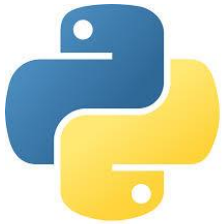
SupplyChain Insights - Understanding Global Supply Chain Operations using Graph RAG

Abhishek Shankar



Introduction

Understanding Supply Chain Operations – Many companies have large datasets either as structured or unstructured data around their products. This project aims to provide a comprehensive solution for accessing and understanding the entire supply chain network.

The final product will facilitate the analysis and visualization of data, enabling organizations to gain insights into their supply chain operations efficiently.

This project demonstrates the application of advanced data science techniques, such as **Graph Data Science and Retrieval-Augmented Generation (RAG)**, to real-world supply chain data, showcasing how these methods can provide actionable insights and enhance decision-making in the industry.

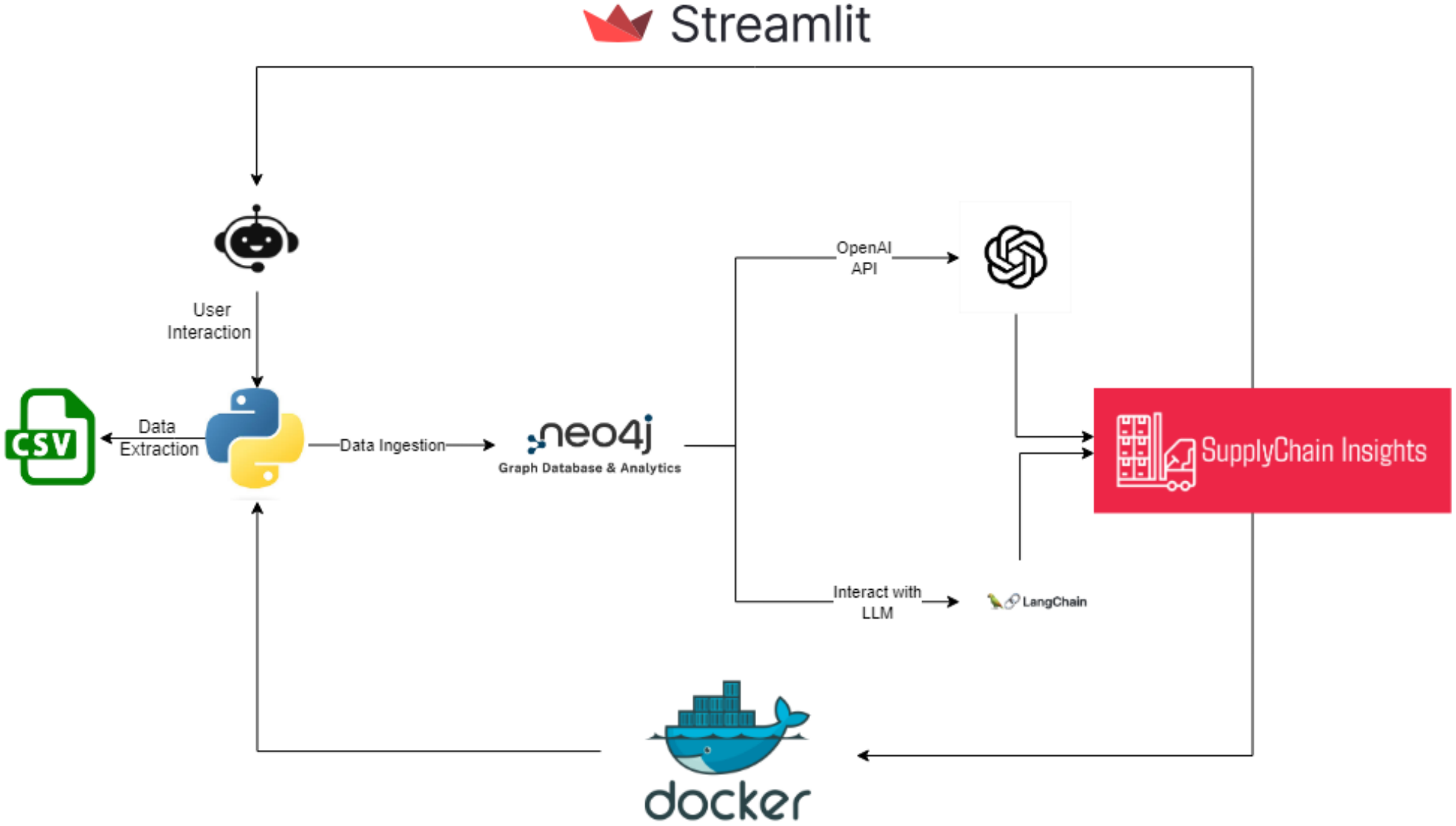
Project Description

Detailed Description: Constructing a Neo4j graph database on a supply chain shipment pricing dataset, using Graph Data Science and RAG for advanced analysis.

