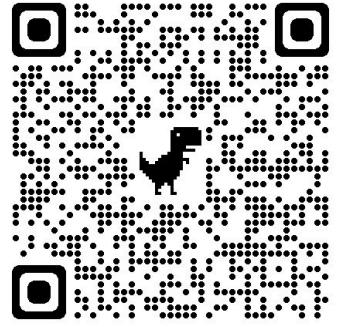




neo4j

GraphRAG Workshop



Workshop Rules

Ask questions straight away, this is an interactive session

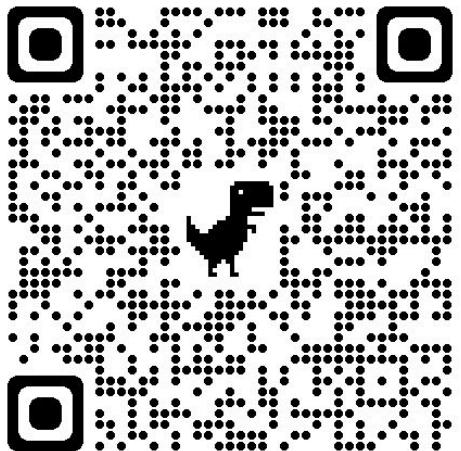
Raise your hand if you are stuck

Have fun

Before We Start

1/ Create a **blank-graph-data-science** Neo4j Sandbox at
sandbox.neo4j.com

2/ Open the notebook in Colab (needs a **google account!**)



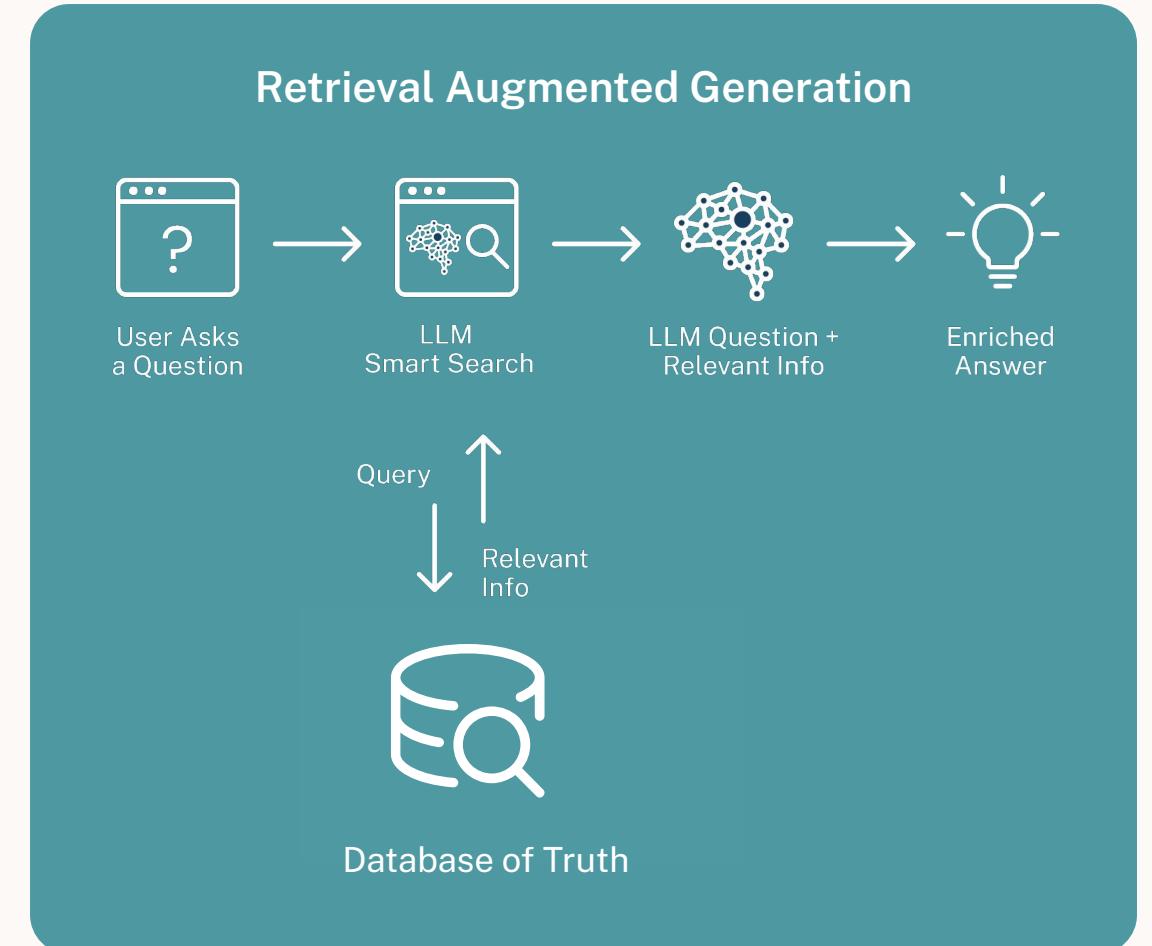
<https://github.com/neo4j-product-examples/genai-workshop>

open **genai-workshop.ipynb**

Retrieval-Augmented Generation Is Becoming an Industry Standard

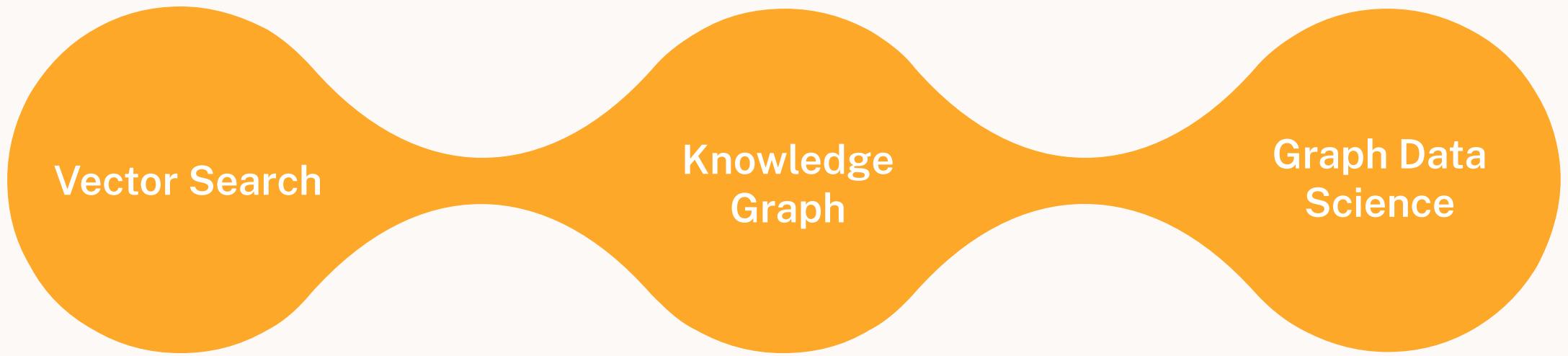
RAG augments LLMs by retrieving up-to-date, contextual external data to inform responses:

- Reduce hallucinations with verified data
- Provide domain-specific, relevant responses
- Enable traceability back to sources



GraphRAG with Neo4j

Unify vector search, knowledge graph and data science capabilities to improve RAG quality and effectiveness



Find similar documents
and content

Identify entities
associated to content and
patterns
in connected data

Improve & enrich GenAI
insights. Discover new
relationships and entities



**Today
We are Choosing
Clothes**

This is Eva

For her birthday, Eva wants
a **Halter Neck Top**

How do you choose what to buy?



This is Eva

For her birthday, Eva wants
a **Halter Neck Top**

How do you choose what to buy?

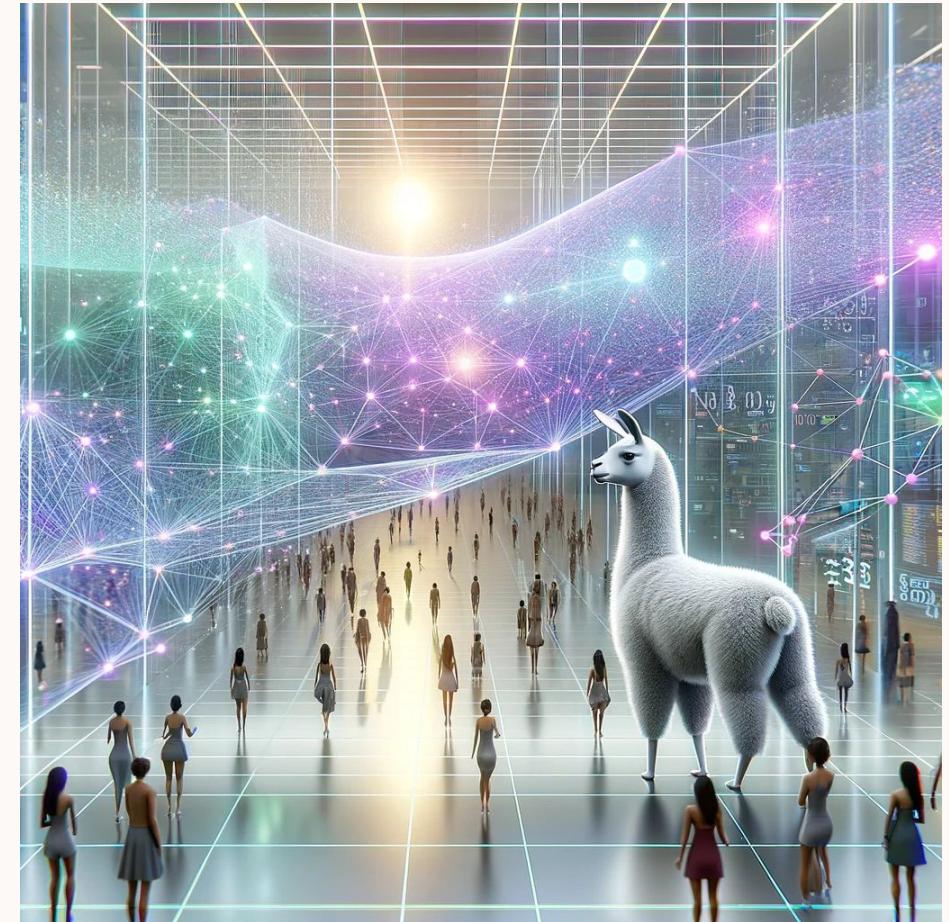
- 1/ Get a list of Halter Neck Top
- 2/ Match the Halter Neck Top for Eva
- 3/ Accessorize or add another item



