Mode of Examination: Online

M.Sc. (Computer Science) Semester-IExamination, 2020

2020

Subject: Computer Science

Paper: CSM102(Advances in Database Management System)

Full Marks: 70

Time: 3:00 Hours

Duration: 12:00 Noon – 3:00 PM

The figures in the margin indicate full marks

Candidates are required to answer in their own words as far as applicable

Each Page of answer scripts should have your examination Roll Number.

The scanned copy of the answer script will be of the following format:

CSM102-Examination-Roll-Number.pdf

(Example: CSM102-C91-CSC-201001.pdf)

The subject of the mail should be answer-script file name

The name of the scanned answer script is to be sent to cucse2020@gmail.com

No answer script will be valid if received after 3:30 PM on Examination Date

Write the answers with black ink ball pen

Answer Question number 1, 2, and any Four from the rest.

---

1. **Answer any five from the questions given below** (5 x 2 =10)

a) Let R = (a,b,c,d,e,f) be a relation with the functional dependencies c→f, e→a, ec→d, a→b. Determine the key.

b) Consider two sets of Functional Dependencies: F = {A→BC, B→A, C→A} and G = {A→B, B→C, C→A}. Are F and G equivalent?

c) Data file contains 1 million records. B+ tree index is used with order equals to 100. What will be the maximum number of nodes to be accessed to search a record?

d) Consider the relation R (A,B,C) which has the Functional Dependency B→C. If A is a candidate key for R, is it possible for R to be in BCNF? Justify your answer.

e) How does the recovery manager ensure atomicity of transaction?

f) Why is cascading in a schedule not desirable?

g) State the reason behind selecting "Number of Secondary memory accesses" as a metric for query optimization.

h) State the importance of serializable schedule in context of concurrency control protocol.

2. **Answer any five from the followings:** (5 x 4 =20)

a) Consider the set of functional dependencies F = {A →BC, CD → E, E→C, D → AEH, ABH→ BD, DH→ BC}. Find the Canonical Cover of F.

b) Consider the relation *R*, which has attributes that hold schedules of courses and sections at a university; *R* = {Course_no, Sec_no, Offering_dept, Credit_hours, Course_level, Instructor_ssn, Semester, Year, Days_hours, Room_no, No_of_students}.
Suppose that the following functional dependencies hold on *R*.
{Course_no} → {Offering_dept, Credit_hours, Course_level}; {Course_no, Sec_no, Semester, Year} → {Days_hours, Room_no, No_of_students, Instructor_ssn}, {Room_no, Days_hours, Semester, Year} → {Instructor_ssn, Course_no, Sec_no}
Determine the sets of attributes that form keys of *R*. How would you normalize this relation?

c) Illustrate, through an example, the role of system log to recover from failure of a transaction.

d) When is it preferable to use a dense index rather than a sparse index? Since indices speed query processing, why might they not be kept on several search keys?

e) Consider the relation X(P, Q, R, S, T, U) with the following sets of functional dependencies.
F = {{P, R} → {S, T}, {P, S, U} → {Q, R}}. Calculate F+.

f) Consider the table as shown in the figurebelow on a sample dentist/patient appointment data. A patient is given an appointment at a specific time and date with a dentist located at a particular surgery. On each day of patient appointments, a dentist is allocated to a specific surgery for that day. The table shown in the figure is susceptible to update anomalies. Provide examples of insertion, deletion and update anomalies.

staffNo	dentistName	patNo	patName	appointment date	time	surgeryNo
S1011	Tony Smith	P100	Gillian White	12-Sep-13	10.00	S15
S1011	Tony Smith	P105	Jill Bell	12-Sep-13	12.00	S15
S1024	Helen Pearson	P108	Ian MacKay	12-Sep-13	10.00	S10
S1024	Helen Pearson	P108	Ian MacKay	14-Sep-13	14.00	S10
S1032	Robin Plevin	P105	Jill Bell	14-Sep-13	16.30	S15
S1032	Robin Plevin	P110	John Walker	15-Sep-13	18.00	S13

g) Justify the statement: Concurrent execution of transactions is more important when data must be fetched from (slow) disk or when transactions are long, and is less important when data is in memory and transactions are very short.

3. a. Consider a relation R(*a,b,c,d,e*) containing 5,000,000 records, where each datapage of the relation holds 10 records. R is organized as a sorted file with dense secondary indexes. Assume that *R:a* is a candidate key for R, with values lying in the range 0 to 4,999,999, and that R is stored in *R:a* order. For each of the following relational algebra queries, state with justification which of the following two approaches is most likely to be the cheapest.

- Access the sorted file for R directly.
- Use a (clustered) B+ tree index on attribute *R:a*.

Queries: 1. Select a<50000(R); 2. Select a=50000(R)

b. Suppose that you have a file with 20,000 pages and you have five buffer pages.
Answer the following questions for each of these scenarios, assuming that our most general external sorting algorithm is used.

