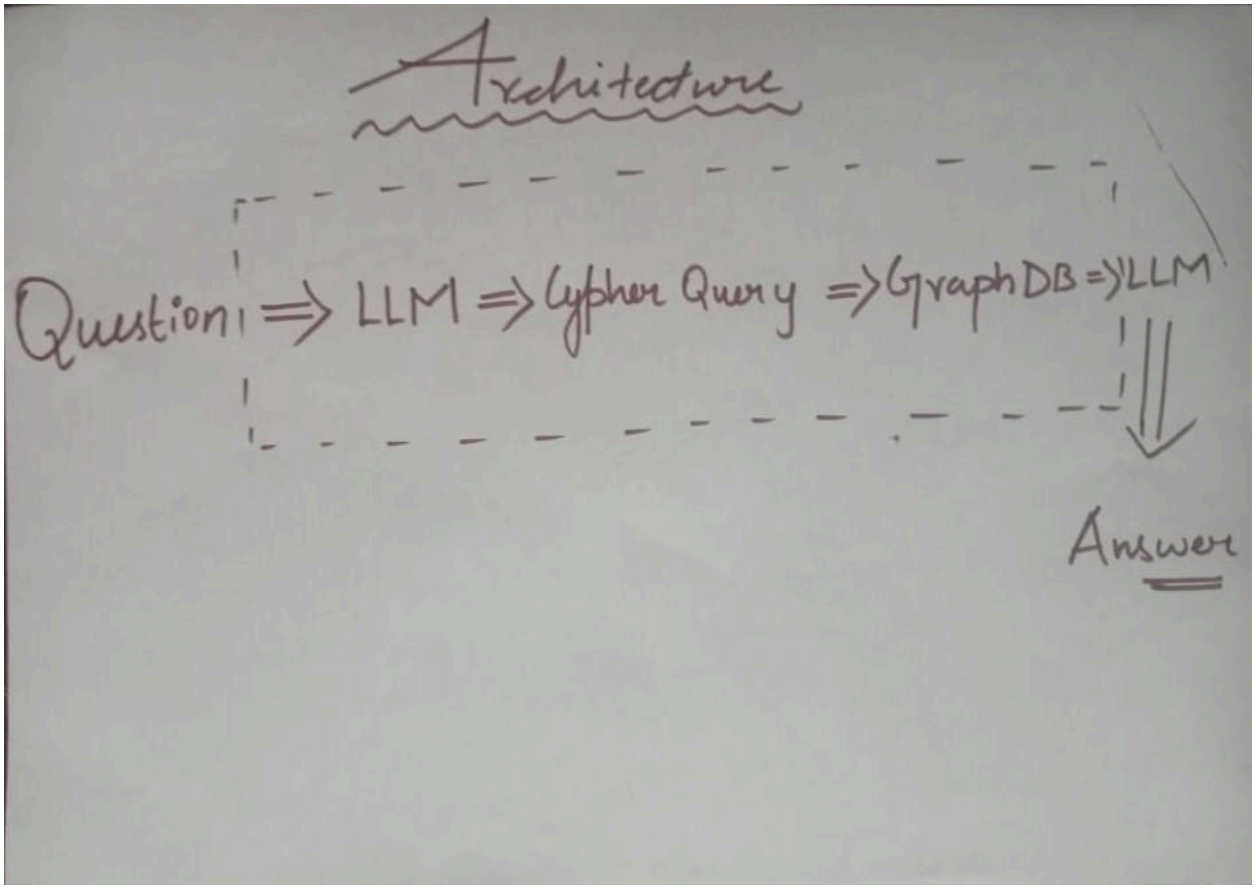
"NLQBase" (Natural Language Query Base)

21PC34 - Shri Meenaakshi S

Problem Statement:

Develop a system that allows users to query a database using natural language, utilizing the Llama2 NLP model to interpret queries and provide responses in natural language. For example, when asked "What is the Genre of Titanic?".

Architecture:



Dataset used:

https://raw.githubusercontent.com/tomasonjo/blog-datasets/main/movies/movies_small.csv

Setup:

- I have used the **Neo4j Graph database**.
- Since running llama2 locally took a huge amount of time , I have used Replicate to deploy llama2 on the cloud.
- Then created an API Token.

Required Packages:

1. Langchain-community
2. Neo4j
3. Pandas
4. Streamlit
5. Replicate

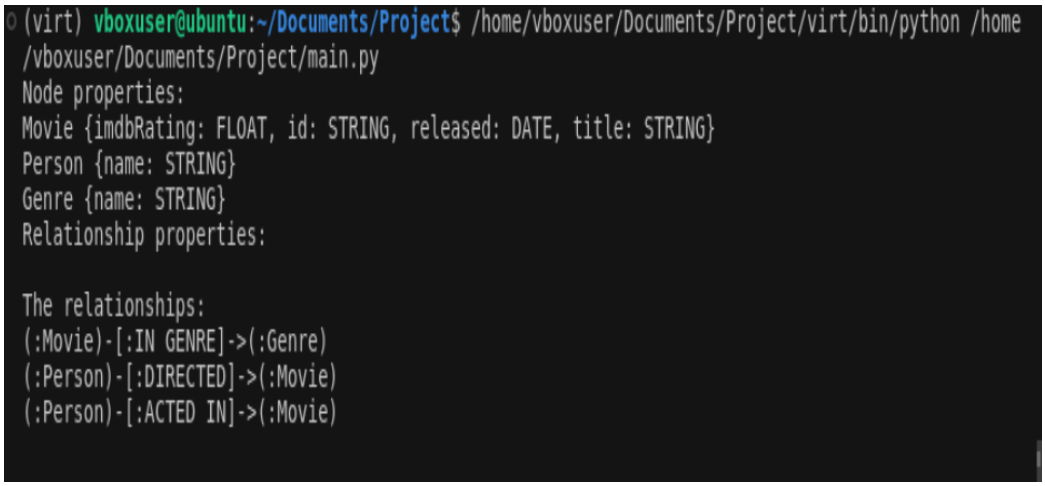
Configuring the environment variables: (To make programs interact with Neo4j)