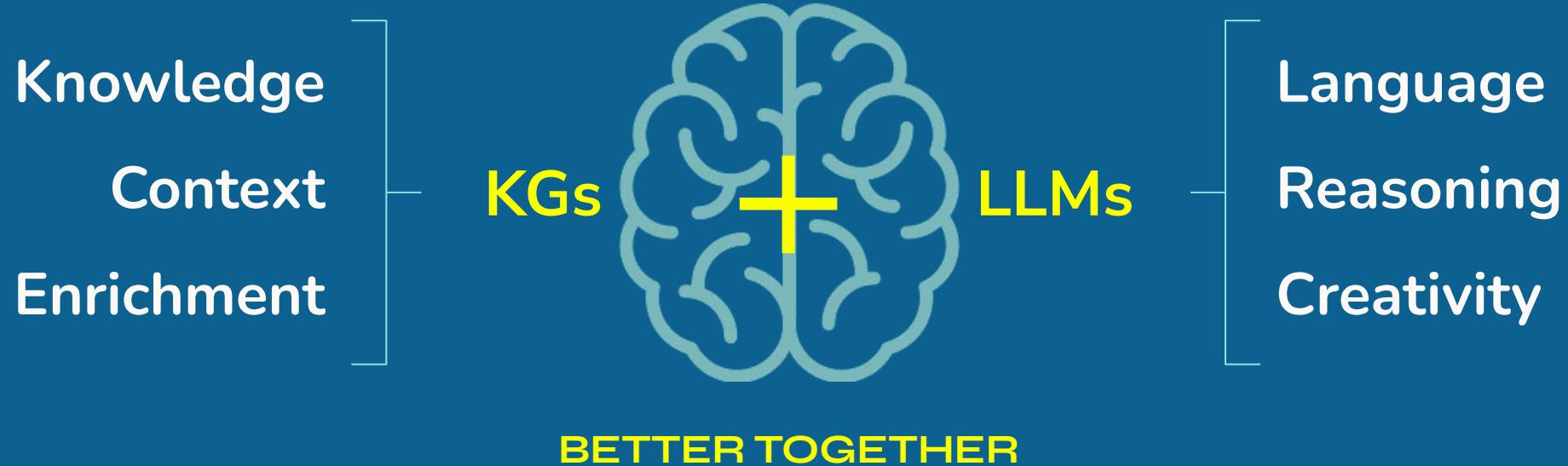
At Scale: Let's Build an AI Fashion Assistant

This Requires:

1. Data: Kaggle H&M
2. Search & Retrieval
3. Context for Personalization
4. Recommendation Engine
5. LLM Powered Content Generator



Solution: LLMs and Knowledge Graphs



We'll Build a Targeted, Personalized Content Generator

Message Generator 😊

Customer ID
daae10780ecd14990ea190ale9917da33fe96cd8cfa5e80b67b4600171aa77e0

Time Of Year
Nov, 2023

Customer Name
Alex Smith

Customer Interest(s)
Oversized Sweaters

[Clear](#) [Submit](#)

Dear Alex Smith,

I hope this email finds you well. As the weather gets cooler, it's the perfect time to update your wardrobe with cozy and stylish oversized sweaters. I wanted to share with you some of our top picks for this season:

1. Queen Sweater: This lightweight sweatshirt fabric sweater features ribbing around the neckline, cuffs, and hem. It's a perfect choice for a casual yet chic look. You can find it [here](#).
2. Jess oversize LS: Made from a soft jersey cotton blend, this oversized top with dropped shoulders and long sleeves is both comfortable and trendy. You can check it out [here](#).
3. Petar Sweater(1): If you're looking for an oversized top in sturdy sweatshirt fabric, this is the one for you. It has dropped shoulders and ribbing around the neckline, cuffs, and hem, with a soft brushed inside. Find it [here](#).
4. Family Crew Ladies: This sweatshirt fabric top is perfect for a cozy and relaxed look. It features dropped shoulders, long sleeves, and ribbing around the neckline, cuffs, and hem. You can find it [here](#).
5. Irma sweater: Add a touch of print to your wardrobe with this top in printed sweatshirt fabric. It has dropped shoulders, long sleeves, and ribbing around the neckline, cuffs, and hem. Check it out [here](#).

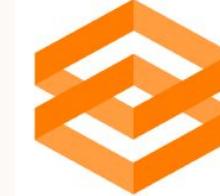
In addition to oversized sweaters, you may also be interested in:

1. GABBE T-shirt: Made from soft, printed slab cotton jersey, this t-shirt features a ribbed neckline. You can find it [here](#).
2. Runar sweater: This oversized top in soft sweatshirt fabric offers a relaxed fit with low dropped shoulders, extra-long sleeves, and ribbing around the neckline, cuffs, and hem. It's perfect for a cozy