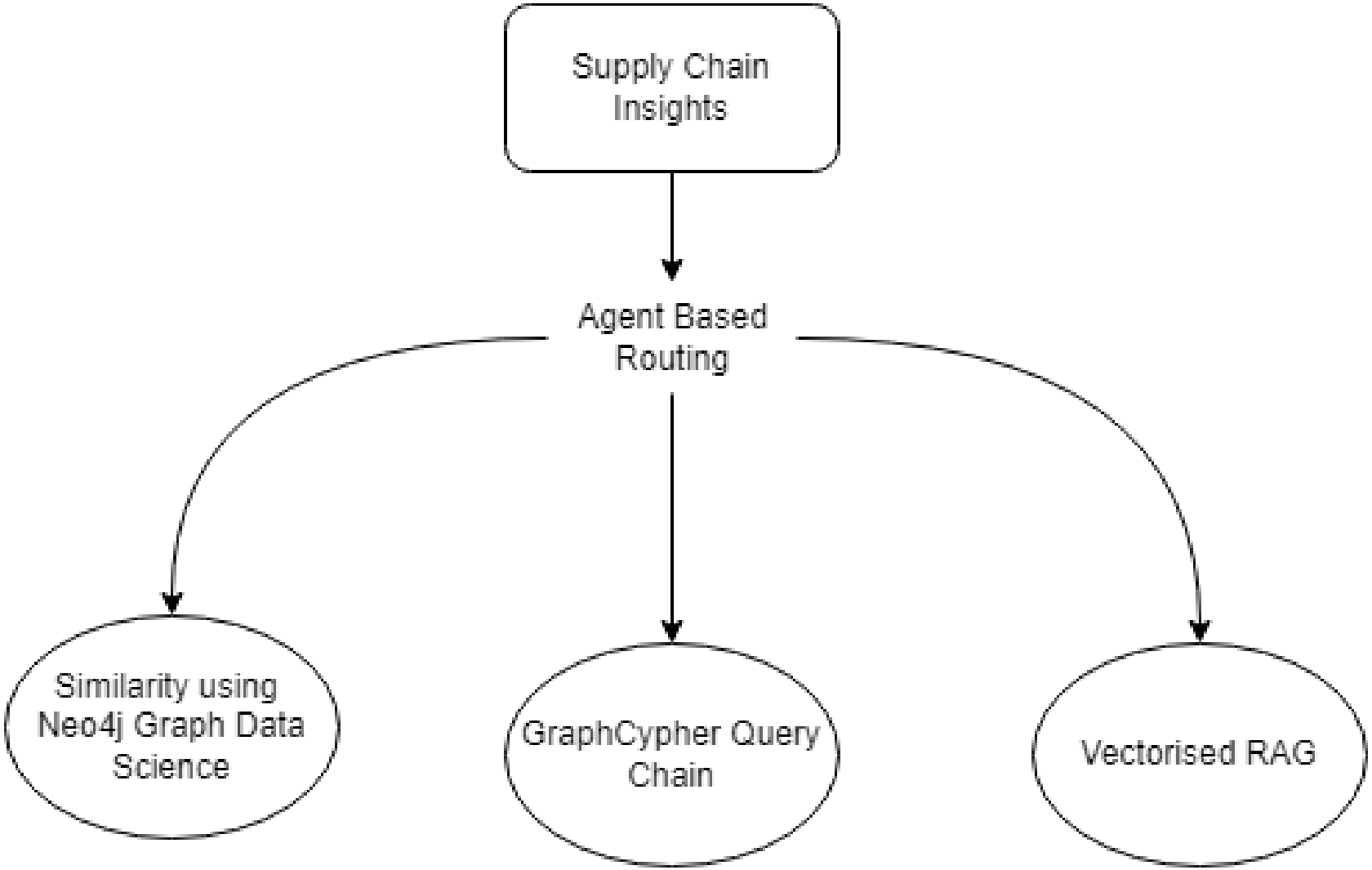
Problem Addressed: Enhances understanding and access to complex supply chain networks, focusing on the health care/pharmaceutical industry.

