CSCI 4180: Big Data Analytics
CSCI/DASC 6010: Big Data Analytics and Management
Spring 2023
Homework 3 – Cypher and SQL
Due Sunday, February 12, at 11:59 PM

Preliminary

The `micro-university.xlsx` file contains the data about a few students, courses, projects, rooms, and the relationships between. Load these data in a graph database (e.g., Neo4j) and in a relational database (e.g., PostgreSQL).

1   Neo4j schema design and data load (50 pts)

1.1   Schema design (20 pts)

Draw the database schema for the graph database, using http://www.apcjones.com/arrows. Include a screenshot of it.

1.2   Data load (30 pts)

List the Cypher queries to load the data in the database.

2   Data analysis (60 pts: 5 pts each per SQL query, 5 pts each per Cypher query)

Write the SQL queries and the Cypher queries to answer the following questions.

a) In which rooms does the course with **Course number** 1 takes place? Retrieve the course name and the names of the rooms in which the course takes place.

b) How many hours and in which projects does student with **Student ID** 1 works on? Retrieve the first name of the student, the project the student works on, and the corresponding number of hours worked on the project.

c) Which students and how many hours do they work on the project with **Project number** 24? Retrieve the project name, the last name of the student, and the corresponding number of hours worked on the project.

d) Which students work in which projects and how many hours? Retrieve the last name of the students, the name of the projects they work on, and the corresponding number of hours. Order the results by the last name of the students. Limit the results to four.

e) Which students work on more than two projects and on how many projects exactly? Retrieve the last name of the students and the corresponding number of projects. Order the results by the number of projects.

f) Which students have the same last name and work on the same projects? Retrieve the first name of the students and the name of projects they share.

Deliverables

Create one ZIP archive with the Rmd, PDF output, and image files, and upload it to Canvas by the due date. Include the following in the Rmd file:

- The Neo4j schema, and the results of the queries from question 2 (screenshots).

- The Cypher queries to load and retrieve the data (for questions 1.2, and 2).

- The SQL queries from question 2.