

Introduction to Graph Databases

Activity 3 - Cypher

You will be querying two Neo4j databases, provided to you. These databases are: (1) A graph representation of the Northwind operational database, denoted **northwindhg.db**; (2) A **rivers** database, with data from the Flanders river system, in Belgium.

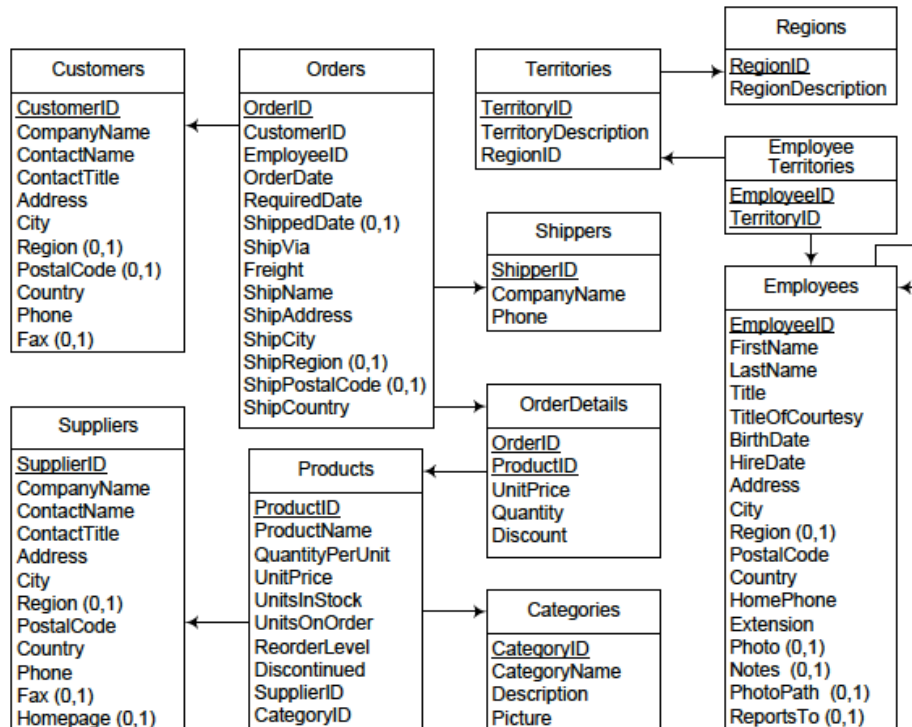
Before starting the Neo4j server, you need to choose the database you will work with. For this, you go to the **conf** folder, and edit the **neo4j.conf** file. You will find something like this:

```
#dbms.default_database=minigraphweb
#dbms.default_database=northwindhg
dbms.default_database=rivers
#dbms.default_database=neo4j
#dbms.default_database=webgraph3
#dbms.default_database=webdb
#dbms.default_database=minigraphweb
```

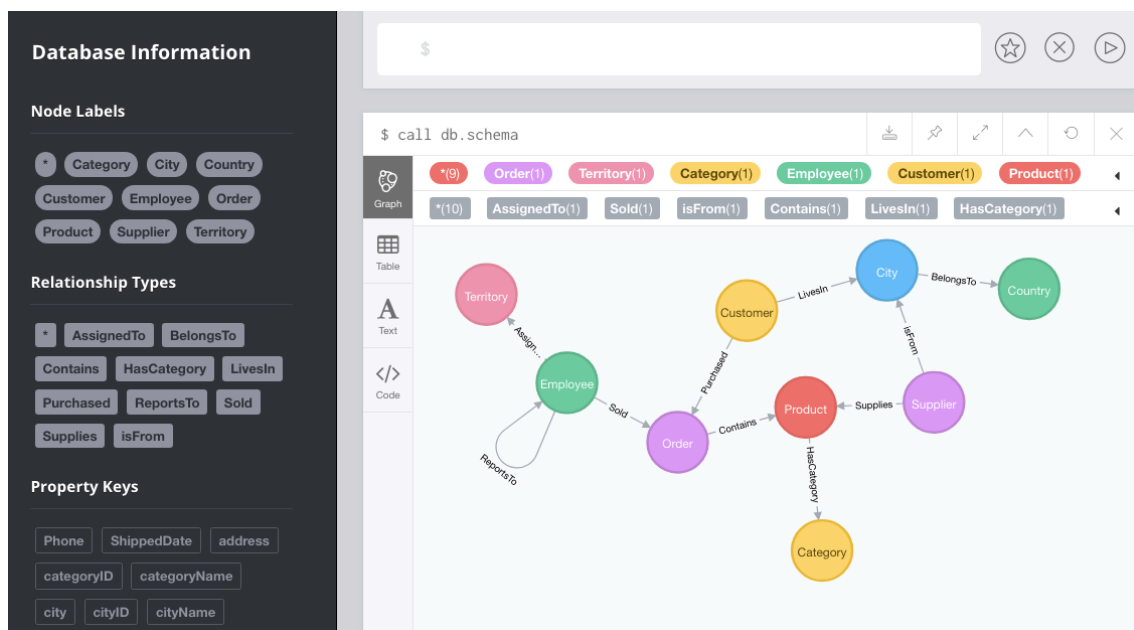
Since `dbms.default_database=rivers` is unmarked, to change the database to `northwindhg`, you mark `#dbms.default_database=rivers`, and unmark `dbms.default_database=northwindhg`. Save the changes, and quit the file. Then you run: `./bin/neo4j console` to start the Server. Then, open a browser, and type the following url: **localhost:7474**. Now you can start writing Cypher queries.

