

Graph Database Topics.

Assignments – Neo4j

You will be querying three Neo4j databases, provided to you together with the software. These databases are: (1) A graph representation of the Northwind operational database, denoted **northwindhg.db**; (2) A database containing information about movies, denoted **movies.db**; (3) A database containing historical information about the world cups, from the start through the 2014 world cup, denoted **worldcup.db**.

Once you login to your account (in Linux), you need to go to the /tmp folder. Once there, open a terminal and type the following commands:

```
cd /tmp
tar xJf /serveur/neo4j.tar.xz
cd neo4j
```

At this point, you are about to start the Neo4j server. First, you need to choose the database you will work with. For this, you go to the **conf** folder, and edit the **neo4j.conf** file. You will find something like this:

```
#dbms.active_database=graph.db
#dbms.active_database=trajectories-NYC-4sq.db
#dbms.active_database=trajectories.db
#dbms.active_database=worldcup.db
#dbms.active_database=web.db
#dbms.active_database=telco.db
dbms.active_database=northwindhg.db
#dbms.active_database=movies.db
#dbms.active_database=calls.db
```

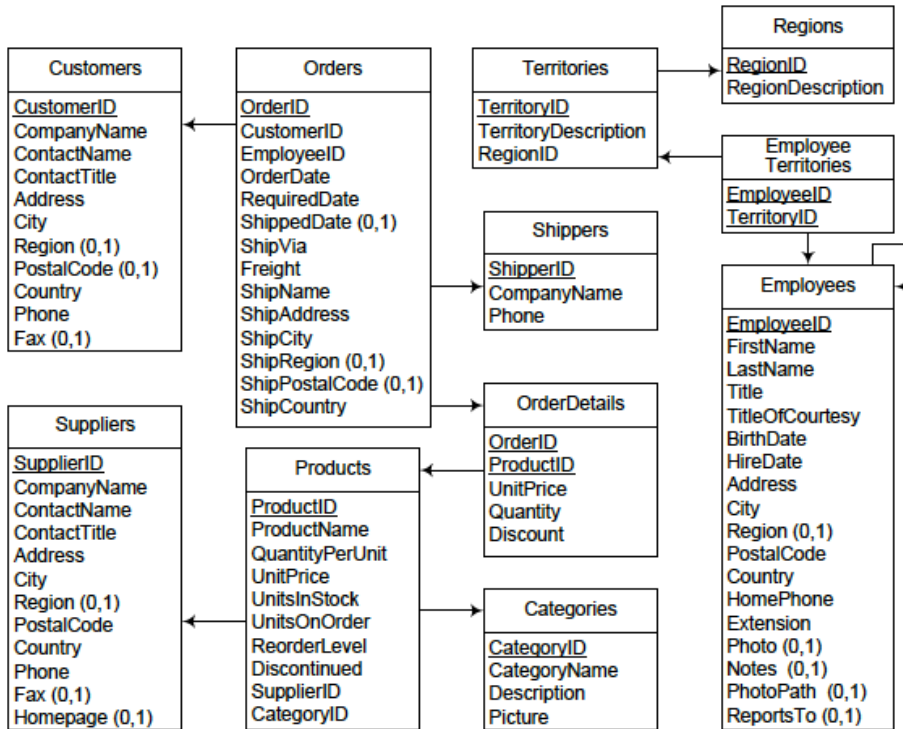
Since `dbms.active_database=northwindhg.db` is unmarked, when you start the server, you will also load the **northwindhg.db** database. To change it to the **movies.db**, you mark `#dbms.active_database=northwindhg.db`, and unmark `dbms.active_database=movies.db`. Save the changes, and quit the file. Then you run: `./bin/neo4j console`

And the server starts. Then, open a browser, and type the following url: **localhost:7474**.

