Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/M.Tech (CSE)/SEM-2/MCS-202/2011

2011

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) What are the advantages of distributed database over centralized database ? 3

   b) Explain the different types of horizontal fragmentation with example. 2

   c) Explain the rule of correctness of fragmentation. 3

   d) Give relation EMP as below :

   Let P1 : TITLE < "Programmer" and

   P2 : TITLE > "Programmer" be two simple predicates. Assume that character strings have an order among them, based on the alphabetical order.

ENO	ENAME	TITLE
E1	Raj	Elec. Engg.
E2	Anu	Syst. Analyst
E3	Anindya	Mech. Engg.
E4	Chhaya	Programmer

i) Perform a horizontal fragmentation based on { P1, P2 }. 2

ii) Dose the above fragmentation obey the rule of correctness ? 2

iii) Modify the predicates so that it obeys rule of correctness of horizontal fragmentation. 2

2. a) Draw the reference architecture of distributed database and explain different levels of transparency. 3 + 3

b) Consider relations EMP and PAY below :

EMP :

ENO	ENAME	TITLE
E1	AA	Elec. Engg.
E2	BB	Syst. Analyst
E3	CC	Mech. Engg.
E4	DD	Programmer
E 5	EE	Syst. Analyst
E 6	FF	Elec. Engg.
E 7	GG	Mech. Engg.
E 8	HH	Syst. Analyst

PAY :

TITLE	SAL
Elec. Engg.	40000
Syst. Analyst	35000
Mech. Engg.	40000
Programmer	30000

EMP and PAY are horizontally fragmented as follows :

EMP 1 = σ_{TITLE} = "Elec. Engg."$^{(EMP)}$

EMP 3 = σ_{TITLE} = "Mech.Engg."$^{(EMP)}$

EMP 2 = σ_{TITLE} = "SYST.ANAL." $^{(EMP)}$

EMP 4 = σ_{TITLE} = "PROGRAMMER" $^{(EMP)}$

PAY 1 = $\sigma_{SAL} \geq 30000$ (PAY);

PAY 2 = $\sigma_{SAL} \geq 30000$ (PAY);

Draw the join graph of EMP X_{TITLE} PAY. Is the glaph simple or partitioned ? If it is partitioned, modify the fragmentations EMP or PAY, so on that the join graph is simple. 2 + 2

c) Define hybrid fragmentation with example and draw the tree. 3

d) Define vertical fragmentation. 1

3. a) EMP relation is given in Question No. (2b). ASG relation is defined as (ENO , PNO, RESP, DUR)

A one to many relationship from EMP to ASG can be indirectly fragmented according to the following rules :

ASG 1 = ASG X_{ENO} . EMP 1

ASG 2 = ASG X_{ENO} . EMP 2

EMP 1 = σ_{TITLE} = "PROGRAMMER" (EMP)

EMP 2 = σ_{TITLE} ≠ "PROGRAMMER" (EMP)

With the above definition reduce the following query :

select *

from EMP, ASG

where ASG·ENO = EMP·ENO

and TITLE = "MECH. ENGG." 6

b) Define parametric query. How CUT operation is introduced in solving parametric query processing ? 2 + 2

c) What is semi-join ? Why is semi-join in some cases is preferred to natural join ? 2 + 2

4. a) Write down two-phase commit protocol. 6

   b) When is "blocking" occurred in case of 2-phase protocol and how is it removed ? 2 + 3

   c) What are the common types of failures that occur in case of distributed database ? 3

5. a) Write down the phases of two-phase locking and describe the functions. 3

   b) Write down the different types of locking in the cases of distributed database to control concurrency. 3

   c) How is potential deadlock detected with centralized and hierarchical controller ? 4

   d) How is log used in database recovery ? 4

6. a) Define Data warehouse. Write down the architecture of data warehouse. 2 + 3

   b) Why is separate storage of data warehouse required instead of combining with OLTP ? 3

   c) Compare OLTP and OLAP. 3

   d) State the difference between star schema and snow flake schema with example. 3

7. a) Draw the data mining architecture and explain different components. 4

   b) Define Association rule. Define support and confidence. 1 + 2

   c) What is frequent item set ? How is association rule extracted from frequent item set ? 1 + 2

   d) Write down the different types of clustering techniques. 4

8. a) Given below is training data tuples from the All electronics customer database :

RID	age	income	student	credit-rating	class : Days-Computer
1	< = 30	high	no	fair	no
2	< = 30	high	no	excellent	no
3	31 40	high	no	fair	yes
4	> 40	medium	no	fair	yes
5	> 40	low	yes	fair	yes
6	> 40	low	yes	excellent	no
7	31 40	low	yes	excellent	yes
8	< = 30	medium	no	fair	no
9	< = 30	low	yes	fair	yes
10	> 40	medium	yes	fair	yes
11	< = 30	medium	yes	excellent	yes
12	31 40	medium	no	excellent	yes
13	31 40	high	yes	fair	yes
14	> 40	medium	no	excellent	no

Apply Naive Bayesian classifier to classify unknown sample X = (age = "< = 30", income = "medium", student = "yes", credit-rating = "fair"). 8

b) What are entropy and gini index ? How are they used in selecting splitting attribute ? 3 + 3

9. Write shot notes on any *two* of the following : 2 × 7

a) ROLAP and MOLAP

b) OLAP operations

c) Partitioned clustering methods.