ourdiu.com | Shared by Abir Mia -231

Daffodil International University

Faculty of Science & Information Technology

Department of Computer Science & Engineering

Midterm Examination, Fall 2026

Course Code: CSE-438, Course Title: Business Intelligence

Level: 4 Term: 3 Batch: 64

Time: 01:30 Hrs **Marks: 25**

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) An e-commerce company uses Hadoop MapReduce to identify its most common customer complaints. The following complaint logs have been collected: **5** **CO1**

Log ID	Complaint Keywords
Log 1	Refund Damage Refund
Log 2	Delay Refund Damage Refund
Log 3	Refund Delay Damage

Illustrate how MapReduce processes these logs to calculate the final Complaint keyword counts and explain how this information could support the company's decision-making.

2. A retail chain is comparing sales from its North and South branches before deciding where to invest. Their reports use different formats and contain repeated records. Management has decided to preprocess both datasets before consolidating them into a single table. They also decided not to discard any potential outliers. **CO2**

Branch North

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Value (thousand Tk.)	21	22	23	NULL	23	24	45

Branch South

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	SUN
Value (Tk.)	23,500	24,000	24,800	-5,000	25,500	26,000	31,000

a) Apply appropriate data preprocessing techniques and present the consolidated table. Use branch-wise median replacement for missing and invalid values. **4**

b) Apply suitable statistical modeling to find out the branch with the most stable performance. Identify the most representative typical sales value for the selected branch based on outlier analysis. **6**

Scanned with CamScanner

ourdiu.com | Shared by Abir Mia -231

3.	*a)*	A hotel group wants to develop a data warehouse to analyze bookings across its properties. Currently, booking details are stored in one table, causing repeated guest, room, and hotel information.	5	CO2

Hotel Booking Records (October 2026)

Booking ID	Date	Guest	Membership	Hotel ID	Hotel	City	Room No.	Room Type	Nights	Bill (Tk.)
B01	Oct 01	Aisha	Gold	H01	Garden Inn	Dhaka	101	Deluxe	2	12,000
B02	Oct 01	Rafi	Silver	H01	Garden Inn	Dhaka	502	Regular	1	4,000
B03	Oct 02	Nabila	Gold	H02	Bay View	Khulna	101	Suite	2	22,000
B04	Oct 03	Aisha	Gold	H02	Bay View	Khulna	101	Suite	1	11,000
B05	Oct 03	Tania	Silver	H02	Bay View	Khulna	102	Regular	2	9,000
B06	Oct 04	Rafi	Silver	H01	Garden Inn	Dhaka	101	Deluxe	3	18,000

Apply appropriate data modeling techniques to construct a Snowflake Schema from the given dataset, clearly indicating the primary keys (PK), foreign keys (FK), and relationships between tables.

4.	*a)*	An e-commerce company stores millions of sales records containing Customer ID, Product, Region, and Revenue. Its BI dashboard frequently calculates total revenue across different regions, but the reports take a long time to load. Moreover, managers access the same revenue reports multiple times daily. Analyze different data storage methods and determine the most suitable approach for improving the dashboard's query performance. What additional technique could be used to handle repeated analytical queries more efficiently? Justify your answer.	5	CO3

Scanned with CamScanner