Assignment 2
Advanced Database Design and Implementation

Purpose and Learning Outcomes

Purpose

The purpose of the assignment is to provide students with the opportunity to apply knowledge and skills developed during the semester with particular reference to:

1. Interpretation of business rules from additional information for case study;
2. Knowledge acquisition with respect to Spatial, Graph and Document Databases:
   a. Creation of a spatial database, graph database and a document database;
   b. Creation of spatial data within the spatial database, graph data within the graph database and document data within the document database;
   c. Querying of the spatial, graph and document data

Learning Outcomes

The learning outcomes directly assessed are:

Knowledge:

K2. Explain the motivation for emerging trends in database technology, and the impact of database technology on organisations.

K3. Discriminate between different types of database systems (e.g. relational, spatial, NoSQL, hierarchical, graph, object)

Skills:

S1. Interpret conceptual level diagrams (e.g. entity relationship) to implement a database.

S2. Demonstrate skills in designing and building a database application using a commercially available database management system development tool.

S3. Use a query language for data manipulation.

Application of Knowledge and Skills:

A1. Design and implement a relational database using a database management system.

A2. Utilise a query language tools and techniques to obtain data and information from a database.

Timelines and Expectations

Percentage Value of Task: 30% of the course marks
Due: Refer to Course Descriptor or Moodle
Minimum time expectation: 25 hours

Students are required to complete the assignment individually.
Students are expected to submit the required report and details (see below) to the submission box in their
Moodle shell.

Assignment Requirements

Case Study – Tourism Track Recommender System

Refer to **Figure 1** for a high-level overview of the system to be developed.

The company for whom you are developing this system wants people to visit local walks that are considered to have
touristic appeal. For instance in Melbourne there are several tourism guide books and websites that detail walks that
interested tourists can use to explore sites of local interest.

In order to find potentially interested people, the company trawls various sources of social media (e.g. Facebook, Twitter,
Instagram etc) for geocoded data – which could include a mention of a certain place, or coordinates tagging postings or
hidden within image EXIF data.

The geographic coordinates are then used to search through a database of *tracklogs* for local touristic walks. The nearest
walk is retrieved and recommended to that person.

Additionally, during the social media trawl, friendship links are gathered, and these are used to recommend the same
retrieved walk to the persons' friends. The links are inserted into a separate database for later use.

Additionally, anyone who wants to submit a review of their walking experience can do so. This is inserted into a separate
database.

