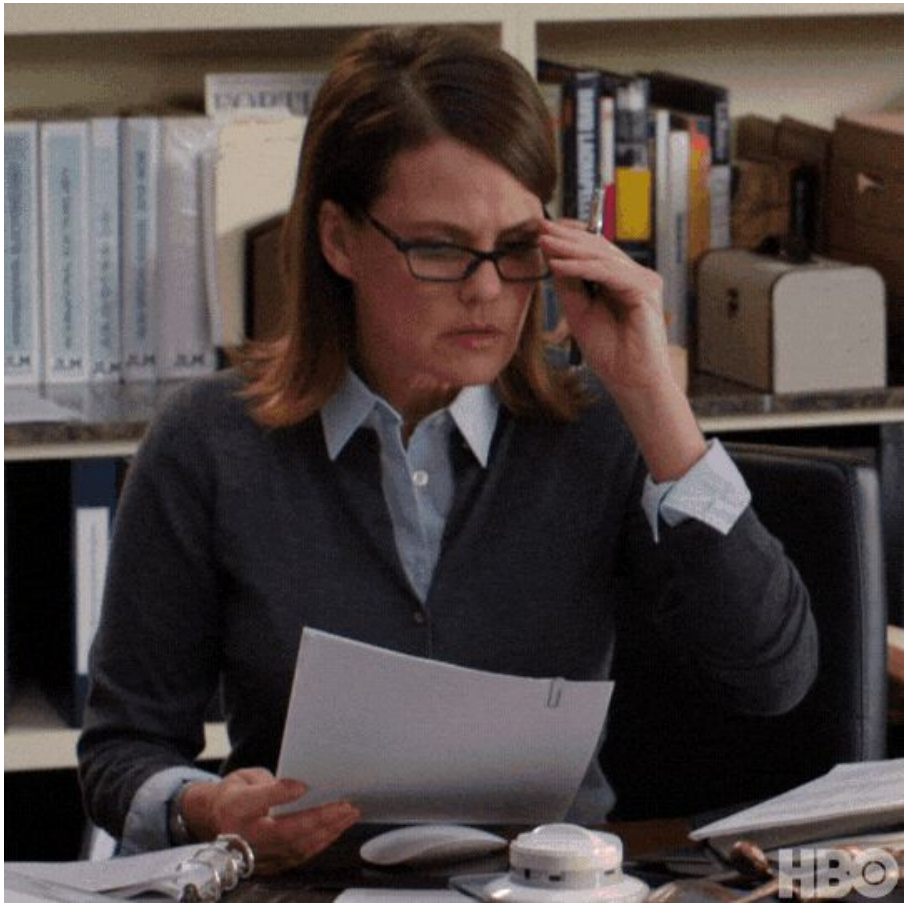
@martinohanlon

Technical Curriculum Developer at Neo4j

Graphs and GenAI

LLMs aren’t perfect!

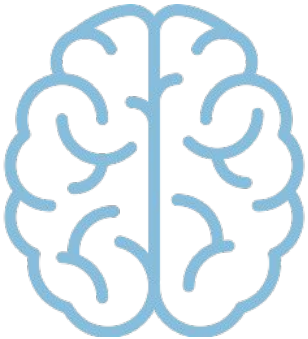

CHATGPT, DUDE.


HBO

Knowledge

Context

Enrichment



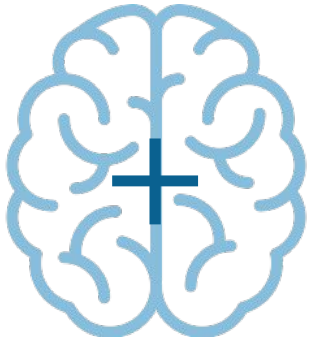
Language

Reasoning

Creativity

Knowledge
Context
Enrichment

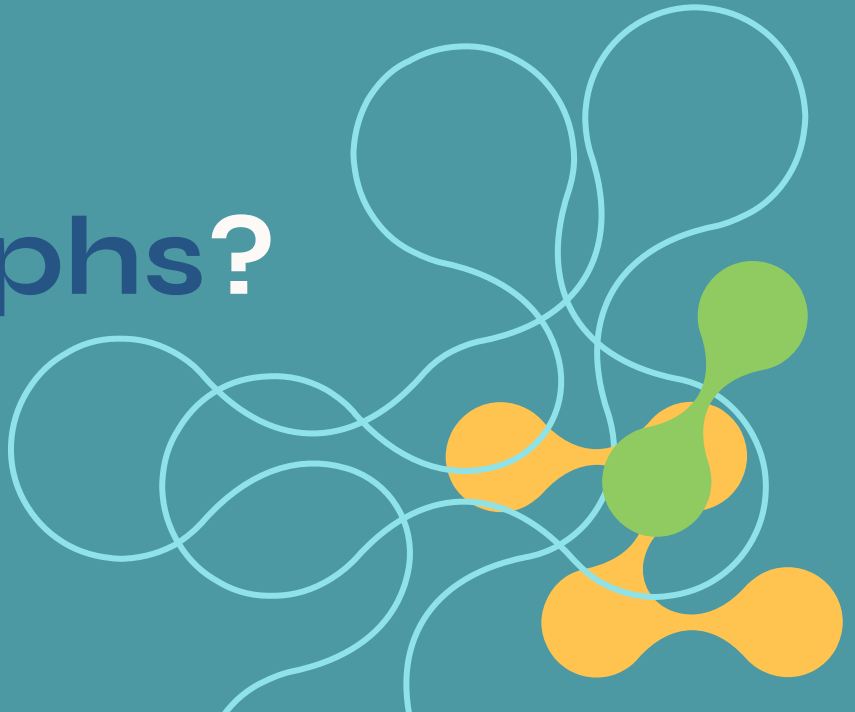
KGs + LLMs



Language
Reasoning
Creativity

BETTER TOGETHER

What are Graphs?



graphs - Google Search

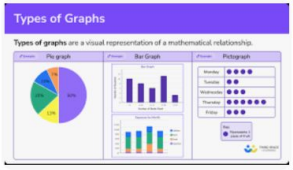
google.com/search?sca_esv=a7dd94b2e88221ce&sca_upv=1&rlz=1C5GCEM_en&sxsrf=ADLYWILO-Kh6XfeDAmQugUguqGhFRfoivA:1717428819204&q=graphs&uds=ADvngMjdDqq...

Google

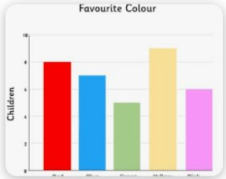
graphs

All | Images | Videos | News | Maps | More | Tools

Math | Data | Line | Bar | Statistics | Blank | Chart | Printable | Function

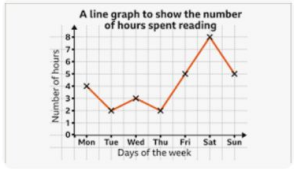


Third Space Learning
Types of Graphs - Math Steps, Examples ...

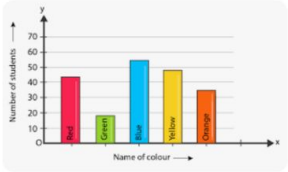


Twinkl
What is a Graph? - Maths - Twinkl

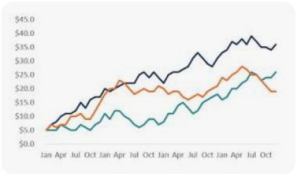
Cambridge Mathematics
Cambridge Mathematics



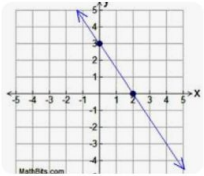
BBC
Line graphs - KS3 Maths - BBC Bitesize



BYJU'S
of Graphs in Mathematics and Statistics ...

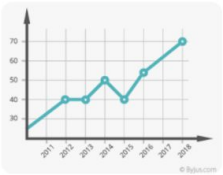


Corporate Finance Institute
Types of Graphs



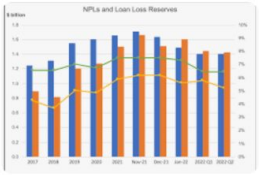
MathBitsNotebook
Interpreting Graphs - MathBit...

Indeed
13 Types of Graphs and Charts (Plus ...

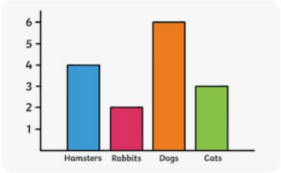


BYJU'S
of Graphs in Mathematics and S...

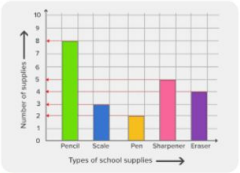
Carolina Knowledge ...
Graphs and Charts



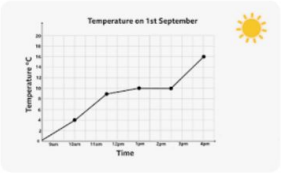
storytelling with data
verbalizing and explaining charts ...



Twinkl
What is a graph? - Twinkl



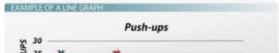
SplashLearn
What Is a Graph in Math? Definitio...



BBC
Introducing line graphs - Maths ...

SplashLearn
What Is a Graph in Math? Definition ...

ChartExpo
Charts and Graphs for Data Visualization



graph theory - Google Search

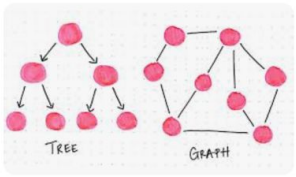
google.com/search?q=graph+theory&sca_esv=a7dd94b2e88221ce&sca_upv=1&rlz=1C5GCEM_en&udm=2&biw=1578&bih=1178&sxsrf=ADLYWIIxiNSux1OcG0VFyelLR4YjhknYZg%3...

Google

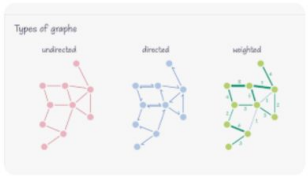
graph theory

All | Images | Books | Videos | News | More | Tools

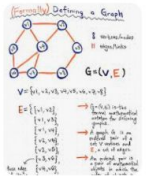
Computer science | Discrete mathematics | Path | Brain | Tree | Simple | Social network | Connectivity



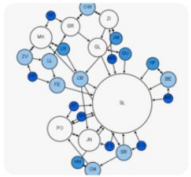
Medium
A Gentle Introduction To Graph Theory ...



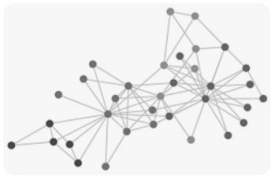
Science in the News - Harvard University
Graph Theory 101 - Science in the News



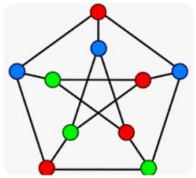
Medium
A Gentle Introducti...



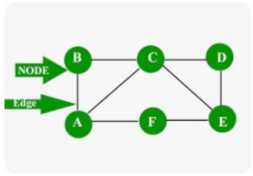
Wikipedia
Graph theory - Wikipedia



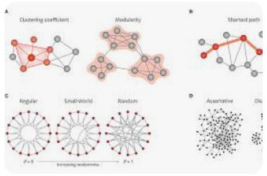
Adatis
Business Insights with Graph Theory ...



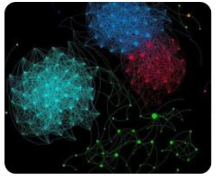
Brilliant
Graph Theory | Brilliant Mat...



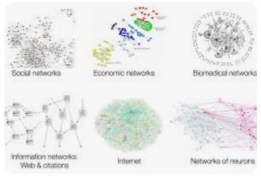
GeeksforGeeks
Mathematics | Graph Theory Basics ...



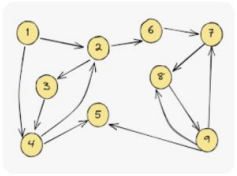
UCF College of Sciences
Dr. Lighthall contributes to new ...



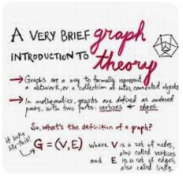
Science in the News - Harvard U...
Graph Theory 101 - Science i...



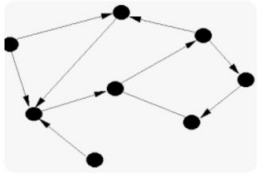
Towards Data Science
A Gentle Introduction to Graph theory ...



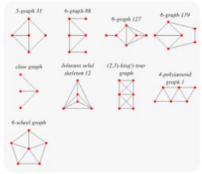
Medium - In Plain English
Graph Theory. 1. Introduction to G...



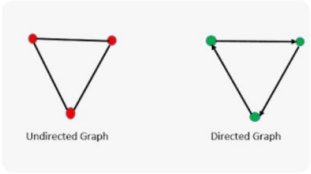
Medium
A Gentle Introduction To ...



people.brunel.ac.uk
Graph theory

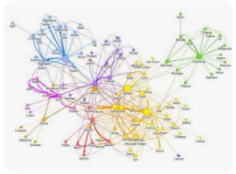


Wolfram MathWorld
Line Graph -- from Wolfram ...

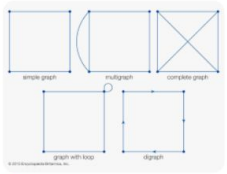


BYJU'S
Graph Theory-Discrete Mathematics ...

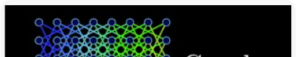
KDnuggets
What is Graph Theory, ...

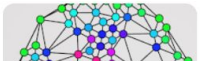


ResearchGate
Graph theory analysis in social ne...



Britannica
Graph theory | Problems & Appli...