Tools



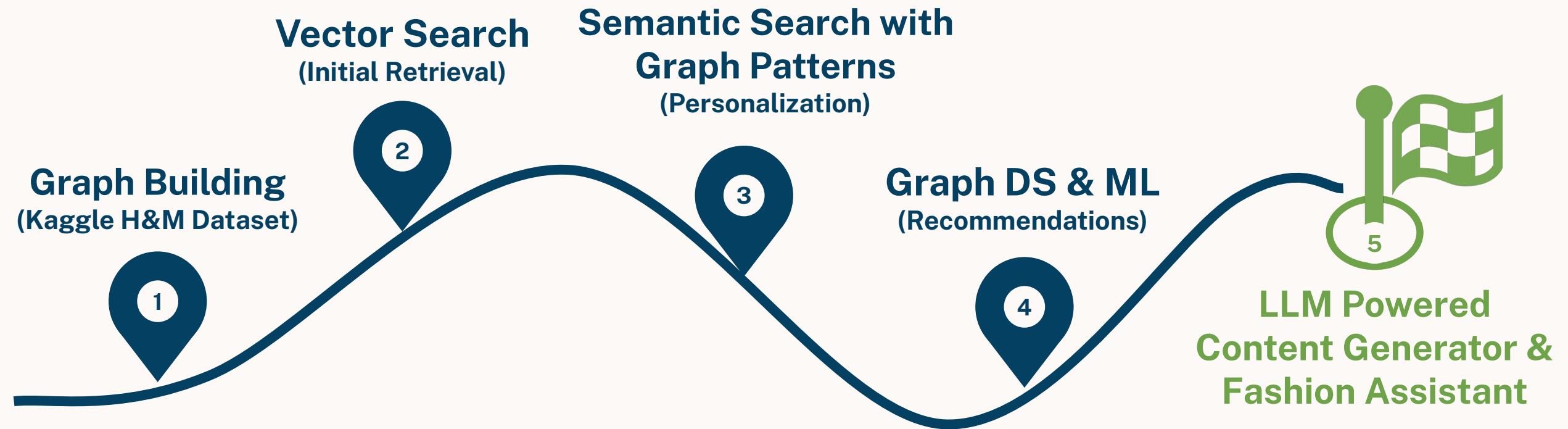
neo4j

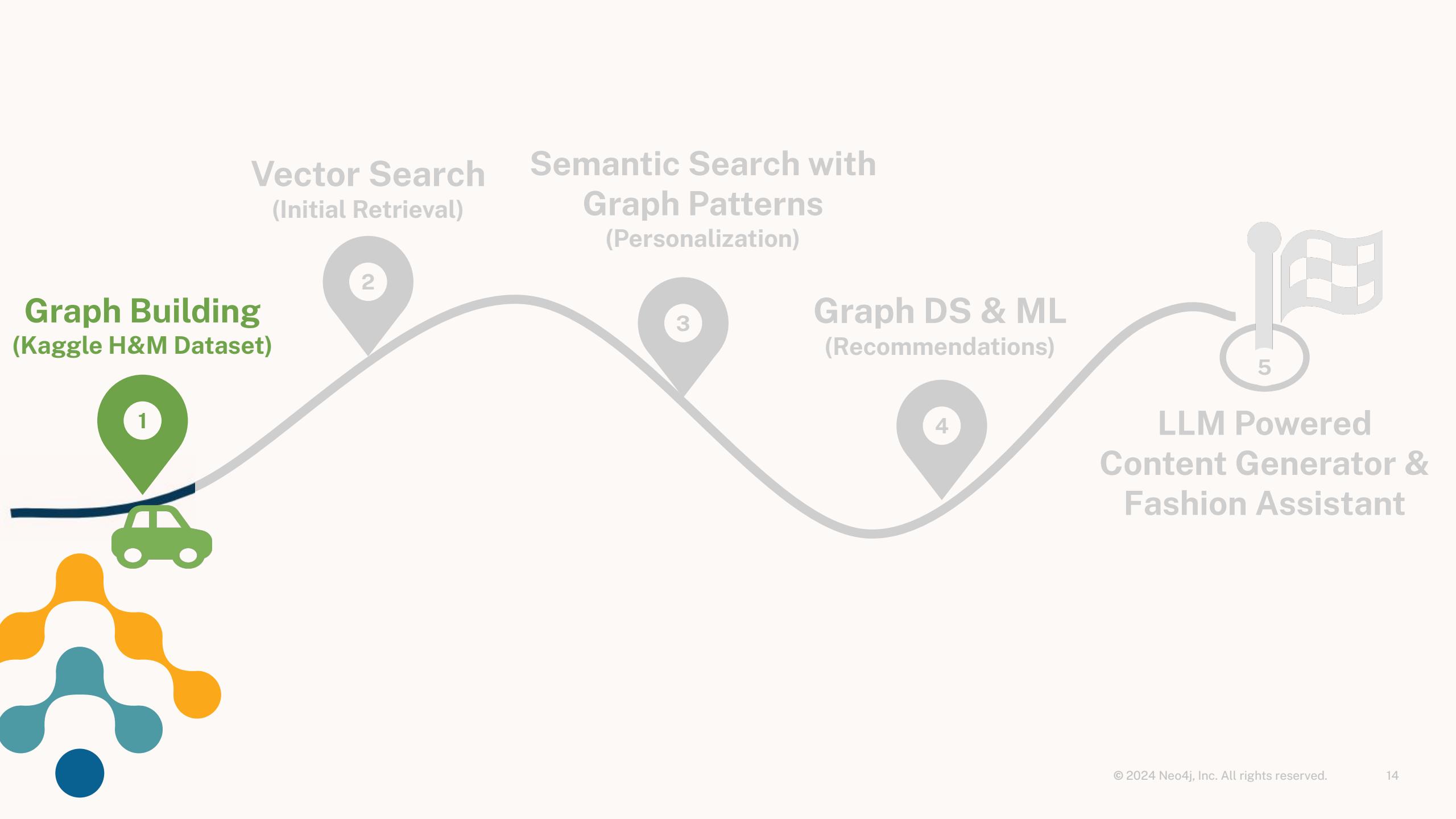


LangChain

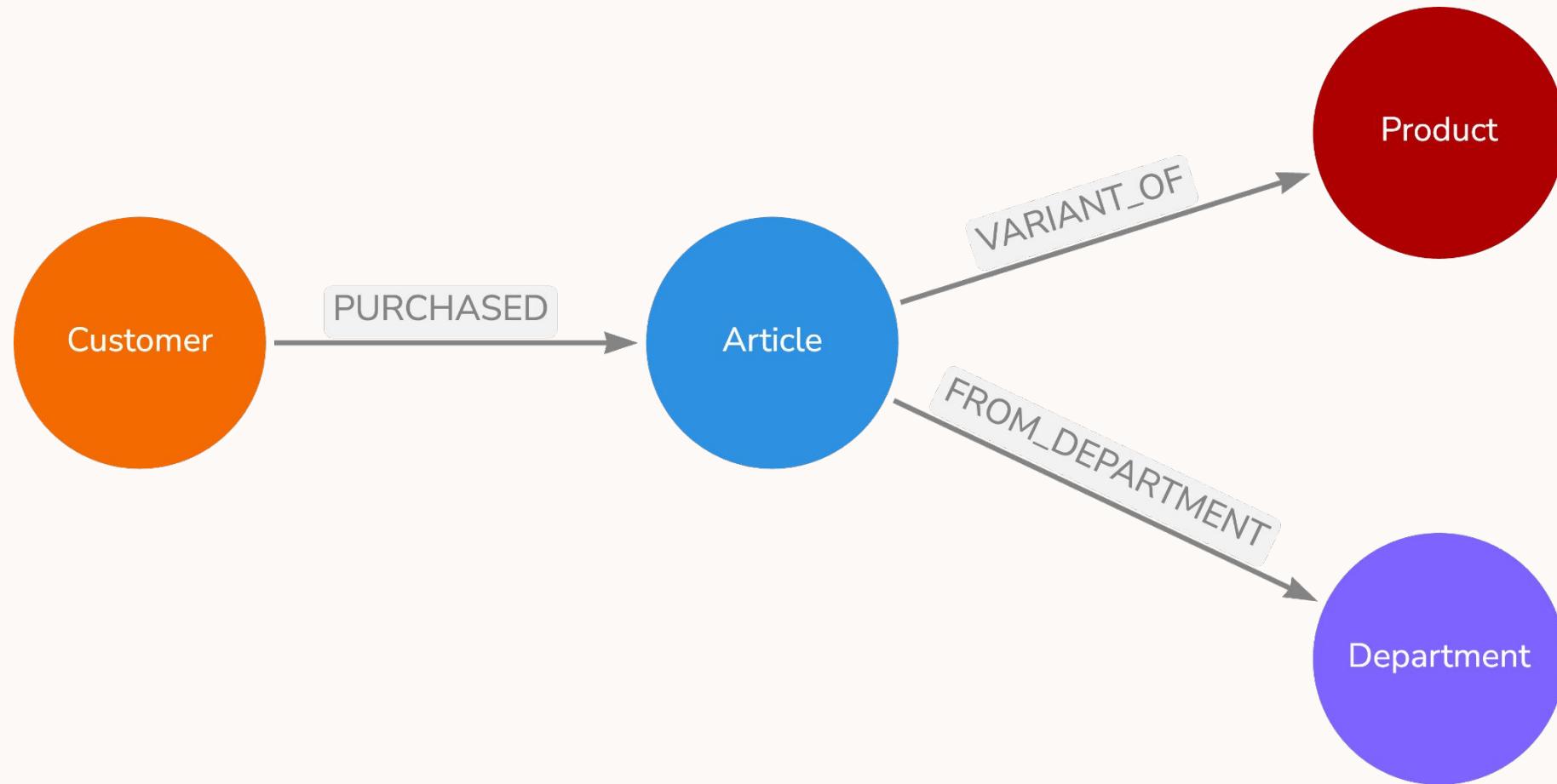


Journey

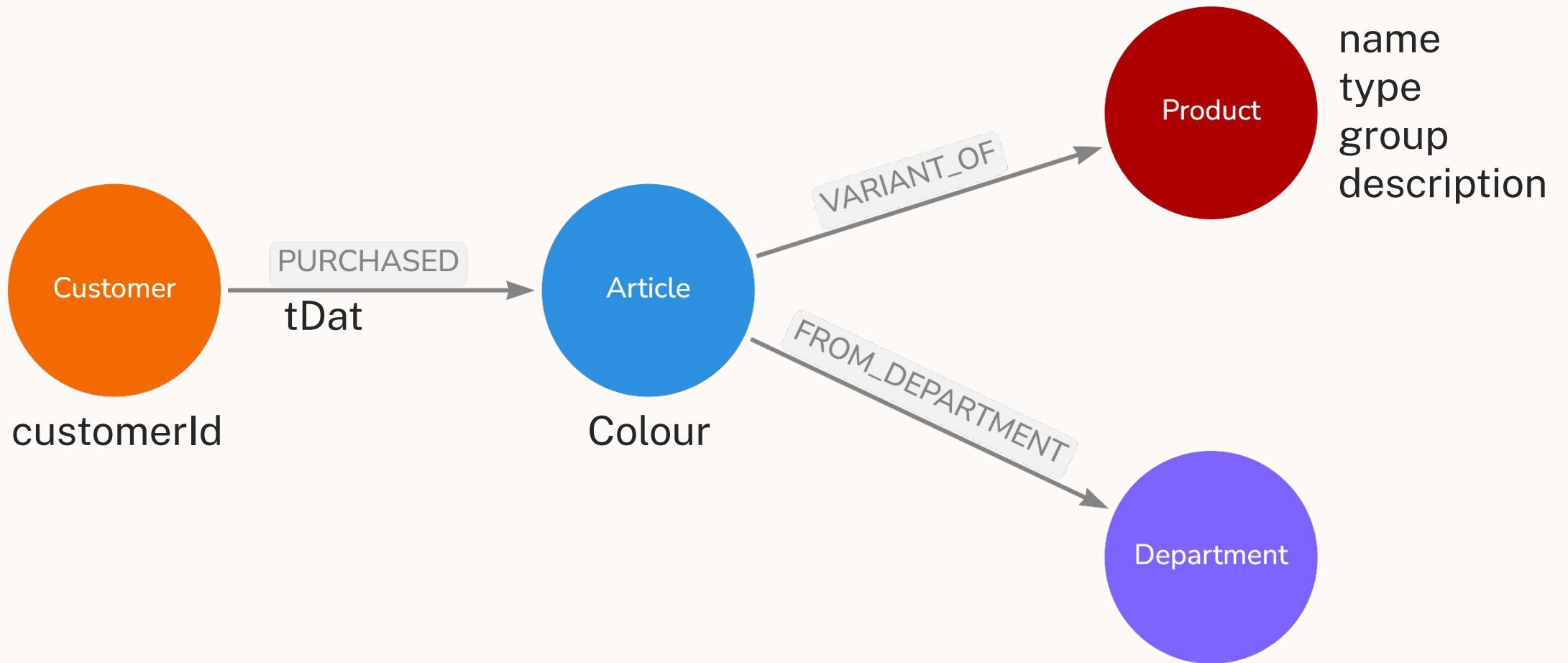




Data Model

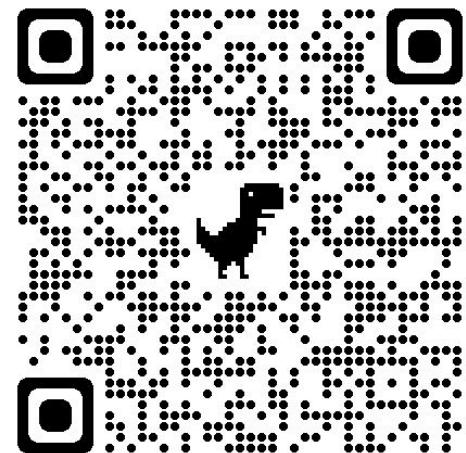


Data Model



Now We Can Start

- 1/ Create a **blank** Neo4j GDS Sandbox at sandbox.neo4j.com
- 2/ Open the notebook in Colab (needs a **google account!**)

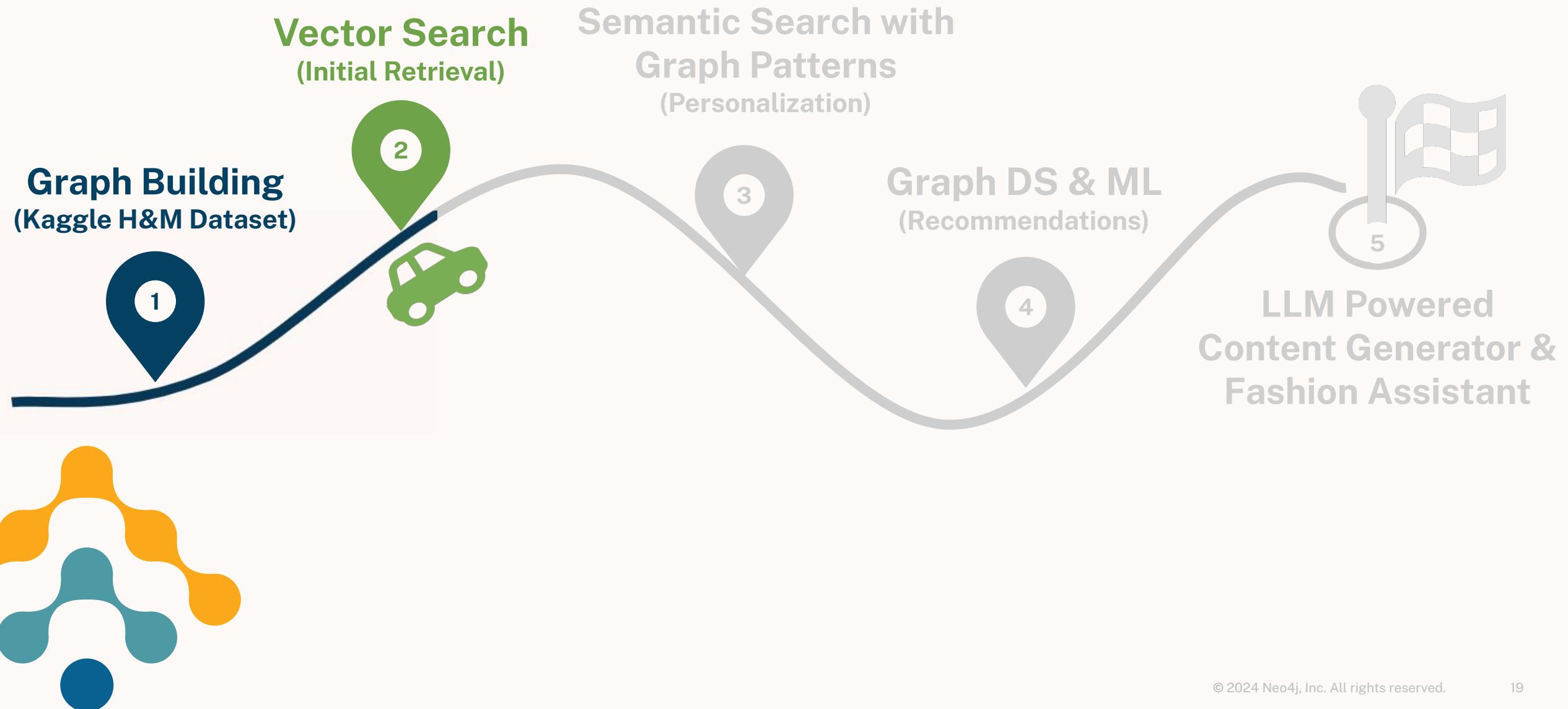


<https://github.com/neo4j-product-examples/genai-workshop>
open **genai-workshop.ipynb**