**NOTE: The complete scenario for the assignment will be discussed in lectures and support provided in tutorials,
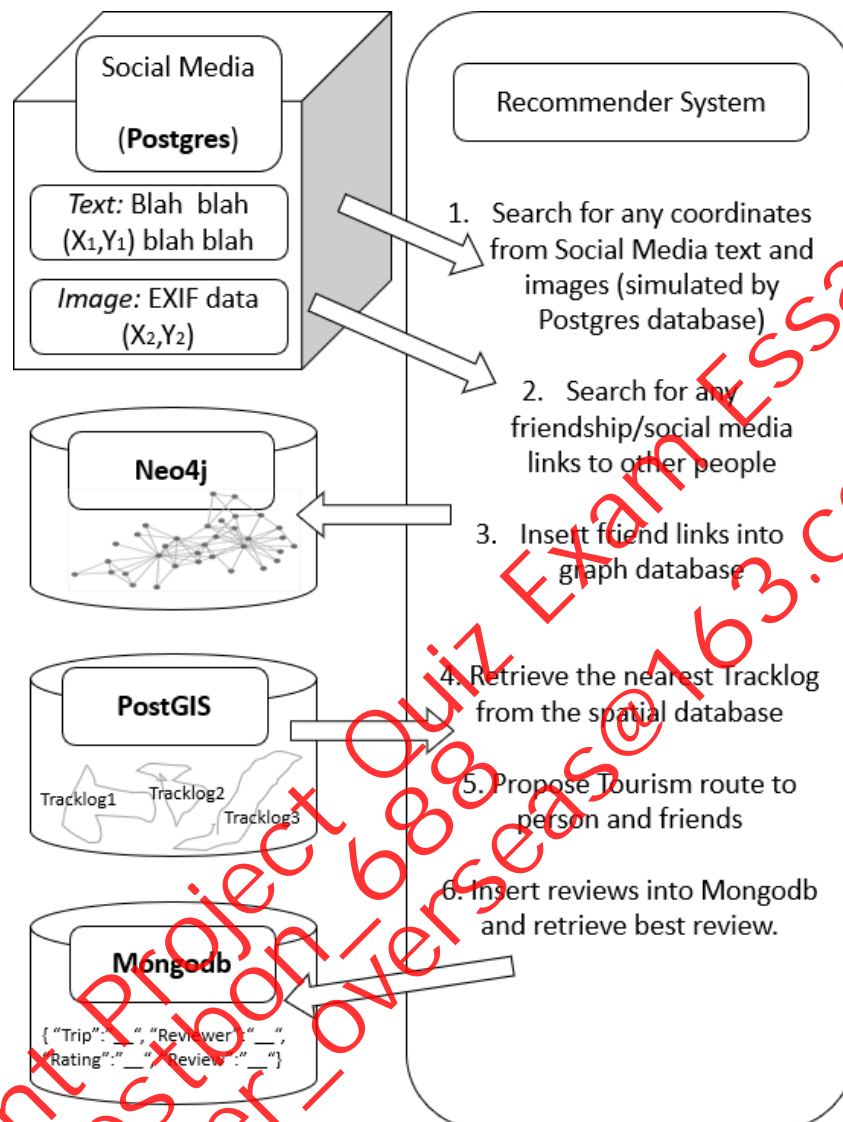including appropriate algorithms for calculating the nearest tracklog.**

Figure 1: Overview of Tourism Track Recommender System

Technical Details

While the system as explained appears to be a single application, you do not need to develop a Graphical User Interface that incorporates all of these different elements. Instead you merely need to demonstrate these different parts. The case study is used to motivate these different types of technology.

To simulate the social media data, you will use a Postgres database, which will include tables of people (at least 5), their connections to other people, and some coordinates related to them - which supposedly originated from their social media profile.

You can use the same Postgres database for the next spatial part but you will need to extend the PostgreSQL relational database to a PostGIS spatial database. You can write SQL commands in pgAdmin4 to create spatial data in the extended database, and write spatial queries using SQL. The PostGIS database will need to have some tracklogs inserted (at least 5), each comprised of at least 10 points. You will need to be able to query this database as follows: given a specific point (set of x,y coordinates), you are to find all tracklogs (as polygons) within which this point is

contained. You are also required to retrieve the nearest tracklog if the point does not lie within a specific tracklog. For the spatial database tasks you should use the Spatial Reference ID (SRID) of 4326 for your spatial data.

The Neo4j database will contain nodes representing people and links representing the relationships between these people. You will need to be able to insert new nodes and links into the database. You will need to be able to retrieve friends of a specific named person. Insert at least 5 people.

The MongoDB document database will contain a collection of reviews, comprised of JSON documents that have appropriate key:value attributes, for instance keys such as "Trip name/ID", "Reviewer name/ID", "Rating", "Review" etc. You will need to write the code to insert at least 5 reviews. Use MongoDBCompass GUI and associated software mongosh to create the document database, create a collection in that document database, create documents within that collection, write MongoDB commands to populate the collection. You will also need to write MQL to query the database to find the best review.

Additional Information

General Comments

The submission must be presented in a professional, clear and concise manner. If you need further system information, please use your initiative and make reasonable and logical assumptions. State your assumptions in your report. Ask your lecturer or tutor for further information.

Readings

Your text, course material and references listed on the Course Description will assist you with this assignment.

