UNIVERSITI MALAYA

EXAMINATION FOR THE MASTERS OF DATA SCIENCE

ACADEMIC SESSION 2023/2024 : SEMESTER I

WQD7009 : BIG DATA APPLICATIONS AND ANALYTICS

January 2024 Duration: 17th Jan 2024 (6.00 pm) till 23th Jan 2024 (5.59 pm)

INSTRUCTIONS TO CANDIDATES:

Candidate is required to answer all questions.

(This question paper consists of 2 questions on 3 printed pages)

SECTION A
Answer **ALL** Questions.

QUESTION 1

Case Study – Big Data Technologies-Based Application for Processing, Analyzing, and Storing Smart Cities Application Data.

The transition to smart cities has introduced numerous challenges, particularly in the collection, processing, analysis, and storage of application data. The growing volume of sensors and connected devices generates vast amounts of data within smart city applications.

Smart city applications gather data from people, devices, and properties, which is then processed and analyzed to monitor and manage various systems, including traffic and transportation, power plants, utilities, water supply networks, waste management, crime detection, information systems, schools, libraries, hospitals, and other community services.

Currently, most city applications still rely on traditional data processing systems. Additionally, many applications do not utilize machine learning algorithms to predict traffic patterns and other city developments based on collected user records.

Based on the above case study, answer the following questions.

a. Discuss any FIVE (5) data processing and analysis problems associated with traditional smart cities applications.

(5 Marks)

b. Based on the above case study, propose an on-premises and cloud-based big data technologies application for smart city data processing, detailing each component of the application in an appropriate diagram. The proposed application must include a generative AI algorithm or relevant AI tools for analyzing smart city application data.

(15 Marks)

c. How can the proposed on-premise and cloud-based smart city application ensure cost-effectiveness during the implementation and deployment processes? Include a cost analysis.

(5 Marks)

QUESTION 2

Entrepreneurship studies have seen significant success in the field of education over the past decades, being widely incorporated into many universities worldwide. This approach aims to nurture entrepreneurial competency among students, fostering the development of young entrepreneurs. However, the investigation into how the presence of entrepreneurship education influences students' entrepreneurial competency is still evolving. Questions arise about whether students in modern technical universities acquire entrepreneurial attitudes and competencies and whether they translate these competencies into entrepreneurial intentions. Entrepreneurial competencies result from multidimensional factors such as perseverance, initiative, competitiveness, self-reliance, a strong desire to achieve, self-confidence, good physical health, mental well-being, key traits, and more. Therefore, a critical examination of entrepreneurial attitudes and competencies among students is essential to identify potential factors influencing their development as successful entrepreneurs.

To harness the potential of the Entrepreneurship dataset, the recommendation is to create a comprehensive dashboard framework for streamlined data visualization using tools like PowerBI or Tableau or Others.
Refer to the Spectrum for the dataset that will aid in developing the dashboard for this inquiry.

Based on the above case study, answer the following questions for data visualization dashboard development for the given dataset.

a. Develop and explain a DataOps architecture diagram for data visualization and analysis, considering data from various sources such as Excel files, SAS files, Postgres files, real-time logs, MySQL, or other sources. Include an analysis of the Entrepreneurship use case in the data ops architecture. The diagram should seamlessly integrate these sources into a streaming data pipeline capable of handling volume, variety, and velocity. Ultimately, it should channel the processed data to visualization tools like PowerBI, Tableau, or any other data visualization software. **(10 Marks)**

b. Develop a comprehensive dashboard using Power BI, or Tableau, or a similar data visualization tool for the provided entrepreneurship dataset. The dashboard should incorporate up to five graphs illustrating key parameters. Attach evidence of the dashboard in screenshots and include the published link in the answer document. **(10 Marks)**

c. Formulate five questions related to the dataset and graphs from the developed dashboard. Provide answers to these questions using the graphs and narrative, constructing a cohesive and informative story that enhances understanding of the entrepreneurship dataset and its implications. **(5 Marks)**

END