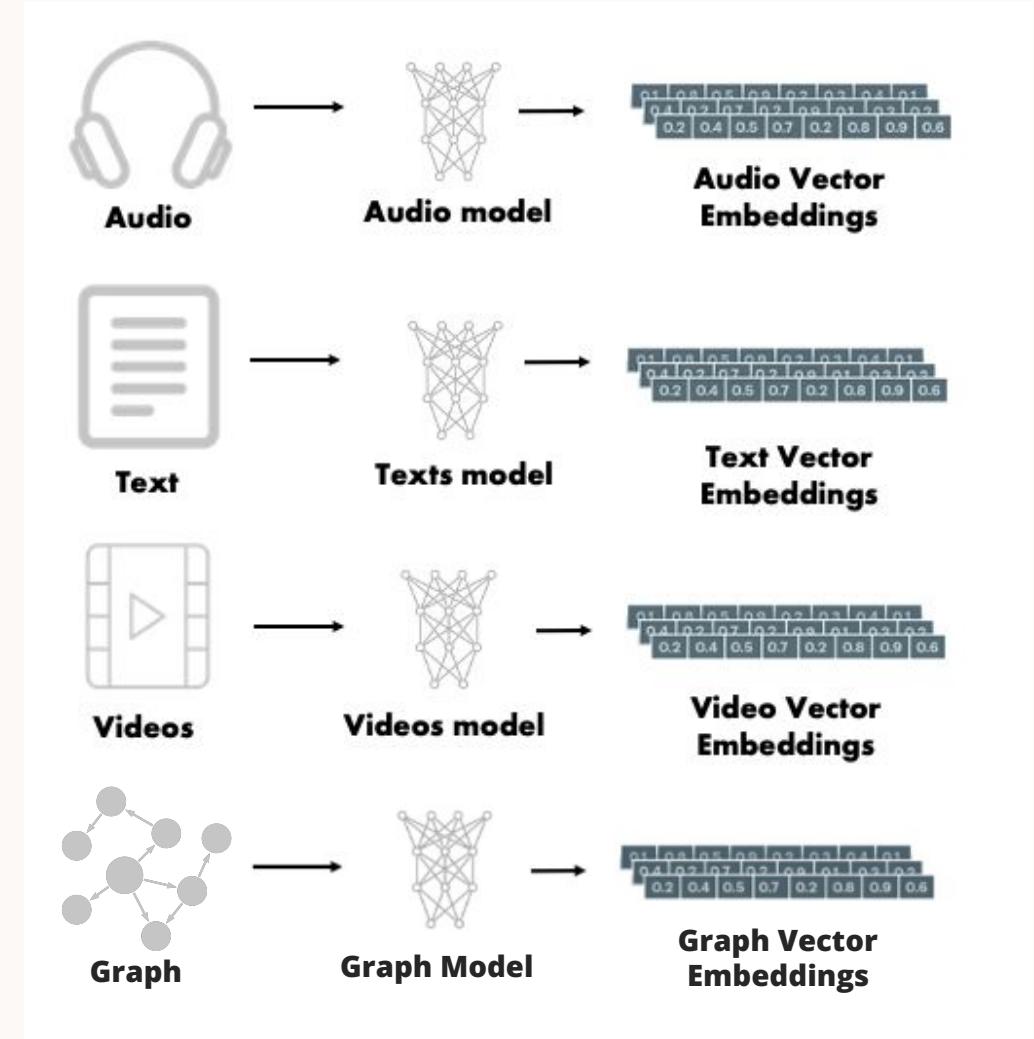
1/ Graph Building

Stop before **Vector Search**



What are Embeddings?

- A type of data compression
- Transform messy data (text, images, audio, etc.) into a compact format for ML algorithms
- Most often numeric **vectors** (a.k.a arrays) with 100s or 1000s of elements
- Preserve information such that “similar” items have proportionally “similar” embedding vectors
- Similarity is measured with vector algorithms (cosine, euclidean, etc.)



What Does “Similarity” Mean?

It Depends:

- Text Embeddings => Semantic Similarity, the meaning behind a text sequence
- Graph Embeddings => similarity in position or structure in a graph - can have semantic meaning too

What will we use embedding for?

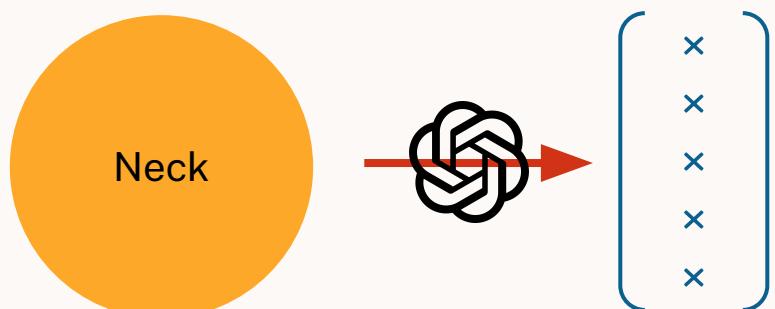
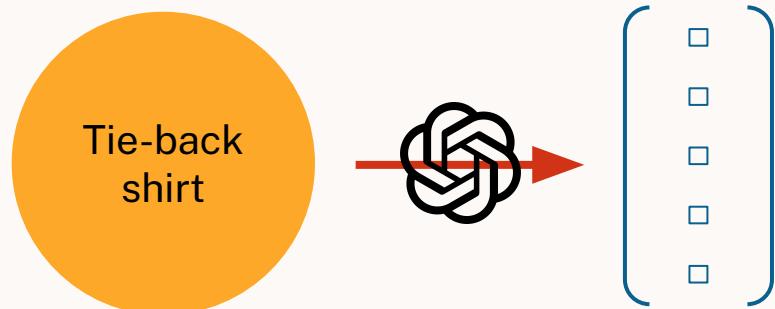
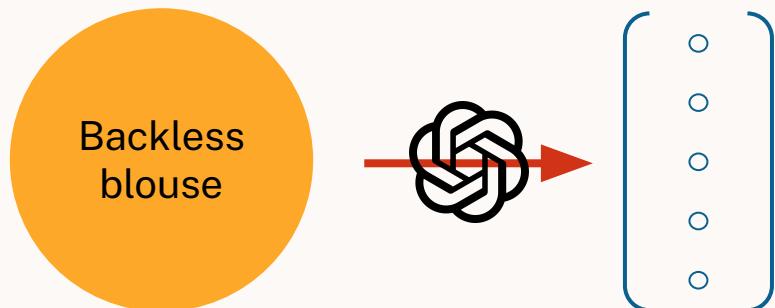
1. **(now) Vector search** using text embeddings
2. **(later) Recommendations** using graph node embeddings

Search & Vectors in Neo4j

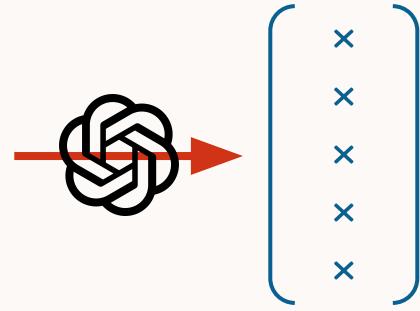
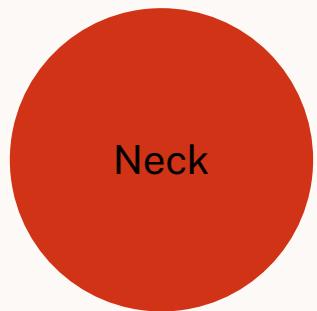
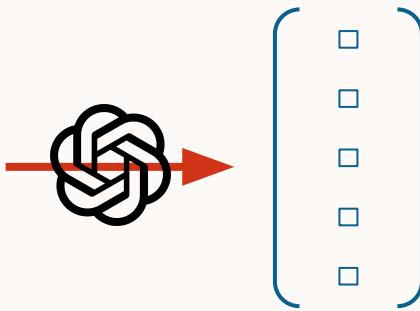
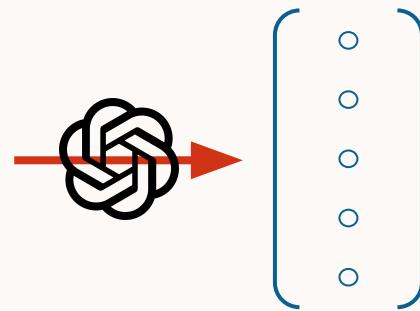
Neo4j makes search efficient through a variety of indexes including

- **Range:** General index for predicates based on equality and range. Numeric, dates, etc.
- **Point:** Predicts on geospatial points like distance bounding boxes, etc.
- **Text:** Predicates on strings like contains, ends with, etc.
- **FullText:** Text search based on tokenization and analyzers
- **Vector:** ANN (Approximate Nearest Neighbor) search on vectors