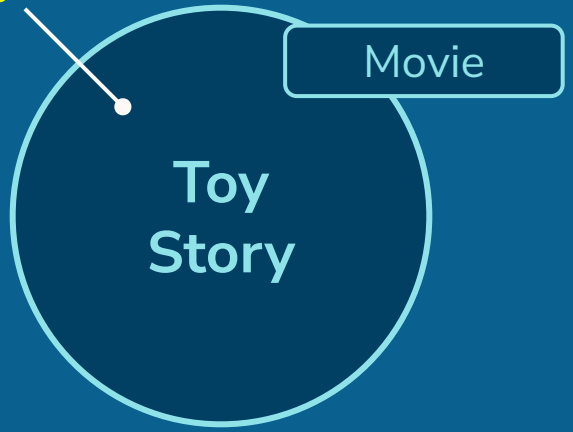





Toy
Story

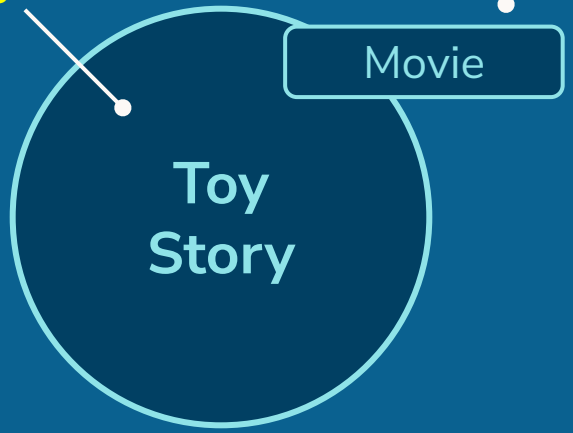
Nodes represent *things*

Nodes represent *things*



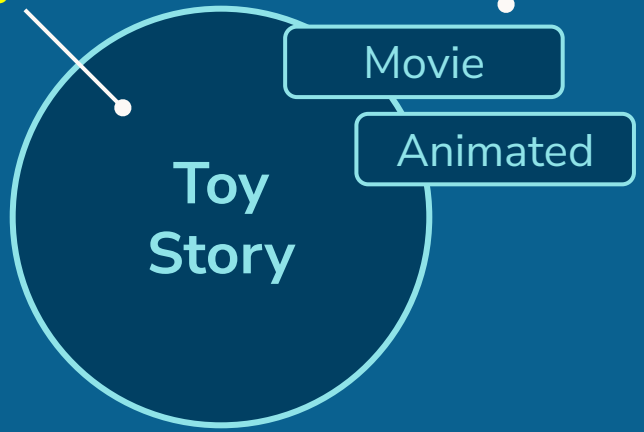
Nodes can be identified by one or more *labels*

Nodes represent *things*



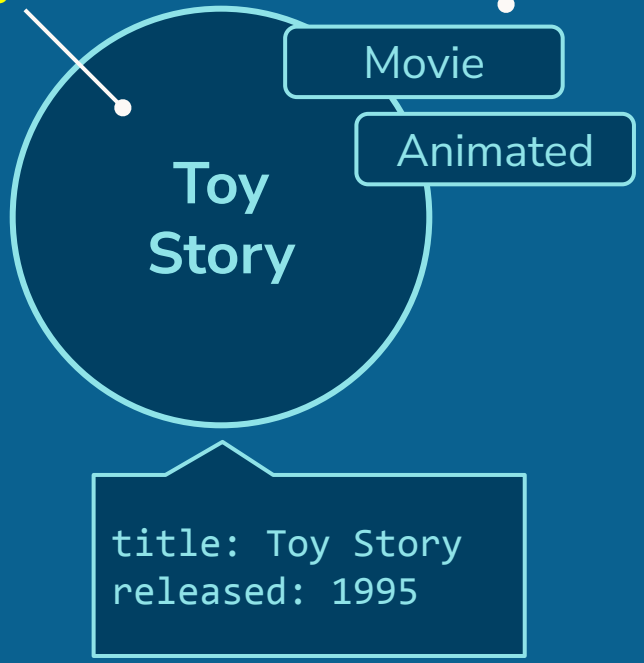
Nodes represent *things*

Nodes can be identified by one or more *labels*



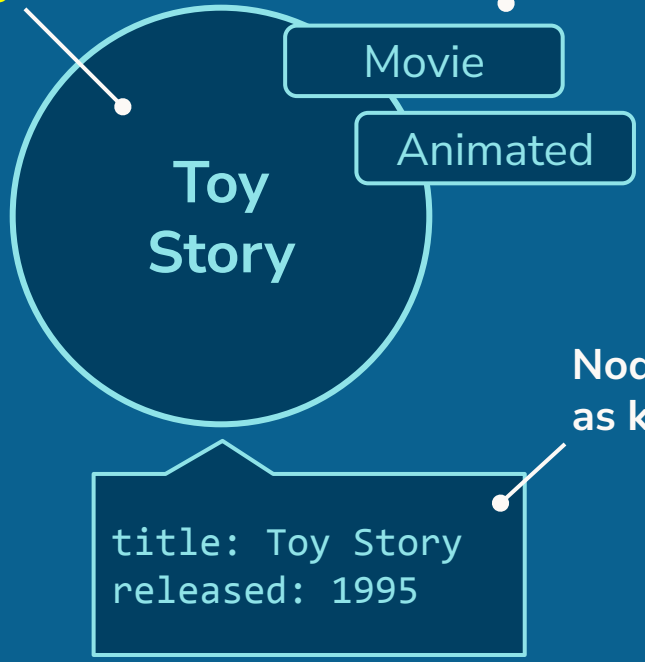
Nodes represent *things*

Nodes can be identified by one or more *labels*

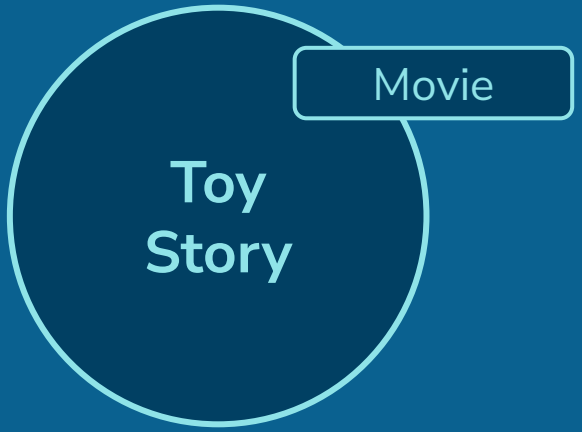


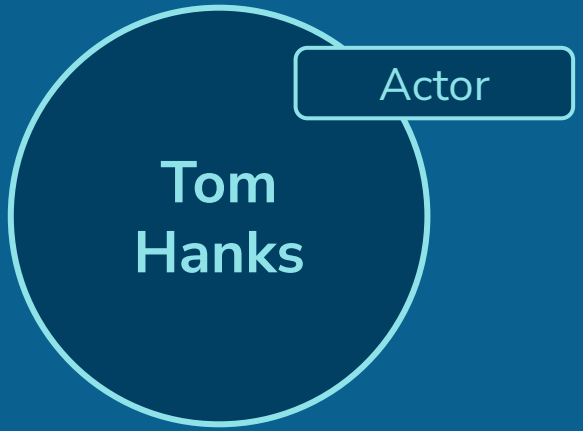
Nodes represent *things*

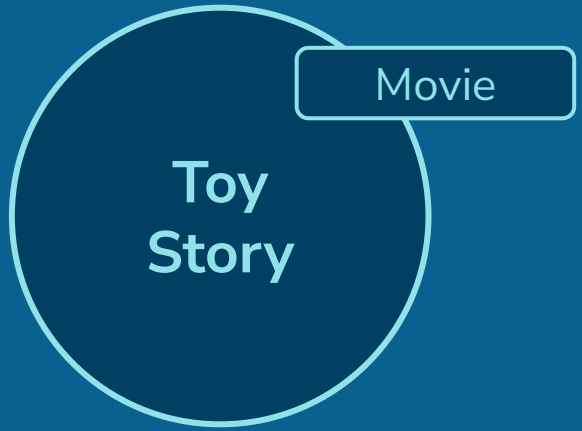
Nodes can be identified by one or more *labels*

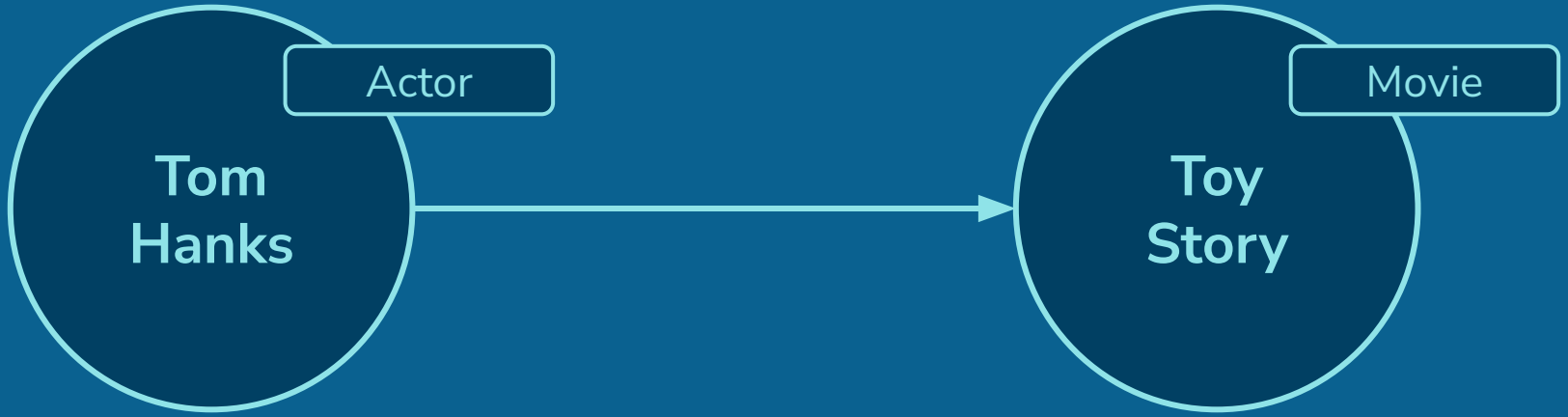


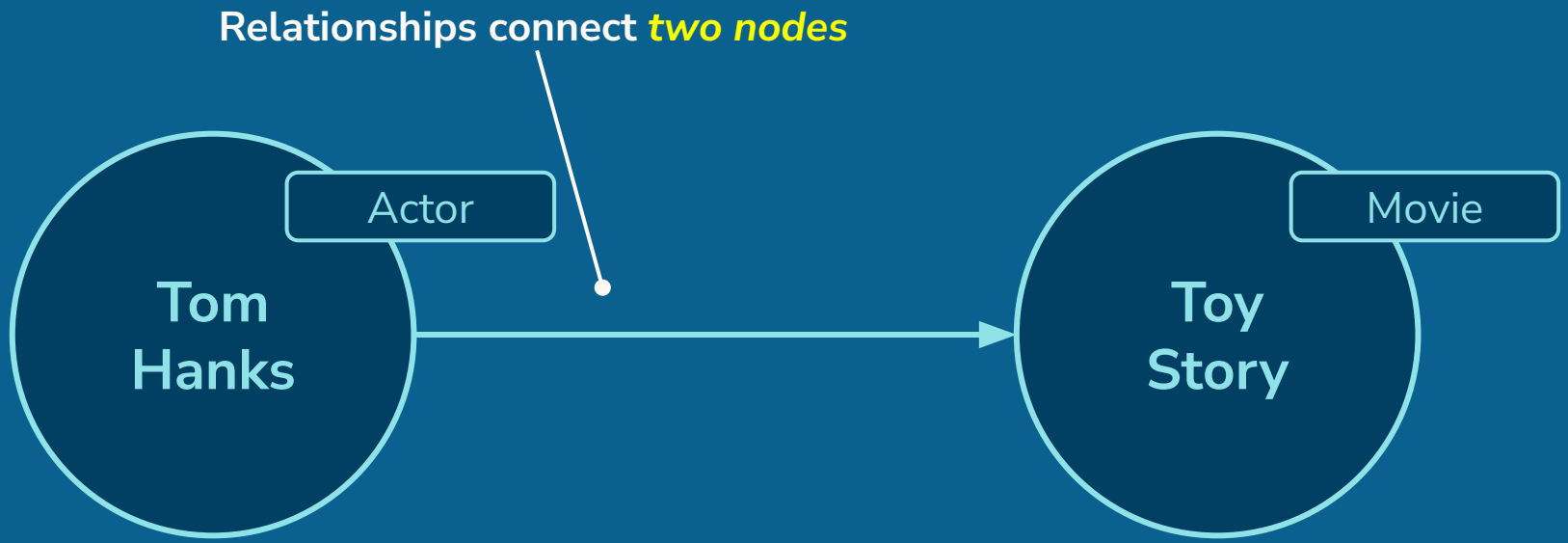
Nodes can hold *properties* as key/value pairs