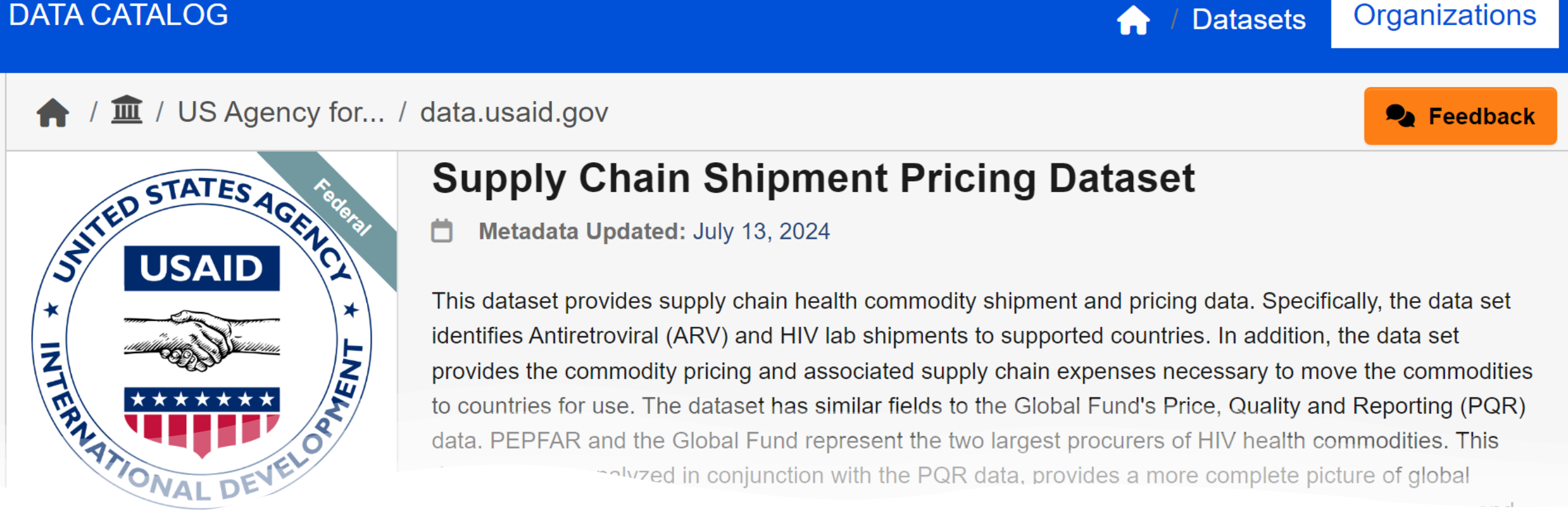
Scope: Integrates data analysis with Graph Data Science and RAG, providing insights into global spending and supply chain trends for informed decision-making.

Project Architecture



Whats Different?

