



Amazon Neptune

DISA TEM Presentation

June 2025

Nicole Moldovan
Principal NoSQL GTM Specialist
Amazon Web Services

Meet Your AWS Account Team

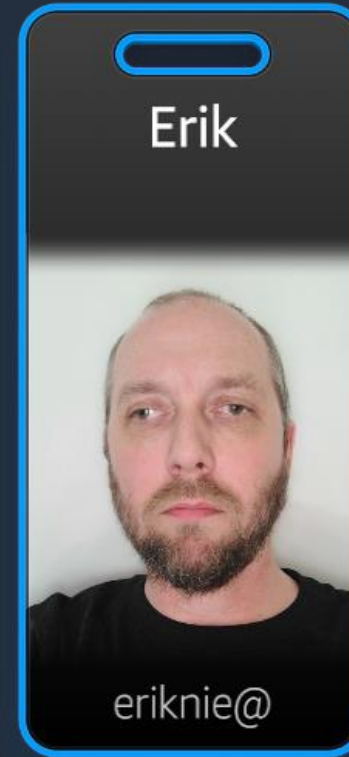
Account Manager



Account Manager



Solutions Architect



Customer Solutions Manager



Enables and facilitates your business account

Provides technical
guidance

Helps you drive
your cloud journey



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.



Most complete set of relational & purpose-built databases



Amazon
RDS



Amazon
Aurora

KEY-VALUE



Amazon DynamoDB

DOCUMENT



Amazon DocumentDB

CACHING



Amazon ElastiCache

GRAPH



Amazon Neptune

TIME-SERIES



Amazon Timestream

MEMORY



Amazon MemoryDB

WIDE COLUMN



Amazon Keyspaces

Reference customers



Amazon Neptune

Customers across different verticals and use cases use Amazon Neptune in production today



FINRA

SIEMENS



intuit

FACTSET

SAMSUNG



Rappi

Uber ATG

HUUUGE

NBCUniversal

NETFLIX

asurion

COX
AUTOMOTIVE™



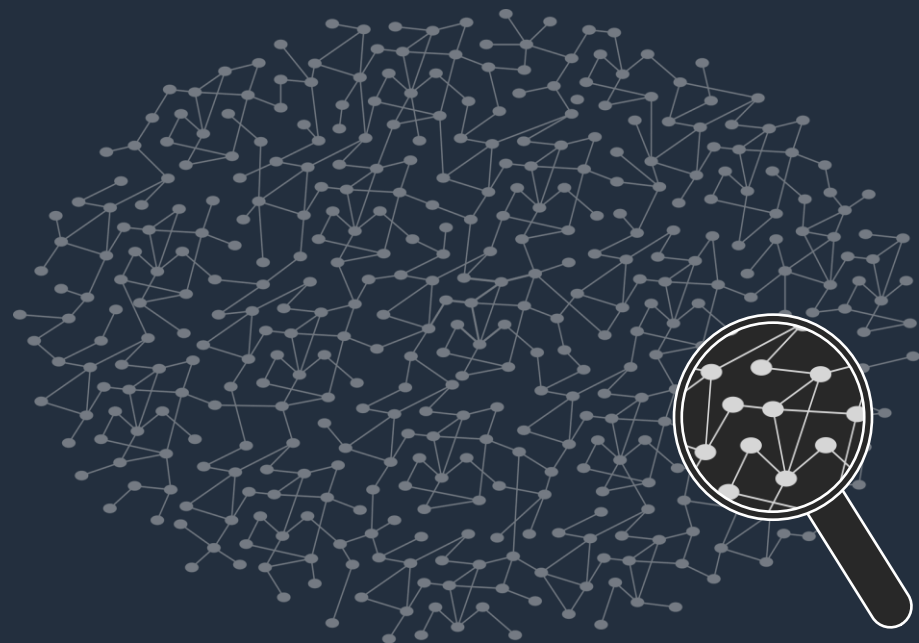
zeta



noonum

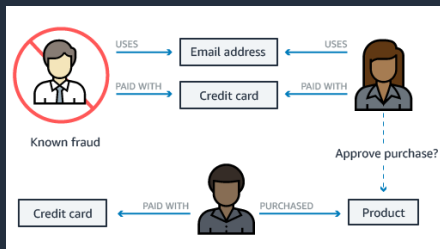
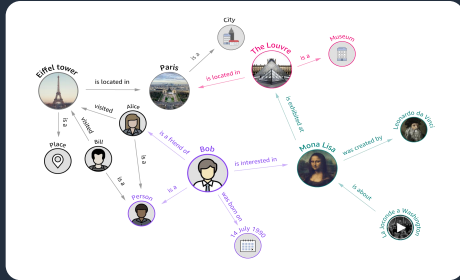


Graphs enable customers to innovate



1. Model data based on relationships
2. Applications explore connections and patterns in connected data
3. Processing graphs is hard due to random data access
4. Generalized graph operations require purpose-built processing

Amazon Neptune Use Cases



Knowledge Graphs

- Link disparate and heterogeneous data sources together to discover hidden connections
- Make data easily accessible through improved search results

Identity Graphs

- Persistent identifier to link all related devices and ids, enabling unified profile creation, targeting, and personalization
- Create audiences based on interests, preferences, and purchases

Fraud Graphs

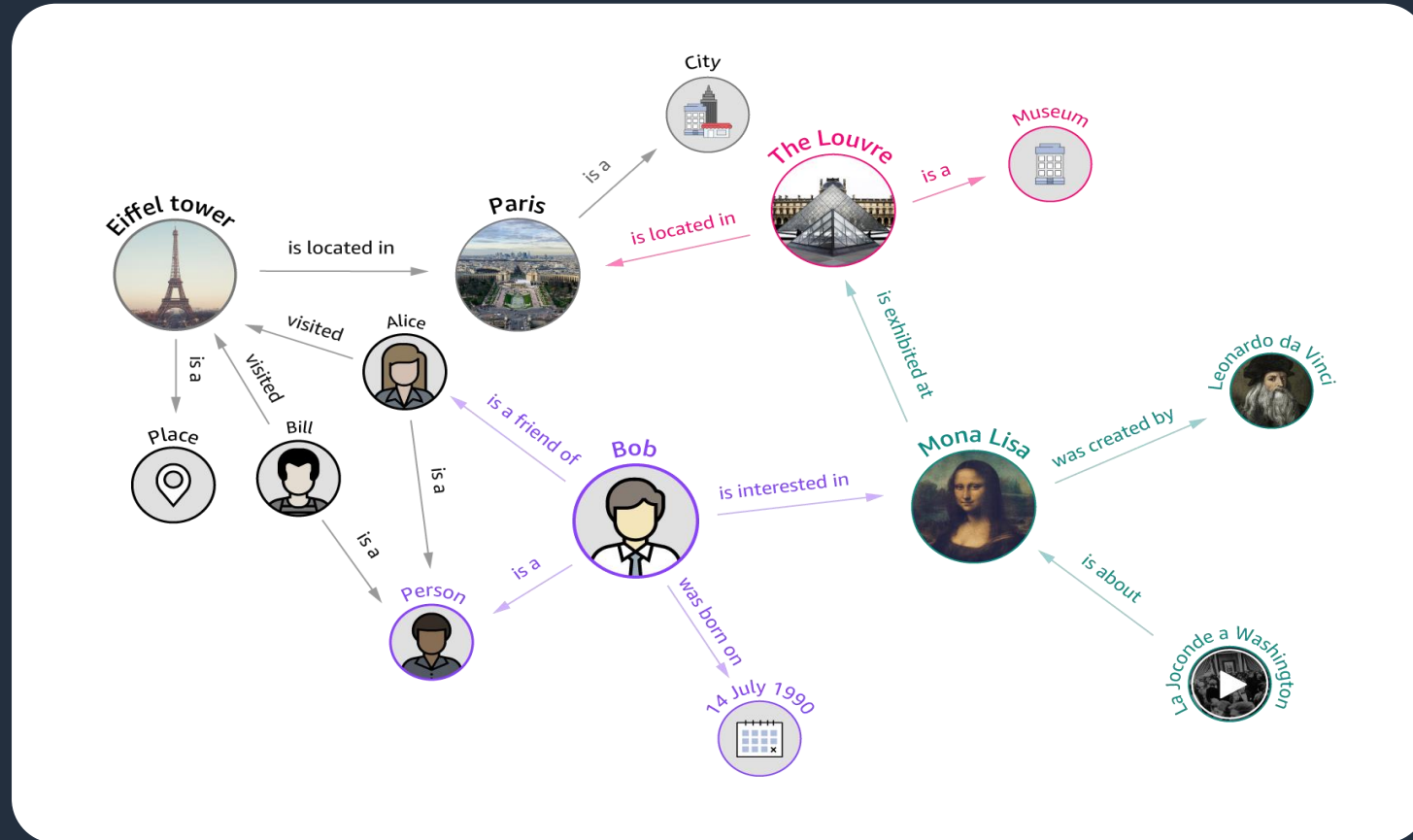
- Locate and prevent fraudulent patterns of transactions as they are occurring
- Derive unique insights commonly used to identify fraud

Security Graphs

- Use graphs to detect threats to your environment such as unwanted user access to applications, or exposed resources