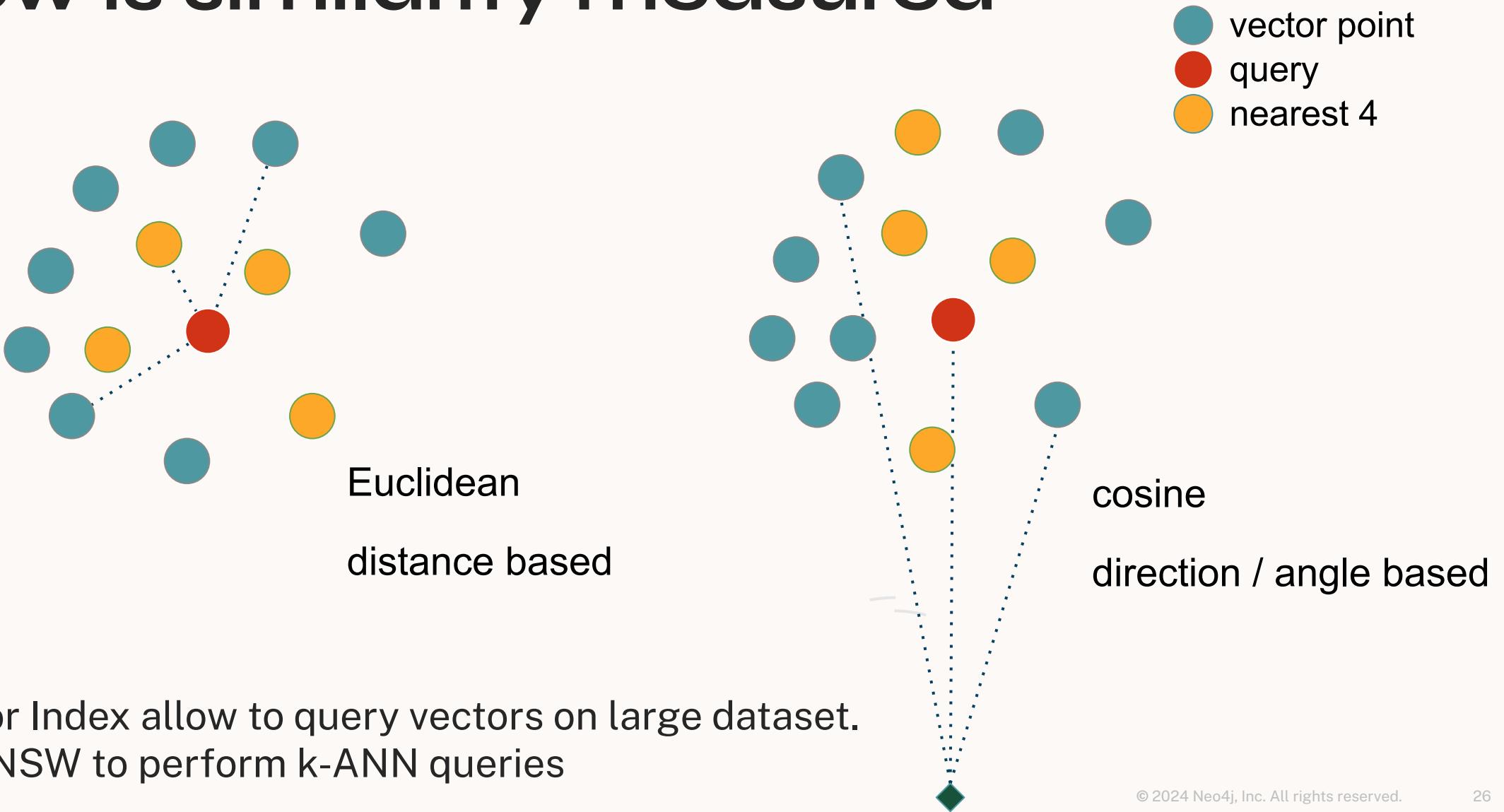
Vector search



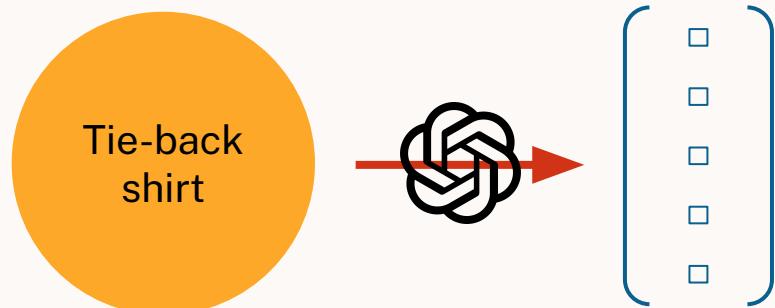
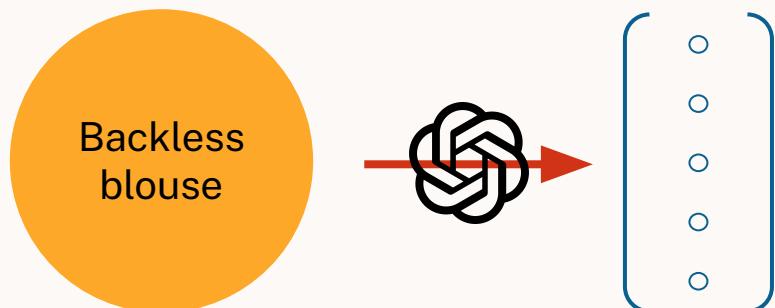
Vector search



How is similarity measured



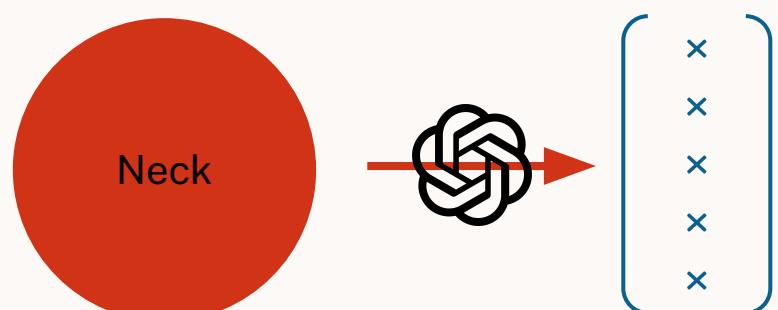
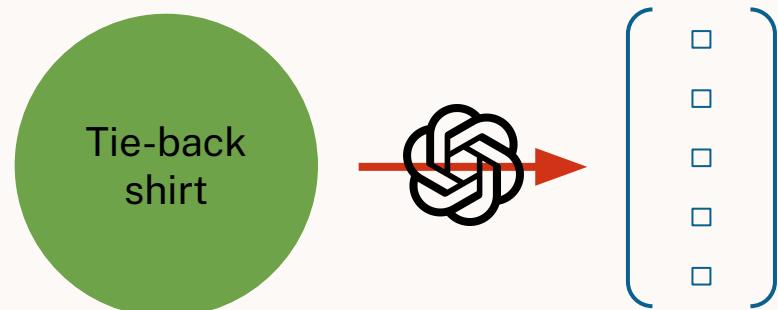
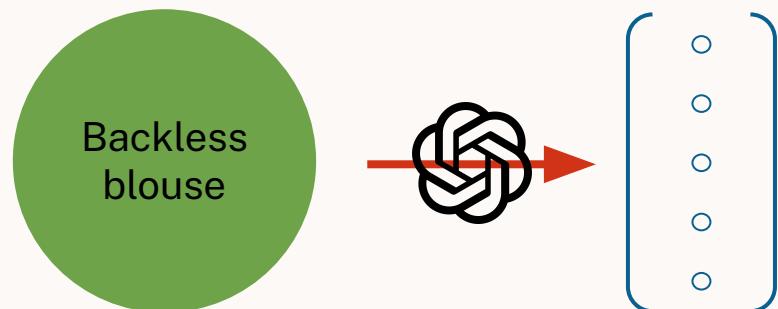
Vector search



