Roll No. _____Name _____ Section ___

National University of Computer and Emerging Sciences, Lahore Campus

Course:	**Data Warehousing & Data Mining**	**Course Code:**	**CS409**
Program:	**BS(Computer Science)**	**Semester:**	**Fall 2017**
Duration:	**60 Minutes**	**Total Marks:**	**25**
Paper Date:	**18-Sep-17**	**Weight**	**12.5%**
Section:	**CS**	**Page(s):**	**5**
Exam:	**Midterm-I**		

Instruction/Notes: Scratch sheet can be used for rough work however, all the questions and steps are to be shown on question paper. No extra/rough sheets should be submitted with question paper.
You will not get any credit if you do not show proper working, reasoning and steps as asked in question statements.

Q1. *(3 points)* Consider the following normalized data structure:

SALES(saleId, storeId, saleDate, …)

SALES_DETAIL(transactionId, saleId, itemId, itemQty, …)

Assume there are 2 million sales and 20 million sales details. Record length of both tables is same i.e. 100 bytes and each column of both tables including PK/FK column is of same size i.e. 10 bytes.
Query: SELECT * FROM sales S JOIN sales_detail D ON S.saleId = D.saleId

You are required to improve the performance of the above query using pre-join de-normalization technique. Show your de-normalized data structure and evaluate increase in additional storage cost (in %age) for the de-normalized data structure.

Normalizes Structure:
Sales: 2m * 100 = 200m
Sales_detail: 20m * 100 = 2000m
Total cost: 2200m
Denormalized Structure:
20m * 190 = 3800m
Increase in additional storage cost: (3800 – 2200)/2200*100 = **73%**

Roll No. _____Name _____ Section ___

Q2. *(8 points)* Give the appropriate answers of the following questions very briefly:

a. What is meant by conforming the dimension? Why is this important in a data warehouse?

b. Roll-up, drill-down, drill-through, drill-across and slice-and-dice are extremely useful features of OLAP systems supporting multidimensional analysis. Briefly discuss drill-across and slice-and-dice. Give an example for each.

c. What is the concept of virtual cube? When will you consider to forming virtual cube?

d. What is the concept of factless fact table? Give an example.

Roll No. _____Name _____ Section ___

Consider the following description for next questions:
Suppose there are 10,000 products sold by the store, 200 sub-categories, 10 categories, there are 100 store locations, 10 cities, 2 countries, there are 2 years sales, also assume fact table row represents exactly one sale per product per store per day.

Q3. *(2 points)* What is the potential cardinality (max rows) of above base fact table?

730,000,000.

Q4. *(4 points)* Estimate the number of rows of fact table retrieved and summarized for following types of queries:

	Product	**Store**	**Time**	**# of Rows retrieved**
Query 1:	5 product	3 store	1 month	**5*3*30= 450**
Query 2:	1 sub-category	1 store	1 month	**50*1*30= 1500**
Query 3:	1 category	1 city	1 month	**(50*20)*10*30= 300,000**
Query 4:	1 category	1 country	1 year	**(50*20)*(5*10)*365= 18,250,000**

Q5. *(3 points)* Suppose you created an aggregate fact table for the above Query2 only … Then how many rows you need to retrieve for Queries 2, 3 and 4 from this aggregate fact table?

	Product	**Store**	**Time**	**# of Rows retrieved**
Query 2:	1 sub-category	1 store	1 month	**1**
Query 3:	1 category	1 city	1 month	**20*10*1= 200**
Query 4:	1 category	1 country	1 year	**20*(5*10)*12= 12000**

Roll No. _____Name _____ Section ___

Q6. *(5 points)* Draw the appropriate star schema that includes base fact table and aggregate fact tables for Query2 & 4 in Question#4. Take appropriate assumption for dimensions and fact tables attributes. Show the primary keys, foreign keys and all the relationships between the dimensions and fact tables.

self

Roll No. _____Name _____ Section ___