CS / M.Tech (CSE) / SEM-2 / MCS-202 / 09

ADVANCED DATABASE MANAGEMENT SYSTEM (SEMESTER - 2)

1. ....
Signature of Invigilator

2. ....
Signature of the Officer-in-Charge

Reg. No.

Roll No. of the Candidate

CS / M.Tech (CSE) / SEM-2 / MCS-202 / 09

ENGINEERING & MANAGEMENT EXAMINATIONS, JULY – 2009

ADVANCED DATABASE MANAGEMENT SYSTEM (SEMESTER - 2)

Time : 3 Hours] [Full Marks : 70

INSTRUCTIONS TO THE CANDIDATES :

1. This Booklet is a Question-cum-Answer Booklet. The Booklet consists of **32 pages**. The questions of this concerned subject commence from Page No. 3.
2. You have to answer the questions in the space provided marked 'Answer Sheet'. Write on both sides of the paper.
3. **Fill in your Roll No. in the box** provided as in your Admit Card before answering the questions.
4. Read the instructions given inside carefully before answering.
5. You should not forget to write the corresponding question numbers while answering.
6. Do not write your name or put any special mark in the booklet that may disclose your identity, which will render you liable to disqualification. Any candidate found copying will be subject to Disciplinary Action under the relevant rules.
7. **Use of Mobile Phone and Programmable Calculator is totally prohibited in the examination hall.**
8. You should return the booklet to the invigilator at the end of the examination and should not take any page of this booklet with you outside the examination hall, **which will lead to disqualification**.
9. Rough work, if necessary is to be done in this booklet only and cross it through.

No additional sheets are to be used and no loose paper will be provided

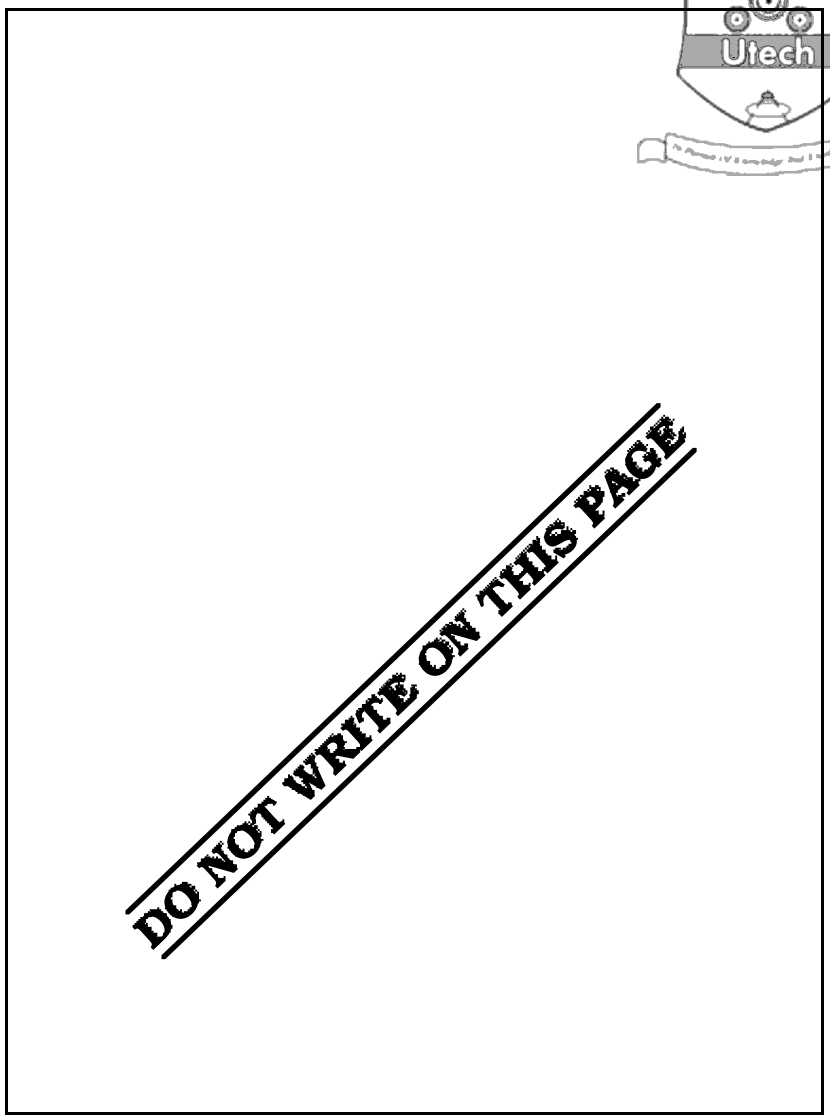
FOR OFFICE USE / EVALUATION ONLY

Marks Obtained

Question Number											***Total Marks***	***Examiner's Signature***
Marks Obtained												