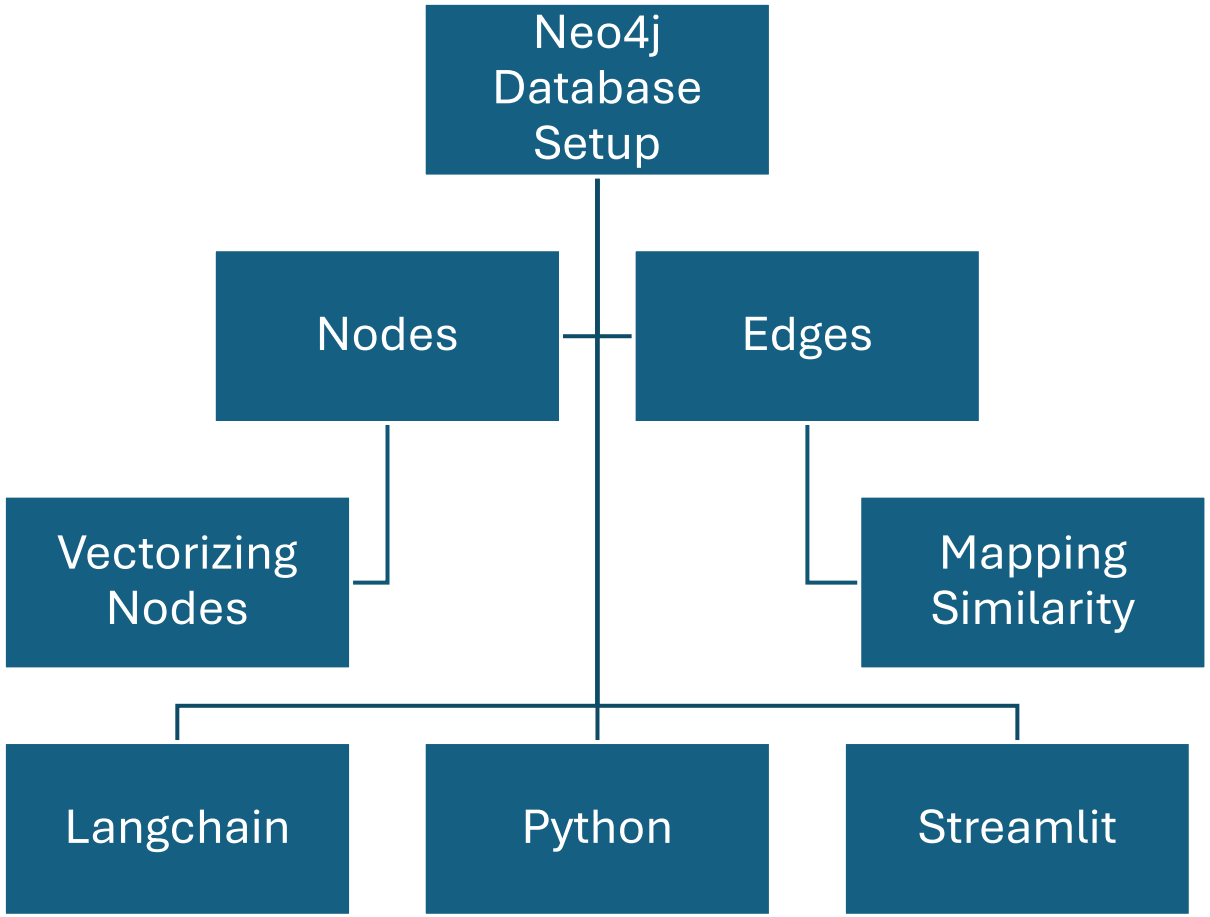




Data Collection and Preprocessing

- Using Python for preprocessing and Cypher for populating my database
- Any data cleaning will be done through Python.

RAG Pipeline Implementation



Performance Metrics

Key Metrics: Accuracy of similarity searches, efficiency of data retrieval, and user satisfaction.

Calculation Methods: Use precision and recall for similarity searches, response time for data retrieval, and user feedback surveys.

Initial Results: High accuracy in similarity searches, fast data retrieval times, and positive initial user feedback.

Methods to Improve Metrics

01

Improvement Strategies: Optimize graph database queries, enhance vector indexing algorithms and refine RAG processes

02

Planned Enhancements: Implement query optimization techniques, refine vector indexing, integrate advanced ML models, and adjust RAG parameters for better context understanding.

03

Expected Impact: Improved search accuracy, faster data retrieval, and increased user satisfaction through more precise and efficient insights.

Deployment Plans

Steps to Deploy: Integrate all components and set up the production environment for the Neo4j graph database and application functionalities.

Tools and Platforms: Use a cloud service provider (e.g., AWS, Azure, GCP) for hosting and Neo4j for database management or containerize the pipeline using docker.

User Testing and Feedback: Launch a beta version using a LinkedIn post or Blog, collect user feedback, and iterate based on insights before full-scale deployment.

Future Scope

Potential Extensions: Integrate additional datasets for broader supply chain analysis and use more unstructured datasets like PDFs to understand product supply chains.

Long-term Vision: Develop a comprehensive supply chain management tool with real-time monitoring and data visualization features.

Further Development: Collaborate with industry stakeholders to refine the tool, incorporate advanced AI techniques, and scale for enterprise use.

Conclusion

Summary of the Project: Developed a Neo4j graph RAG Pipeline to understand global supply chains, utilizing advanced data science techniques for analysis.

Key Takeaways: Improved understanding of supply chain networks, enhanced data retrieval, and actionable insights for better decision-making.

Final Thoughts: The project demonstrates the potential of integrating graph databases and advanced analytics to solve complex industry problems and provides a foundation for future enhancements.

ANY QUESTIONS ?????? ☺