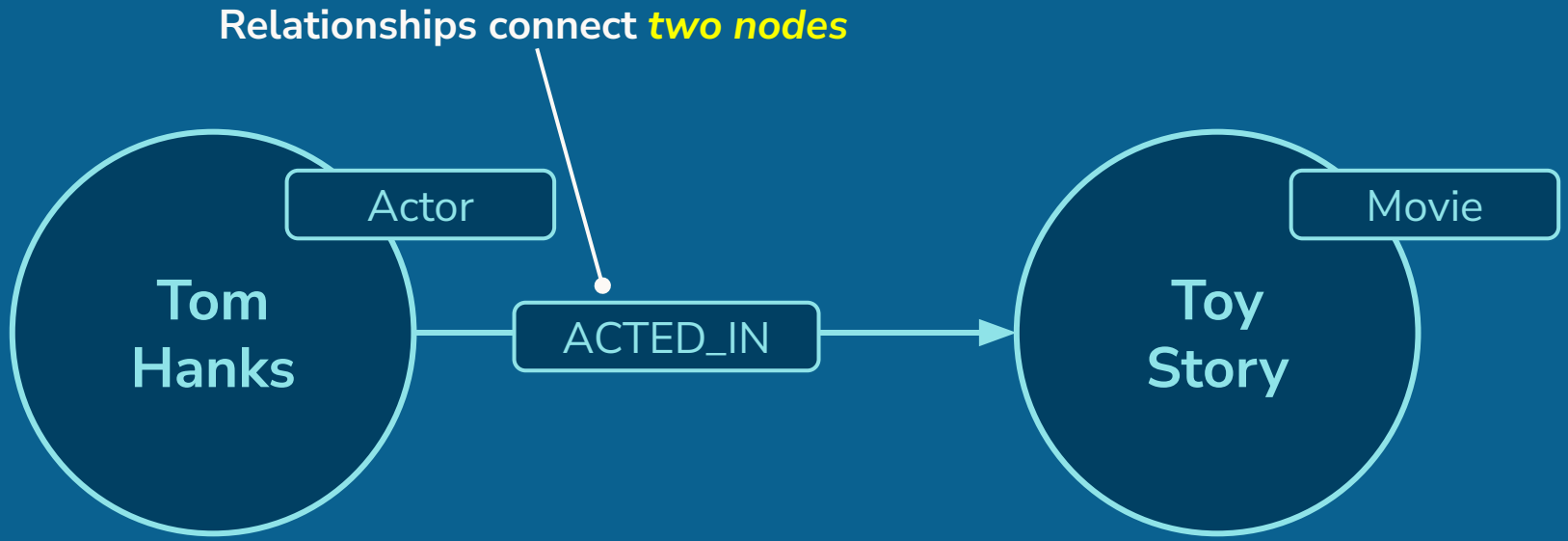

Movie
Toy
Story

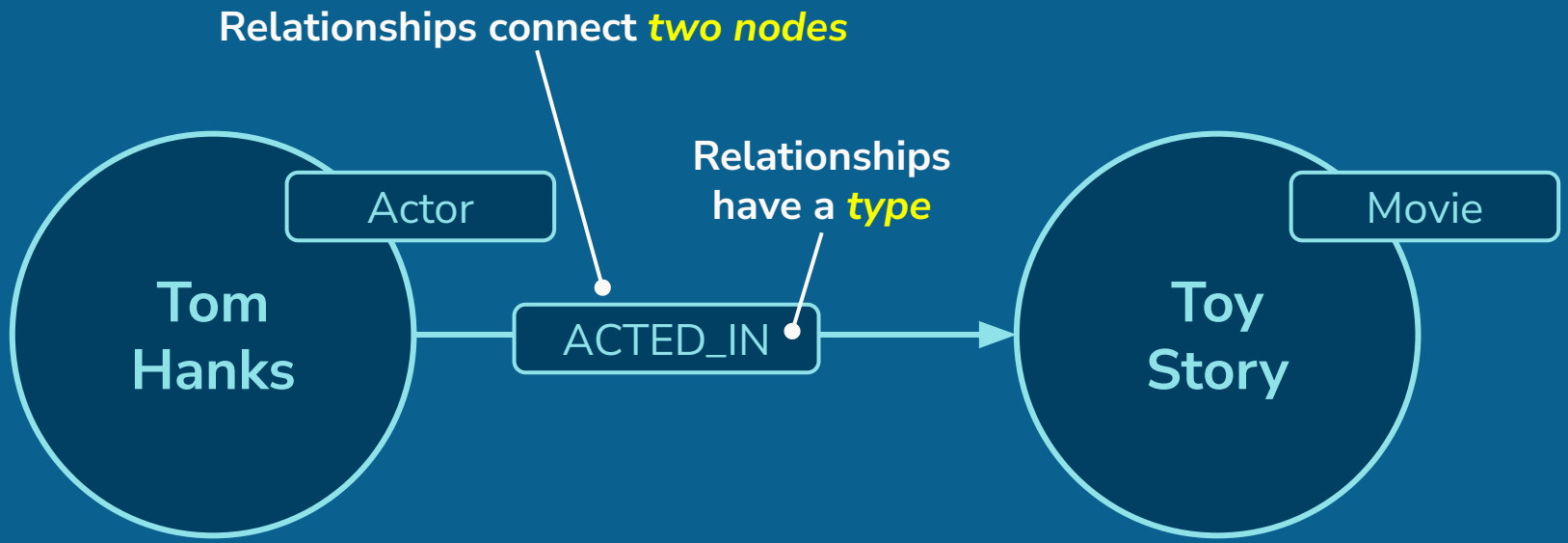

Actor
Tom
Hanks

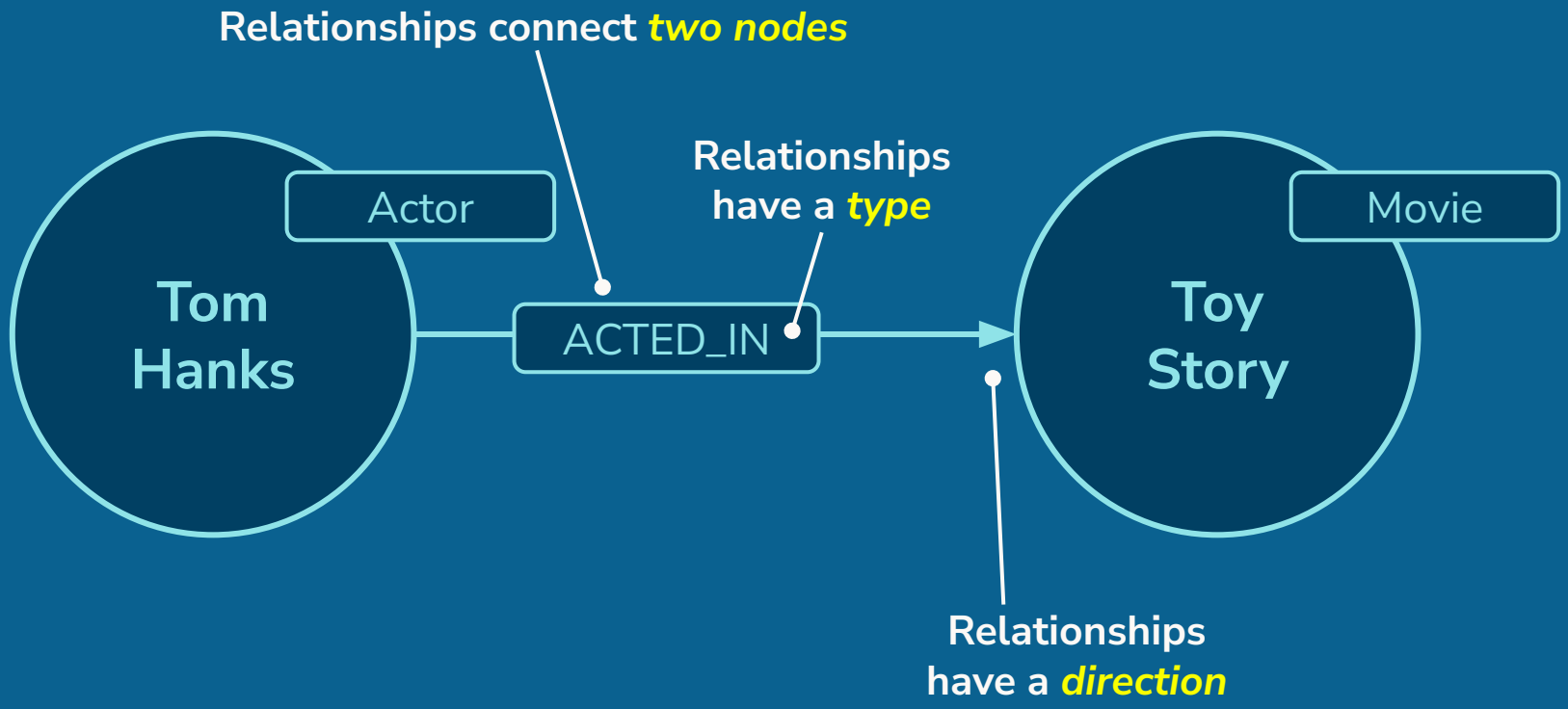

Movie
Toy
Story

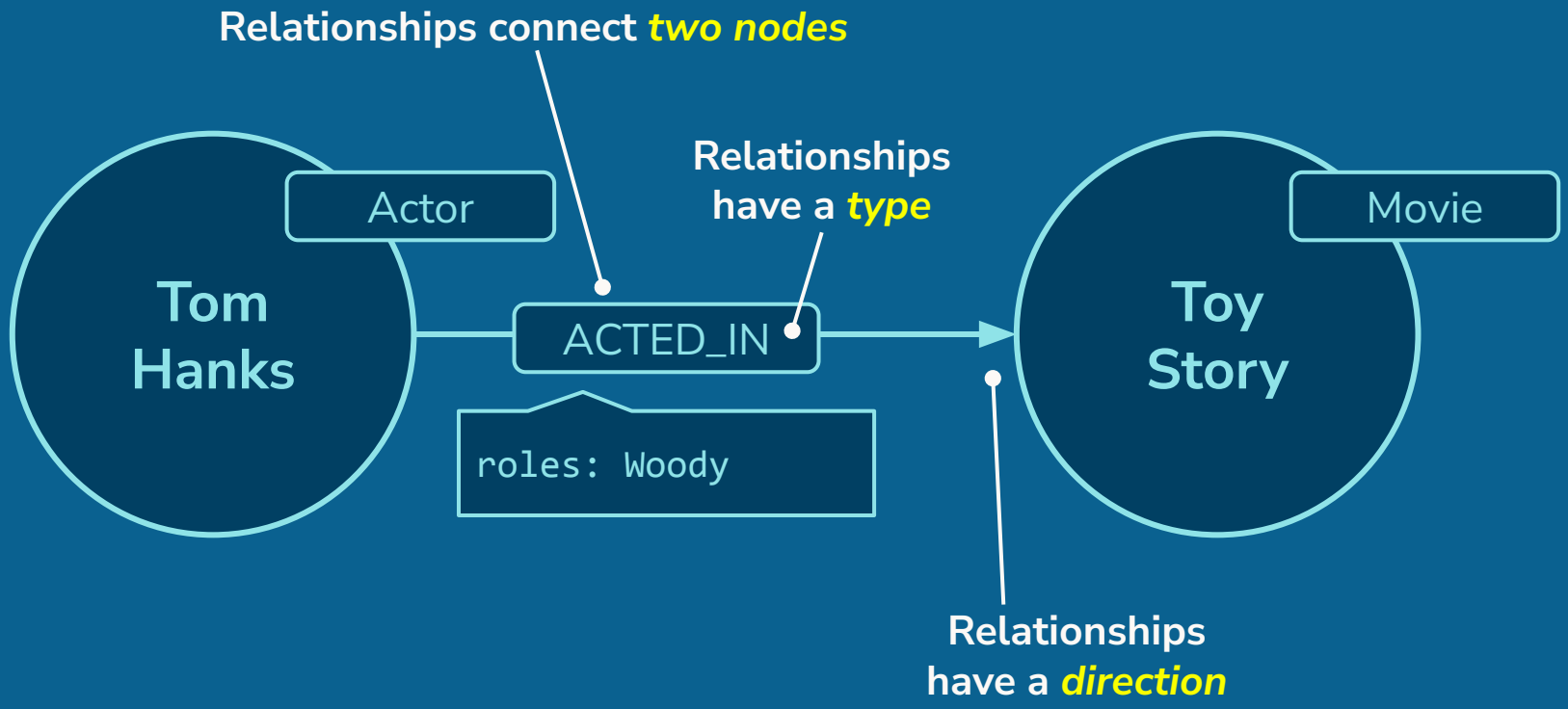

Actor
Tom Hanks
Movie
Toy Story

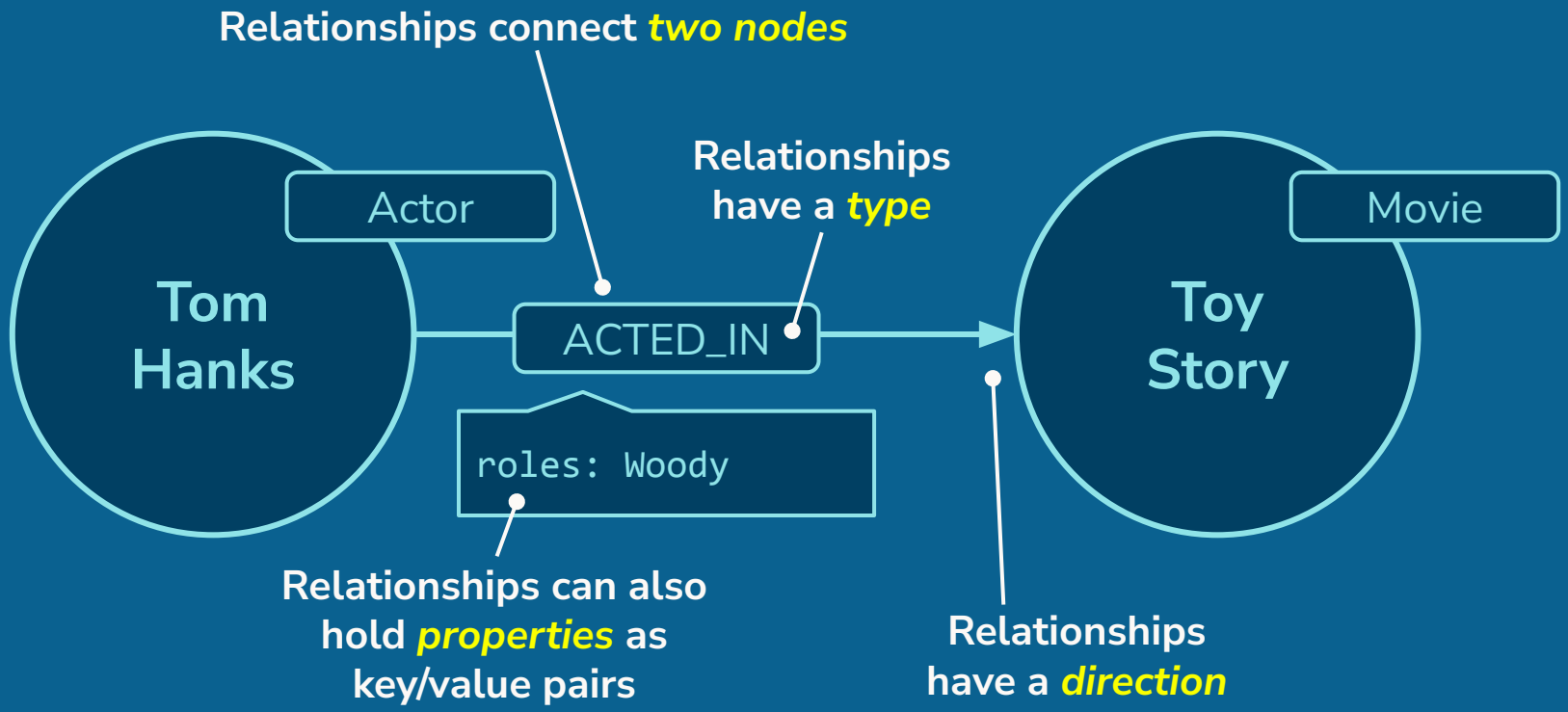

Relationships connect *two nodes*
Actor
Tom Hanks
Movie
Toy Story


Relationships connect *two nodes*
Actor
Tom Hanks
ACTED_IN
Movie
Toy Story


Relationships connect *two nodes*
Relationships have a *type*
Actor
Movie
Tom Hanks
ACTED_IN
Toy Story


