DBMS Practical

MONDIAL Database - Assignment

Deadline: DD-MM-YYYY HH:MM

For each task, there is a maximum allotted time. Please do not work longer than the specified time on a task! If you think that you will not be able to finish the task in the given maximum time, stop working on it 10 minutes before the end, and provide an explanation containing the following information:

- Whether you think that the task is solvable with the current system at all, and why (not)?
- If you think that is solvable with more time: which approach, would you try out next?

TASK 1 Use Case: Equi Join

1.1 List of cities with their countries
For each city, show the name of the country it belongs to. The output should contain the city's name and the country's name it is located in.

1.2 Number of cities
For each country: Find out how many cities are located in each country. The output per country shall contain the corresponding number of cities.

1.3 Number of organizations being part of
For the country ‘United States’: Count the number of organizations it is a member of. The output should contain the country’s name and the number of organizations.

TASK 2 Use Case: Theta Join

2.1 Correlation between GDP and organization memberships

Are countries with a higher GDP than the average GDP are part of more international organizations than countries with a lower GDP?
Query for countries that have a higher GDP than the average GDP and check whether they are members of more organizations than those countries that have a lower GDP (equal is not considered).
First, compute the average GDP and display it along with its country's GDP.
Then, produce two separate outputs:
- One for all countries with GDP above the average
- One for all countries with GDP below the average
Finally, output the number of organizations joined by countries above and below the average GDP.

TASK 3 Use Case: Schema Evolution

3.1 Update the Schema
Introduce a new element (attribute) to the country's entity set. Find the general syntax to do that.

3.2 Add information for entity set
Use the syntax from the previous task and add the new column ‘density’ to the table/node ‘country’.

Calculate and insert for each country its population density (per square kilometer) using the existing columns area and population.

TASK 4 Use Case: Missing Values

4.1 Find missing values
Find missing values for each attribute of the politics table/node. Which attribute has the most missing values?

TASK 5 Use Case: Range queries

5.1 Countries with a population between two values
Select all countries that have a population between one million and 6 million inhabitants. Return the name and population of each country in that range.

5.2 Countries in organizations founded between two dates

Select all countries that are members of at least one organization that was established between 01.01.1945 and 31.12.1960. First, find out which date and time format is used. Return the country's name and the names of the corresponding organizations.

TASK 6 Use Case: Network analysis

Using borders between countries
Network analysis can be done to understand geopolitical structures. In this context, each country is considered a node in the network, and the edges are represented by land borders connecting them. To determine how far apart two countries are in this network, we count the number of hops, that is how many borders must be crossed to travel from one country to another over land. To calculate the number of hops, we use the formula m + 1, where m is the number of intermediate countries (nodes) between two countries. In the Mondial dataset, the network consists of countries connected by shared borders, forming a geographical network that can be analyzed just like a social or communication network.

6.1 Network size by border crossings
If the network nodes are fully connected, how many hops are needed to reach all other countries using the border table/relationship, starting from Germany? Count how far each country is from Germany and identify the maximum number of hops required to reach any reachable country.
(Hint Neo4j: Constrain the path length for smaller paths until you find the right query.)

6.2 2-Hop Border Network of Greece
Which countries are in the 2-hop border network of Greece? Use only the border table/relationship, and return all countries that are reachable in exactly two hops (excluding Greece itself).

6.3 Border Network Crossings from Portugal to Greece
How many land borders must be crossed at minimum to get from Portugal to Greece overland?
Use the border table/relationship to compute the shortest border path.

6.extra Count outgoing edges
Find all countries that are exactly 5 borders away from Uganda. The output shall contain the name(s) of the countries.

(does not work properly, as countries are round and are typically connected to the same neighbors in both directions which causes multiple overlapping paths and often more countries to be counted than truly expected in a strict one-way path model)

Using river estuaries
Network analysis can also be done to natural systems such as river networks. In this context, each river is modeled as a node in the network, and the edges between them indicate that one river flows into another.
To determine how far apart two rivers are in this network, we count the number of hops, which corresponds to the number of estuary connections between them. To calculate the number of hops, we use the formula m + 1, where m is the number of intermediate rivers in the path between the source and the destination river. Further, in the Mondial dataset, cities are connected to rivers via a relationship. This enables river-based routing between cities by traversing from the river of the starting city to the river of the destination city.

6.4 River Path from one to another city
How many river hops are needed to travel from Luxembourg to Mainz via connected rivers?
(Neo4j: Consider “ESTUARY_IN_RIVER” links to compute the number of hops.)

TASK 7 Use Case: User defined function (UDF)

7.1 Categorize Countries by Population Size
Take the population value of each country and classify it into categories such as "Small", "Medium", or "Large". You might need to use an external programming language to implement a UDF (User Defined Function) that returns a population size category based on a given threshold (e.g., small < 5 million, medium < 50 million, otherwise large). Apply the UDF to the population attribute of all countries and return each country along with its assigned category.