Knowledge graphs

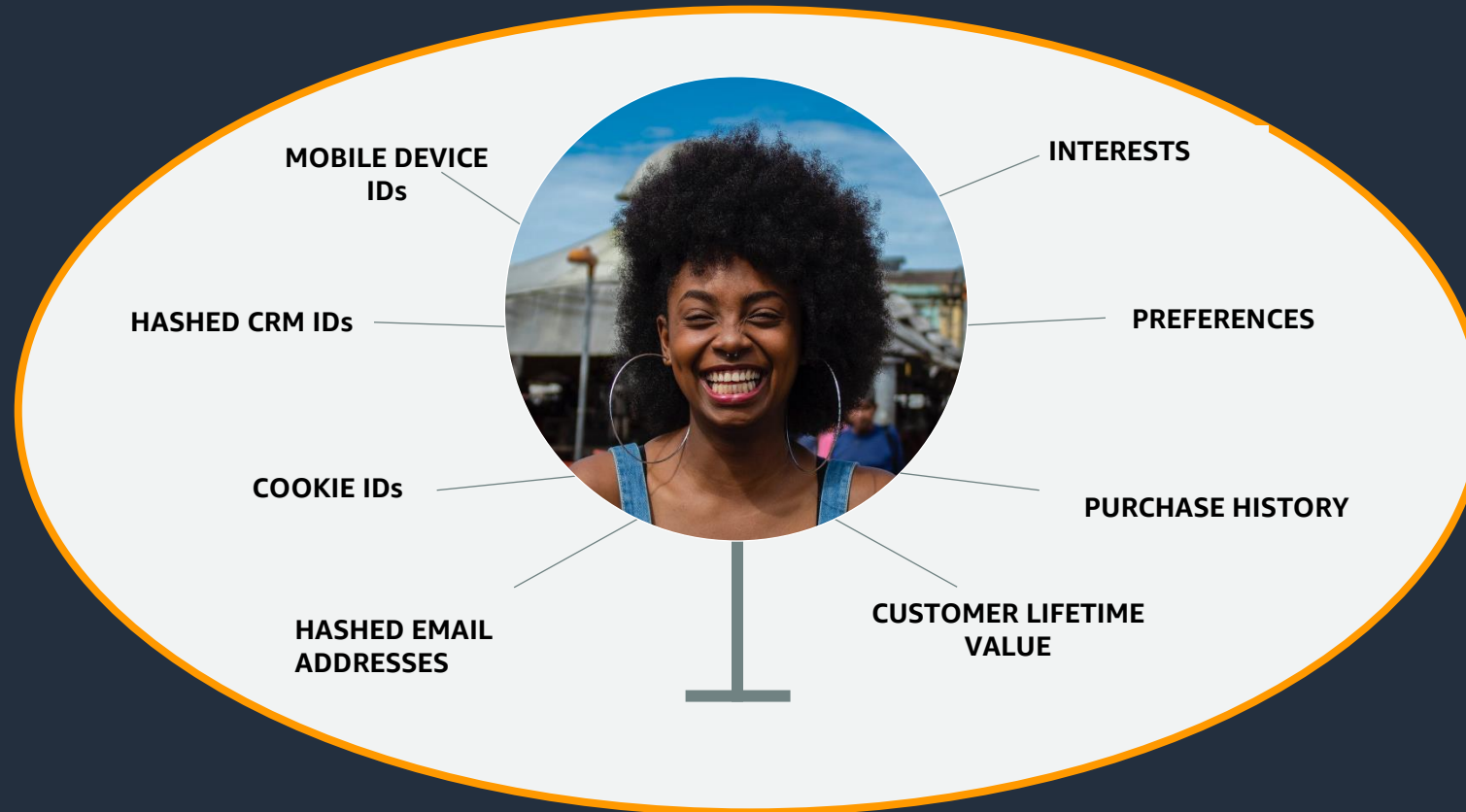
UNDERSTANDING THE WHO, WHAT, WHEN, AND WHERE



<https://aws.amazon.com/neptune/knowledge-graphs-on-aws/>

Identity graphs

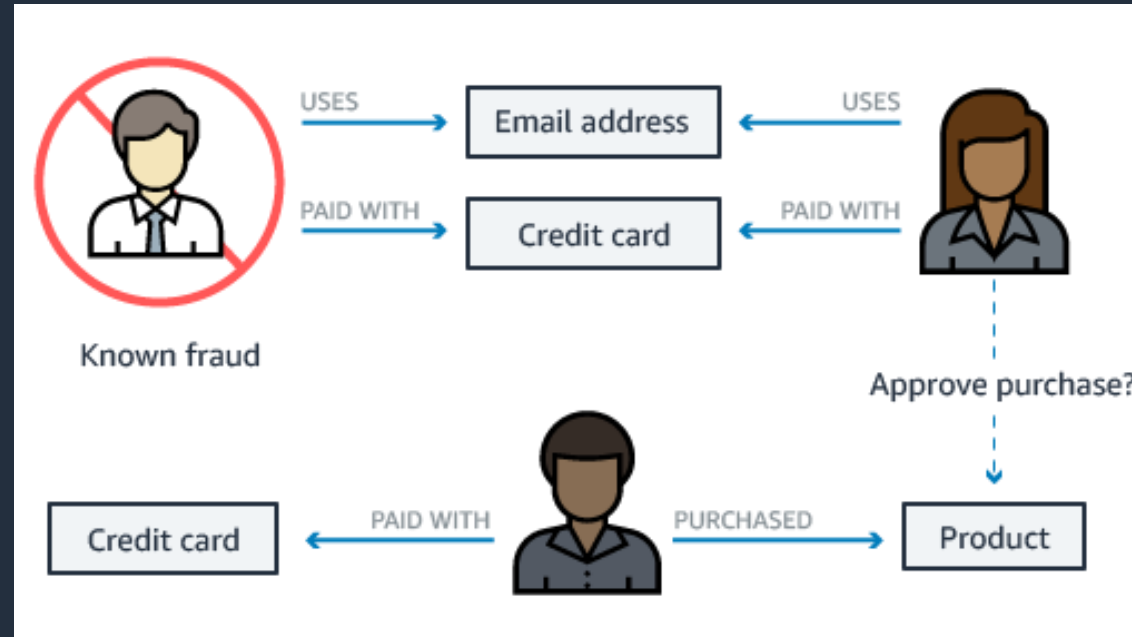
UNIFIED 360° VIEW OF THE CUSTOMER



<https://aws.amazon.com/neptune/identity-graphs-on-aws/>

Fraud graphs

DETECTING FRAUD AS IT HAPPENS USING RELATIONSHIPS



<https://aws.amazon.com/neptune/fraud-graphs-on-aws/>



As India's leading gaming company, Games24x7 is known for its flagship products like RummyCircle, which offers online rummy, and My11Circle, which offers fantasy sports.

Challenge:

As the game of Rummy involves real money, Games24x7 has to stay vigilant to prevent fraud and collusion during tournaments.

Solution:

It uses the Amazon Neptune graph database to detect if two players in a game are colluding to beat the other four players. This is accomplished by assigning a table in the database to each player when they log in.



Security graphs

UNDERSTAND SECURITY VULNERABILITIES ACROSS LAYERS



1. Cloud Security Posture Management

2. Data Flow/Exfiltration

3. Identity and Access Management

<https://aws.amazon.com/neptune/security-graphs-on-aws/>

Wiz: Redefining cloud security

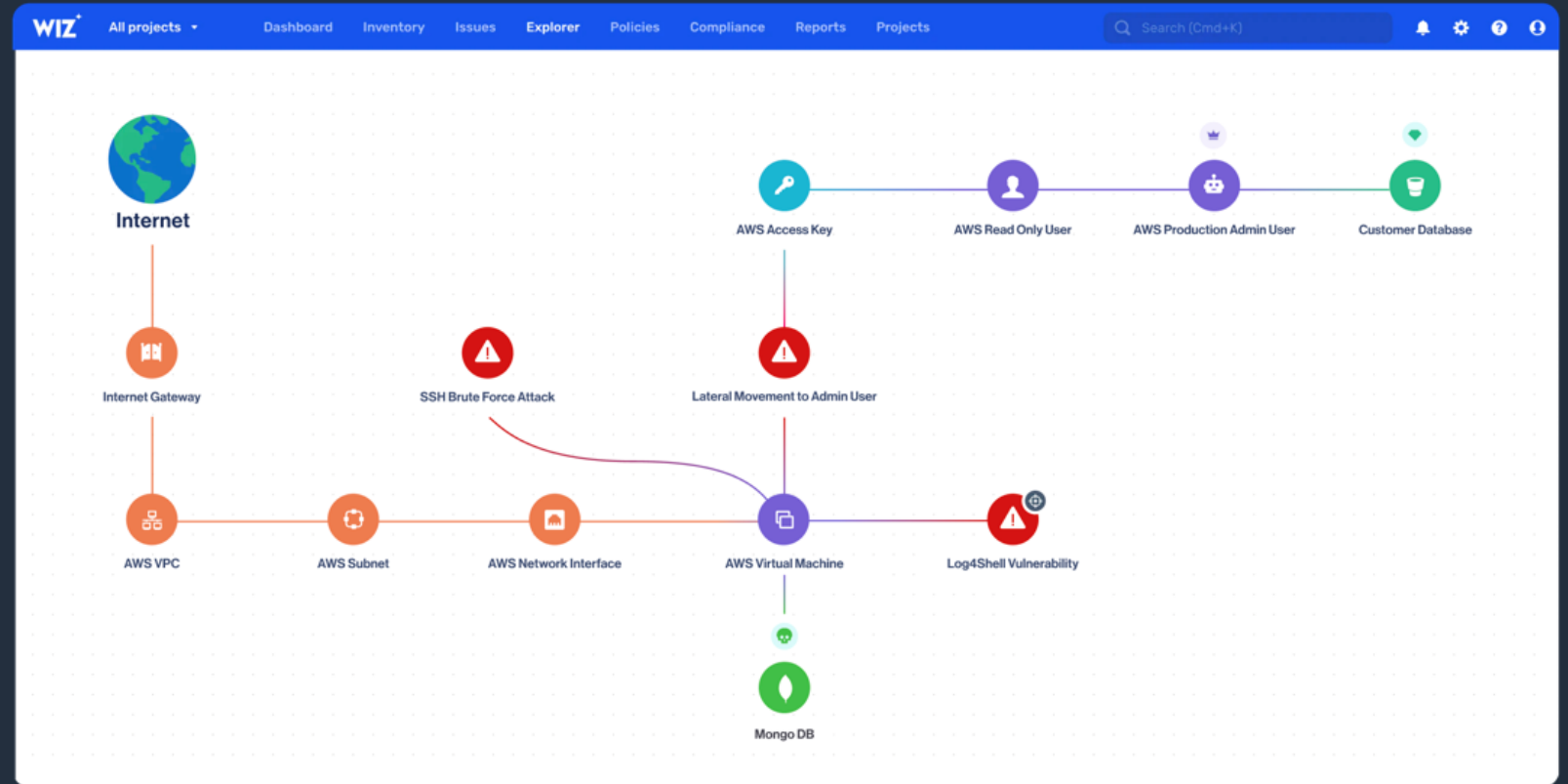
WIZ CONTINUES TO GROW AT AN ASTONISHING RATE*