CREATE VECTOR INDEX

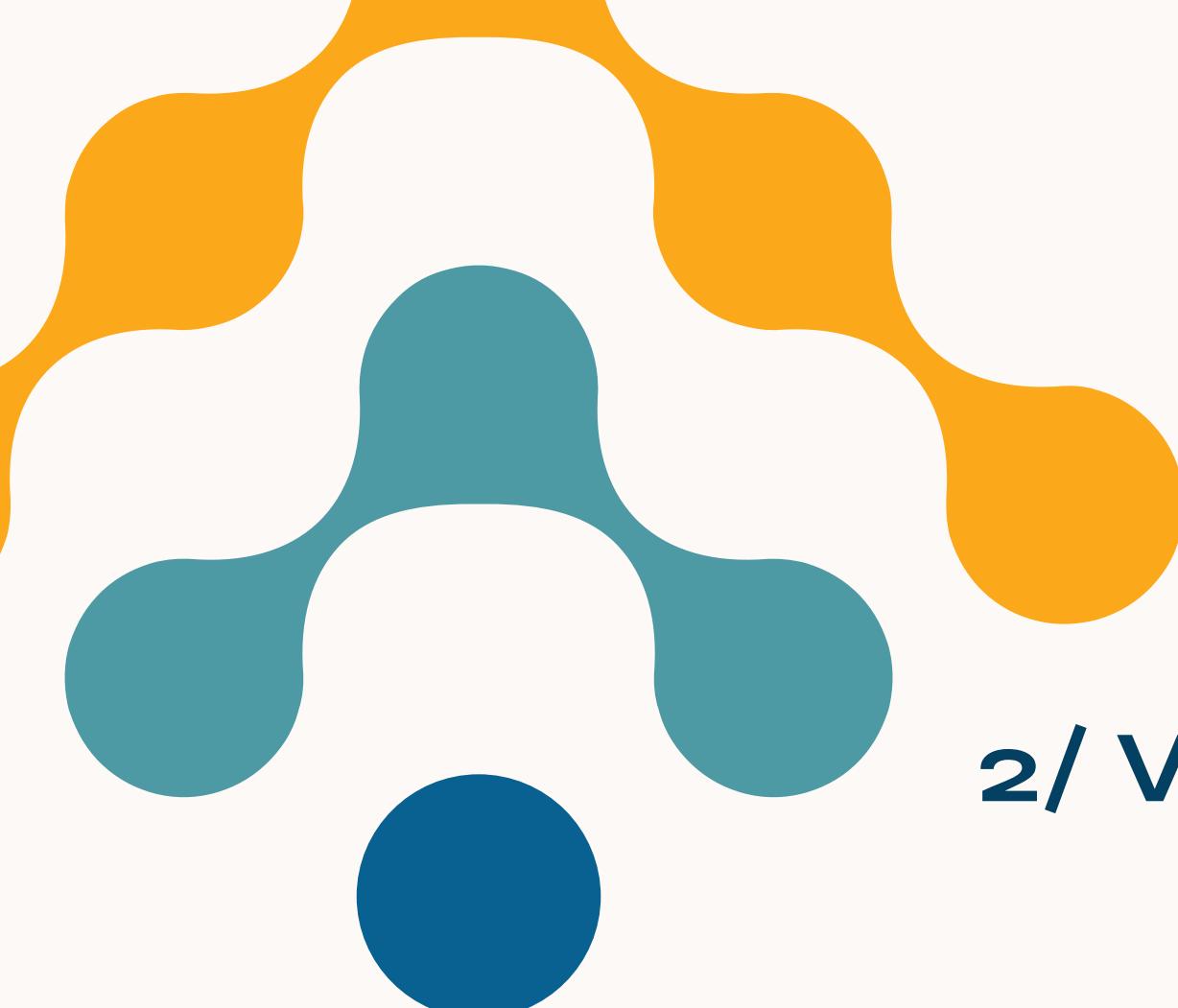
`db.index.vector.queryNodes`

Vector search



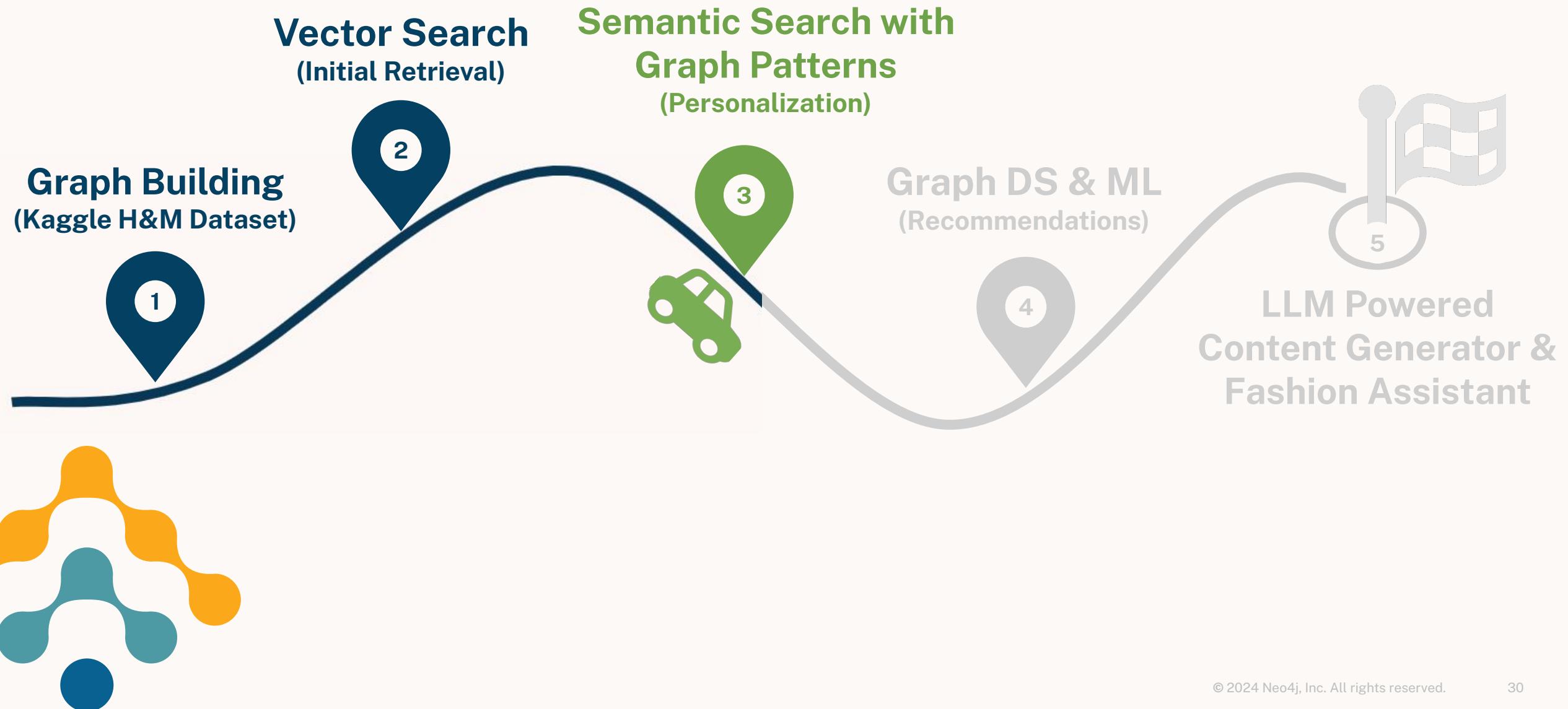
CREATE VECTOR INDEX

`db.index.vector.queryNodes`



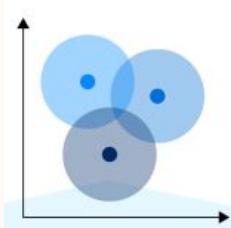
2/ Vector Search

Stop before **Semantic Search with Graph Patterns**
(Personalization)



Neo4j & Semantic Search

Vector Similarity Search



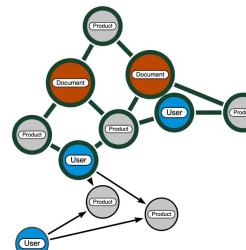
Find relevant documents and content for user queries

Vector Search

Graph Database

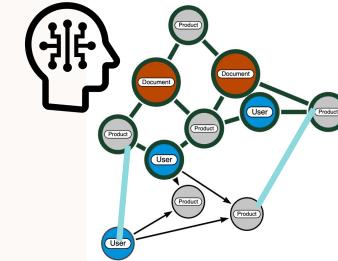


Graph Pattern Matching



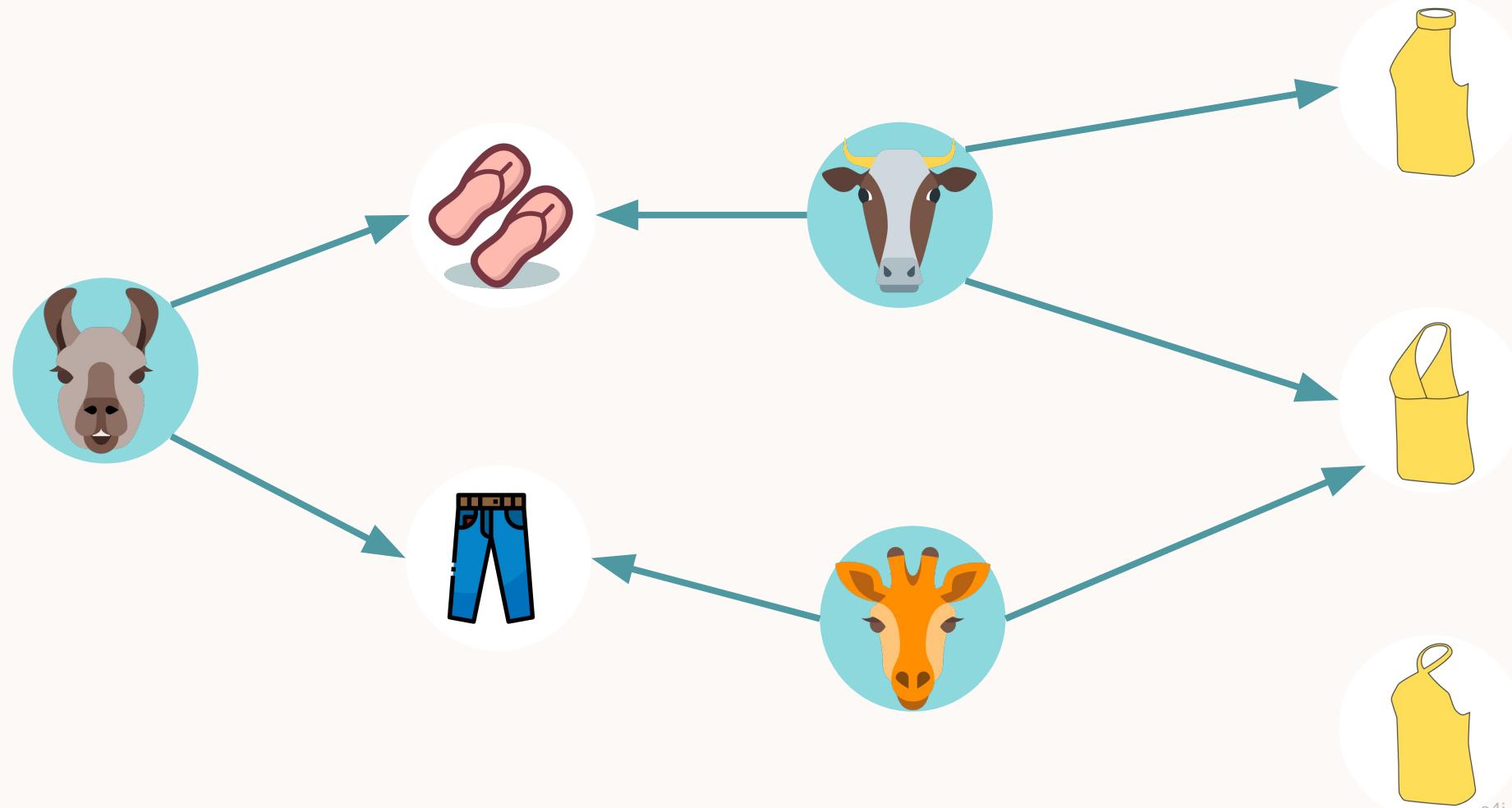
Find entities associated to content and patterns in connected data.

Knowledge Graph DS/ML

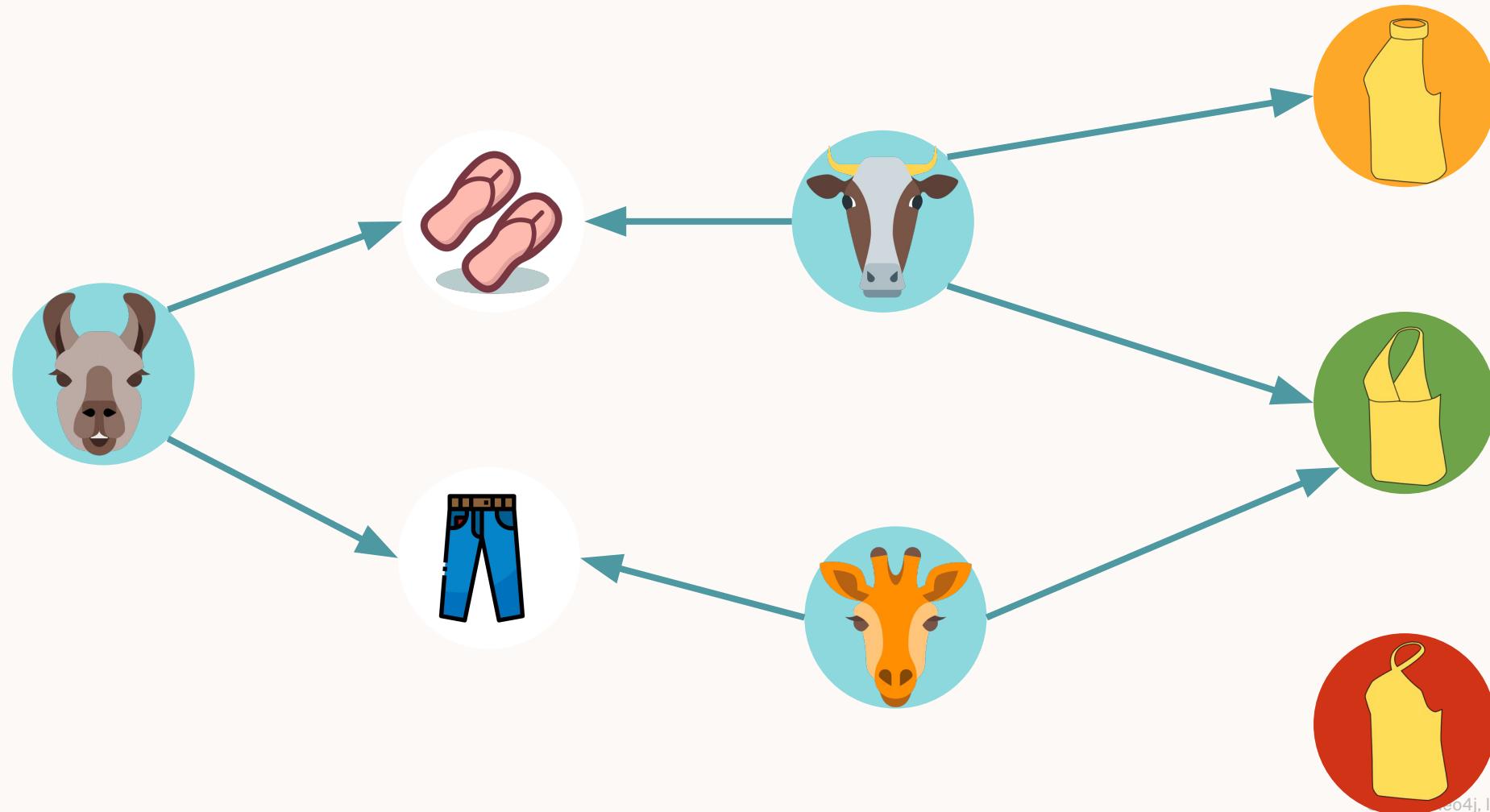


Improve search relevance & insights by enhancing a Knowledge Graph. Use graph algorithms and ML to discover new relationships, entities, and groups.

