ourdiu.com | Shared by SUPAN ROY

Daffodil International University

Faculty of Science & Information Technology

Department of Software Engineering

Final Examination, Spring 2026

Course Code: DS 331;

Course Title: Introduction To Data Science and Data Management & Analysis

Sections & Teachers: ALL

40

Time: 2:00 Hrs | Marks: 40

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1	A large online retail company collects data from many sources such as customer purchases, website visits, payment gateways, and delivery systems. In the beginning, the company used a process where data was collected from all sources, processed in a separate server to clean and organize it, and then stored in the main data warehouse for reporting. This system worked well when the data volume was low. As the company grew, the amount of daily data increased significantly. Millions of records were generated every day, and the separate processing server became slow. Because of this delay, business reports were not updated quickly, and managers could not make timely decisions. To improve performance, the company changed its system so that raw data was first stored directly into a cloud-based data warehouse, After loading the data, cleaning and transformation tasks were performed inside the warehouse using cloud computing resources. This reduced the delay and improved the system's ability to handle larger amounts of data. Later, the company wanted to provide live sales dashboards, instant fraud detection, and real-time product recommendations for customers. To achieve this, they introduced a new system where data was captured continuously from user activities, sent to a streaming platform, processed immediately, and stored in a real-time analytics system. Based on the above scenario, **interpret** how the company's data architecture evolved from **ETL to ELT and then to real-time data processing**. Compare the advantages of each stage and discuss the main challenges of maintaining scalability and high performance in real-time data systems.	Mark 10	CLO2 Level-5
2	At Hogwarts School, Harry Potter is asked by Professor Snape to build a neural network model to classify magical mushrooms as poisonous or edible using 10 numerical features. Harry creates a model with one hidden layer, sigmoid activation, and an SGD optimizer. He experiments with learning rates of 0.1, 0.01, and 0.001. During training, he observes that with 0.1 the loss quickly becomes infinite, with 0.001 the loss decreases very slowly and stays around 0.69, and with 0.01 the training loss falls to 0.32 with 91% training accuracy. However, the validation accuracy remains only 71%, indicating weak performance on unseen data. Even after tuning several parameters, Harry cannot improve the model's validation performance and asks Professor Snape to explain the behavior of the model and suggest improvements.	Mark 10	CLO2 Level-5

ourdiu.com | Shared by SUPAN ROY

	a. **Evaluate** how the different learning rates influence the model's training behavior, and explain why the learning rate of 0.01 provides better convergence compared to the other two values. b. Based on the training and validation results, **evaluate** the model's generalization performance and recommend suitable changes to improve its effectiveness on unseen data.		
3	You are a data analyst at an AI learning platform. The dataset contains the following attributes: **student_id, course_id, enrollment_date, total_watch_time, quiz_attempts, forum_posts, course_status** The platform tracks detailed student engagement metrics to help instructors improve course quality. A. The academic director wants to understand how much time students typically spend watching course content in each course. Write an SQL query to **assess** the average total watch time for each distinct course B. To evaluate course popularity, the marketing team needs to know which courses have the highest number of enrollments. Write an SQL query which can **rank**, counts the number of students enrolled in each course and orders the results in descending order. C. Instructors want to identify the students who have successfully completed their courses. Write an SQL query to **justify** the list details of students whose course status is 'Completed'.	Mark 6	CLO2 Level-5
4	You are a data analyst working for a power distribution company. The company has recorded the monthly electricity consumption (in megawatt-hours, MWh) for a residential area during the year 2025:	Mark 14	CLO3 Level-4

Month	Consumption (MWh)	Month	Consumption (MWh)
January	420	July	520
February	400	August	500
March	380	September	470
April	360	October	430
May	390	November	410
June	450	December	440

A. **Analyze** the overall trend of electricity consumption throughout the year.
B. **Detect** the seasonal patterns (high or low).
C. **Identify** the monthly percentage change in consumption.
D. **Examine** the moving average for the sales data with a window size of 4 months.