1. How many runs will you produce in the first pass?
2. How many passes will it take to sort the file completely?
3. What is the total I/O cost of sorting the file?
4. How many buffer pages do you need to sort the file completely in just two passes?

[(3+3)+4 =10]

4. a. Consider the following relational schemas:
EMPLOYEE (Ename, SSN, Bdate, Addr, Dnum); DEPARTMENT (Dname, Dnum, Dmgr_SSN); PROJECT (Pname, Pnumber, Ploc, Dnum); WORKS_ON (SSN, Pnumber, Hrs)
Write down the following query in relational algebra
"Find the last names of employees born after 1984 and before 1990 who work on a project named 'Installation of Hydro Power Plant' ".
Draw the initial query tree. Apply the Heuristic rules for query optimization to obtain a optimized query execution plan. Show the query trees for each step.

b. Discuss the scheme for preventing deadlock in concurrent transactions execution.
[(2+5)+3 = 10]

5. a. Consider the following schedule:

$$S : r_1(X); r_2(X); w_1(X); r_2(Y); r_1(Y); w_2(X); r_1(Z); w_2(Y); w_1(Z); r_3(z); r_2(Z); w_3(Z); w_2(Z);$$

Draw the precedence graph for the schedule and justify whether the schedule is conflict serializable or not. Here suffix of each operation indicates the transaction.

b. Compare the performance of nested loop, block nested loop and indexed nested loop join in worst cases.[5+5 = 10]

6.

a. A PARTS file with Part# as the hash key includes records with the following Part# values: 2369, 3760, 4692, 4871, 5659, 1821, 1074, 7115, 1620, 2428, 3943, 4750, 6975, 4981, and 9208. The file uses eight buckets, numbered 0 to 7. Each bucket is one disk block and holds two records. Load the records into expandable hash files based on extendible hashing. Show the structure of the directory at each step, and the global and local depths. Use the hash function $h(K) = K \bmod 128$.

b. Explain the utility of Thomas Write rule.

[7+3 = 10]

7.a. Suppose that a disk unit has the following parameters: seek time s = 20 msec; rotational delay rd = 10 msec; block transfer time btt = 1 msec; block size B = 2400 bytes; interblock gap size G = 600 bytes. An EMPLOYEE file has the following fields: Ssn, 9 bytes; Last_name, 20 bytes; First_name, 20 bytes; Middle_init, 1 byte; Birth_date, 10 bytes; Address, 35 bytes; Phone, 12 bytes; Supervisor_ssn, 9 bytes; Department, 4 bytes; Job_code, 4 bytes; deletion marker 1 byte. The EMPLOYEE file has r = 30,000 records, fixed-length format, and unspanned blocking. Write appropriate formulas and calculate the following values for the above EMPLOYEE file:

- Calculate the record size R (including the deletion marker), the blocking factor bfr, and the number of disk blocks b.
- Calculate the wasted space in each disk block because of the unspanned organization.
- Calculate the average number of block accesses needed to search for an arbitrary record in the file, using linear search.

b. Discuss, through an example, the contribution of lock upgradation scheme.

[8+2 =10]

8. a. In timestamp ordering,W-timestamp(Q) denotes the largest timestamp of anytransaction that executed write(Q) successfully. Suppose that, instead, we definedit to be the timestamp of the most recent transaction to execute write(Q)successfully.Would this change in wording make any difference? Explain youranswer.

b. In multiple-granularity locking, what is the difference between implicit andexplicit locking? Explain through an example.

c. If deadlock is avoided by deadlock avoidance schemes, is starvation still possible?Explain your answer.

[3+4+3 = 10]

9. a. Describe, with examples, the types of problem that can occur in a multi-user environment when concurrent accessto the database is allowed.
b. For each of the following schedules, state whether the schedule is serializable (conflict or view) and recoverable or not.
(i) read(T1, x), read(T2, X), write(T1, x), write(T2, x), commit(T1), commit(T2)
(ii) read(T1, x), read(T2, y), write(T3, x), read(T2, x), read(T1, y), commit(T1), commit(T2), commit(T3)
(c) write(T1, x), read(T2, x), write(T1, x), commit(T2), abort(T1)
(e) read(T1, x), write(T2, x), write(T1, x), read(T3, x), commit(T1), commit(T2), commit(T3)

(c) Produce a wait-for graph for the following transaction scenario and determine whether deadlock exists:

TRANSACTION	DATA ITEMS LOCKED BY TRANSACTION	DATA ITEMS TRANSACTION IS WAITING FOR
T1	x2	x1, x8
T2	x3, x10	x7
T3	x8,	x4
T4	x7	x1
T5	x1, x5	x3
T6	x4, x9	

[3+4+3 = 10]