Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/M.Tech (CSE)/SEM-2/MCS-202/2010

2010

ADVANCED DATABASE MANAGEMENT SYSTEM

Time Allotted : 3 Hours | Full Marks : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer any *five* questions. 5 × 14 = 70

1. a) State the correctness rule of fragmentation. 2

b) Consider the following PROJECT Relation. This relation is fragmented into two fragments namely P1 and P2 as given below :

P No.	E.No.	P TYPE	P DESC
P 001	E 001	DEVELOPMENT	Proj 1
P 002	E 005	DEVELOPMENT	Proj 2
P 003	E 0014	DEVELOPMENT	Proj 3
P 004	E 002	MAINTENANCE	Proj 4

$P1 = \sigma_{PTYPE = \text{"DEVELOPMENT"}}(PROJECT)$

$P2 = \sigma_{PTYPE = \text{"MAINTENANCE"}}(PROJECT)$

Show the correctness of the fragmentation. 5

c) How can you differentiate between tightly coupled system and loosely coupled system ? 2

d) Write 2-phase commit protocol. When Blocking problem will take place ? 5

2. a) Describe the objectives of Data Distribution.

State the advantages of DDBMS over centralized DBMS.

3 + 2

b) Consider the following schemas :

Global schema : Student (regno, name, coursename, year-of-admission, DOB)

Fragmented Schema :

STU 1 = $\sigma_{\text{course name = "MBA"}} \left(\pi_{\text{regno, name, coursename}} (\text{student})\right)$

STU 2 = $\sigma_{\text{course name = "MBA"}} \left(\pi_{\text{regno, year-of-admission DOB}} (\text{student})\right)$

STU 3 = $\sigma_{\text{course name = "PGDBM"}} \left(\pi_{\text{regno, name, coursename}} (\text{student})\right)$

STU 4 = $\sigma_{\text{course name = "PGDBM"}} \left(\pi_{\text{regno, year-of-admission DOB}} (\text{student})\right)$

Assume that "MBA" and "PGDBM" are only possible values for the course name.

i) Write an application that requires the student_id (regno) from the terminal and outputs the name, coursename, DOB and year-of-admission at levels fragment, location and local mapping transparency.

ii) Write an application that moves the student no 102 from the coursename "MBA" to the course-name "PGDBM" at different levels of transparency.

iii) Write an application that moves a student whose regno and coursename is given at the terminal to the other courses site at location transparency.

3 + 4 + 2

3. a) Describe distributed serializability. Describe distributed 2 PL. 2 + 2

b) Simplify the following query using the idempotency results :

Select ENO from ASG where RESP = "Analyst"

AND NOT (PNU = "P2" OR DUR = 12)

AND PNO ≠ "P2"

AND DUR = 12 considering

ASG (ENO, PNO, RESP, DUR) 2

c) Optimize the following query.

List the flats that are for rent along with the corresponding branch details :

Relations : BRANCH (Branch no, Street, City, Post code)

PROP FORRENT (Prop no, Addr, Type, Rent Amount, Owner no., Branch no.)

Partitions :

P1 : $\sigma_{\text{Branch no = 'B003' ^ TYPE = "HOUSE"}}$ (PROPFORRENT)

P2 : $\sigma_{\text{Branch no} > \text{'B003' ^ TYPE = "FLAT"}}$ (PROPFORRENT)

P3 : $\sigma_{\text{Branch no} \neq \text{'B003'}}$ (PROPFORRENT)

B1 : $\sigma_{\text{Branch no = 'B003'}}$ (BRANCH)

B2 : $\sigma_{\text{Branch no} \neq \text{'B003'}}$ (BRANCH)

Write the SQL. Write the corresponding relational algebraic expression. Then optimize the Query. 8

4. a) Define Datawarehouse. Explain each term. 3

b) Compare between OLTP and OLAP systems. 3

c) Define ROLAP, MOLAP, HOLAP. 2

d) Suppose that the datawarehouse consists of the three dimensions time, doctor and patient and two measures count and charge, where fee is that a doctor charges a patients for a visit. Draw a fact constellation scheme for the above datawarehouse. 6

5. a) Discuss different OLAP operations. 5

b) Explain support & confidence. 2

c) There is a sample database shown in the following table :

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
Sunny	hot	high	false	no
Sunny	hot	high	true	no
Overcast	hot	high	false	yes
Rainy	mild	high	false	yes
Rainy	cool	normal	false	yes
Rainy	cool	normal	true	no
Overcast	cool	normal	true	yes
Sunny	mild	high	false	no
Sunny	cool	normal	false	yes
Rainy	mild	normal	false	yes
Sunny	mild	normal	true	yes
Overcast	mild	high	true	yes
Overcast	hot	normal	false	yes
Rainy	mild	high	true	no

Suppose we want to classify the data set into two classes, C_1 and C_2 with play 'yes' or 'no'. This classification will be based on the values of attributes 'outlook', 'temperature', 'humidity' and 'windy'.

i) Calculate $P(C_1 \mid x = (\text{sunny, hot, high, false}))$. How would the naive Bayes classifier classify the data instance x = (sunny, hot, light, false) ? 3

ii) Consider a new data instance x = (overcast, cool, high, true). How would the naive Bayes classifier classify x ? 4

6. a) Differentiate between classification and clustering. 2

b) Discuss the advantages and short comings of Decision Tree Classification. 5

c) What are the different methods of computing best split ? What is the gini index ? What are entropy gain and gain ratio ? 3 + 2 + 2

7. a) Discuss the importance of similarity metrics in clustering ? 2

b) Why is it difficult to handle categorical data for clustering ? 2

c) Describe the working of PAM algorithm. 5

d) Compare its performance with CLARA and CLARANS. 5

8. Write short notes on any *two* of the following : 2 × 7

i) Show flake schema

ii) Data mart

iii) 3-phase commit

iv) Distributed deadlock.