Birla Institute of Technology & Science-Pilani
Hyderabad Campus
2nd Semester 2016-17
Database Systems (CS F212)
Comprehensive Examination (Regular)

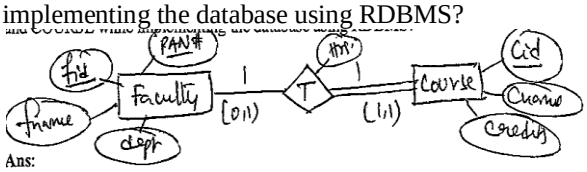
Dt: 09.05.2017 AN Weightage: 35% Time: 3 Hrs. Type: Closed Book

==========

Note: (i) Answer all parts of a question together in the same order, else will not be evaluated.
(ii) Start answering a question from fresh page. Allocate time according to the weightages.
(iii) Perfection in presentation of solution/concept, and neatness carry some weightage.

==========

Q.1 (a)What is the purpose of Conceptual level schema in three-schema architecture of DBMS? What data model is used here (high-level/physical/representational)? (2M)
(b) What is *total specialization* constraint (total participation) in EER. Give an example. (2M)
(c) Map the following ER to relational schemas. How can we enforce the min-max conditions between FACULTY and COURSE while implementing the database using RDBMS? (3M)



Ans:

Q.2 Look at the following Database schema.

Faculty(*fid, fname, age, dept****)***
Course(*cid, cname, credits****)*** *//info about courses*
FacultyCourse (*fno, cno, hrs****)*** *//to capture faculty course details; fno is FK to fid of Faculty table and cno is FK to cid of course table.*

Relationship between Faculty and Course, is a many-to-many relationship.

(i) Write a relational algebraic expression to get the *cid*, and *cname* for those courses which are not taught by any faculty from EEE department whose *age* >50.
(ii) Write an SQL query to get *fid* and *fname* for those who are involved in teaching all 5 credit courses and not teaching any 3 credit course.
(iii) Write an SQL query to get *fid* and *fname* for those who are spending more than 30 Hrs (in total) on teaching courses having credits greater than or equal to 4.
(iv) Write a Tuple relational calculus expression to get *fid*, *fname* for those faculty whose *age* is greater than 50 and teaching at least one course. (2+2+2+2=8M)

Note: You don't have to rename the attributes in the result, use only DML statements, and don't use outer joins. Do not define new tables or views.

Q.3. (a) Brief when a functional dependency (FD) will be in 2NF? Why FDs are so important?
(b) For the following SQL query, give the ***query graph***.

SELECT S.sid, S.age, C.cname, C.ceo, P.salary,
FROM Student as S, Company as C, Placement as P
WHERE S.sid=P.sid and C.cid=P.cid and S.cgpa>8.0 and C.city='DELHI';

(c) List and brief on the Cost components for Query execution. (2+3+3=8M)

Q.4 (a) Assume that we have two relations *Emp*(*eid, ename, esal, city*) and *Dept*(*dnum, dname, loc*).
Explain how *single-loop join* algorithm works to join the above tables on condition Emp.city=Dept.loc. Assume necessary access structures as suitable. (2M)
(b) With a simple example explain what is cascading rollback, and its disadvantages. (2M)
(c) Look at the following partial schedule involving four transactions T1, T2, T3 and T4; and the data items A, B, and C.

Partial Schedule: T2_lock_X(B); T2_R(B); T4_lock_X(A); T4_R(A); T4_W(A); T2_W(B); T1_lock_X(A); T1_R(A); T3_lock_S(C); T3_R(C); T3_lock_S(B); T4_lock_X(B); T2_releaseLock(B); T4_R(B); T1_W(A); T4_W(B); T4_releaseLock(A); T1_lock_S(C); T3_R(B); T3_releaseLock(B); T1_R(C); T2_lock_X(B); T2_lock_X(A); T2_W(A);

Note: T2_lock_S(A) – means T2 locks data item A in Share mode.
T1_lock_X(B) – means T1 locks data item B in Exclusive mode.
T2_R(A)- means T2 reads data item A
T3_W(A) - means T3 writes data item A

T2_releaseLock(B)- means T2 release lock it holds on data item B

(i) Will this result in deadlock?
(ii) Give wait-for graph resulting, after T2_releaseLock(B) operation is executed.
(iii) Give the final wait-for graph and give comments that supports your answer for (i). (5M)