For the spatial database component:
PostGIS documentation and in particular;
Introduction to PostGIS Tutorial;
PostGIS 3.1.5dev Manual – Chapter 5 PostGIS Reference;
IISS ITECH2004 PostGIS Installation Guide V1.0;
Week 7 laboratory sheet;
Week 7 lecture;

For the document database component:
Bradshaw, S., Brazil, E. & Chodorow, K. (2020). MongoDB: The Definitive Guide. O'Reilly Media Inc.
Coronel, C., & Morris, S. (2018). Database systems: Design, implementation & management (13th ed.). Cengage Learning;
https://docs.mongodb.com/manual/tutorial/
https://docs.mongodb.com/guides/
Phaltankar, A., Ahsan, J., Harrison, M. & Nedov, L. (2020). MongoDB Fundamentals. Packt Publishing
SEITPS ITECH2004 MongoDB Installation Guide V1.0;
Week 8 laboratory sheet;
Week 8 lecture and Additional Mongo Details slides

Submission

Each student should submit an electronic copy of their report via Moodle. Please refer to the Course Description for information regarding late assignments, extensions, special consideration, and plagiarism. A reminder all academic regulations can be accessed via the university's website, see:

http://federation.edu.au/staff/governance/legal/feduni-legislation/feduni-statutes-and-regulations

Students are reminded that there are supports available regarding writing, researching and general academic skills. Various sources of help are available at:

- https://federation.edu.au/current-students#Learning_and_study;
- https://studyskills.federation.edu.au/student-skills/; and
- https://federation.edu.au/library/student-resources.

Documentation requirements

1.   Prepare a report (single Word or PDF document) which contains the following:

   a). Title page

   b). Introduction and table of contents

   c). Details of all Tasks 1-6 of the Recommender System.

   d). A copy of your code – **you must also provide associated output (as screengrabs, inserted into the report)**

   e). Statement of any resources used. These includes full disclosure of assistance from all sources including tutors and other students. Full APA referencing of any resources used.

Report file name: **ITECH2004_Assignment2_Report_*yourname_studentID***

A separate copy of the SQL code statements, file name:
ITECH2004_Assignment2_SQL_*yourname_studentID*.txt

Feedback

Assessment marks will be made available in fdlMarks, Feedback to individual students will be provided via Moodle or as direct feedback during your tutorial class.

Plagiarism

Plagiarism is the presentation of the expressed thought or work of another person as though it is one's own without properly acknowledging that person. You must not allow other students to copy your work and must take care to safeguard against this happening. More information about the plagiarism policy and procedure for the university can be found at:

 http://federation.edu.au/students/learning-and-study/online-help-with/plagiarism

Any support material must be compiled from reliable sources such as the academic resources in Federation University library which might include, but not be limited to: the main library collection, library databases and the BONUS+ collection as well as any reputable online resources (you should confirm this with your tutor).

Marking Criteria

Work will be assessed according to the details provided in the following Marking Rubric.

Assessment Criteria	Marking Scale
Postgres database • Set up tables and data (5) • Retrieve coordinates for a specific person (5) • Retrieve friendship/social media links for a specific person (5)	/15
PostGIS Spatial database • Set up tables and data (15) • Retrieve a Tracklog for a given set of point coordinates – within polygon (5) • Retrieve a Tracklog for a given set of point coordinates – nearest polygon (5)	/25
Document NoSQL database • Set up database, collection and documents (10) • Insert new review into database (5) • Retrieve highest rated review (5)	/20
Graph database • Set up database and data (10) • Insert new person and links (5) • Retrieve all connected people for a specific person (5)	/20
Critical reflection • Proposed extensions to the system – Spatial (5), NoSQL (5), Graph (5)	/15
Report style and presentation • Report is well written using professional language and adheres to guidelines given for this assessment (Any assumptions must be clearly stated and appropriate)	/5
Comments	
Total Mark **[100 marks]**	**0.0**
Course Mark **[30%]**	**0.0**