5M workloads

Scanned daily

Over 30%

of the Fortune 100



*Forbes article by Peter Cohan: <https://bit.ly/3EZGJtq>

Neptune's Family



Neptune Databases

- **Store and manage graph data sets**
- A serverless graph database designed for superior scalability and availability.
- For graph database workloads that need to scale to 100,000 queries per second, multi-AZ high availability, and multi-region deployments.
- Example use cases: Social Networking Applications, Fraud Alerting, and Customer 360



Neptune Machine Learning

- **Creates, trains, and applies ML models on your graph data**
- An integration with SageMaker that uses the Deep Graph Library ([DGL](#)) and graph neural networks (GNNs) to make easy, fast, and more accurate predictions using graph data
- For automating the selection and training of ML models on graph data and for using Neptune Database APIs to make predictions on live data.
- Example use cases: Link prediction for recommendations, fraud detection, or identity resolution.

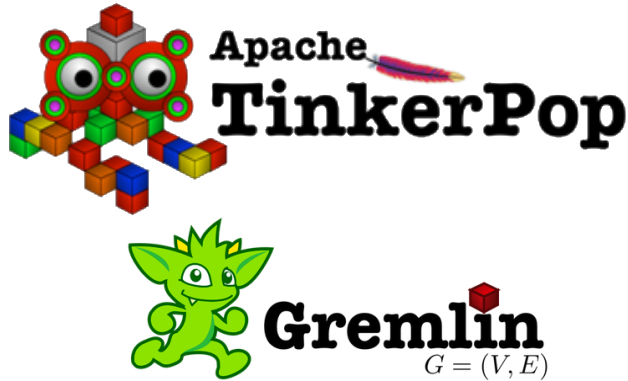


Neptune Analytics

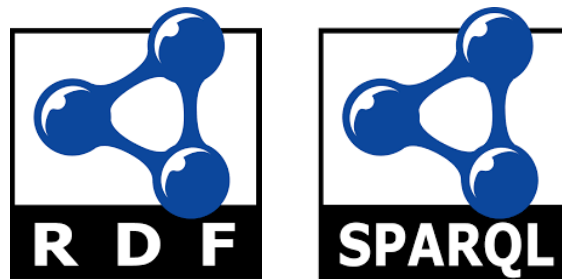
- **Analyze graph data sets in-memory**
- A graph analytics database engine to quickly analyze large amounts of graph data to get insights and find trends.
- For quickly analyzing existing graph databases or graph data sets stored in a data lake using popular graph analytic algorithms and low-latency analytic queries.
- Store vector embeddings as properties of nodes. Add explainability to semantic search results
- Example use cases: Targeted content recommendations, Fraud investigation, and Network threat detection

Leading graph models and frameworks

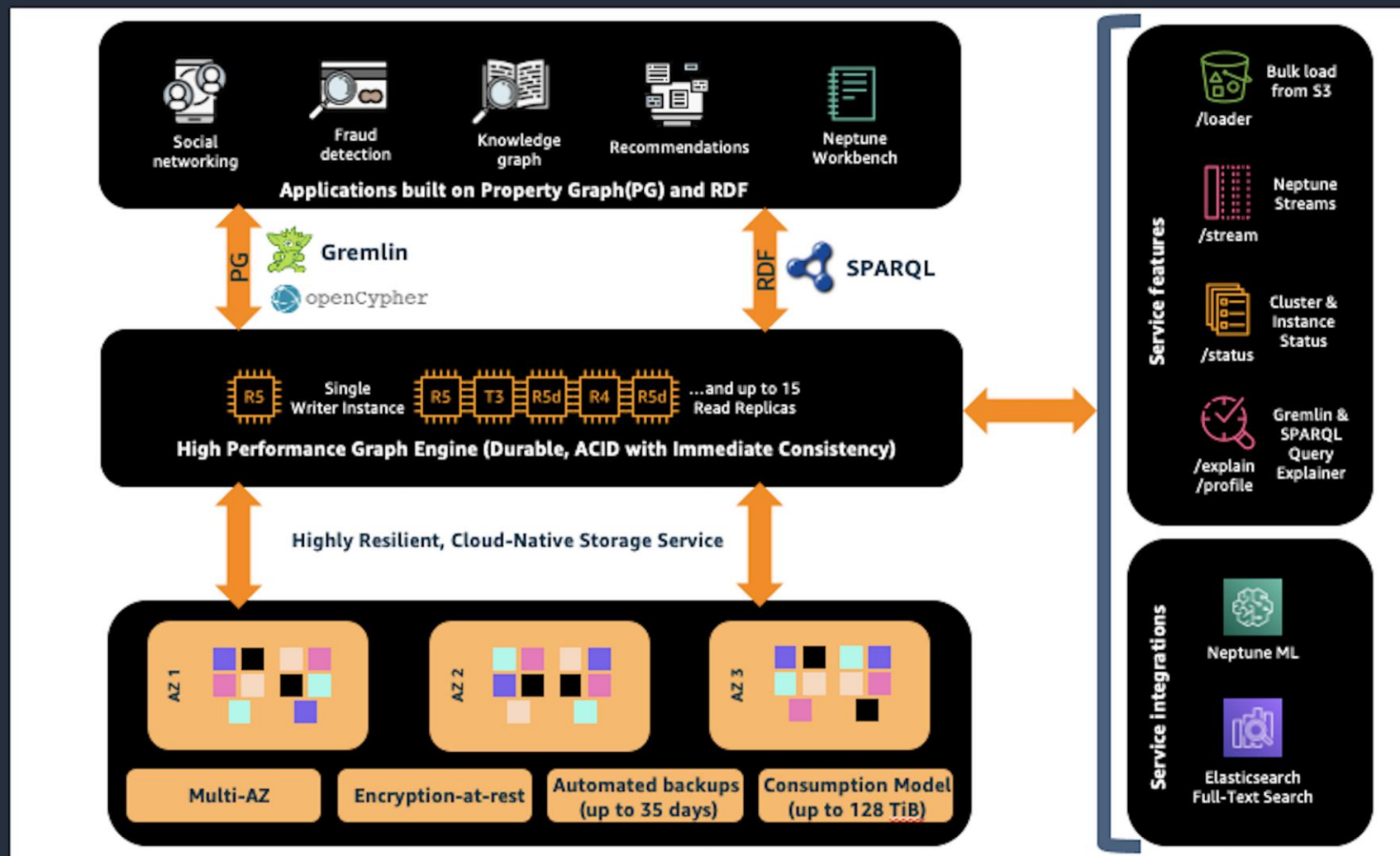
PROPERTY GRAPH



RESOURCE DESCRIPTION FRAMEWORK (RDF)



Amazon Neptune Database High Level Architecture

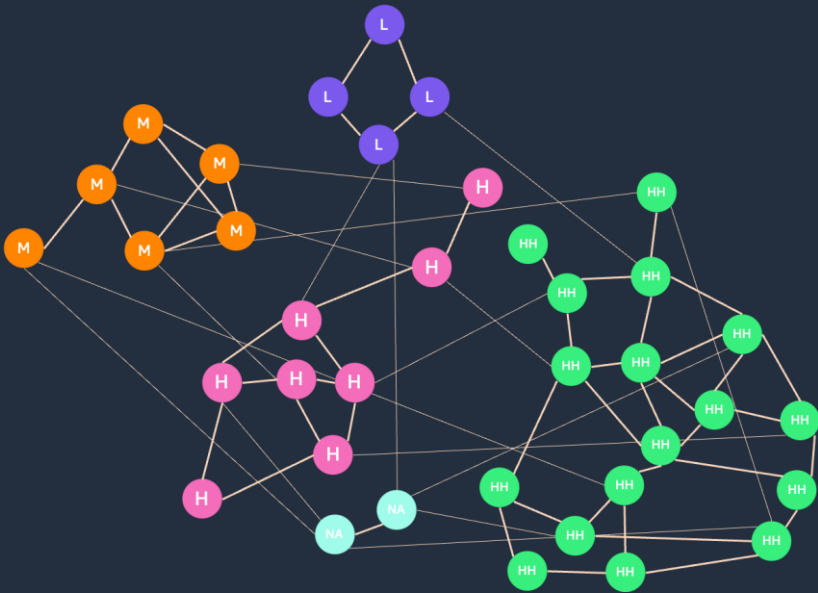


Amazon Neptune Analytics



Amazon Neptune Analytics

New analytics database engine for Amazon Neptune helps data scientists and application developers make data discoveries faster by analyzing graph data with tens of billions of connections in seconds