Semantic search with graph pattern



Semantic search with graph pattern

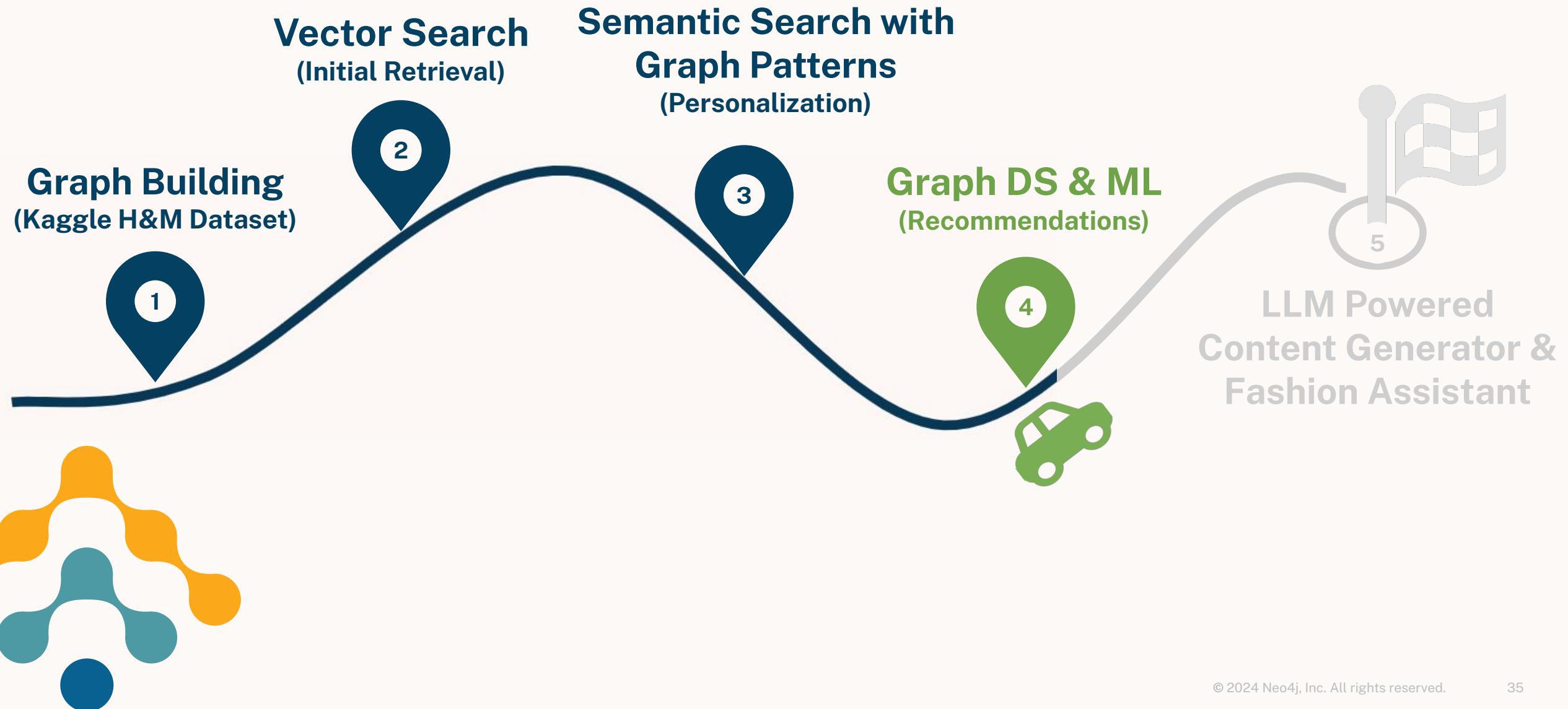




3/ Semantic Search with Graph Patterns

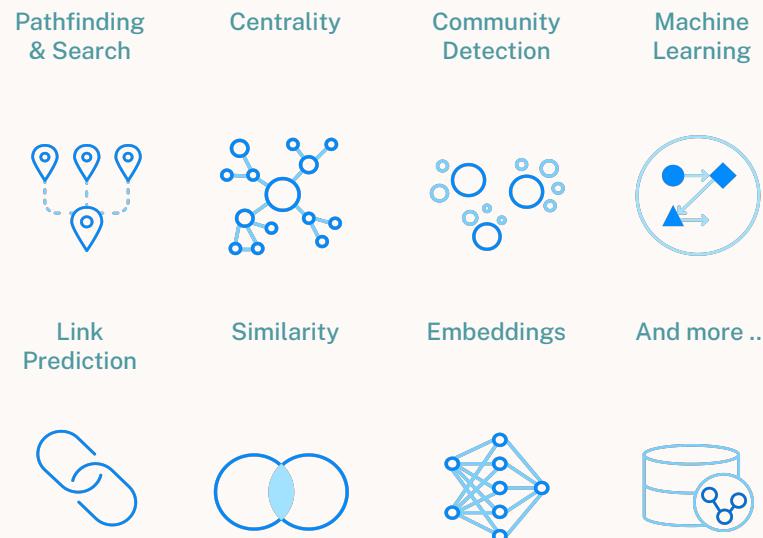
Personalization & Context

Stop before **Graph Data Science & ML (Recommendations)**



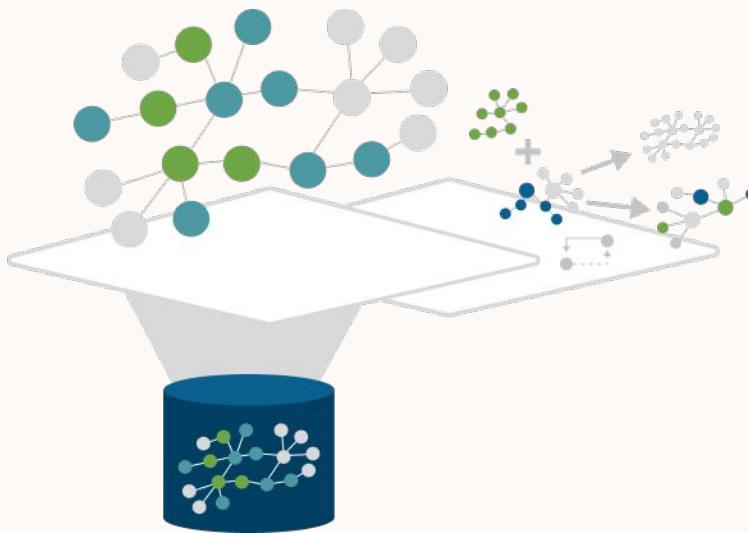
Neo4j Graph Data Science

The Largest Catalog of Graph Algorithms



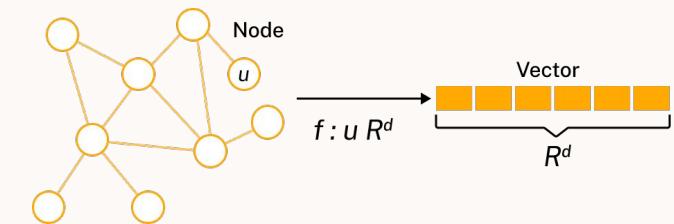
Over 65 pretuned, parallelized algorithms.

Native Graph Catalog and Analytics Workspace



Iterate fast with different data sets, models, and version trained models.

Graph Embeddings for Machine Learning



Bring the context of your connected data into a format that other pipelines can ingest.

Graph Algorithms

Version 0.1

based on neo4j Graph Data Science 2.5

by Hari Gurumoothi
@weetofhari

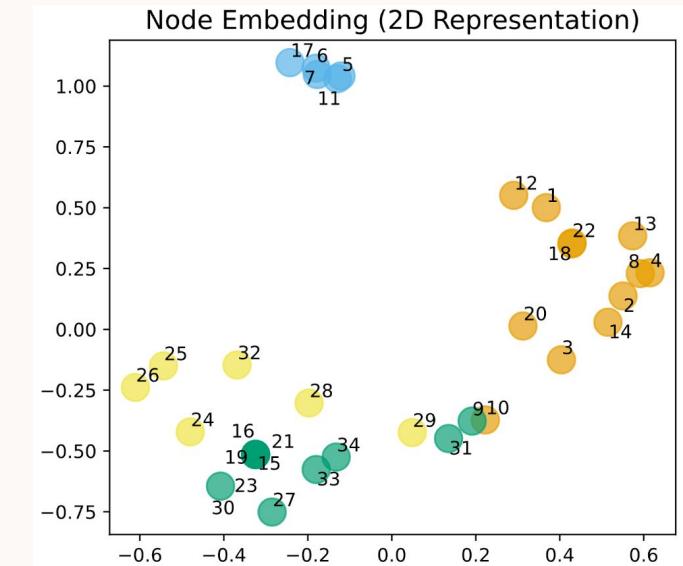
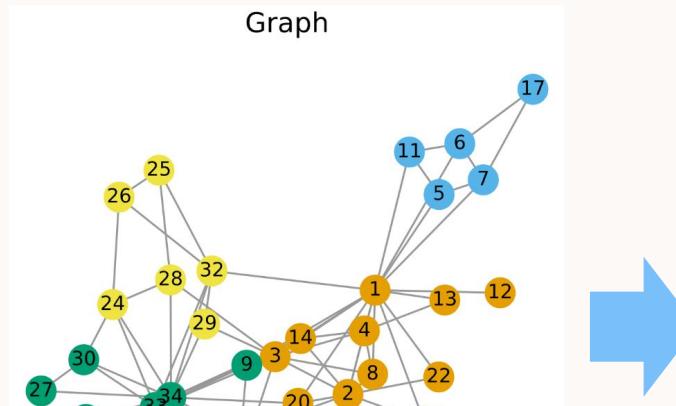
α Indicates that the feature is experimental and might be changed or removed at any time.

β Indicates that the feature is a candidate for the production-quality tier.

Lu Louvain												
Km K-Means Clustering	Kc K-Core Decomposition	Dt Dijkstra Source-Target Shortest Path	Dj Dijkstra Single Source Shortest Path									
Tc Triangle Count	K1 K-1 Coloring	Ds Delta Stepping	Df Depth First Search									
Cm Conductance metric	Mm Modularity metric	Bf Breadth First Search	Ys Yen's Shortest Path	Pr Page Rank								
Mo Modularity Optimization	Sc Strongly Connected Components	As A* Shortest Path	Bp Bellman-Ford Short Path	Ar Article Rank	Cc Closeness Centrality	Ja Jaccard Similarity				Fr FastRP		
Lp Label Propagation	Lc Local Clustering Coefficient	Ms Minimum Spanning Tree	Rw Random Walk	Bc Betweenness Centrality	Dc Degree Centrality	Op Overlap Similarity	Ed Euclidean Distance	Cn Common Neighbors	Ad Adamic Adar	Hg HashGNN		
Le Leiden	Wc Weakly Connected Components	Md Minimum Directed Steiner Tree	Mw Min Weight k-Spanning Tree	Ce CELF	Ec Eigenvector Centrality	Co Cosine Similarity	Kn K-nearest Neighbors	Pa Preferential Attachment	Tn Total Neighbors	Nv Node2Vec	Lo Longest Path	
Sp Speaker-Listener LP	Ak Approx. Max. k-cut	Ap All Pairs Shortest Path	Ld Longest Path for DAG	Hi HITS	Hc Harmonic Centrality	Pe Pearson Similarity	Ns Node Similarity	Ra Resource Allocation	Sc Same Community	Gs GraphSAGE	Ts Topological Sort	
Community Detection	Path Finding & Search	Centrality	Similarity	Topological Link Prediction	Node Embeddings	DAG Algorithms						

