National University of Computer and Emerging Sciences, Lahore Campus

Course Name:	Data Warehousing and Business Intelligence	Course Code:	DS3003
Degree Program:	BS (Data Science)	Semester:	Fall 2023
Exam Duration:	60 Minutes	Total Marks:	30
Paper Date:	Sat 30-Sep-2023	Weight	15%
Section:	All	Page(s):	5
Exam Type:	Midterm-1	Total Questions:	3

Name: _____ **Roll No:** _____

Instructions: 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. *You may use a calculator*.

Q1. *(15 points)* Briefly answer the following questions.

a. What is the main difference between dependent data marts and independent data marts.

Ans: See Ch#2.

b. Differentiate between pre-join denormalization and column replication denormalization techniques.

Ans: See lecture notes on LDM & PDM.

c. Briefly describe the following OLAP analytical operations: Drill-down, Drill-across, slice-and-dice.

Ans: See Ch#15 OLAP.

d. What are Partitioned cubes? When would you use partitioned cubes in multidimensional online analytical processing (MOLAP)?

Ans: See lecture notes on OLAP Implementations techniques.

e. How does a snowflake schema differ from a star schema? Name one advantage and one disadvantage of the star schema.

Ans: See Ch#10 Dimensional Modeling.

Roll No: _____

Consider the following three dimensions and a base fact table for the next Questions:

Semester (SemID, SemDescription, AcademicYearID, AcademicYearDescription)
Course (CourseCode, CourseDescription, OfferingSchoolID, OfferingSchoolDescription)
Student (StudentID, StudentDescription, BatchID, BatchDescription)
Registration (SemID, CourseID, StudentID, GPA, LetterGrade, RegistrationCount (always=1)

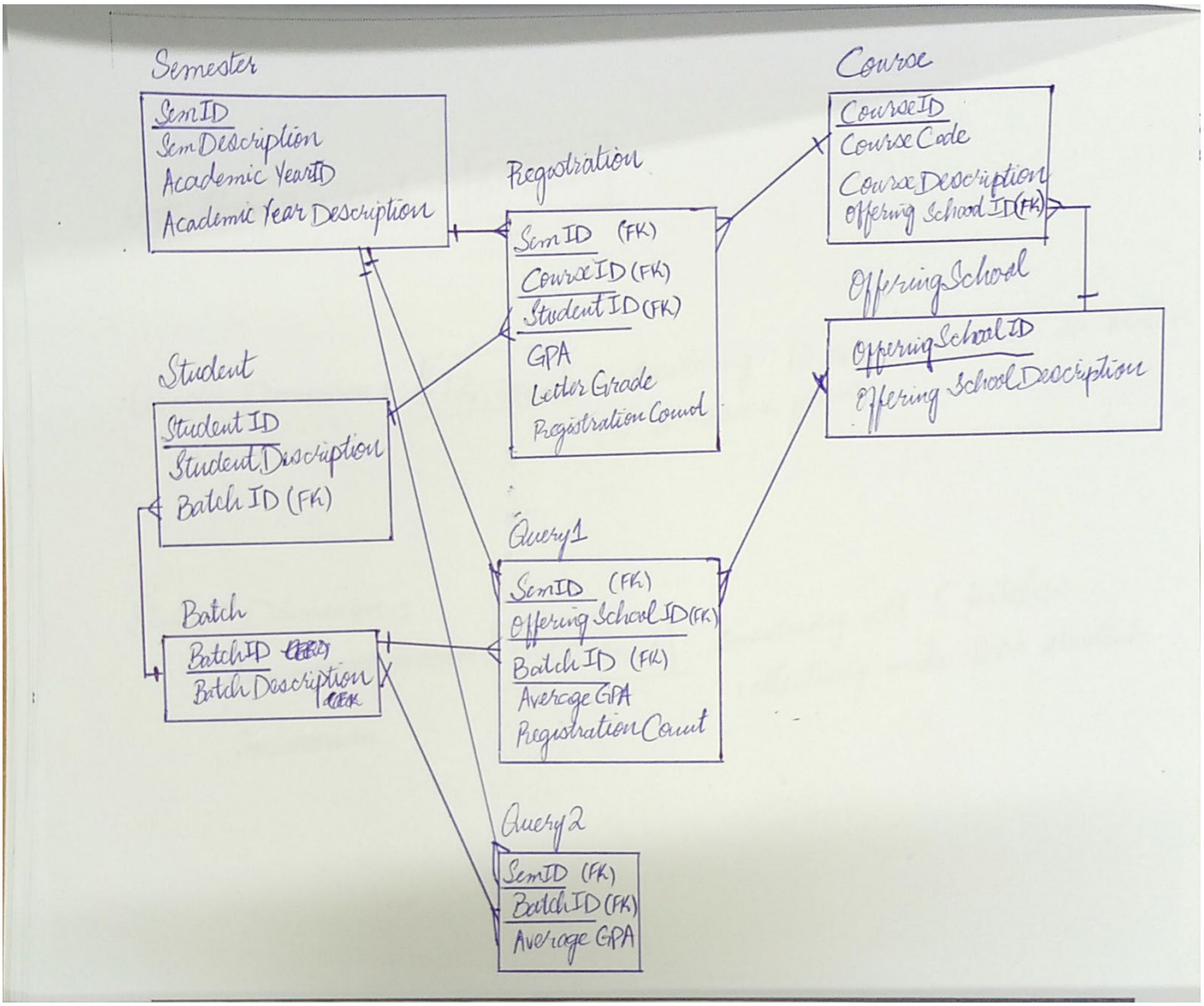
Following queries are made most frequently:

Query 1. Average GPA and total number of registered students by semester by offering school by batch.
Query 2. Average GPA by semester by batch.

Assume: 12 semesters, 6 academic years, 400 courses, 10 schools, 5000 students, and 5 batches.

Q2. *(10 points)* Draw a star schema that includes registration base fact table and aggregate fact tables for the above requirements. Take appropriate assumption, if required. Show the primary keys, foreign keys and all the relationships between the dimensions and fact tables. Note: Draw only a single diagram that includes base fact table as well as aggregate fact tables.

Ans:



Q3. *(5 points)* Estimate the size (in number of rows) of the above base fact table, both aggregate fact tables, semester dimension, and student dimension.

Ans:
Base Fact Table: 12 x 400 x 5,000 = 24,000,000 rows
Aggregate Fact Table1: 12 (semester) x 10 (offering schools) x 5 (batch) = 600 rows
Aggregate Fact Table2: 12 (semester) x 5 (batch) = 60 rows
Semester dimension: 12 rows
Student dimension: 5000 + 5 = 5005 rows