One Endpoint to orchestrate your graph workloads

- Use one API endpoint to create a graph, load data, invoke queries, and perform vector similarity search
- Popular OSS query language: openCypher

High-performance graph analytic queries and graph algorithms

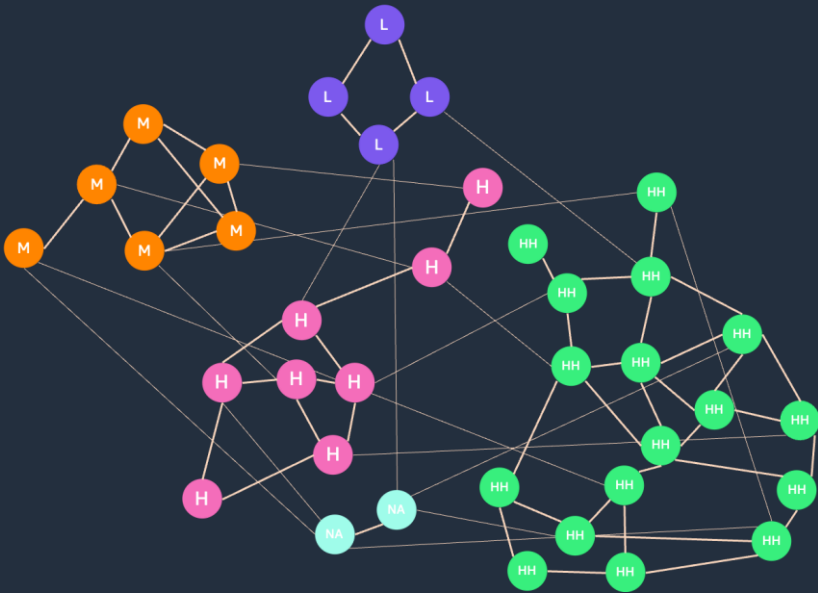
- 80X faster than previous AWS solutions (10M/edges/sec)
- 5 families of algorithms with 30 variants
- Via openCypher, integrated query syntax with mutations

Store and Search Vectors for Generative AI Applications

- Store and search embeddings trained in graph queries
- Combine graph algorithms, pattern matching traversals, and vector search into one query to reduce services required for generative AI workloads.

Amazon Neptune Analytics

Example Use Cases



Ephemeral Analytics

Customers load and analyze large datasets quickly for hourly/daily metric computations, near real-time analytics, risk modeling, and situations requiring quick insights.

Low Latency Analytical Queries

With low-latency graph queries, you can extract graph features from connected data to fuel predictive models in real-time to fuel personalized AI models.

Vector Search with Graph Data

Best of both worlds: build AI applications using techniques like retrieval augmented generation (RAG) that combine graph traversals and vector search for context augmentation.

GenAI & GraphRAG with Amazon Neptune



RAG is a powerful architecture pattern but has complex data challenges

CONNECTEDNESS

data spread across multiple disparate documents is hard to retrieve

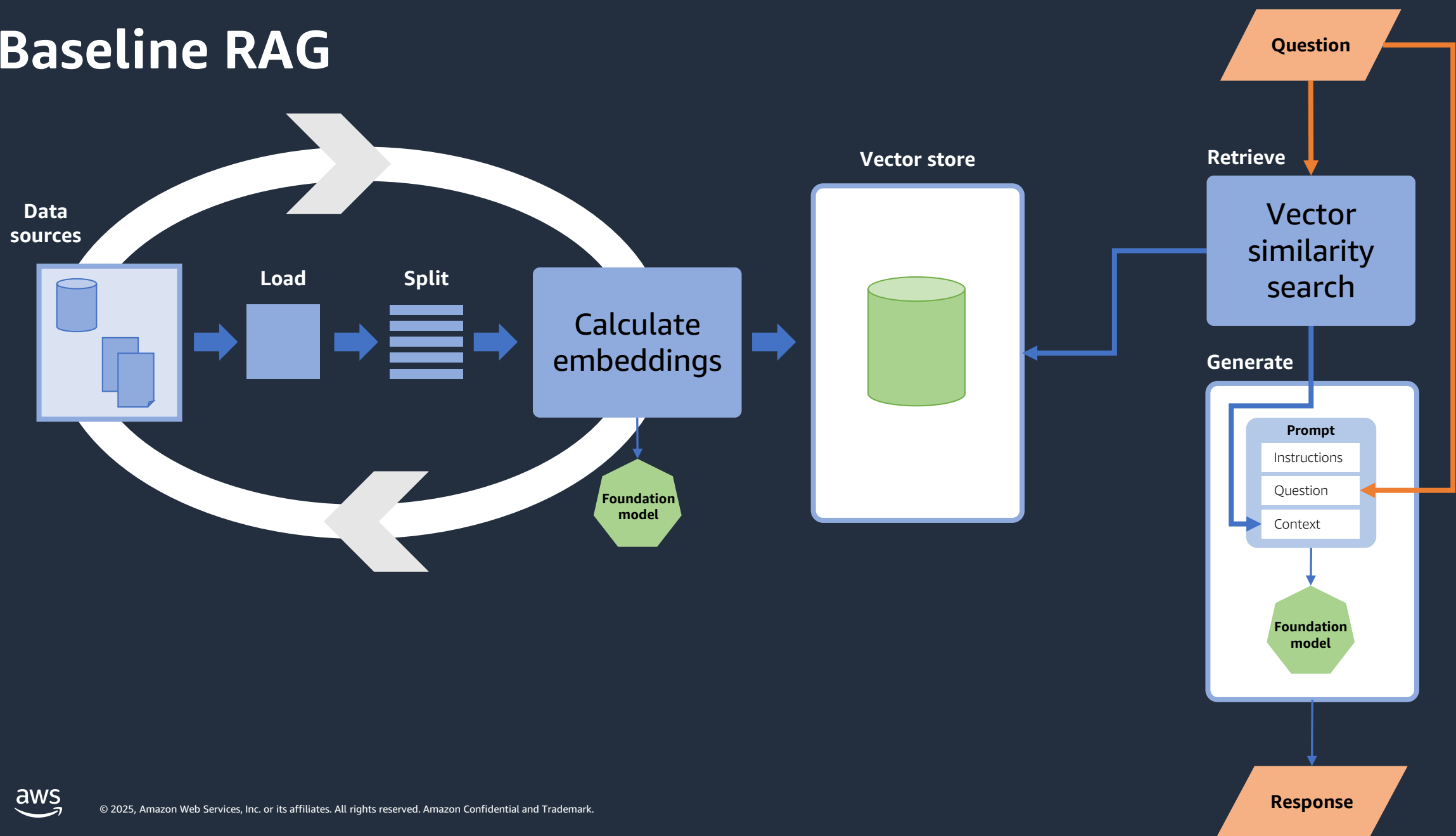
SPECIFICITY

embeddings are sparse representations of data which may lack crucial details

EXPLAINABILITY

explaining the relevance of data retrieve is demanding

Baseline RAG



Sometimes the most relevant information to answering a question is not the closest in meaning.

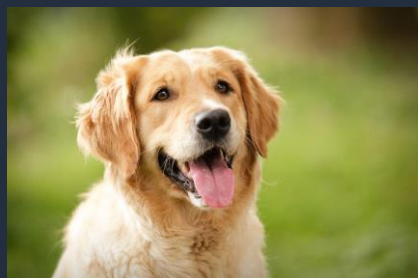
Searching in Vector space vs. Graph space?

Similarity in vector space
compares mathematical closeness

Zucchini is similar to *summer squash* and *courgette*

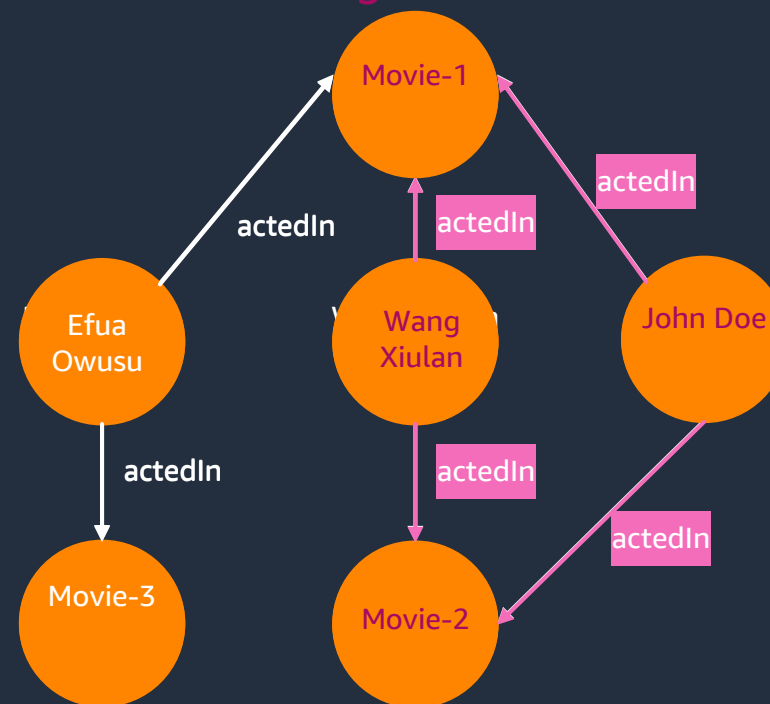
Vectors can represent ...