Relationships connect two nodes
Relationships have a type
Actor
Movie
Tom Hanks
ACTED_IN
Toy Story
Relationships have a direction

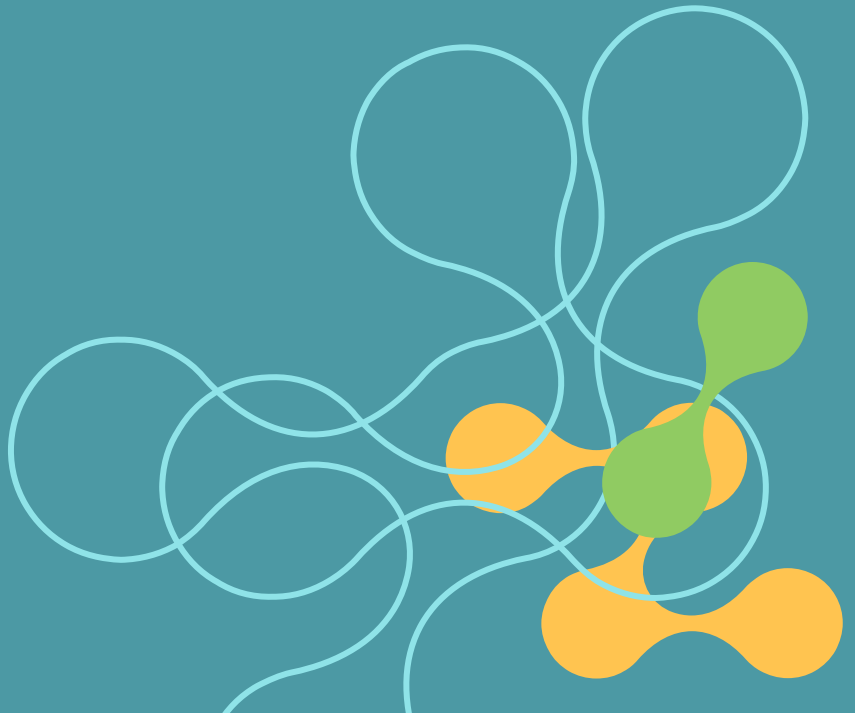

Relationships connect two nodes
Relationships have a type
Actor
Tom Hanks
ACTED_IN
roles: Woody
Movie
Toy Story
Relationships have a direction


Relationships connect two nodes
Relationships have a type
Actor
Movie
Tom Hanks
ACTED_IN
Toy Story
roles: Woody
Relationships can also hold properties as key/value pairs
Relationships have a direction

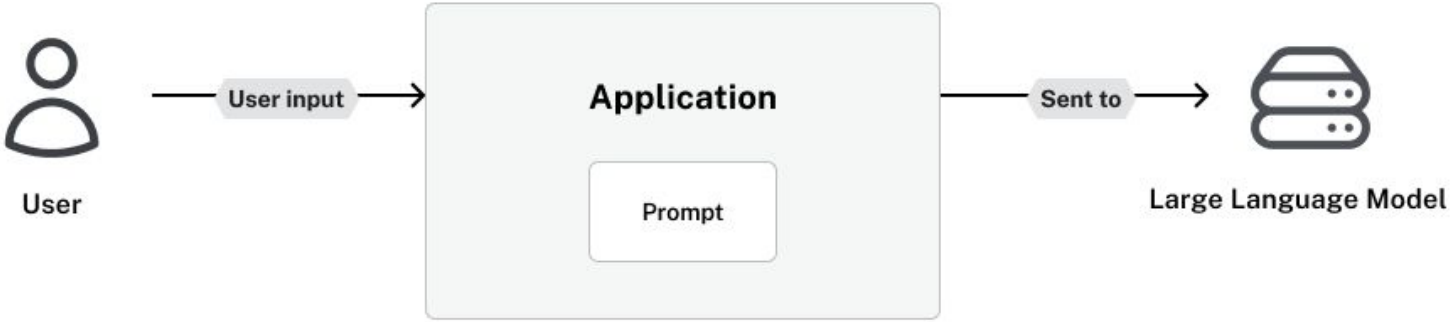
Graphs are everywhere

Graph databases are designed to efficiently store and query complex networks of nodes and relationships

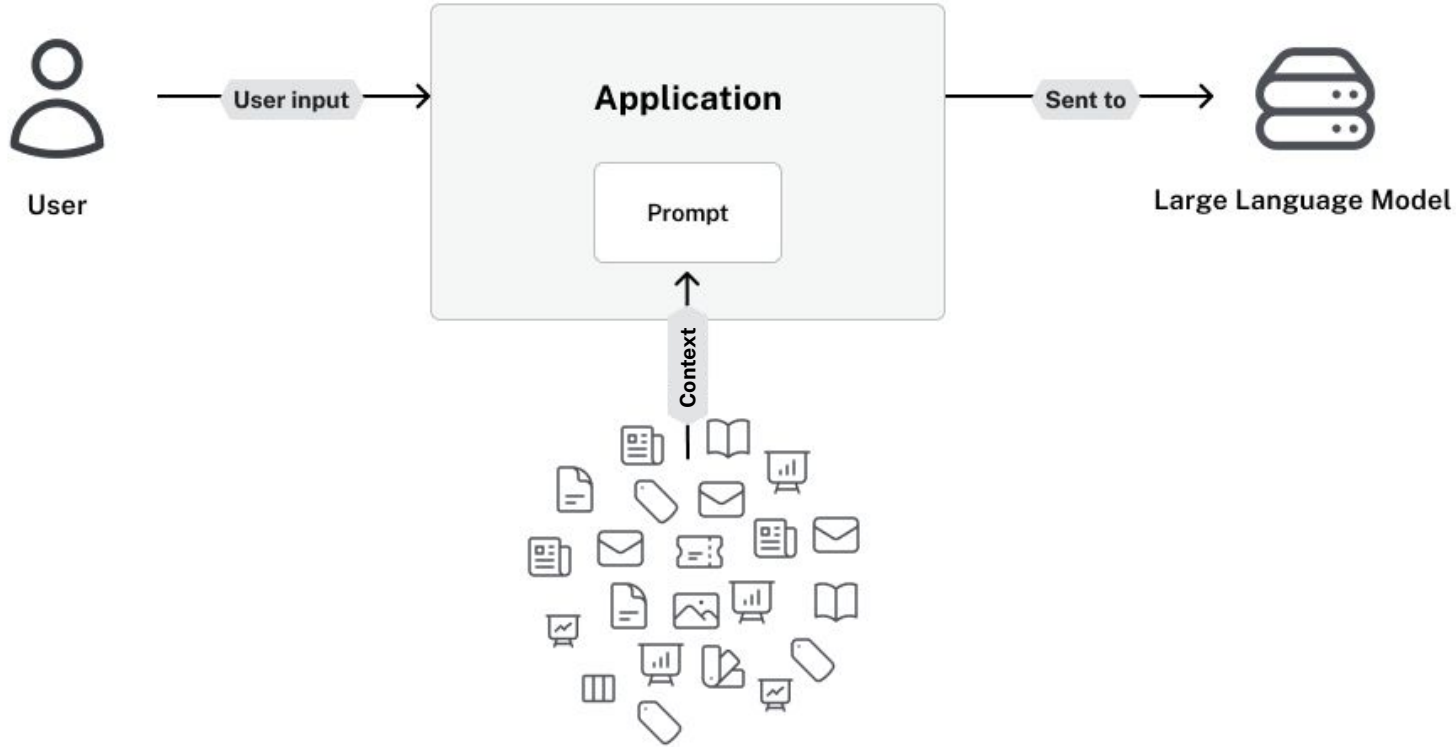
Retrieval augmented generation



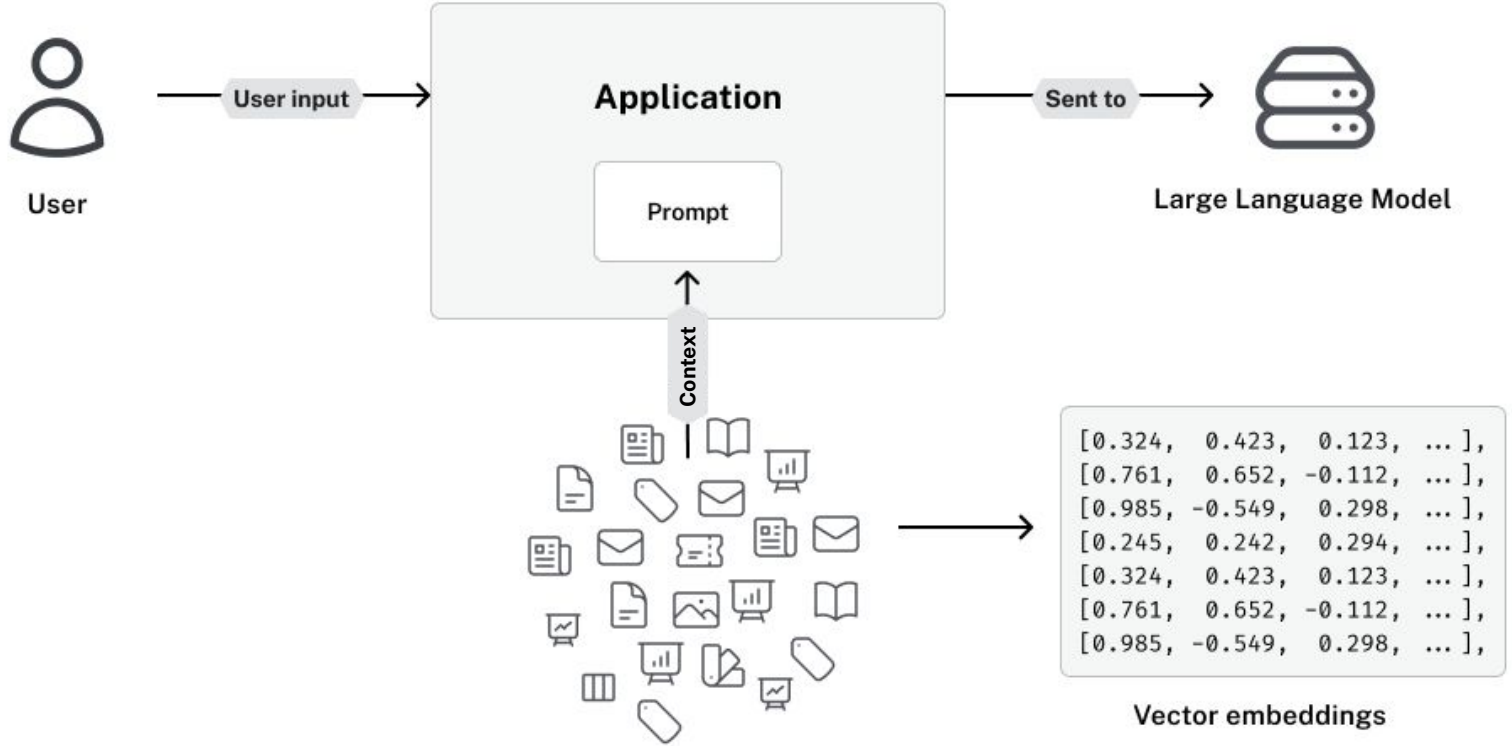
Interacting with LLMs



Interacting with LLMs



Interacting with LLMs



Vector embeddings

Text	Vector embedding
She’s just a small town girl	`[0.12, -0.34, 0.56, 0.78, ... , -0.91]`
Living in a lonely world	`[0.22, 0.45, -0.67, 0.11, ... , 0.33]`
She took the midnight train	`[-0.55, 0.89, 0.12, -0.44, ... , 0.67]`
Going anywhere	`[0.78, -0.23, 0.45, 0.91, ... , -0.12]`

Where vector search fall short

Vectors work well for:

- Contextual or Meaning Based Questions
- Synonyms or Paraphrasing
- Fuzzy or Vague queries
- Broad or Open-Ended questions
- Complex queries with multiple concepts

What does Paul Graham think about Generative AI?