....
Head-Examiner / Co-Ordinator / Scrutineer

38007 (02/07)


DO NOT WRITE ON THIS PAGE

CS/M.Tech (CSE)/SEM-2/MCS-202/09
ADVANCED DATABASE MANAGEMENT SYSTEM
SEMESTER - 2

Time : 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* of the following. 5 ∞ 14 = 70

1. a) Discuss the reasons why distributed databases are developed. 4

   b) Consider the following relations :

   EMP(ENO, ENAME, TITLE) PAY (TITLE, SAL)

   Let p1 : SAL < 3000 and p2 : SAL > = 3000 be two simple predicates.

   i) Perform a horizontal fragmentation of PAY with respect to these predicates to obtain PAY_1 and PAY_2 .

   ii) Using the fragmentation of PAY, perform further derived horizontal fragmentation for EMP.

   iii) Show completeness, reconstruction and disjointness of the fragmentation of EMP. 2 + 2 + 2

   c) Draw and explain the reference architecture for distributed databases. 4

2. a) Consider the following relations :

   EMP = { ENO, ENAME, TITLE }

   ASG = { ENO, PNO, RESP, DUR }

   PROJ = { PNO, PNAME, BUDGET, LOC }

PROJ relations has following data :

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Kolkata
P2	Database Development	135000	Mumbai
P3	CAD/CAM	250000	Mumbai
P4	Maintenance	310000	Delhi

Assume the relation PROJ is horizontally fragmented in

$PROJ_1 = SEL_{PNO<="P2"} (PROJ)$

$PROJ_2 = SEL_{PNO>="P2"} (PROJ)$

The relation ASG is indirectly fragmented as

$ASG_1 = ASG \; SJ_{PNO} \; PROJ_1$

$ASG_2 = ASG \; SJ_{PNO} \; PROJ_2$

And relation EMP is vertically fragmented as

$EMP_1 = PJ_{ENO, ENAME} (EMP)$

$EMP_2 = PJ_{ENO, TITLE} (EMP)$

i) Transform the following query into a reduced query on fragments :

SELECT ENAME

From PROJ.PNO = ASG.PNO

AND PNAME = "Instrumentation"

AND EMP.ENO = ASG.ENO

ii) Draw the intitial Query tree.

iii) Show how the query tree is optimized by heuristic rules. 2 + 2 + 4

b) Discuss briefly the distributed deadlock prevention mechanism. 4

c) Why is data replication necessary for Distributed DBMS ? 2

3. a) What is a timestamp ? How does systems generate timestamp ? Discuss the timestamp ordering protocol for concurrency control. 1 + 2 + 5

b) Discuss different methods of deadlock detection using centralized or hierarchical controller. 3

c) Discuss 3-phase commit method. 3

4. a) What are the different steps for finding out association rules ? 3

b) A database has four transactions. Let min-supp = 60% and min-conf = 80% :

TID	*Date*	*Items bought*
T100	10/5/99	{ *K*, *A*, *D*, *B* }
T200	10/5/99	{ *D*, *A*, *C*, *E*, *B* }
T300	10/6/99	{ *C*, *A*, *B*, *E* }
T400	10/7/99	{ *B*, *A*, *D* }

Find all frequent item sets using FP-growth algorithm. 6

c) Write K-means algorithm for clustering techniques. 5

5. a) Write down the apriori algorithm. 5

b) Discuss the different types of clustering techniques. 4

c) Describe the three phases of construction of a decision tree. 5

6. a) What is splitting attribute ? Discuss the methods for selecting splitting attribute for numerical and categorical attributes. 1 + 3 + 3

b) Briefly discuss the major ideas of naïve Bayesian classification. 5

c) What is prediction accuracy ? 2

7. a) Draw and explain datawarehouse architecture. 6

b) Briefly compare snowflake schema, star schema and fact constellation. 5

c) Distinguish between ROLAP and MOLAP. 3

8. a) Suppose that a datawarehouse consists of the four dimensions date, spectator, location, game and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given data. Spectators may be student, adults or seniors, with each category having its own charge rate.

Draw a star schema for the datawarehouse. 5

b) Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM-place. 3

c) What are the methods for filling in the missing values for the attributes ? 3

d) What are the methods to remove noise from data ? 3

9. Write short notes on any *two* of the following : 2 ∞ 7

a) Web mining

b) OLTP *vs*. OLAP

c) 2-phase commit.

END