... A text embedding model
... An image embedding model

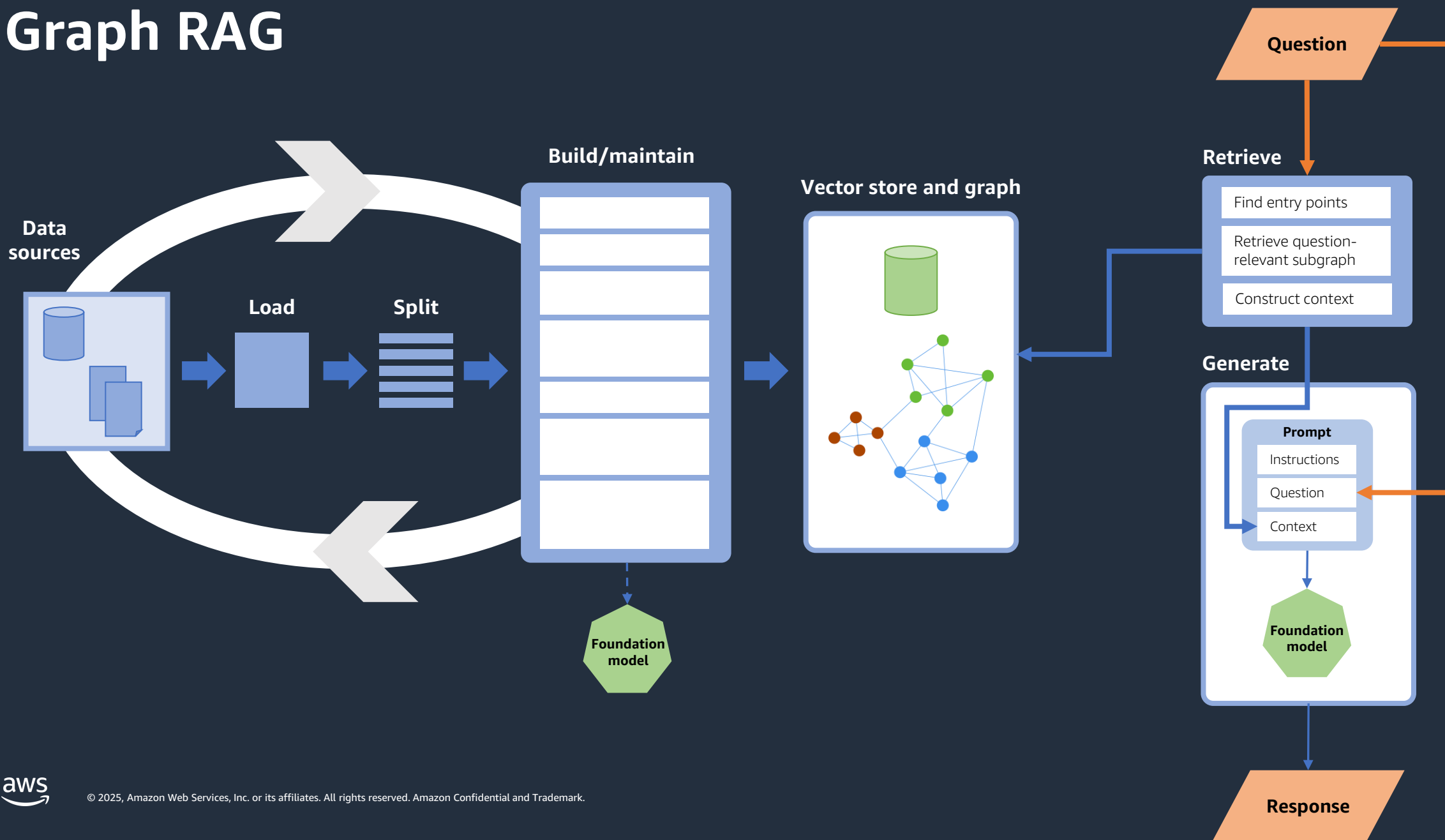


Relatedness in graph space
compares shared connections

Wang Xiulan and *John Doe* are related
because they've starred in multiple movies
together



Graph RAG



When to use GraphRAG

Multiple documents and data sources: When relevant information is dispersed across multiple sources or documents and the relationships between entities are not clear or hidden

Summarization: When a summarization question targeting a specific domain across multiple documents

GraphRAG with Amazon Neptune in Amazon Bedrock Knowledge Bases

This will create a graph and embeddings and store them in provided Neptune Analytics







- Step 1
● Provide Knowledge Base details
- Step 2
● Configure data source
- Step 3
● **Select embeddings model and configure vector store**
- Step 4
○ Review and create

Select embeddings model and configure vector store

Choose an embeddings model to convert the data that you will provide in the next step, and provide details for a vector data store in which Bedrock can store, manage, and update your embeddings. The embeddings model and vector store cannot be changed after creation of Knowledge Base.

Embeddings model

Select an embeddings model to convert your data into an embedding. Pricing depends on the model. [Learn more](#)

 Titan Text Embeddings v2 [?] By Amazon	<input checked="" type="radio"/>	 Titan Embeddings G1 - Text v1.2 [?] By Amazon	<input type="radio"/>
 Embed English v3 [?] By Cohere	<input type="radio"/>	 Embed Multilingual v3 [?] By Cohere	<input type="radio"/>

► Additional configurations

Vector database

Create a new vector or choose an existing store so that Amazon Bedrock can store, update, and manage embeddings. [Learn more](#)

Vector store creation method

<input checked="" type="radio"/> Quick create a new vector store - Recommended If you've used a vector store before, we recommend that you continue using it. Otherwise, choose the vector store that best fits your use case.	<input type="radio"/> Choose a vector store you have created Select Amazon OpenSearch Serverless, Amazon Aurora, Pinecone, Amazon Aurora, MongoDB Atlas, Neptune Analytics (GraphRAG) or Redis Enterprise Cloud and provide field mappings.
--	---

Vector store

Choose an engine that you are familiar with to combine with vector embeddings.

- ☐ **Amazon OpenSearch Serverless**
Select to optimize and provide contextually relevant responses across billions of vectors in milliseconds. Combine search with text-based keywords for hybrid requests. [Learn more](#)
- ☐ **Amazon Aurora PostgreSQL Serverless - new** | [Info](#)
Select to store and index vector embeddings for fast similarity search. Use SQL queries to combine search results with your business data. [Learn more](#)
- ☒ **Amazon Neptune Analytics (GraphRAG) - new**
Select to optimize for high-performance graph analytics and enable graph-based Retrieval Augmented Generation (GraphRAG) solutions. [Learn more](#)

Foundation model

By default, Amazon Bedrock chooses the foundation model below to automatically build graphs for your knowledge base. This automatically enables contextual enrichment.

Claude 3 Haiku v1

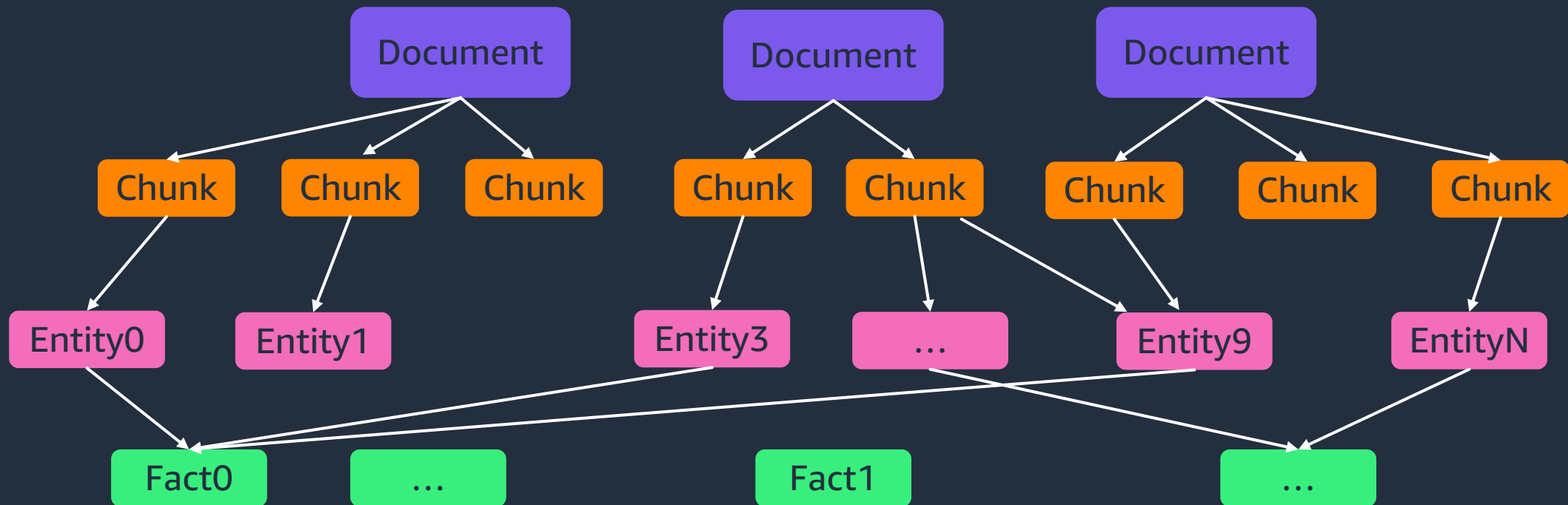
► Additional configurations

[Cancel](#)