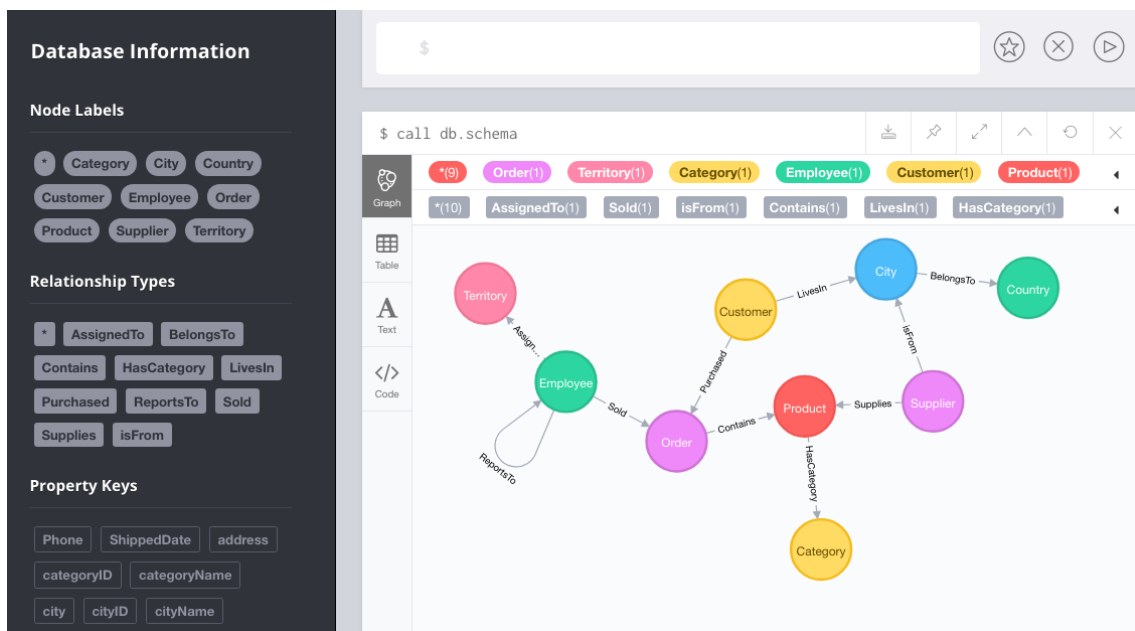
Now you can start writing Cypher queries.

Assignment 1.

Consider the Northwind database, whose schema is:



This database has been exported to Neo4j, and you can find it at: /...../data/databases/northwindhg.db. The graph schema is:

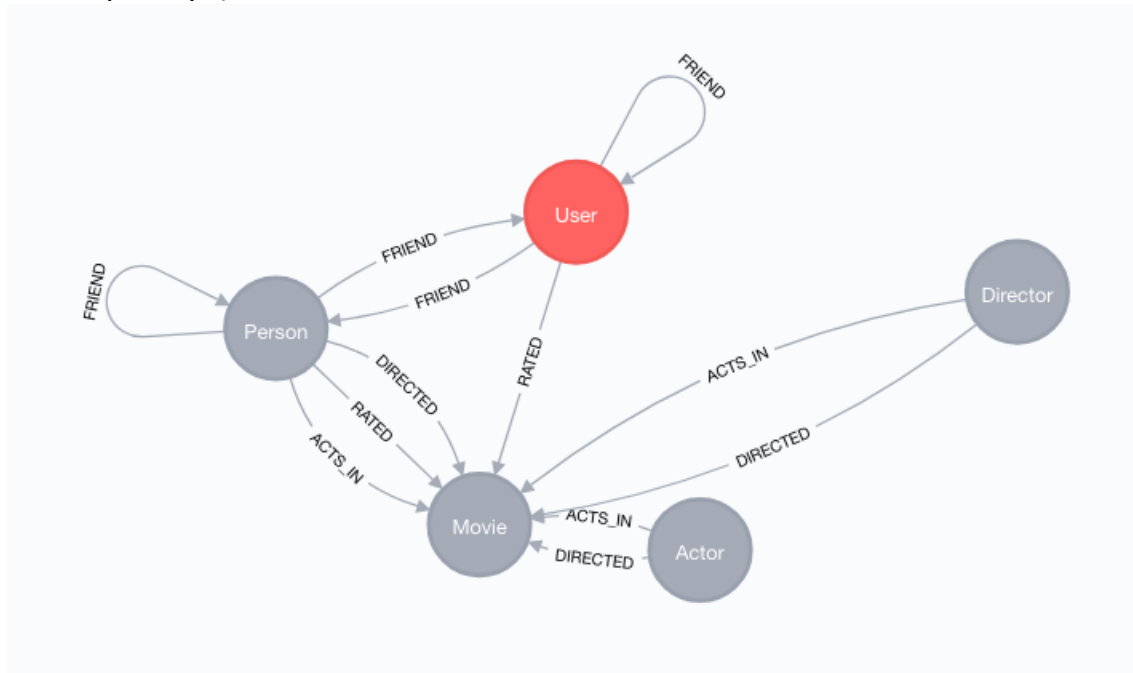


Write in Cypher the following queries over the northwindhg.db database:

- 1 - List products and their unit price.
- 2 - List information about products 'Chocolate' & 'Pavlova'.
- 3 - List information about products with names starting with a "C", whose unit price is greater than 50.
- 4 - Same as 3, but considering the sales price, not the product's price.
- 5 - Total amount purchased by customer and product.
- 6 - Top 10 employees, considering the number of orders sold.
- 7 - For each employee, list the assigned territories.
- 8 - For each city, list the companies settled in that city.
- 9 - How many persons an employee reports to, either directly or transitively?
- 10 - To whom do persons called "Robert" report to?
- 11 - Who does not report to anybody?
- 12 - Suppliers, number of categories they supply, and a list of such categories
- 13 - Suppliers who supply beverages
- 14 - Customer who purchases the largest amount of beverages
- 15 - List the 5 most popular products (considering number of orders)
- 16 - Products ordered by customers from the same country than their suppliers

Assignment 2.

