SULIT

FACULTY OF COMPUTING
UTM Johor Bahru

UNIVERSITI TEKNOLOGI MALAYSIA
FINAL EXAMINATION SEMESTER II 2022/2023

SUBJECT CODE	**:**	**SECP 3843**
SUBJECT NAME	**:**	**SPECIAL TOPIC DATA ENGINEERING**
SECTION	**:**	**01**
TIME	**:**	**10.00 AM**
DATE/DAY	**:**	**25 JUNE 2023 (SUNDAY)**

INSTRUCTIONS:
1. Read the question carefully.
2. Answer all questions.
3. Any form of plagiarism is not allowed.

Submission must be done in 2 stages:
1. Interim Submission (consists of question 1, 2 and 3). Submission date: **27 JUNE 2023** at 5:00 PM.
2. Full Submission (All questions). Submission date: **29 JUNE 2023** at 5:00 PM.

Submission must be done through GitHub. Submission later than due date and time is not accepted.

(Please Write Your Lecturer Name And Section In Your Answer Booklet)

Name	
I/C No.	
Year / Course	
Section	
Lecturer Name	

This question paper consists of **Three (3)** printed pages excluding this page.

Building a Portal using Django, JSON, MySQL, and MongoDB: A Case Study

Your major role as an IT consultant overseeing the creation of a new portal is to assure the project's effective implementation. The major goal of the portal is to provide users with a platform for viewing data and executing dashboard visualizations based on JSON-provided data. The Django web framework was picked by the management team as the base for building the portal. To do so, you must define the critical stages and concerns involved in developing the portal using Django, JSON, MySQL, and MongoDB while adhering to the requirements of storing data in both MySQL and MongoDB databases. Your job is to provide a comprehensive overview of the development process, including system architecture, Django interface with both databases, JSON data handling, and dashboard.

You are granted access to a dataset for analysis. The dataset's details can be found at the following URL: https://github.com/drshahizan/dataset. It is essential that you download the specific dataset assigned to you, which will be in JSON file format. Your task is to answer the given case studies based solely on the provided dataset. Please carefully examine the dataset and provide comprehensive answers to the following case studies, utilizing the information contained within the dataset. Please answer the following questions:

QUESTION 1

a. You have been assigned the task of providing instructions to the technical staff on implementing a configuration using five servers used in this project. This task is to ensure seamless integration between the Django web framework, a JSON dataset provided in this project, and the MySQL and MongoDB databases. This integration will enable the creation of dynamic web pages for the portal. To successfully accomplish this, please provide a comprehensive explanation of the steps required to integrate Django with the JSON dataset, ensuring efficient data storage and retrieval from both MySQL and MongoDB databases.

[10 marks]

b. Design a detailed and comprehensive system architecture. Focusing on the seamless integration between the web server (Django), dataset (JSON), and databases (MySQL and MongoDB). Give detailed explanations for each component, utilizing precise terminology and suggesting acceptable terms to describe your understanding of the scenario.

[15 marks]

QUESTION 2

a. You have been given a JSON file that contains data that must be imported into a MongoDB database. The JSON file must follows the appropriate structure for MongoDB documents. Your task is to outline the step-by-step process to add the data from the JSON file into MongoDB. Provide a detailed explanation of each step involved in the process with screenshots, including any necessary commands. Ensure that your answer covers the preparation of the JSON file, starting the MongoDB server, accessing the MongoDB shell, selecting the target database, choosing the collection, and executing the import command.

[10 marks]

b. Create FIVE (5) MongoDB queries that demonstrating various Create, Read, Update, and Delete (CRUD) operations on the documents stored in the database. Ensure that you use different types of functions provided by MongoDB for each query. The breakdown for the number of queries per operation is as follows:

   i. Create – 1 query
   ii. Read – 1 query
   iii. Update – 2 queries
   iv. Delete – 1 query.

   For each query, include the MongoDB query itself and provide screenshots of the results obtained from executing the queries.

[10 marks]

QUESTION 3

a. On this project, the module for managing user registration and login will be placed in the mySQL database. This system is used by three types of people: customers, technical workers, and senior management. Describe in detail the creation of this module on the web server (Django) and the database (mySQL).

[10 marks]

b. When working with two different databases, the challenge of Data Replication and Synchronization arises between the MySQL and MongoDB databases. This challenge involves ensuring that any changes made in one database are accurately reflected in the other, thereby

maintaining data consistency across both systems. To overcome this issue, it is recommended to explore database-specific replication techniques or leverage external tools that facilitate real-time updates and seamless interaction between the databases. In your response, provide a detailed description of the steps involved in addressing this challenge. You may include relevant code snippets and screenshots that illustrate the solution implemented.

[10 marks]

QUESTION 4

Machine learning can be used to improve the functioning of the portal as well as provide advanced data analysis and visualisation tools. Please choose an appropriate machine learning approach or algorithm for your case study and describe how it will be used in your case study. You may include relevant code snippets and screenshots that illustrate the solution implemented.

[15 marks]

QUESTION 5

a. How can the performance of the portal be optimized when dealing with large volumes of JSON data from the dataset, especially during dashboard visualizations? Please provide an illustrative solution with code and screenshots.

[10 marks]

b. Create a dashboard utilizing a JSON dataset, and provide a comprehensive description of its functionalities. You may include relevant code snippets and screenshots that illustrate the solution implemented.

[10 marks]