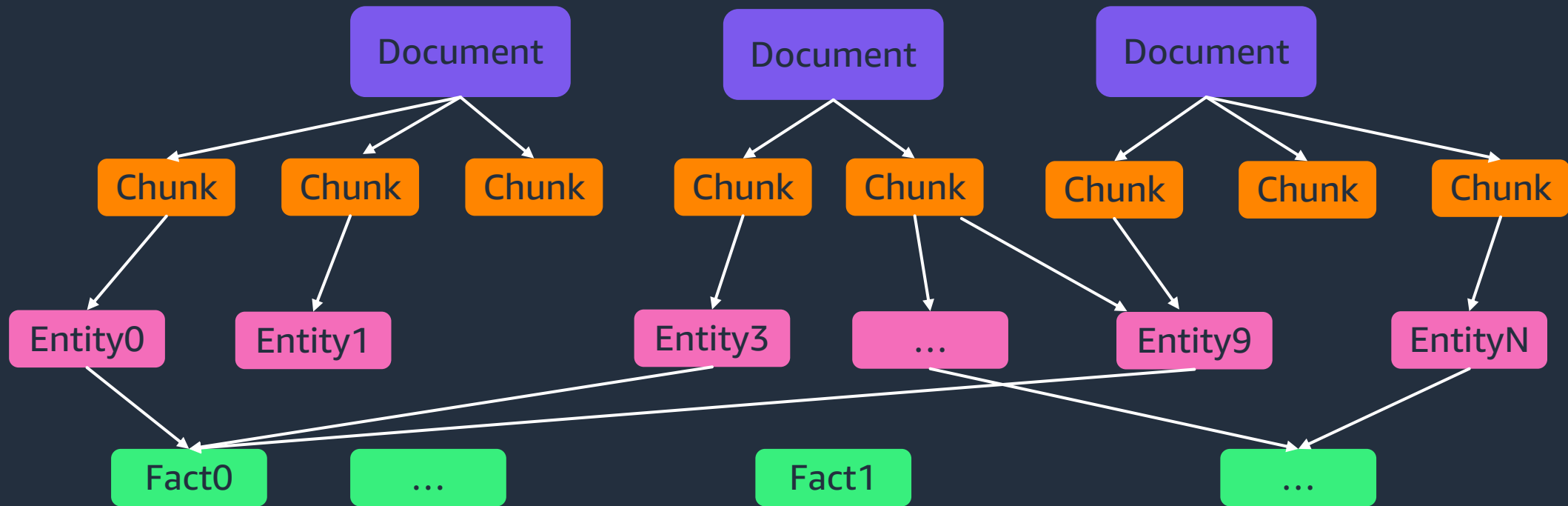
[Previous](#)

[Next](#)

Automatically generates a graph from your data that is updated as you add new content

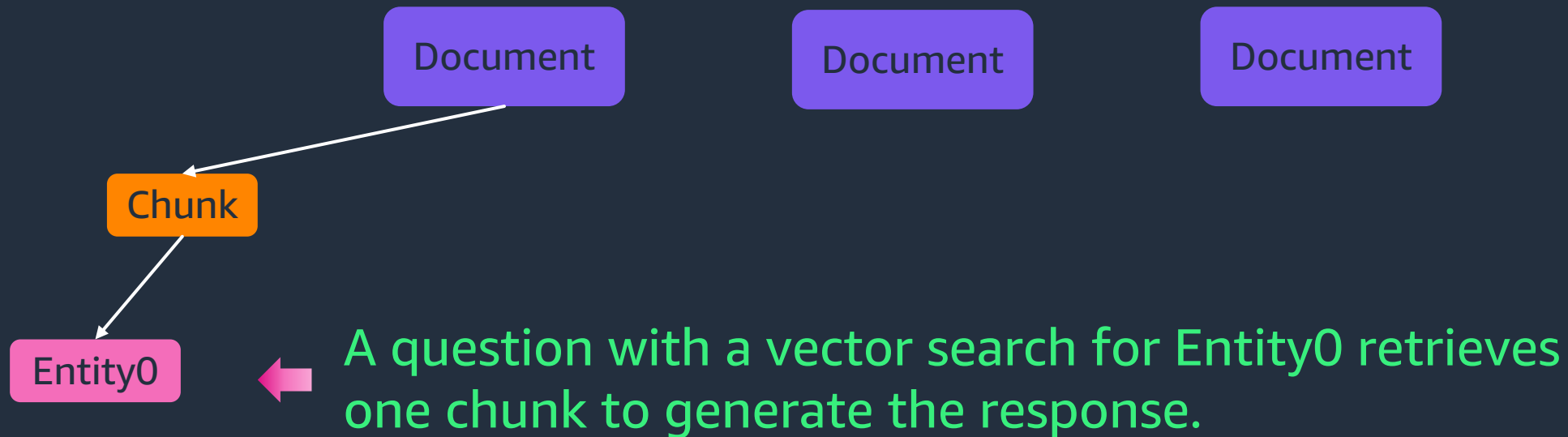


Uses the graph within the retrieval process to improve the responses

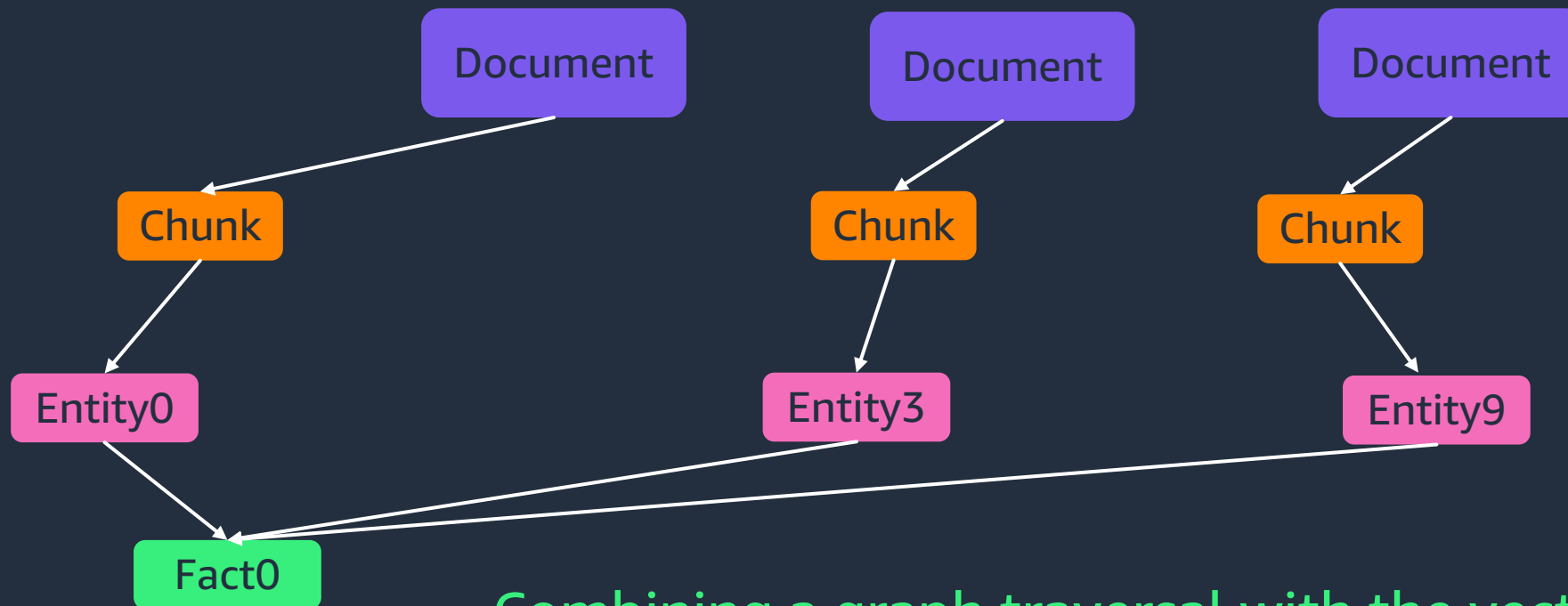


Responses can be more accurate and comprehensive, especially for multi-document retrieval, because the graph can be used to identify relevant chunks that are not top-K vector search results.