os.environ["NEO4J_URI"] = "bolt://localhost:7687"

os.environ["NEO4J_USERNAME"] = "neo4j"

os.environ["NEO4J_PASSWORD"] = "password"

Database Schema:



Implementation:

- This code demonstrates a streamlined approach to enable users to interact with a Neo4j graph database using natural language queries.
- By leveraging the Llama-2 model via Replicate, it provides accurate and intuitive responses to user queries, enhancing the user experience.
- Through Streamlit, it offers a user-friendly interface for users to input queries and visualize query results.
- Overall, this application simplifies the process of querying graph databases, making it accessible to users without expertise in database query languages.

Code:

```
import os
from langchain_community.llms import Replicate
from neo4j import GraphDatabase
import replicate
import streamlit as st
import pandas as pd

os.environ["REPLICATE_API_TOKEN"] = "r8_c9xueVedSGsoJOcpAjSlMMZEKW7Qeua34VqBt"

llm = Replicate(
  model="meta/llama-2-70b-chat",
  model_kwargs={"temperature": 0.75, "max_length": 500, "top_p": 1},
)

df =
pd.read_csv("https://raw.githubusercontent.com/tomasonjo/blog-datasets/main/movies/movies_small.csv")
```

```
st.title(" NLQBase (Natural Language Query Base) ")
st.write("The Dataset which was choosen for the demonstration is : ")
st.write(df)

# Initialize session state to keep track of the application's state
if 'queries' not in st.session_state:
  st.session_state.queries = []

# Define the prompt template with the schema
prompt_input = """
Given the following schema:
Node properties:
Movie {imdbRating: FLOAT, id: STRING, released: DATE, title: STRING}
Person {name: STRING}
Genre {name: STRING}
Relationship properties:

The relationships:
(:Movie)-[:IN_GENRE]->(:Genre)
(:Person)-[:DIRECTED]->(:Movie)
(:Person)-[:ACTED_IN]->(:Movie)

Provide me only the Cypher query not any other text for the following natural language request.
Provide me only the correct and only query.
"""

# Function to generate Cypher query from natural language query using LLM
def get_cypher_query(user_query):
  full_prompt = prompt_input + f"\n{user_query}"
  cypher_query = llm(full_prompt)
```

```
    return cypher_query

# Function to run Cypher query on Neo4j
def run_cypher_query(query):
    neo4j_uri = "bolt://localhost:7687"
    neo4j_user = "neo4j"
    neo4j_password = "password"

    driver = GraphDatabase.driver(neo4j_uri, auth=(neo4j_user, neo4j_password))
    with driver.session() as session:
        result = session.run(query)
        return result.data()

# Display previous queries and results
for query, results, nl_response in st.session_state.queries:
    st.write(nl_response)

# User input for natural language query
user_query = st.text_input("Enter your query")

if user_query:

    cypher_query = get_cypher_query(user_query)

    query_results = run_cypher_query(cypher_query)

    prompt_to_nl = f"The query by the user is: {user_query}\nThe exact answer
is:\n{query_results}\nGive me the above response in natural language only like an answer to a
question."
    natural_language_response = llm(prompt_to_nl)
```

```
st.write("The answer for the given query is:")
st.write(natural_language_response)

# Store the query and results in session state for repeated display
st.session_state.queries.append((user_query, query_results, natural_language_response))
```

Output:

NLQBase (Natural Language Query Base)

The Dataset which was choosen for the demonstration is :

	released	title	actors
0	1995-11-22	Toy Story	Jim Varney|Tim Allen|Tom Hanks|Don Rickles
1	1995-12-15	Jumanji	Robin Williams|Bradley Pierce|Kirsten Dunst|Jonathan Hyde
2	1995-12-22	Grumpier Old Men	Walter Matthau|Ann-Margret|Jack Lemmon|Sophia Loren
3	1995-12-22	Waiting to Exhale	Whitney Houston|Lela Rochon|Angela Bassett|Loretta Devine
4	1995-12-08	Father of the Bride Part II	Steve Martin|Kimberly Williams-Paisley|Diane Keaton|Martin Short
5	1995-12-15	Heat	Al Pacino|Robert De Niro|Val Kilmer|Jon Voight
6	1995-12-15	Sabrina	Julia Ormond|Harrison Ford|Nancy Marchand|Greg Kinnear
7	1995-12-22	Tom and Huck	Jonathan Taylor Thomas|Brad Renfro|Eric Schweig|Charles Rocket
8	1995-12-22	Sudden Death	Jean-Claude Van Damme|Powers Boothe|Raymond J. Barry|Whittni W
9	1995-11-17	GoldenEye	Pierce Brosnan|Famke Janssen|Sean Bean|Izabella Scorupco

Enter your query

What is the genre of Tom and Huck

The answer for the given query is:

Sure, I'd be happy to help! The genre of Tom and Huck is Adventure and Children's literature.