Vectors are ineffective for:

Where vector search fall short

Vectors work well for:

- Contextual or Meaning Based Questions
- Synonyms or Paraphrasing
- Fuzzy or Vague queries
- Broad or Open-Ended questions
- Complex queries with multiple concepts

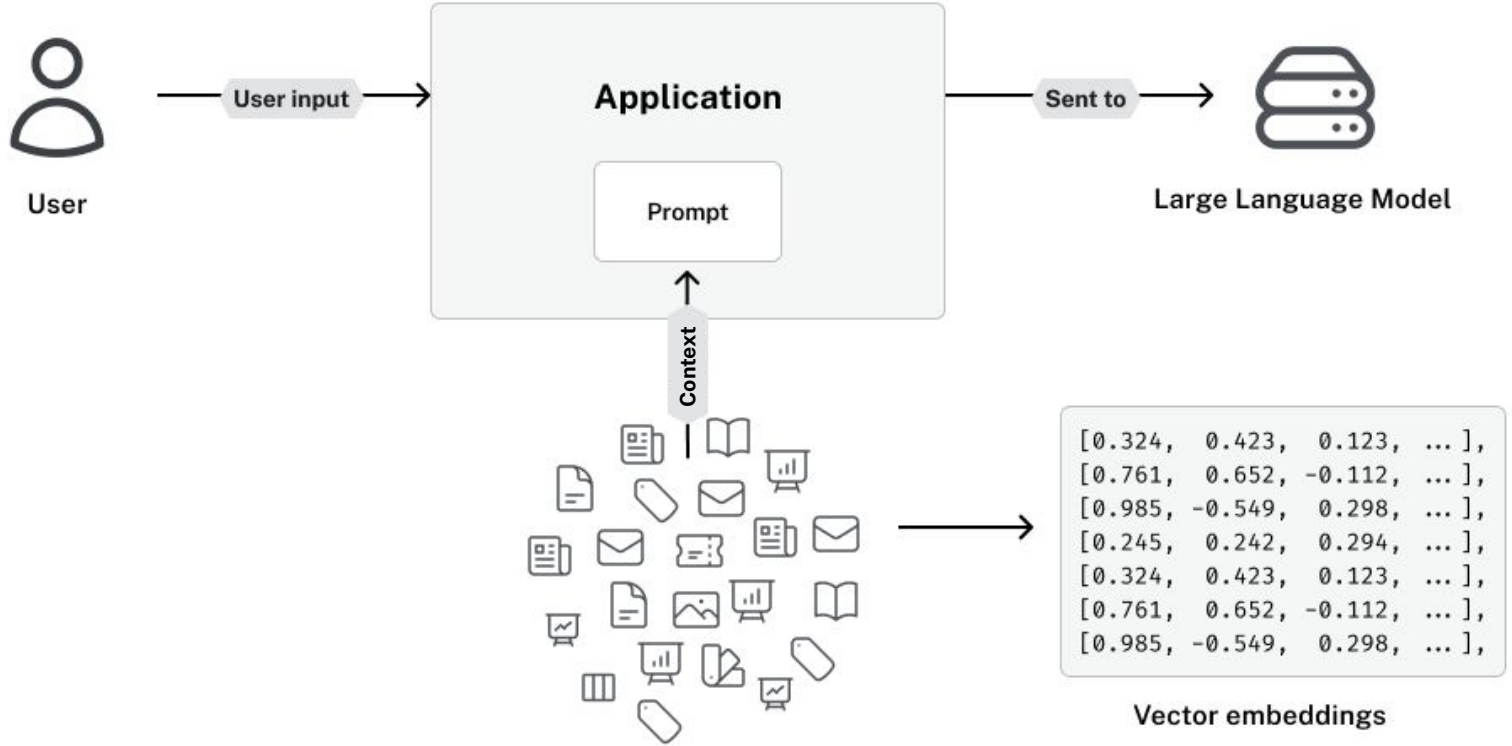
What does Paul Graham think about Generative AI?

Vectors are ineffective for:

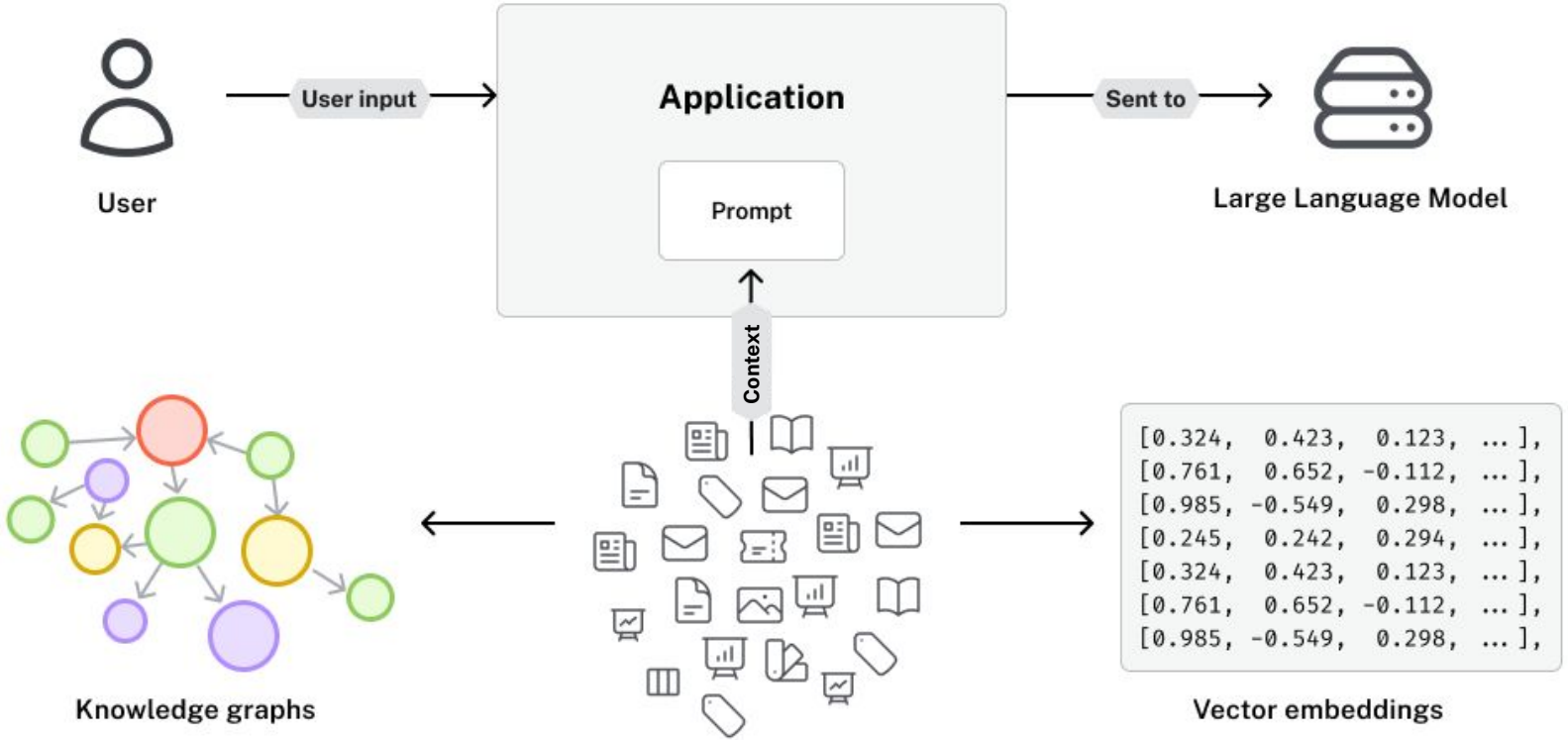
- Highly Specific or Fact-Based Questions
- Numerical or Exact-Match Queries
- Boolean or Logical Queries
- Ambiguous or Unclear Queries without Context
- Specialised Knowledge

How many Generative AI Startups has Paul Graham invested in?

Interacting with LLMs



Interacting with LLMs

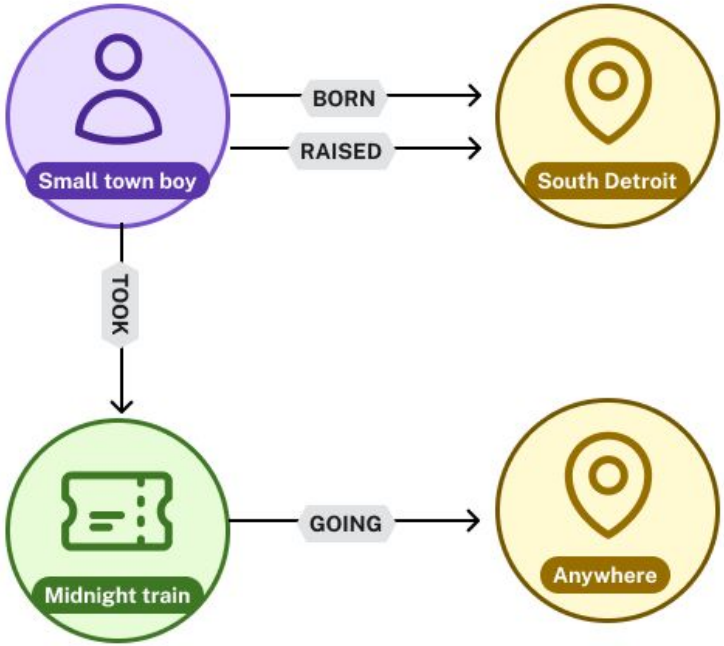


Vector embeddings and a knowledge graph

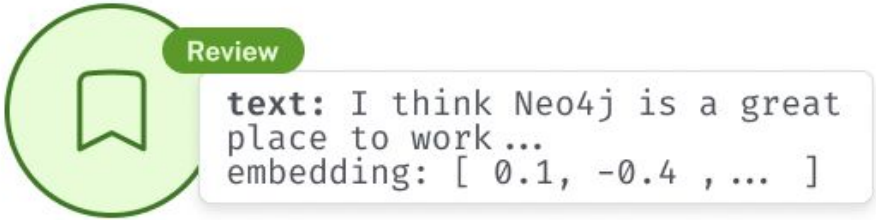
Text	Vector embedding
She’s just a small town girl	`[0.12, -0.34, 0.56, ... ,-0.91]`
Living in a lonely world	`[0.22, 0.45, -0.67, ... , 0.33]`
She took the midnight train	`[-0.55, 0.89, 0.12, ... , 0.67]`
Going anywhere	`[0.78, -0.23, 0.45, ... ,-0.12]`

Vector embeddings and a knowledge graph

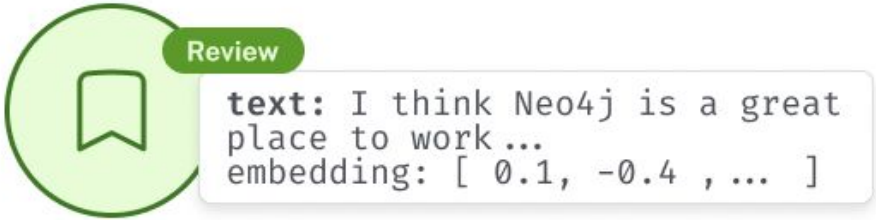
Text	Vector embedding
She's just a small town girl	`[0.12, -0.34, 0.56, ... ,-0.91]`
Living in a lonely world	`[0.22, 0.45, -0.67, ... , 0.33]`
She took the midnight train	`[-0.55, 0.89, 0.12, ... , 0.67]`
Going anywhere	`[0.78, -0.23, 0.45, ... ,-0.12]`



Context is key



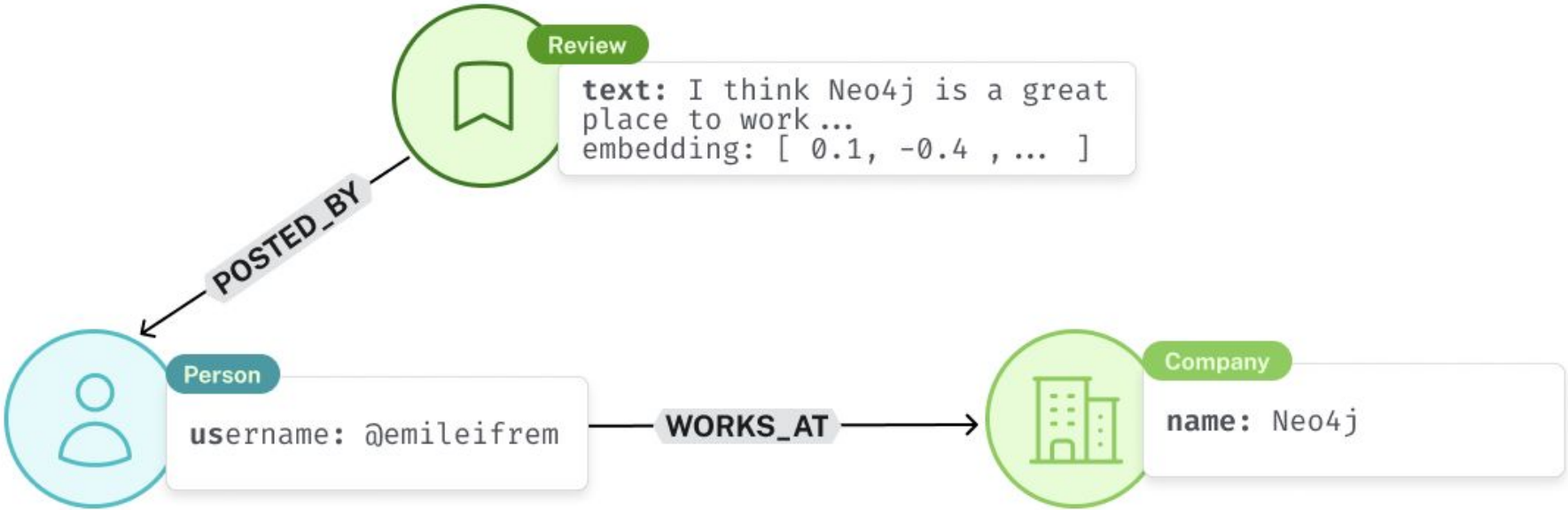
Context is key



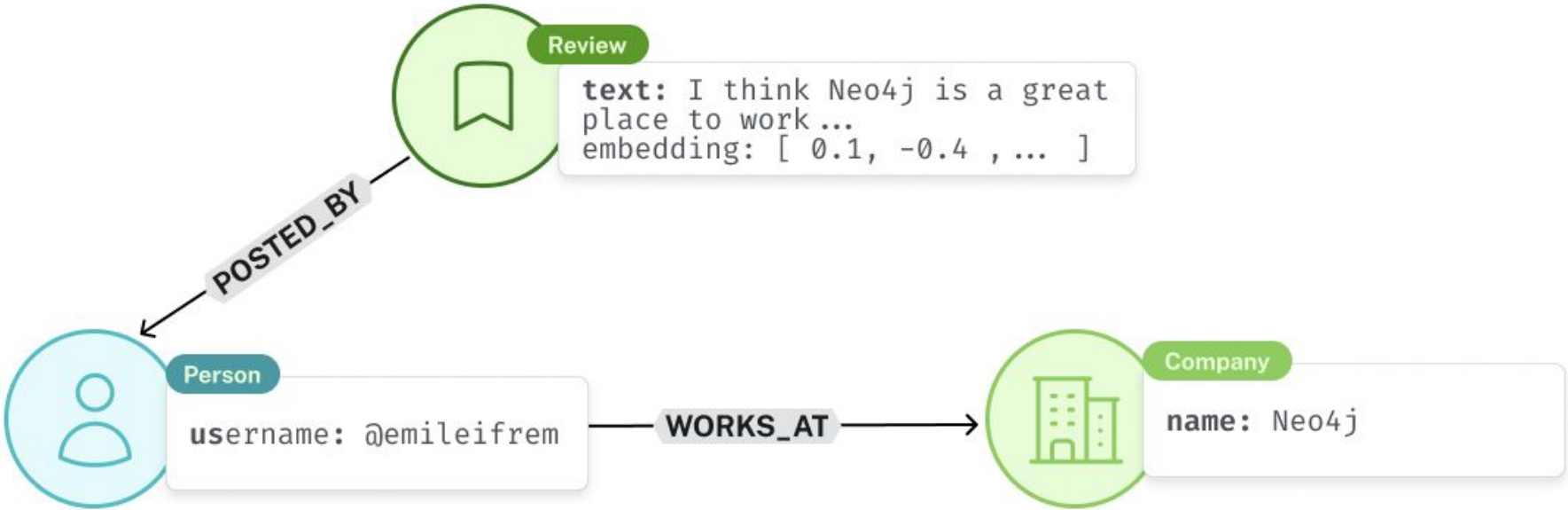
According to the provided context,
Neo4j is a great place to work…