Node Embeddings

The representation of nodes as low-dimensional vectors that summarize their graph position, the structure of their local graph neighborhood as well as any possible node features



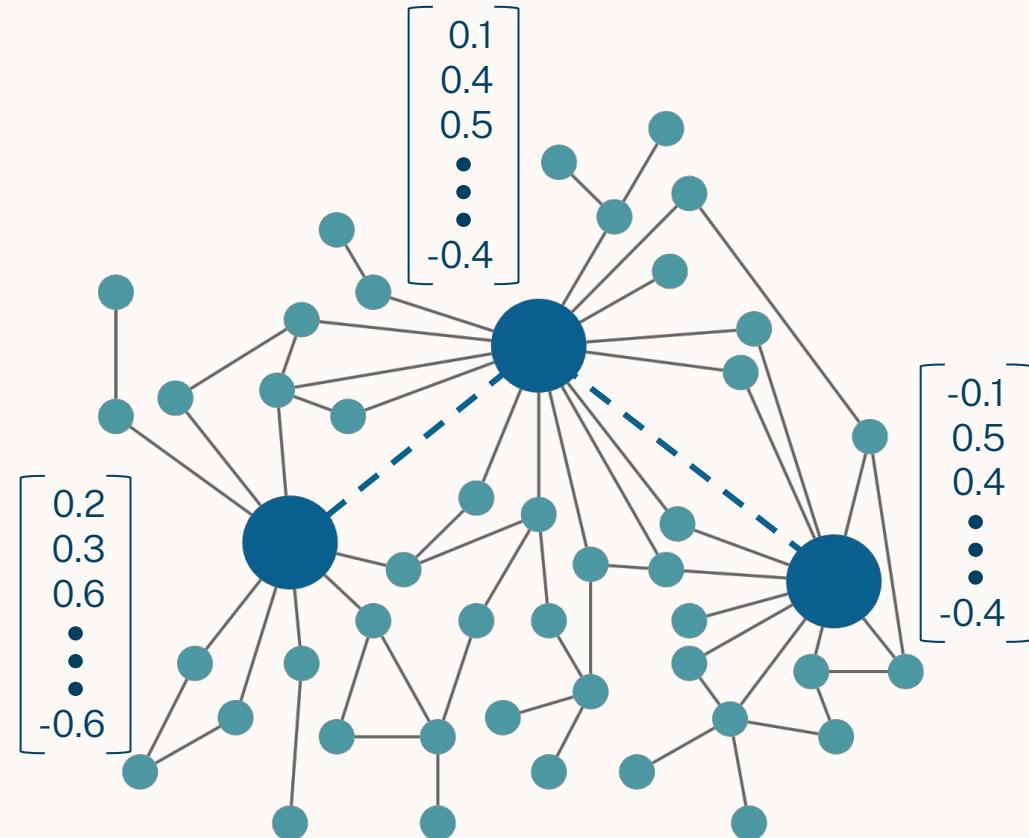
1	0	0	0	1	0	0	1	1	0	0
0	1	1	0	1	0	1	0	0	0	0
0	0	1	1	0	1	0	0	0	0	1
0	0	0	1	0	1	0	0	0	0	1
0	0	0	0	1	1	0	0	1	0	1
⋮										
0	0	0	0	1	0	0	0	1	0	1
0	1	0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0	1	0	0



0.91	0.43
0.65	1.10
0.50	0.57
0.91	0.89
0.14	1.09
⋮	
0.91	1.01
0.06	0.75
0.01	1.40

K-Nearest Neighbor (KNN) w/ Node Embeddings

*Draw connections between
highly interconnected nodes
and/or those that have similar
roles in the graph*

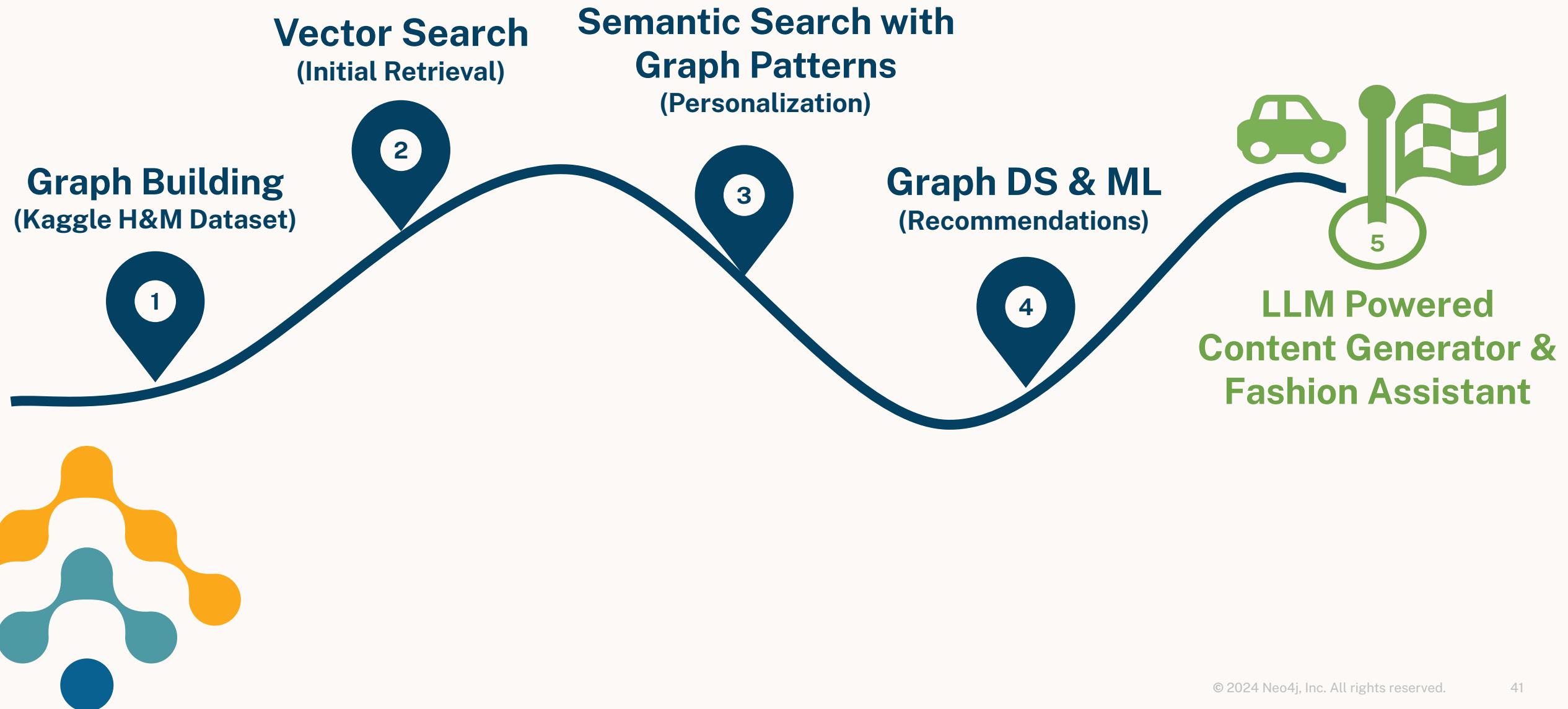




4 / Graph Data Science & ML

Recommendations

Stop before LLM Powered Content Generator



Prompting the LLM

You are a personal assistant named Sally for a fashion, home, and beauty company called HRM. Write an engaging email to **{customerName}**, one of your customers, to promote and summarize products relevant for them given:

- The current season / time of year: **{timeOfYear}**
- Recent searches/interests: **{customerInterests}**

Please only mention the products listed below. Do not come up with or add any new products to the list. Each product comes with an https url field. Make sure to provide that https url with descriptive name text in markdown for each product.

RelevantProducts:

These are products from the HRM store the customer may be interested in based

on their recent searches/interests: **{customerInterests}**

{searchProds}

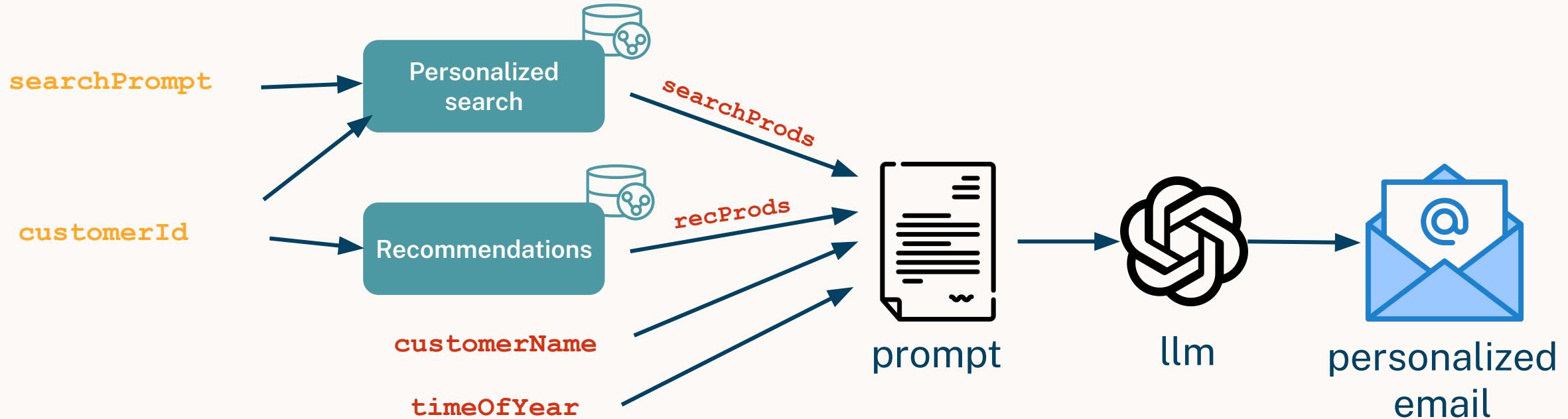
Customer May Also Be Interested In the Following

The below candidates are recommended based on the shared purchase patterns of other customers in the HRM database.

Select the best 4 to 5 product subset from the context that best match the time of year: **{timeOfYear}** and to pair with the RelevantProducts above. For example, even if scarfs are listed here, they may not be appropriate for a summer time of year so best not to include those.

{recProds}

LangChain Chain



```
{  
  searchProds: (searchPrompt , customer_id)| personalizedSearch  
  recProds:customer_id | recommendations  
  customerName:  
  timeOfYear:  
}
```

prompt | llm | outputParser



4/ LLM Powered Fashion Assistant

LLM For Generating Grounded Content

Real-World GenAI Breakthroughs Powered by Neo4j



Leveraging AI
for customized
content at scale



Integrating AI with
knowledge graphs for
smarter supply chain
management



Merging structured
and unstructured data
for efficient operations



Converting vast amounts
of unstructured data into
actionable knowledge

1,700+ Organizations Use Neo4j

Banking & Financial Services



Technology



Telecommunications



Energy



E-Commerce



Health & Life Sciences





Thank you!