Week 13 Assignment

Week 13 assignment is due end of day on Sunday May 3rd.. *You may work in a small group on the assignment.*

Instructions

In this assignment, you'll load a data set into Neo4j and manipulate it by constructing appropriate Cypher queries. Note that we will prefer the Load CSV feature, but you can (and probably should!) test your queries individually by hand first.

Preliminaries

The data we will work with come from the two CSV files attached to this assignment. The data contain records for (fictional) students enrolled at a local college, the courses in which they are enrolled, and their residence arrangement. The data structure will be as follows (with each pattern's attributes given in parentheses):

Nodes:

- Student (firstname, lastname, id, gender, address, city, state, zipcode, phone)
- Course (department, number, title)
- Dormitory (name)

Relationships:

- Enrolled (section, instructor, grade [value of IP for in progress])
- Completed (section, instructor, grade)
- Housed (room)

Deliverables

1. Your first job is to produce the code to read in the CSV files to create the proper output. Note that student nodes will have relationships with both courses (enrolled or completed) and dormitories (housed). Submit the code as a plain text file. Note that you may wish to preprocess the CSV files to make your life easier. Explain any steps you take to do so. (For instance, you may wish to create a separate CSV file for distinct courses, or separate files for completed versus enrolled.)

2. In addition, please provide a short paragraph commenting on whether a graph database is a better choice or a worse choice than a SQL database for this task. Back up your claim succinctly with explanation. You **do not** need to implement the SQL solution.

3. Write the Cypher query that will find all of the roommates of the student Richard Kowalski.

4. Finally, **s**uppose you were told Richard Kowalski, who was enrolled in section 12136 of Math 120: Finite Mathematics, completed the course with a grade of B. Show the query that would find the appropriate relationship and update both the label (enrolled to completed) and the grade.

Bonus Deliverable

We have instructor as a property of the relationship "enrolled" in our model. Describe a data model that might improve on this setup by making instructor a type of node rather than an attribute. Which way do you think might make more sense? Does the use case affect your opinion? Explain. (You may wish to sketch a picture of what this new model would look like.)