Context is key

Context is key



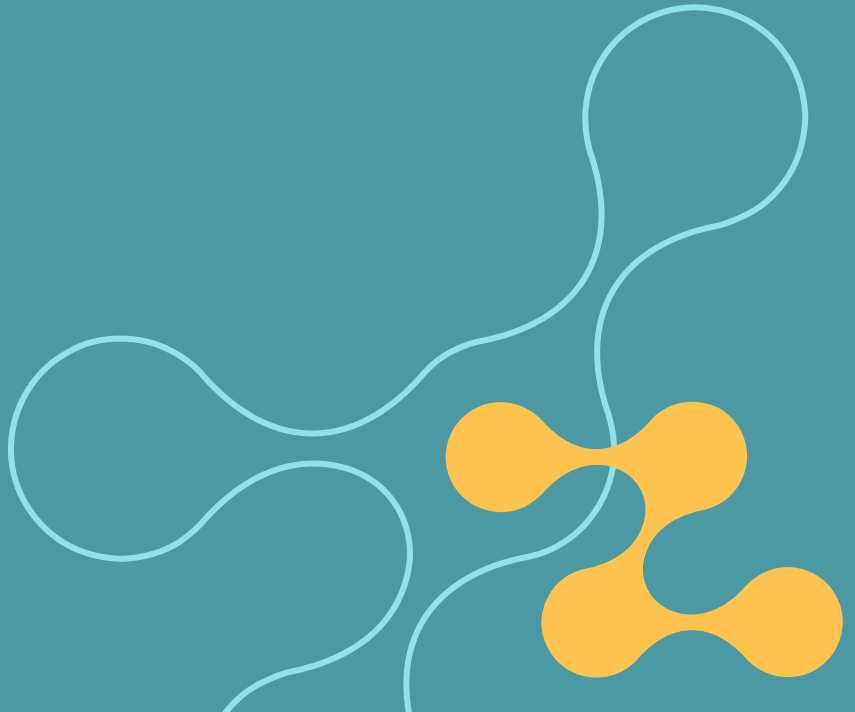
Context is key



*According to **Emil Eifrem, the CEO**,*
Neo4j is a great place to work…

Knowledge Graphs

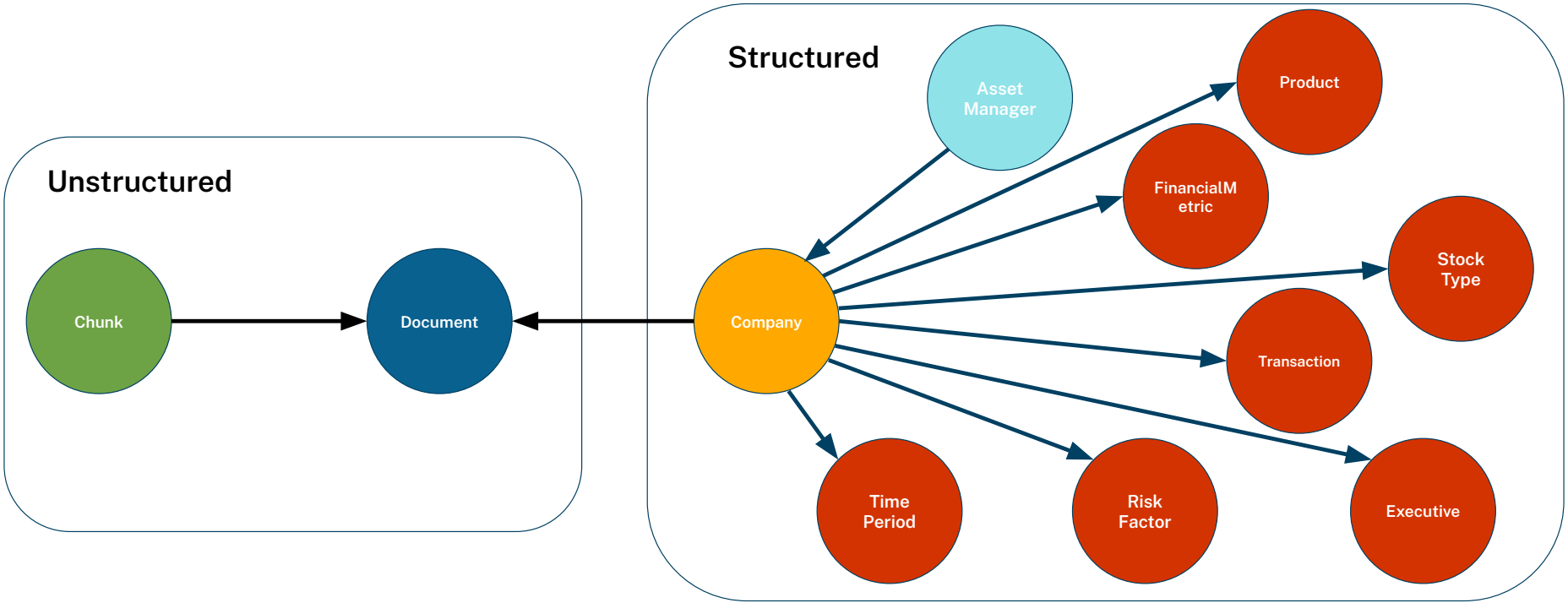
Building Graphs from unstructured data

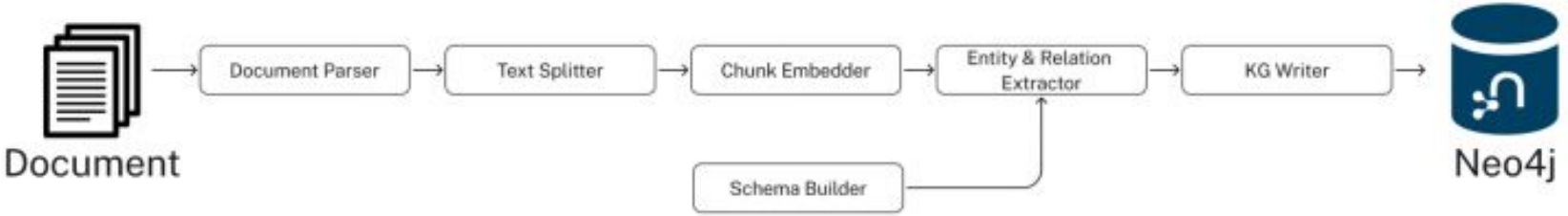


continual improvement in product price and performance characteristics, rapid adoption of technological advancements by competitors, and price sensitivity on the part of consumers and businesses. Many of the Company's competitors seek to compete primarily through aggressive

pricing and very low cost structures, and by imitating the Company's products and infringing on its intellectual property.

The Company's ability to compete successfully depends heavily on ensuring the continuing and timely introduction of innovative new products, services and technologies to the marketplace. The Company designs and develops nearly the entire solution for its products, including the hardware, operating system, numerous software applications and related services. Principal competitive factors important to the Company include price, product and service features (including security features), relative price and performance, product and service quality and reliability, design


Unstructured
Chunk
Document
Structured
Company
Asset Manager
Product
FinancialMetric
Stock Type
Transaction
Time Period
Risk Factor
Executive




```
pipeline = SimpleKGPipeline(
    driver=driver, # Neo4j connection driver
    llm=llm, embedder=embedder,  # OpenAI llm and embeddings
    entities=entities, relations=relations,  # Define schema
    enforce_schema="STRICT",
    prompt_template=prompt_template,
)
# Process the SEC filing documents
pdf_documents = [
    "apple-10K-2023.pdf", "microsoft-10K-2023.pdf",
    # ... more company filings
]
# Run the pipeline to transform PDFs into knowledge graph
for pdf_file in pdf_documents:
    pipeline.run(file_path=pdf_file)
```