Exercise 1.

Consider the Northwind database, whose schema is:



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



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

Query 1 - List products and their unit price.

Query 2 - List information about products 'Chocolade' & 'Pavlova'.

Query 3 - List information about products with names starting with a "C", whose unit price is greater than 50.

Query 4 - Same as 3, but considering the sales price, not the product's price.

Query 5 - Total amount purchased by customer and product.

Query 6 - Top ten employees, considering the number of orders sold.

Query 7 - For each employee, list the assigned territories.

Query 8 - For each city, list the companies settled in that city.

Query 9 - How many persons an employee reports to, either directly or transitively?

Query 10 - To whom do persons called "Robert" report to?

Query 11 - Who does not report to anybody?

Query 12 - Suppliers, number of categories they supply, and a list of such categories

Query 13 - Suppliers who supply beverages

Query 14 - Customer who purchases the largest amount of beverages

Query 15 - List the five most popular products (considering number of orders)

Query 16 - Products ordered by customers from the same country than their suppliers

Exercise 2.

We will query the Flanders river system depicted in Figure 1. The schema and properties are shown in Figures 2 to 4. Segments are represented as nodes, with label `:Segment` (and their corresponding properties), and the relation between the nodes is called `:flowsTo`, defined as follows: there is a relation `:flowsTo` from node A to node B if the water flows to segment B from segment A. This is stored in the **rivers database**.