Uses the graph within the retrieval process to improve the responses



Uses the graph within the retrieval process to improve the responses



Combining a graph traversal with the vector search for Entity0, retrieves three chunks to generate the response

Scenario: Example Corp's Quarterly Financial Statement

Example Corp's quarterly report had four paragraphs:

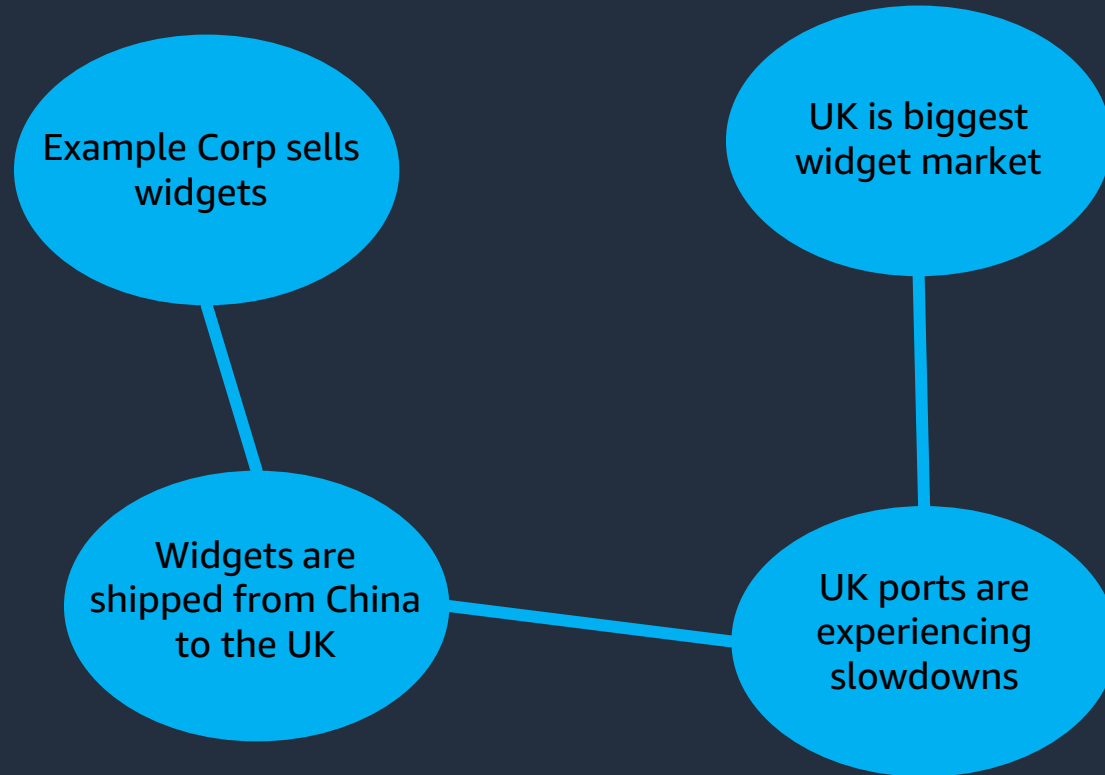
- Paragraph 1 discussed how "Example Corp sells widget"
- Paragraph 2 discussed that the "UK is the biggest widget market"
- Paragraph 3 discussed that the "Widgets are shipped from China to the UK"
- Paragraph 4 discussed that "UK ports are experiencing slowdowns"

Each of paragraph has the text and associated embedding stored

Question: What is the outlook for widget sales in the UK?

STEP 1: AN EMBEDDING IS CREATED OF THE QUESTION BEING ASKED

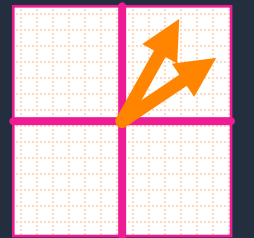
question embedding



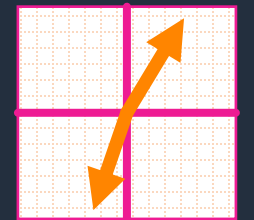
Question: What is the outlook for widget sales in the UK?