Schema driven extraction

Company

Executive

Product

FinancialMetric

RiskFactor

StockType

Transaction

TimePeriod

Company HAS_METRIC FinancialMetric

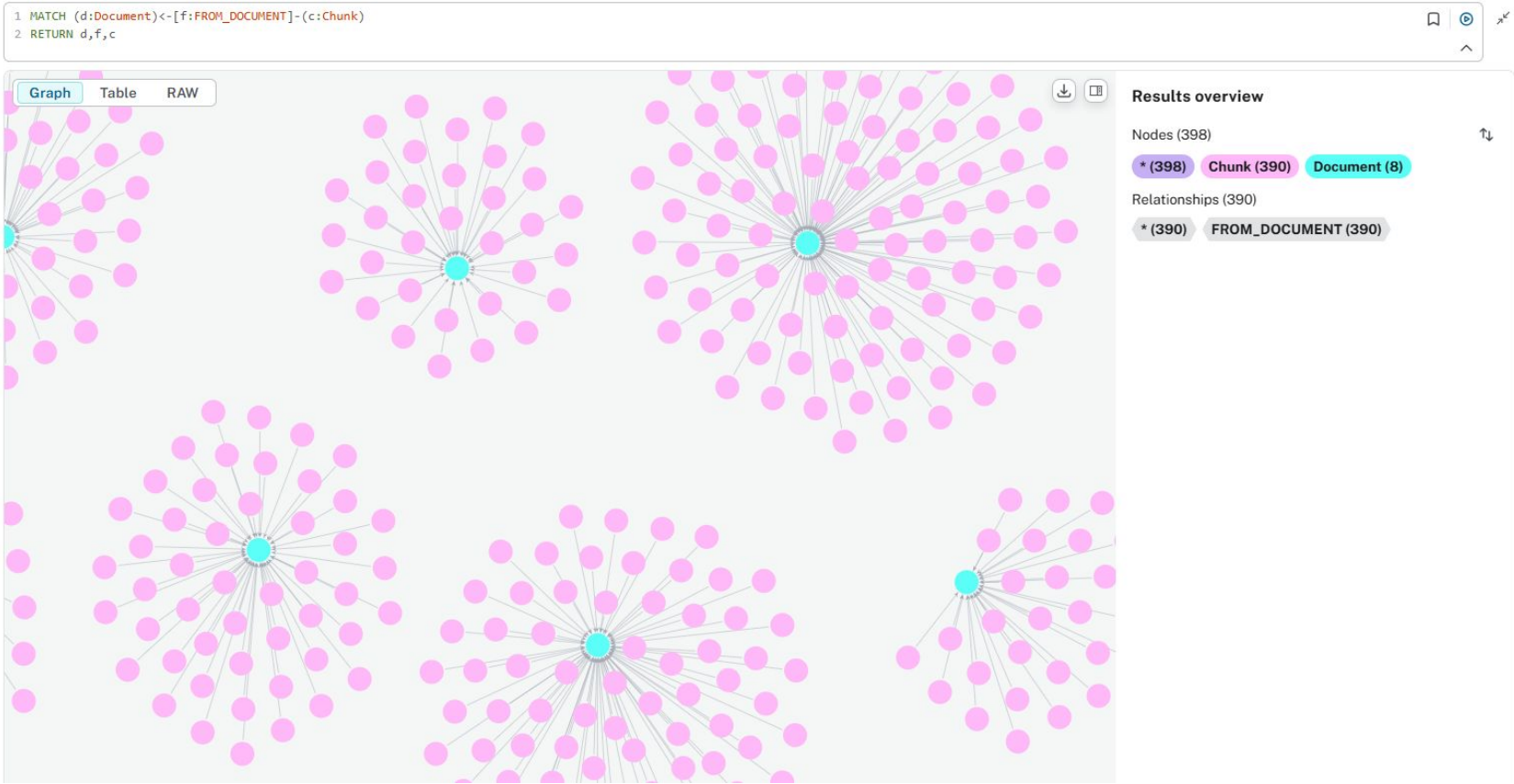
Company FACES_RISK RiskFactor

Company ISSUED_STOCK StockType

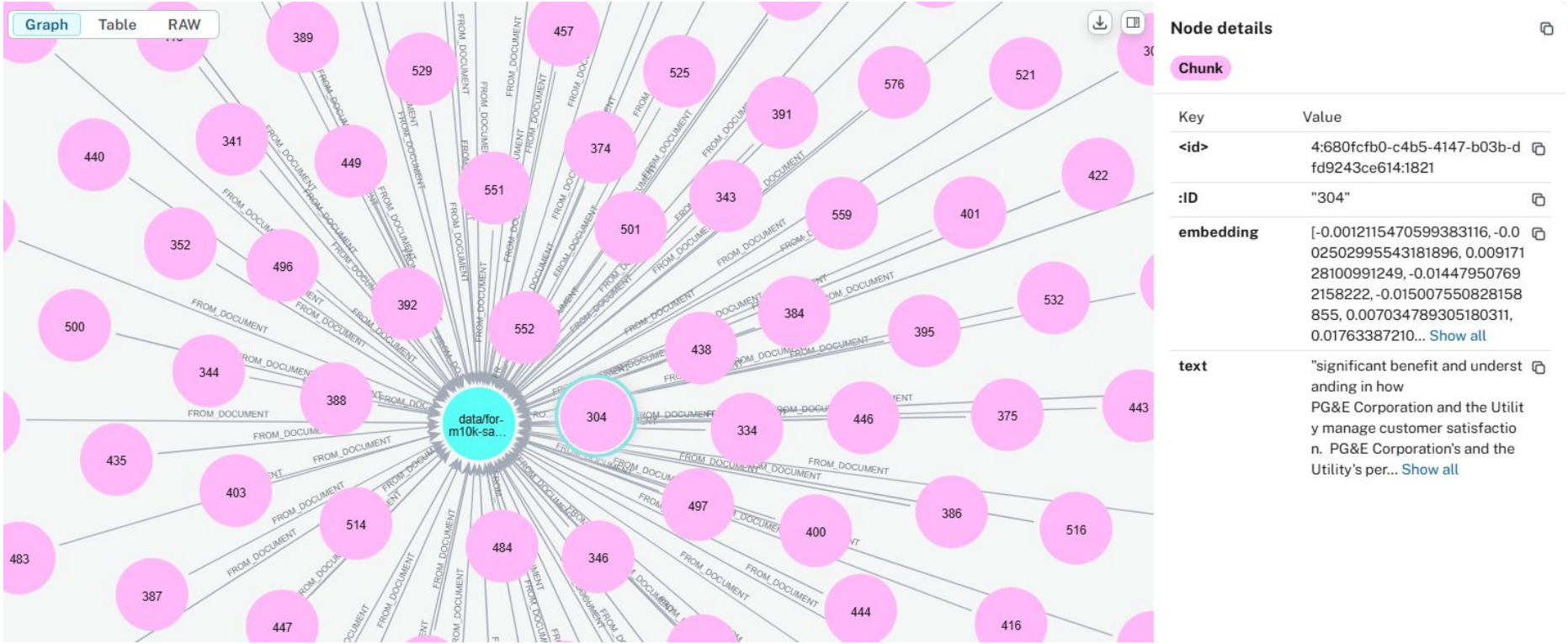
Company MENTIONS Product

DEMO

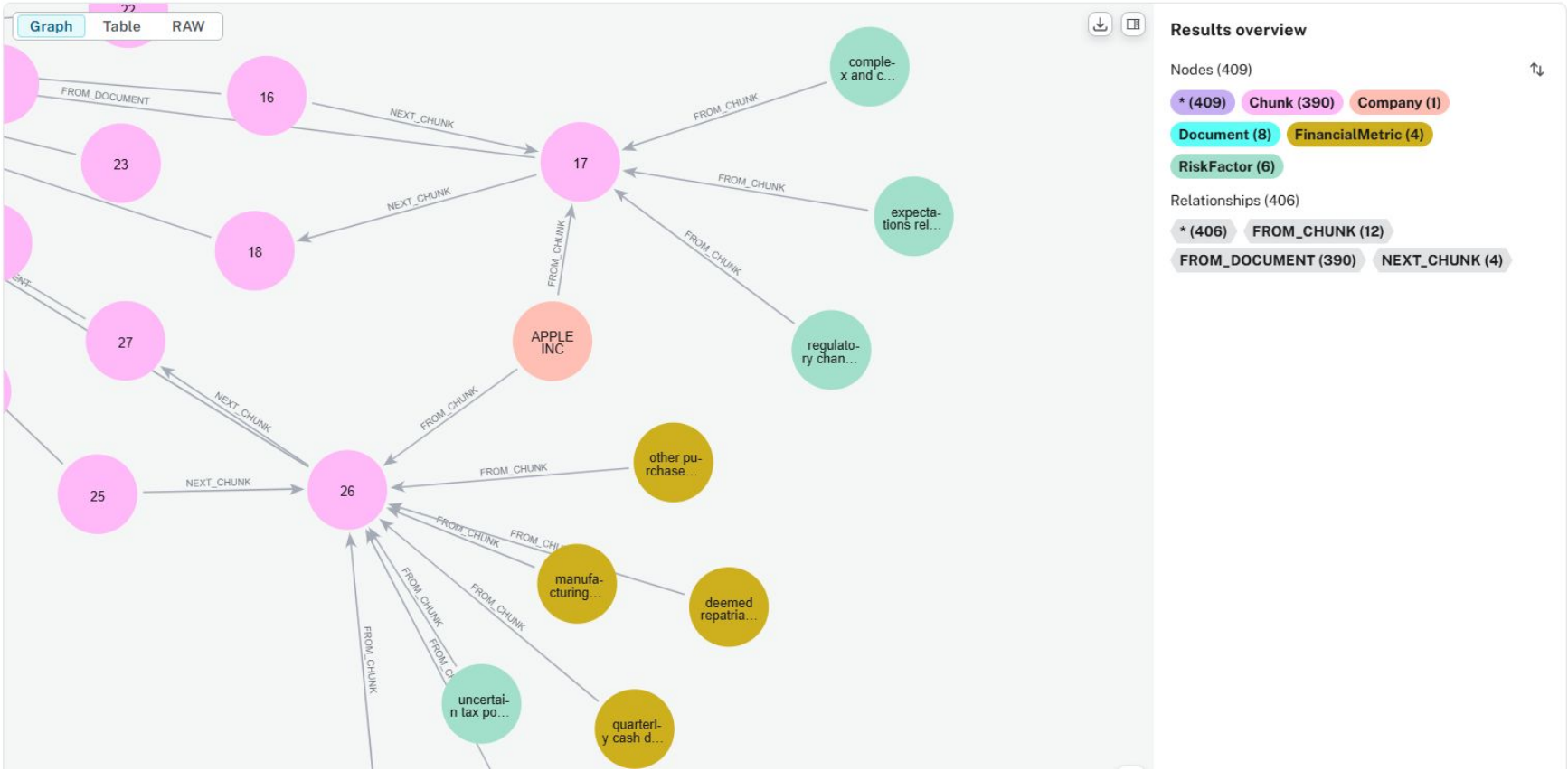
```
MATCH (d:Document)<-[f:FROM_DOCUMENT]-(c:Chunk)
RETURN d,f,c
```



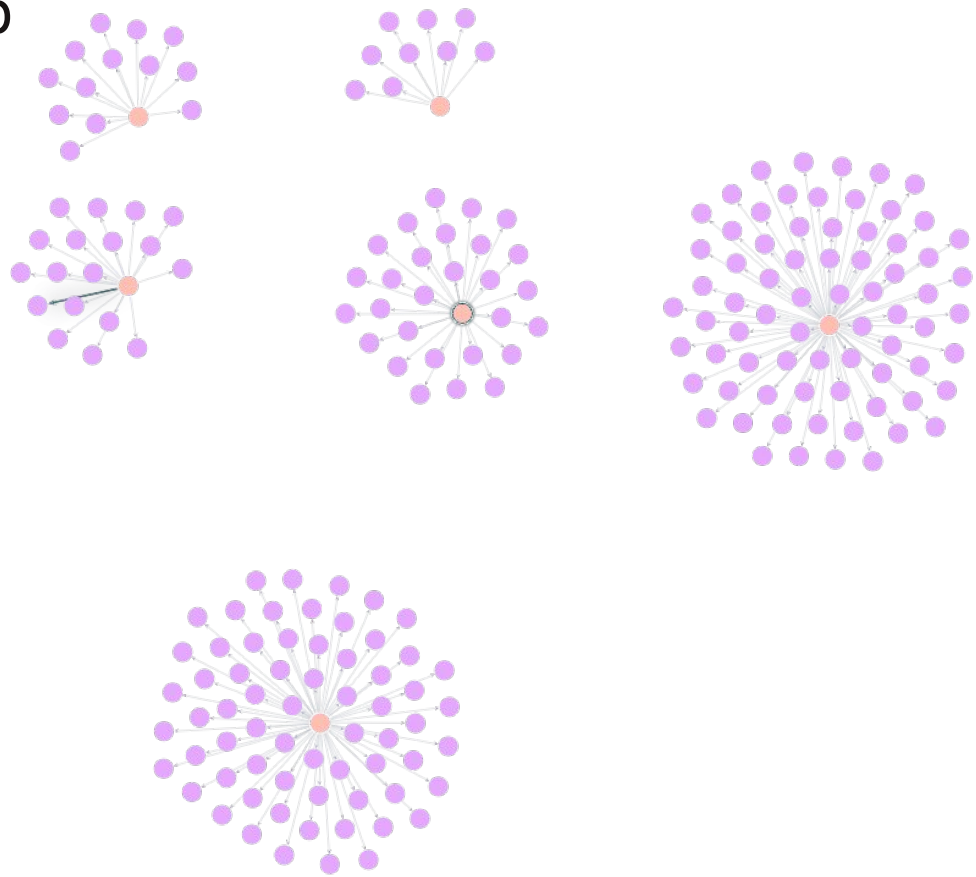
```
MATCH (d:Document)<-[f:FROM_DOCUMENT]-(c:Chunk)
RETURN d,f,c
```



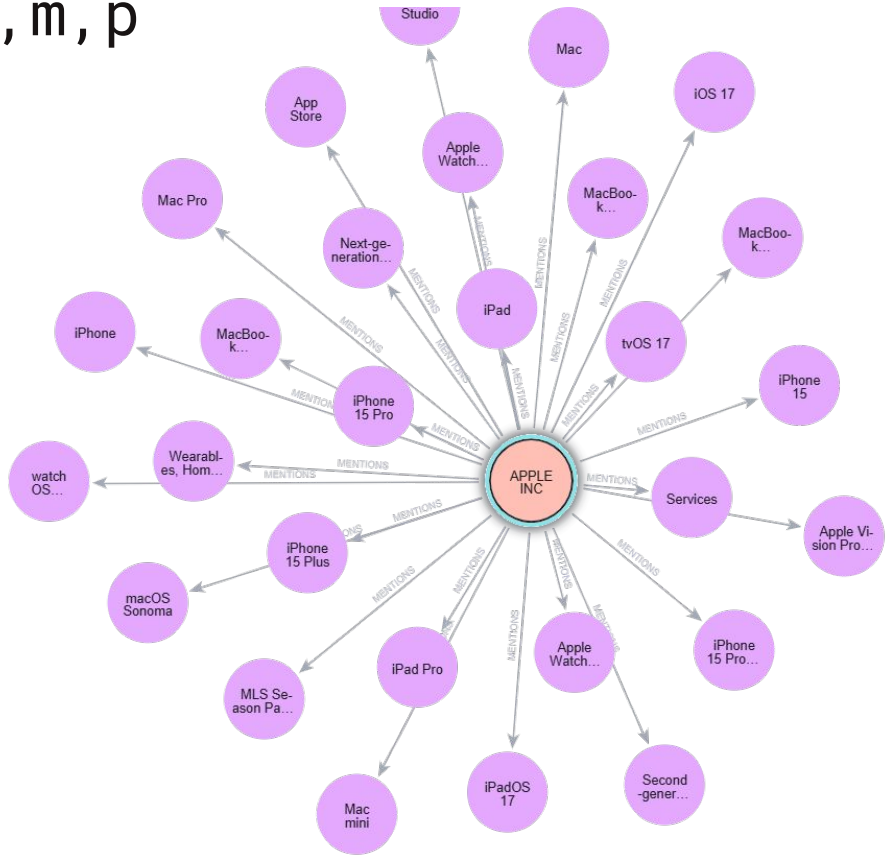
```
MATCH (d:Document)<-[f:FROM_DOCUMENT]-(c:Chunk)
RETURN d,f,c
```



```
MATCH (c:Company)-[m:MENTIONS]->(p:Product)
RETURN c,m,p
```



```
MATCH (c:Company)-[m:MENTIONS]->(p:Product)
RETURN c,m,p
```



```
MATCH (e) WHERE NOT e:Document AND NOT e:Chunk
RETURN
  labels(e) as entityType,
   count(e) as count
ORDER BY count DESC
```

	entityType	count
1	["RiskFactor"]	820
2	["FinancialMetric"]	470
3	["Product"]	241
4	["TimePeriod"]	102
5	["Transaction"]	46
6	["Executive"]	29
7	["AssetManager"]	15
8	["Company"]	12
9	["StockType"]	9

```
MATCH (c:Company {name: 'MICROSOFT CORP'})
-[:HAS_METRIC]->(m:FinancialMetric)
RETURN c.name, m.name
```

	c.name	m.name
1	"MICROSOFT COR P"	"Provision for Income Taxes"
2	"MICROSOFT COR P"	"Scope 1 emissions"
3	"MICROSOFT COR P"	"Scope 2 emissions"
4	"MICROSOFT COR P"	"Scope 3 emissions"
5	"MICROSOFT COR P"	"transaction volumes with Black- and African American-owned financial institutions"
6	"MICROSOFT COR P"	"$100 million program focused on mission-driven banks"

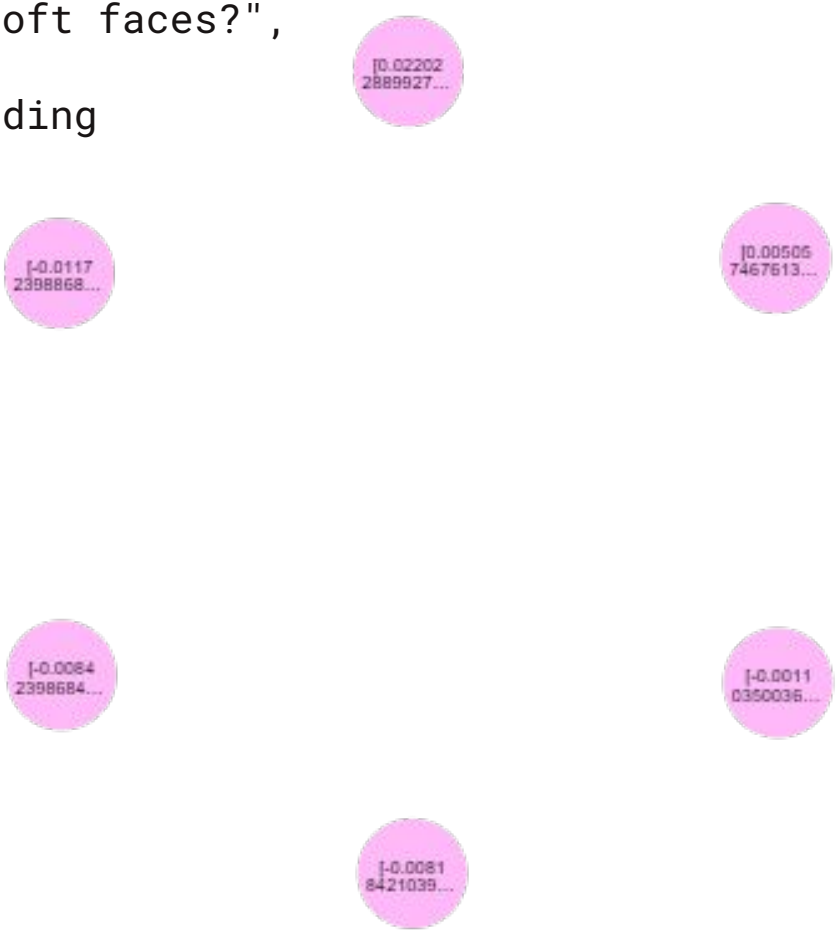
```
MATCH (c:Company)-[:FACES_RISK]->(r:RiskFactor)
RETURN c.name, r.name
```

	c.name	r.name
1	"APPLE INC"	"Geography"
2	"APPLE INC"	"Aggressive price competition"
3	"APPLE INC"	"Frequent introduction of new products"
4	"APPLE INC"	"Short product life cycles"
5	"APPLE INC"	"Evolving industry standards"
6	"APPLE INC"	"Commodity pricing fluctuations"
7	"APPLE INC"	"Industry-wide shortage and significant commodity pricing fluctuations"
8	"APPLE INC"	"Initial capacity constraints when new technologies are used"
9	"APPLE INC"	"Availability of components at acceptable prices"

```
WITH genai.vector.encode(
    "What are the risks that Microsoft faces?",
    "OpenAI",
    { token: $token }) AS userEmbedding

CALL db.index.vector.queryNodes(
  'chunkEmbeddings',
  6,
  userEmbedding)

YIELD node, score
RETURN node
```