Q.5 (a) Assume that we use ***Linear hashing*** technique in some situation and we use the hash Functions- h0, h1, h2, ... as (K mod 2), (K mod 4), (K mod 8) and so on.
Assume that a bucket (one block) can accommodate 2 records. Now insert the records with following keys in same order and show the dynamic structure of the hashing scheme after each insertion. Note that a split occurs whenever the *File Load Factor* (*f*) exceeds 0.75. Do not consider overflow buckets while calculating the *f*. But consider records inserted into Overflows.
Keys to be inserted are: 16, 33, 74, 56, 47, 29, 3, 12.
(Note: use the conventions taught in the class; complete working is to be given including all steps). (4M)

(b) Assume that we have a relational ***R*** with schema ***R(A,B,C,D,E)*** , with the following set (***F)*** of functional Dependencies. ***F={AB→CD; D→E; A→C }*** . If ***R*** is decomposed into three relations- ***R1(A,B,D), R2(A,C,D,E), R3(A,B,E)***. Now check if this decomposition is dependency preserving or not.
[*Note*: Give complete working in steps]. (3M)

(c) Why 3NF to BCNF decomposition is not always preferable? (2M)

Q.6. (a)Assume a situation where we have 8,45,000 records to be stored in a file. The record length is 120 Bytes and the block size is 512 Bytes. The address of any disk block needs 4 Bytes. We need to build a 2-level index on a non-ordering key field *k* (3 Bytes length) of the file.
Now, design a 2-level index for the above file on the non-ordering key attribute *k*. Now answer the following.

(i) How many data blocks are needed to store data records.
(ii) How many index blocks are needed at 1st and 2nd Level.
(iii) Give the number of block accesses needed (worst case) to retrieve a record with given key (*k*) value from the file using two level indexing structure.
(iv) Give the number of block accesses needed (worst case) to retrieve a record with given key (*k*) value from the file if there is no indexing available.
Note: assume un-spanned record organization. (1+2+1+1=5M)

(b) Assume that we need to build Bit-map indices on *dno* attribute of EMP table, which has 68000 records. Each record takes 78 Bytes and attribute *dno* needs 3 Bytes. Further, we have 8 distinct values for *dno* attribute in the table. Now compute the space required for building Bit-map index on *dno*, in Kilo Bytes. (2M)
(c) If there are 128 cylinders in a disk pack having 8 double sided disks with uniform surface configuration. The total capacity of each cylinder is 128MB, calculate the capacity of: (i) each track and (ii) each surface, in KBs. (1MB=1024KB). (3M)

Q.7. (a)Assume that we have two relations R(A, B, C, D) and S(B, E, F) where B in R is FK to PK of S, and the domain of D and F are same. Further, R has 200 tuples and S has 80 tuples. Now give the minimum and maximum number of tuples that can result from following operations on R and S. (i) R * S (ii) R (equi-join) S on R.D=S.F.
(b) Look at the following schedule-S1 for the concurrent transactions T1, T2 and T3. The data items are A and B.
S1 : {r_1(A); r_1(B); r_2(B); r_3(B); w_2(B); w_1(A); r_3(A); w_3(A); r_2(A); w_2(A)}
Now, give a schedule-S2 which is serial and conflict equivalent to S1.

(c) Look at the following schedule S1 involving two transactions T1 and T2, and the data items A and B.

S1: {T1_R(A); T2_R(B); T1_R(B); T1_W(B);T1_W(A); T2_W(B) }
Note: T2_R(A)- means T2 reads data item A
T2_W(A) - means T2 writes data item A

(i) is this conflict serializable?
(ii) Now give another schedule S2 that introduces lock/unlock operation in S1 observing basic two-phase locking scheme and makes S1 conflict serializable. (2+3+4=9M)

Q.8. Brief on the following.

(a) What is the principle behind *immediate modification* approach to DB recovery? Also mention about usage of logs and required log entries in this scheme (brief how it is done).
(b) What is a *recoverable schedule*? Are the schedules in Time-stamp based protocol (used in concurrency control) always recoverable? If yes, how? If no, under what circumstances they are not recoverable? Explain with simple example.
(c) Give concurrent Schedules- S1 and S2 such that S2 is view equivalent to S1 , and S1 is not conflict serializable.(you can assume transactions, read/write operations and data items as needed).
(d) What is *conservative 2-phase locking* scheme?
(e) If the Write Time-stamp of a data item *A* is 167 and Read Time-stamp is 135; if a transaction ***Tx*** with timestamp 146 arrives and wants to write(update) the data item *A*, then what happens according to *Thomos's Write-rule* ? Brief.

******

(2+2+2+2+2=10M)