STEP 2: SIMILARITY SEARCH IS RUN TO FIND THE STARTING NODES

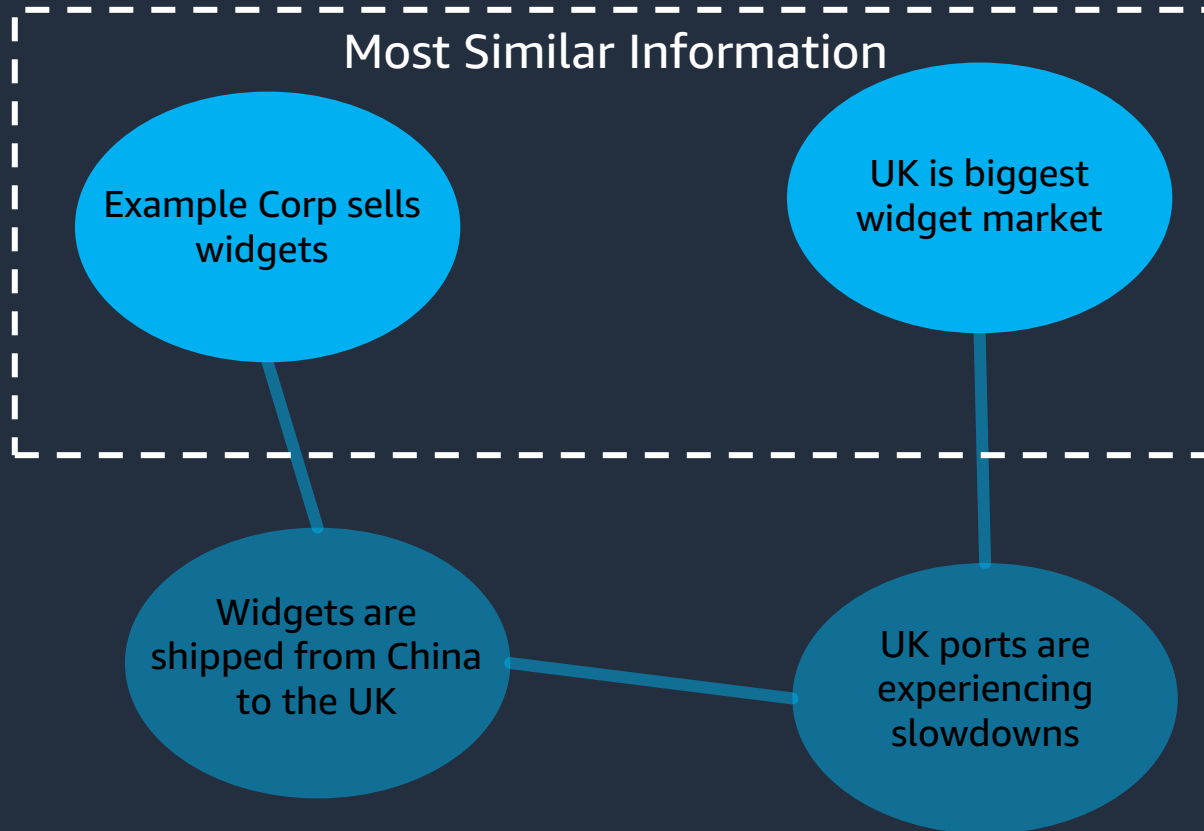
question embedding



More Similar

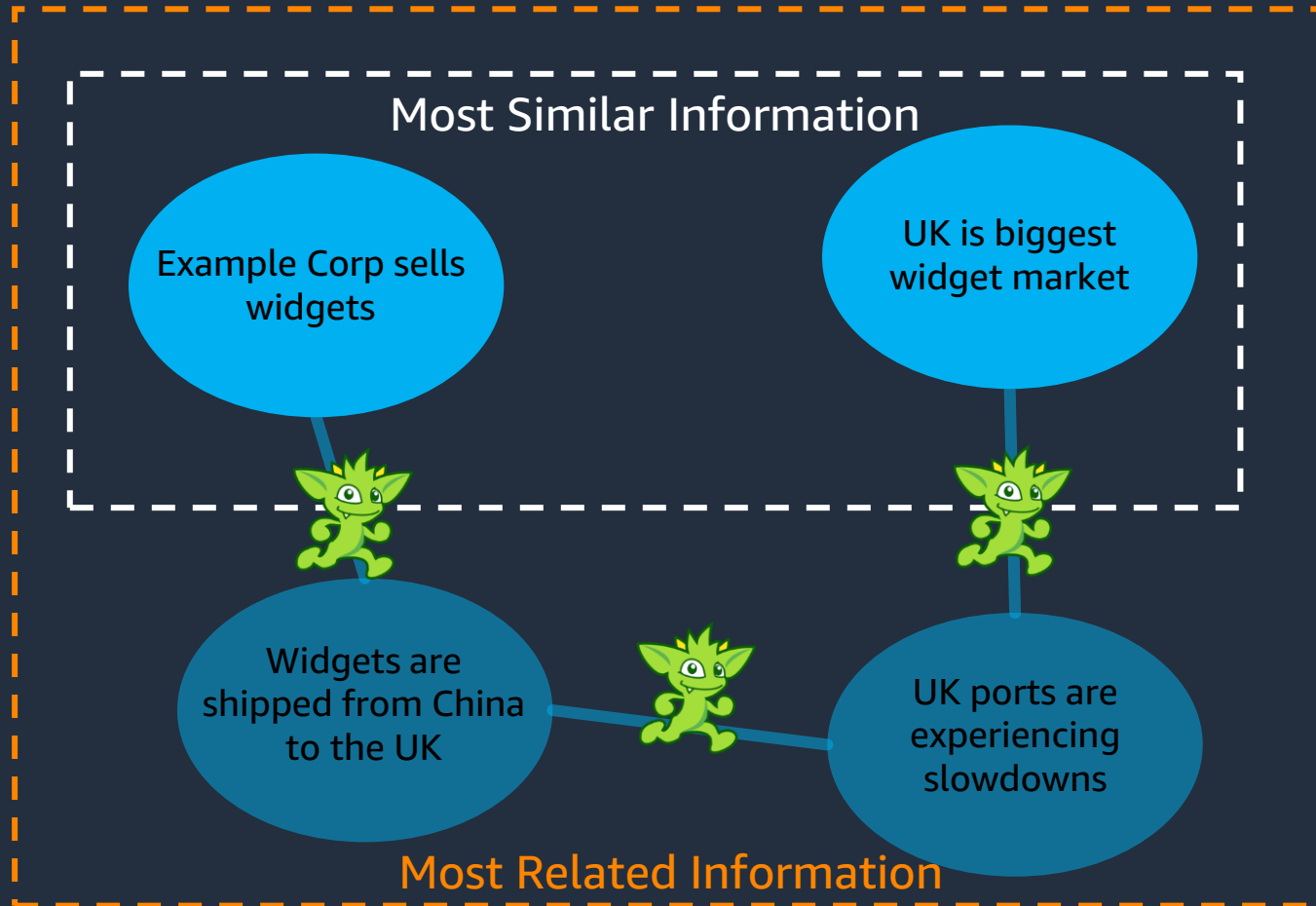


Less Similar



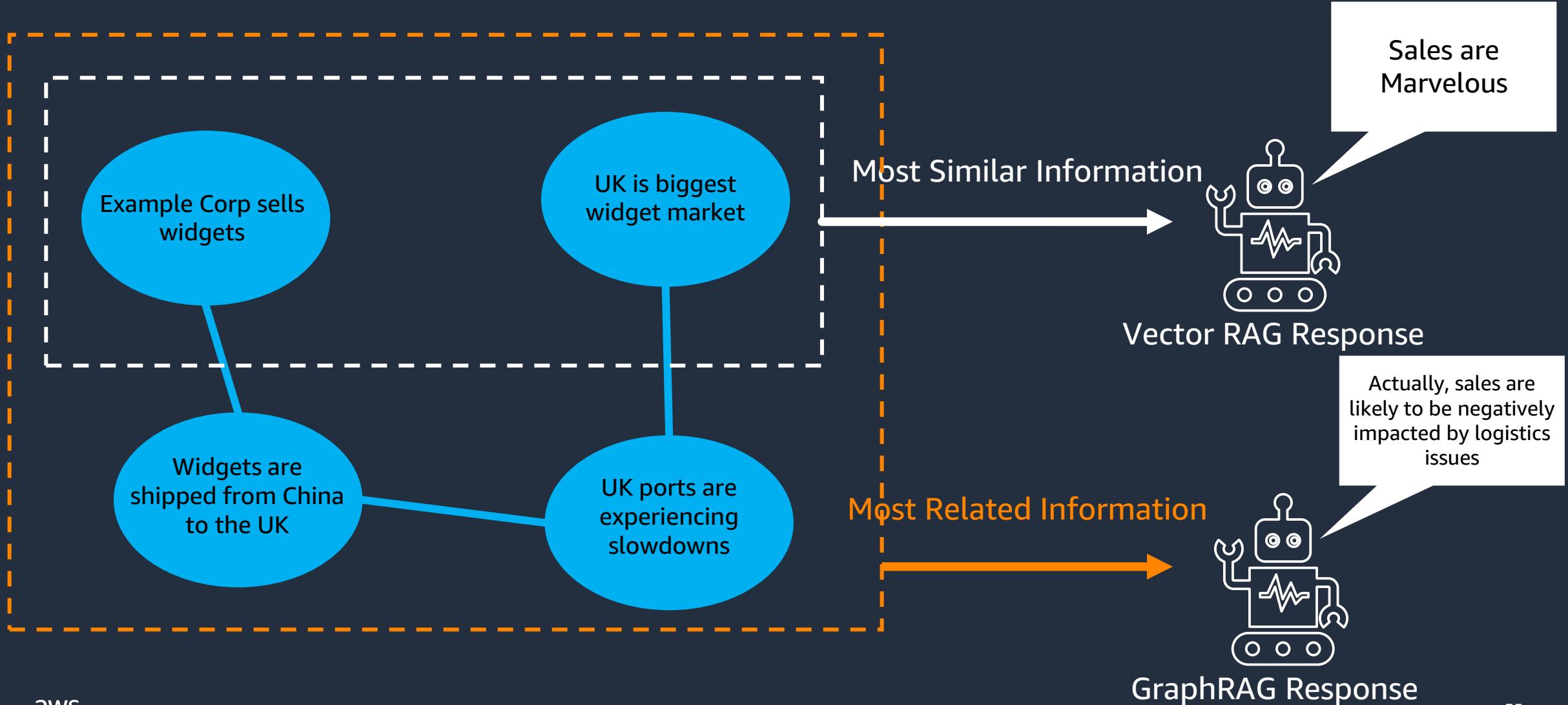
Question: What is the outlook for widget sales in the UK?

STEP 3: IN GRAPH RAG THE GRAPH IS ALSO USED TO LOCATE RELATED IDEAS



Question: What is the outlook for widget sales in the UK?

STEP 4: MOST RELEVANT INFORMATION PASSED TO GENERATE A RESPONSE



The **combination of graph and vector search** produces responses that can be more **accurate and comprehensive**, especially for **multi-document, multi-hop questions**

Leading players in many industries are already building GraphRAG applications

Finance

Various documents such as research reports, projections, analyst projections are leveraged to provide recommendation for better financial investments

Cyber Security

Analyzing logs to interpret alerts, summarize security reports, do threat intelligence, and make recommendations

Automotive

Increasing efficiency for internal processes – quickly pull the right datasets/databases for further analysis such as spare parts analysis

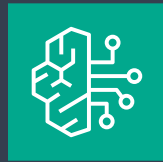
Pharmaceuticals

Improving Drug trials processes with better analysis, and performing publication analysis

Amazon Web Services announces support for GraphRAG with Amazon Neptune in Amazon Bedrock Knowledge Bases



Amazon Neptune Analytics



Amazon Bedrock Knowledge Bases

Fully managed AWS solution
Seamless integration within the AWS ecosystem for high security and performance

Now Generally Available

<https://aws.amazon.com/neptune/graphrag/>

AWS Labs Open Source GraphRAG toolkit



LlamaIndex



Amazon Neptune



Amazon Bedrock

Flexible, adaptable solution for organizations that prefer open-source options and customizability

Now on GitHub!

<https://github.com/aws-labs/graphrag-toolkit>

Open-Source GraphRAG Python Toolkit



graphrag-toolkit

Private

Unwatch 9

Fork 0

Star 1

main

1 Branch

0 Tags

Go to file

Add file

Code

ianrob

Update query doc

fa6c80b · 7 minutes ago

103 Commits

docs	Update query doc	7 minutes ago
examples	Update query examples	1 hour ago
images	Delete unnecessary readme	2 days ago
src	Update query examples	1 hour ago
CODE_OF_CONDUCT.md	Initial commit	2 weeks ago
CONTRIBUTING.md	Initial commit	2 weeks ago
LICENSE	Initial commit	2 weeks ago
NOTICE	Initial commit	2 weeks ago
README.md	Update docs	last week
pyproject.toml	Update CFN	5 days ago

Readme

Apache-2.0 license

Code of conduct

Security policy

Activity

Custom properties

1 star

9 watching

0 forks

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Languages

Python 100.0%

README

Code of conduct

Apache-2.0 license

Security

GraphRAG Toolkit

The graphrag-toolkit is a Python toolkit for building GraphRAG applications. It provides a framework for automating the construction of a graph from unstructured data, and composing question-answering strategies that query this graph when answering user questions.

The toolkit uses low-level [LlamaIndex](#) components – data connectors, metadata extractors, and transforms – to implement much of the graph construction process. By default, the toolkit uses [Amazon Neptune Analytics](#) or [Amazon Neptune Database](#) for its graph store, and Neptune Analytics or [Amazon OpenSearch Serverless](#) for its vector store, but it also provides extensibility points for adding alternative graph stores and vector stores. The default backend for LLMs and embedding models is [Amazon Bedrock](#); but, as with the stores, the toolkit can be configured for other LLM and embedding model backends using LlamaIndex abstractions.

```
from graphrag_toolkit import LexicalGraphQueryEngine
from graphrag_toolkit.storage import GraphStoreFactory
from graphrag_toolkit.storage import VectorStoreFactory

import nest_asyncio
nest_asyncio.apply()

graph_store = GraphStoreFactory.for_graph_store(
    'neptune-db://my-graph.cluster-abcdefghijkl.us-east-1.neptune.amazonaws.com'
)

vector_store = VectorStoreFactory.for_vector_store(
    'aoss://https://abcdefghijkl.us-east-1.aoss.amazonaws.com'
)

query_engine = LexicalGraphQueryEngine.for_traversal_based_search(
    graph_store,
    vector_store
)

response = query_engine.query('What are the differences between Neptune Database
                               and Neptune Analytics?')

print(response.response)
```

\$ pip install https://github.com/aws-labs/graphrag-toolkit/releases/latest/download/graphrag-toolkit.zip





Thank you!

Nicole Moldovan

nictune@amazon.com

linkedin.com/in/nicolemoldovan

Chris Walker

cwalkcw@amazon.com

linkedin.com/in/christopherawalker/

Neptune Team

Neptune-gtm@amazon.com

Product Page

<https://aws.amazon.com/neptune/>