
King Fahd University of Petroleum and Minerals

Information and Computer Science Department

ICS 426: Data warehousing and Data Mining

DO NOT OPEN UNTIL INSTRUCTED TO DO SO!!!!

Write clearly, precisely, and briefly!!

ID:	
Name:	

Grades		
Section	Max	Scored
A	30	
B	30	
TOTAL	60	

A. Questions from Chapter 2 (Data Preprocessing) (30)

1. Briefly explain each of the following two measures. Also, explain what the benefit of understanding them is. [4]

- a. Algebraic measure

- b. Holistic measure

2. Using smoothing by bin boundary, remove the noise from the following data:

5, 9, 16, 22, 22, 25, 26, 29, 35 [3]

3. Briefly explain the two main problems that are resolved during data integration. [3]

4. How is correlation analysis used for data integration? [3]

5. Data transformation involves normalization. Using an example, explain the use of normalization for clustering. [3]

6. Explain the following data reduction techniques: **[8]**

a. cluster sample

b. stratified sample

c. clustering

d. Stepwise forward selection algorithm

7. Explain how discretization is done using the following techniques [6]

a. ChiMerge

b. Entropy-based discretization

c. 3-4-5 Rule

B.

Questions from Chapter 3 (DW & OLAP)**(30)**

1. What is a data warehouse

[2]

2. Suppose that a data warehouse consists of 3 dimensions, namely, TIME, PRODUCT, and BRANCH, and two measures namely, COUNT, and SALES. Assume the distinct values in TIME, PRODUCT, and BRANCH are 100, 10, and 5 respectively **[15]**

a. Draw a star schema diagram for the data warehouse

3. Why the star schema gives better performance than the snow flake schema? [3]

4. Using an example, explain the following OLAP operations [4]

a. Drill down

b. Dice

5. Briefly explain the architecture of a data warehouse. [3]

6. Compare ROLAP and MOLAP in terms of flexibility and performance. [3]