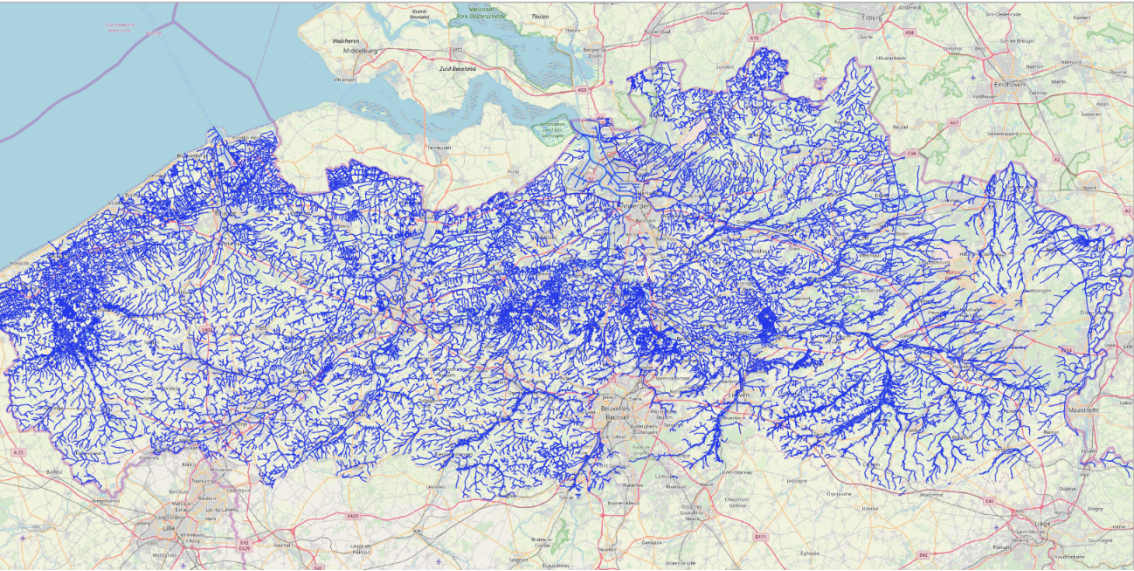


Figure 1

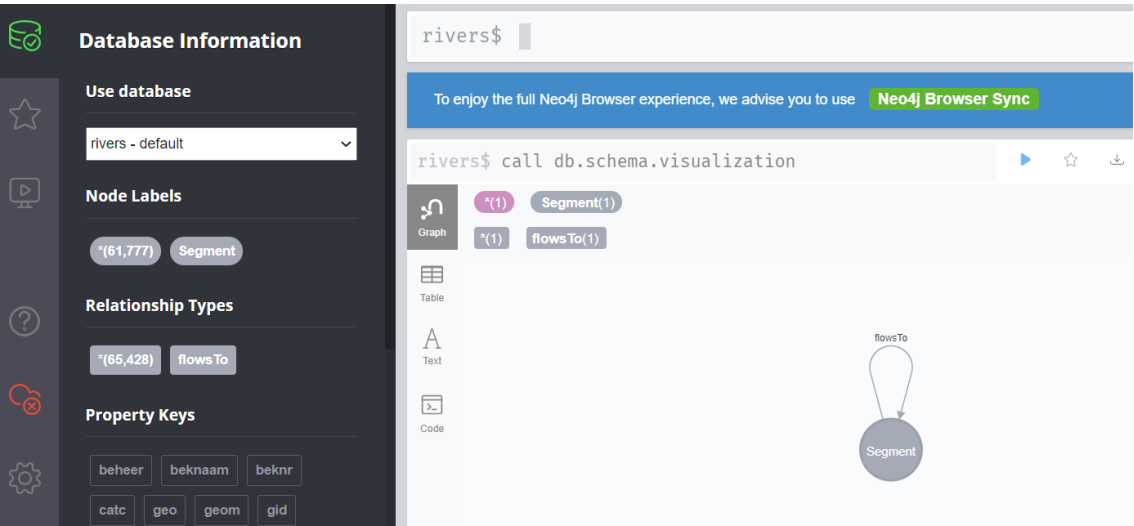


Figure 2. Schema

	n
1	<pre>{ "identity": 23715, "labels": ["Segment"], "properties": { "kwaldoel": 110, "gid": 45346, "wtrlichc": "NG_L217_0601", "source": 45686, "geom": "SRID=31370;MULTILINESTRING((91163.005400002 213959.5757,91164.0419000015</pre>

Figure 3. Properties

rivers\$ MATCH (n:Segment) RETURN n LIMIT 25	
	n
	<pre>-----,----- 214144.799400002,91215.6618999988 214145.390700001))", "lblkwal": "Productie drinkwater", "source_long": 3.5263787509275333, "oidn": 117936, "geo": 1, "vhas": 4520093, "target_long": 3.527102449351701, "beheer": "P4.045", "beknr": 2, "vhazonenr": 84, "catc": 9, "uidn": 635422, "lengte": 193.33,</pre>

Figure 4. Properties

Property vhas is the segment iD.
Property lengte is the length of the river
Property geom is the geometry of the segment
Property catc is the category of the segment

Query 1. Compute the average segment length.

Query 2. Compute the average segment length by segment category

Query 3. Find all segments that have a length within a 10% margin of the length of segment with ID 6020612.

Query 4. For each segment find the number of incoming and outgoing segments.

Query 5. Find the segments with the maximum number of incoming segments.

Query 6. Find the number of splits in the downstream path of segment 6020612

Query 7. Find the number of in-flowing segments in the downstream path of segment 6020612.

Query 8. Determine if there is a loop in the downstream path of segment 6031518.

Query 9. Find the length, the # of segments, and the IDs of the segments, of the longest branch of upstream flow starting from a given segment.

Query 10. How many paths exist between two given segments X and Y?

Query 11. Find all segments reachable from the segment closest to Antwerpen's Groenplaats