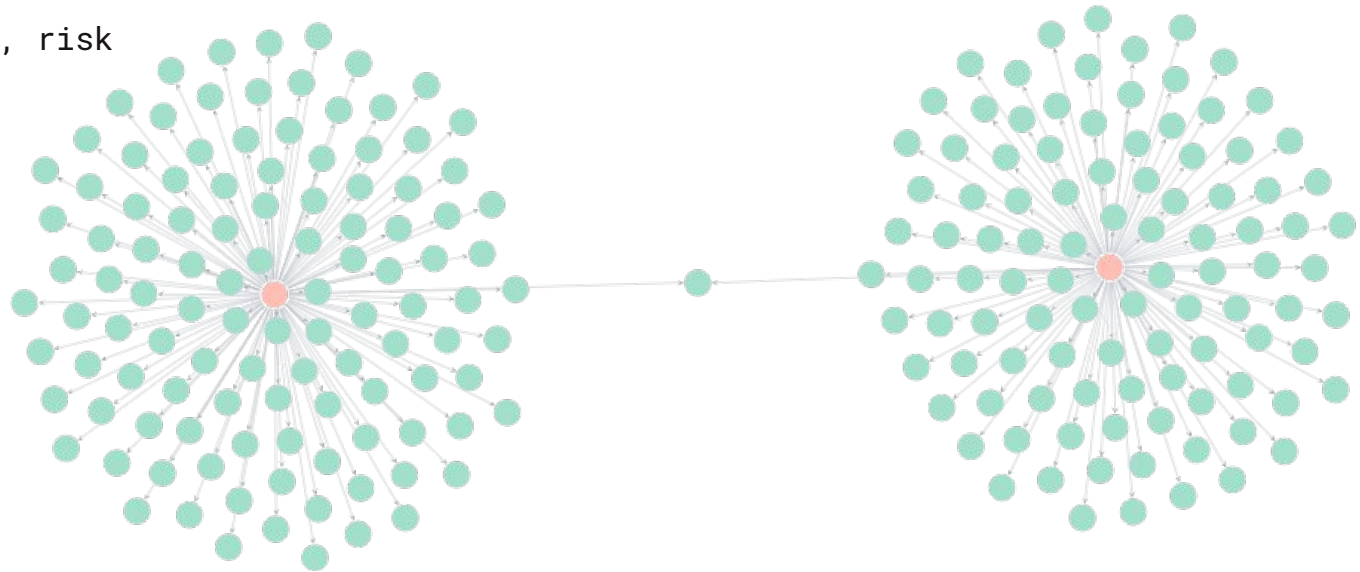


node.text	score
"time that investors might find useful or interesting. - Opportunities to sign up for email alerts to have information pushed in real time. We publish a variety of reports and resources related to our Corporate Social Responsibili ty programs and progress on our Reports Hub website, www.microsoft.com/corporate-responsibility/reports-hub, including reports on sustainabilit y, responsible sourcing, accessibility, digital trust, and public policy engagement. The information found on these websites is not part of, or incorporated by reference into, this or any other report we file with, or furnish to, the SEC. In addition to these channels, we use s ocial media to communicate to the public. It is possible that the information we post on social media could be deemed to be material to investors. We encourage investors, the media, and others interest	0.9173431396484 375

```
WITH genai.vector.encode(
    "What are the risks that Microsoft faces?",
    "OpenAI",
    { token: $token }) AS userEmbedding
CALL db.index.vector.queryNodes('chunkEmbeddings', 6, userEmbedding)
YIELD node, score

MATCH (node)-[:FROM_DOCUMENT]->(doc:Document)-[:FILED]
-(company:Company)-[fr:FACES_RISK]->(risk:RiskFactor)

RETURN company, fr, risk
```



DEMO

Vector

```
# Create Vector
chunk_vector = Neo4jVector.from_existing_index(
    embedding_model,
    graph=graph,
    index_name="chunkEmbeddings",
    embedding_node_property="embedding",
    text_node_property="text"
)
```

```
# Define a tool to retrieve financial documents
@tool("Retrieve-financial-documents")
def retrieve_docs(query: str):
    """Find details about companies in their financial
documents."""
    # Use the vector to find relevant documents
    context = chunk_vector.similarity_search(
        query,
        k=3,
    )
    return context

# Add the tools to the agent
tools = [retrieve_docs]

agent = create_react_agent(
    model,
    tools
)
```

Microsoft faces a variety of risks that could impact its operations and financial results. These risks are categorized into strategic and competitive risks, operational risks, legal, regulatory, and litigation risks:

1. **Strategic and Competitive Risks**:
   - Intense competition across all markets, including technology and platform-based ecosystems.
   - Competitors range in size and capabilities, creating barriers and rapid changes in technology.
   - Challenges in maintaining scale and profitability in marketplace ecosystems against competitors who control both hardware and software elements.

2. **Operational Risks**:
   - Risks related to product design, manufacturing defects, and reliability issues affecting customer trust and potential legal liabilities.
   - Supply chain dependencies, including limited suppliers for certain components and potential disruptions that may affect sales and margins.
   - Security vulnerabilities in third-party components, leading to potential costs, liability claims, and reputational damage.
   - Challenges in protecting personal data from breaches, misuse, or regulatory liabilities. Increased data breaches heighten the risk environment despite security controls.

3. **Legal, Regulatory, and Litigation Risks**:
   - Scrutiny from government agencies and regulatory actions related to competition laws, which are actively enforced, particularly in the EU, U.S., and China.
   - Potential regulatory restrictions could limit the design and marketing of products or result in fines.
   - Risks related to data and privacy, where improper disclosure could harm reputation and lead to legal exposure.
   - Challenges in protecting information and content from unauthorized use, which may weaken product value.

These risks highlight several pressures across different dimensions of Microsoft's operations, from competitive challenges and technological innovations to regulatory scrutiny and data security issues.

Vector

```
# Define the retrieval query
retrieval_query = """
MATCH (node)-[:FROM_DOCUMENT]-(doc:Document)-[:FILED]-(company:Company)
RETURN
    node.text as text,
    score,
    {
        company: company.name,
        risks: [ (company:Company)-[:FACES_RISK]->(risk:RiskFactor) |
risk.name ]
    } AS metadata
ORDER BY score DESC
"""

# Create Vector
chunk_vector = Neo4jVector.from_existing_index(
    embedding_model,
    graph=graph,
    index_name="chunkEmbeddings",
    embedding_node_property="embedding",
    text_node_property="text",
    retrieval_query=retrieval_query,
)
```

```
# Define a tool to retrieve financial documents
@tool("Retrieve-financial-documents")
def retrieve_docs(query: str):
    """Find details about companies in their financial
documents."""
    # Use the vector to find relevant documents
    context = chunk_vector.similarity_search(
        query,
        k=3,
    )
    return context

# Add the tools to the agent
tools = [retrieve_docs]

agent = create_react_agent(
    model,
    tools
)
```

Microsoft faces a variety of risks that can potentially impact its operations and financial health. Key risks include:

1. **Competition Risks**: Intense competition across all markets, especially in the technology sector. Competitors include both global diversified companies and small specialized firms. Competition among platform-based ecosystems and business models also poses a strategic risk.

2. **Technological Risks**: Risks related to the cost of converting ideas into software products, investment in AI development, and open source competition. There are also execution risks in cloud-based services and enhanced risks in specific industries.

3. **Security and Cybersecurity Risks**: Threats include cyberattacks, security vulnerabilities in products and services, and data breaches. Specific concerns are nation-state attacks, supply chain cyberattacks, and emerging cybersecurity regulations.

4. **Legal and Regulatory Risks**: Includes compliance issues, such as non-compliance with data privacy laws (like GDPR), legal challenges to data transfer frameworks, anti-corruption law violations, and increased costs or fines from trade laws.

5. **Geopolitical and Economic Risks**: These include geopolitical instability, regional epidemics or global pandemics, climate change, and adverse economic or market conditions.

6. **Operational Risks**: Risks associated with supply chain disruptions, quality or supply problems with hardware products, and design defects leading to product liability.

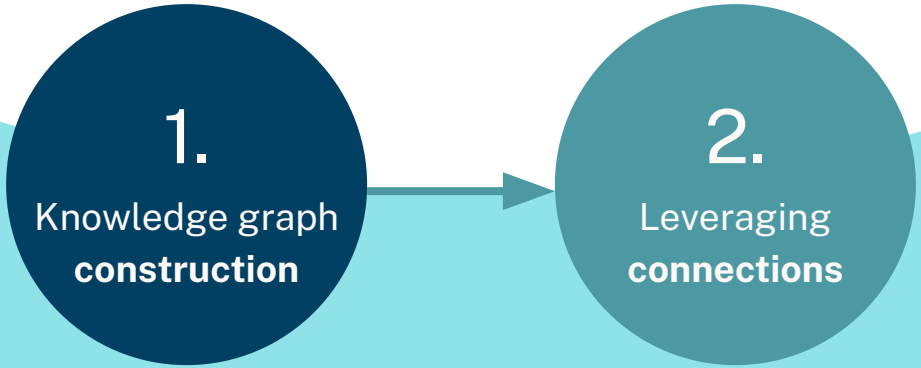
7. **Financial Risks**: Currency risks, interest rate risks, market price risks, and credit exposures. There are unresolved IRS audit issues related to transfer pricing and income taxes in various jurisdictions.

8. **AI and Innovation Risks**: Risks from AI scenarios include reputational or competitive harm, ineffective AI development or deployment practices, ethical issues, and broad societal impacts.

9. **Brand and Reputational Risks**: Reputational harm from GDPR violations, infringement claims, or data privacy and security concerns.

These risks are detailed in Microsoft's financial documents and reflect the comprehensive challenges the company faces in today's dynamic and complex environment.

GraphRAG



Extracting entities and relationships from unstructured text

Using relationships to provide context and answer questions

NODES 2025 - Neo4j Online Developer Conference

NODES is back for its seventh year! Developers and data pros, get ready to learn the latest graph best practices.

When: Thursday, **November 6, 2025**

Format: Live and unique sessions from community and Neo4j experts
24 hours of technical talks **across all time zones**

Speakers: Check out the Featured Speakers. Full Agenda coming up Sept 3

Themes:
AI Engineering: GenAI, RAG and Agents
App Dev : Libraries, Frameworks, and Platforms
Data Intelligence: ML, Algorithms, and Models
Knowledge Graphs: Data Integration, Management, and Visualization.

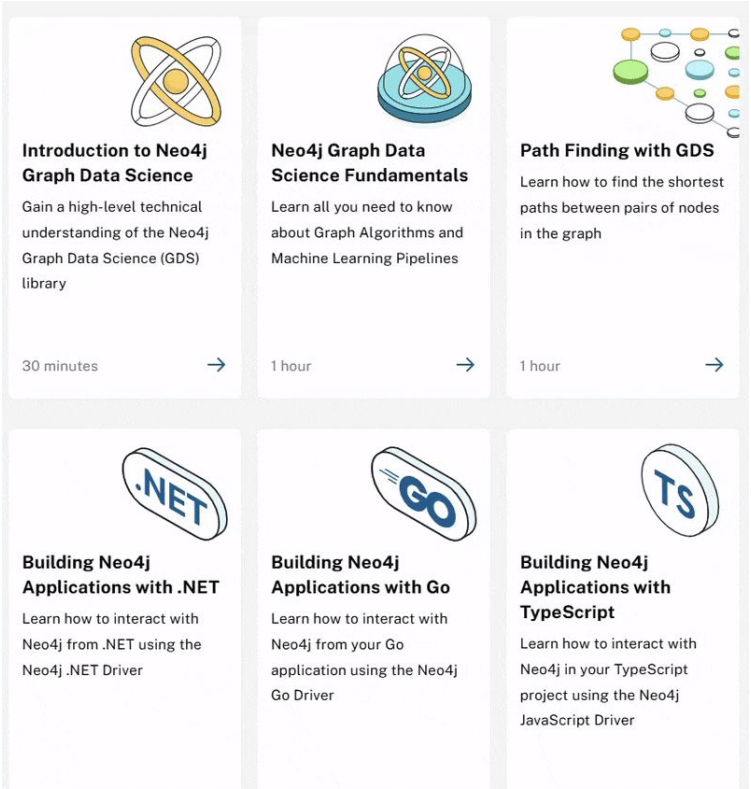
Learn More: neo4j.com/nodes

Learn more with

neo4j | GraphAcademy

- **Completely Free**
- **Hands-on Courses** teaching Neo4j Fundamentals, Cypher, Drivers and Graph Data Science
- **Curated Learning Paths** catering for everyone from beginners to experts
- **Free Certifications**

graphacademy.neo4j.com



graphacademy.neo4j.com