Switch to the movies.db database. For this, **stop the server**, using the ctrl-c command in the terminal. Then, edit the neo4j.conf as explained, and unmark the movies.db line. Then, star the server again. When you open the browser and type the url localhost:7474, you'll have the neo4j database available. The schema is (you can get this writing call db.schema at the prompt):

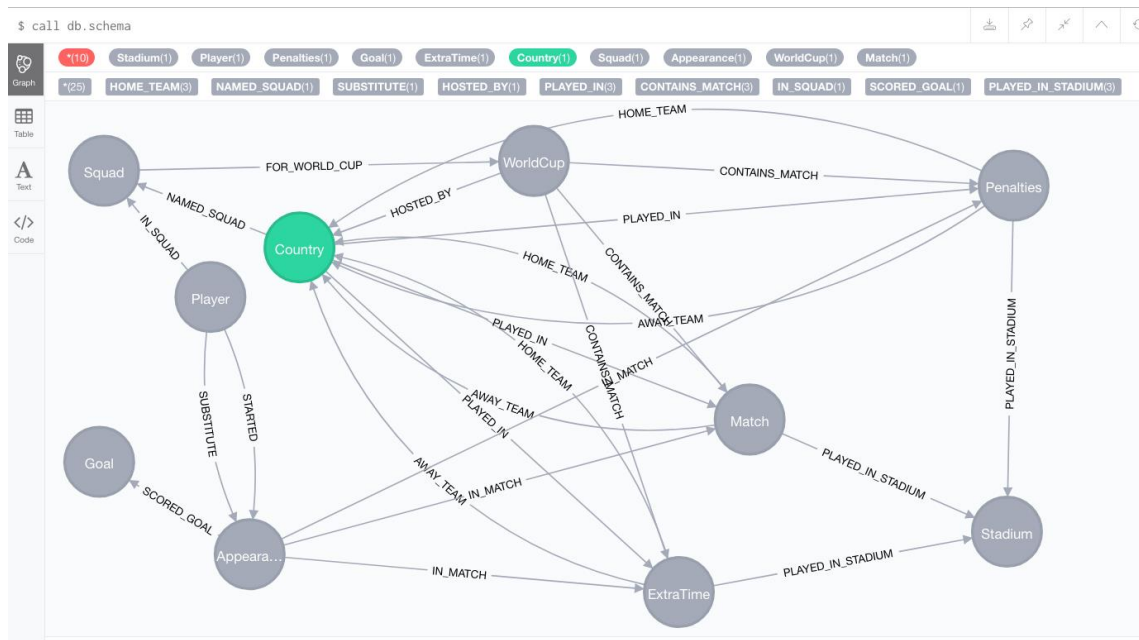


Write in Cypher the following queries over the movies.db database:

- 1 - Actors who played in two movies directed by the same director. Return the actor, the films, and the director.
- 2 - Actors who played in the same film with Kevin Bacon.
- 3 - Actors who played in a movie directed by Robert De Niro.
- 4 - For each actor, list the number of actors she played with in a movie.
- 5 - Actors who played in a movie with Samuel L. Jackson.
- 6 - Shortest path between Robert De Niro and Kevin Bacon.
- 7 - Shortest path between Kevin Bacon and Stephen Lang.
- 8 - Shortest path between Kevin Bacon and any other actor.

Assignment 3.

Switch to the **worldcup.db** database, doing the same steps as in Assignment 2. Now, the database is **worldcup.db**. The schema is:



Write in Cypher the following queries over the **worldcup.db** database:

- 1 - Who hosted the world cup?
- 2 - Who hosted the World Cup more than once, and when.
- 3 - Hosts that won the World Cup, and the result of the final match.
- 4 - Top scorers per world cup.
- 5 - Top scorer playing in the 2018 World Cup.
- 6 - Which hosts won the World Cup that they hosted?
- 7 - Which countries have never won a match at a World Cup?
- 8 - What's the highest number of goals scored in a World Cup match?
- 9 - Which stadium has hosted the most World Cup matches?
- 10- Which country has scored the most goals across all World Cups?
- 11 - Which country has participated in the most World Cups?

12 - Which hosts won the World Cup that they hosted?

13 - Which countries have never won a match at a World Cup?

14 - What's the highest number of goals scored in a World Cup match?

15 - Which stadium has hosted the most World Cup matches?

16 - Which country has scored the most goals across all World Cups?

17 - Which country has participated in the most World Cups?