UNIVERSITY OF RUHUNA

BACHELOR OF COMPUTER SCIENCE (GENERAL) DEGREE LEVEL II
(SEMESTER II) - EXAMINATION
NOVEMBER/DECEMBER 2016

SUBJECT : COMPUTER SCIENCE
COURSE UNIT: CSC 2242 (Advanced Database Systems) TIME: 2 hours

Answer all four (04) questions.

1.

a.

i. Write an advantage and a disadvantage for each of the heap, sequential and hash file organization methods.

ii. Write a situation where bucket overflow can occur in Hash file organization.

iii. Explain a method to handle the situation you have written above in a.(ii).

b. Consider a file named ***Orders*** with ***sales*** as the hash key. The hash key includes records with the following ***sales*** values:

2369, 3760, 4692, 4871, 5659, 1821, 1074, 7115, 1620, 3943, 4750, 6975, 4981, 9208,5491,7677,8855

The file uses 10 buckets, numbered 0 to 9. Each bucket is one disk block and holds two records. The hash function is **h(K)= K mod n**

i. What is the suitable value for "***n***" in the above function?

ii. Insert these records into the file in the given order using the given hash function.

iii. List the overflow bucket numbers.

iv. Calculate the average number of block accesses required for a random retrieval of ***sales*** records from the ***Orders*** file.

c. Explain why indexing is needed in database management systems.

d. Consider a block with block size B=512 bytes. A block pointer is P=6 bytes long, and a record pointer is PR=7 bytes long. A file has r=25,000 STUDENT records of fixed length. Each record has the following fields.

SID (7 bytes), NAME (30 bytes), ADDRESS (40 bytes), COURSE (10 bytes), BIRTHDATE (8 bytes), SEX (1 byte), CONTACT (10 bytes), ACADEMIC_YEAR (4 bytes). An additional byte is used as a deletion marker.

i. Calculate the record size **R** in bytes.

ii. Calculate the blocking factor ***bfr*** and the number of file blocks ***b*** by assuming that the above has an unspanned file organization.

iii. Suppose the file is ordered by the key field SID and it is required to construct a primary index on SID. Calculate the index blocking factor ***bfri*** and the number of first level index entries.

Continued...

2.

a.

i. Write two (02) advantages and two (02) disadvantages of B+ trees.

ii. Consider the following values.

2, 5, 7, 10, 13, 16, 20, 22, 23, 24

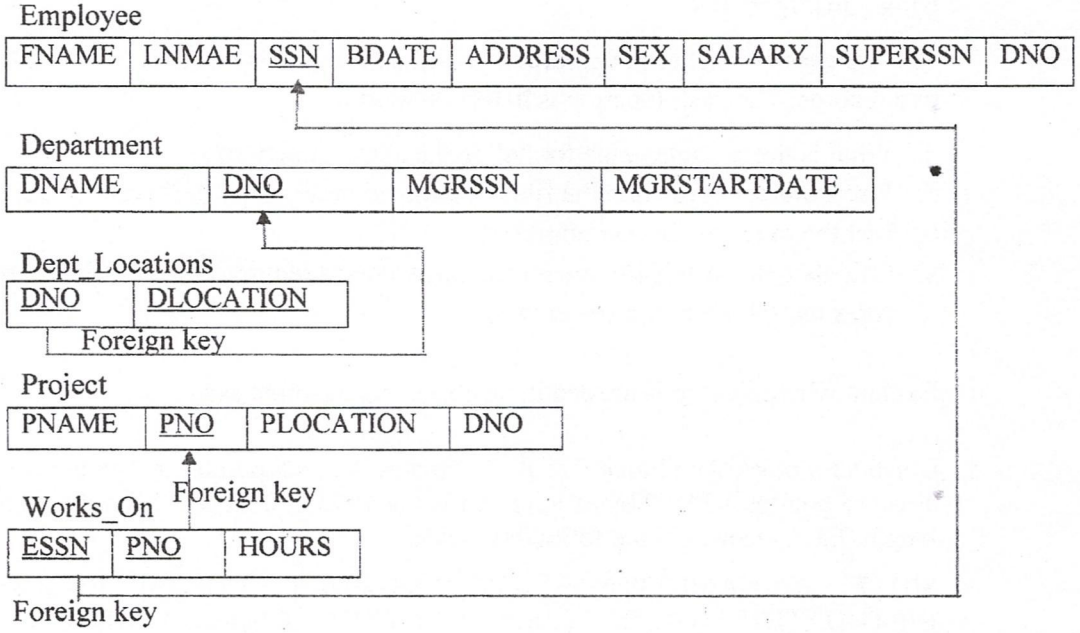
Insert the above values into a B+ tree of order p = 4. Show the intermediate steps clearly.

b. Write the difference between materialization and pipelining in query evaluation process.

c. Briefly explain the concept of Heuristic optimization.

d. Consider the following SQL query related to the relations given below.

SELECT ***FNAME,LNAME***
FROM ***Employee e, Project p, Works_On w***
WHERE ***e.SALARY>25000*** AND ***p.PLOCATION='COLOMBO'*** AND ***p.PNO=w.PNO*** and ***w.ESSN=e.SSN***



Optimize the above SQL query using ***Heuristics optimization***. Clearly show the intermediate query trees.

Continued...

3.

a. Atomicity is one of the ACID properties which is used to explain the important characteristics of a transaction.

i. What is meant by atomicity?
ii. Explain the atomicity using an example.

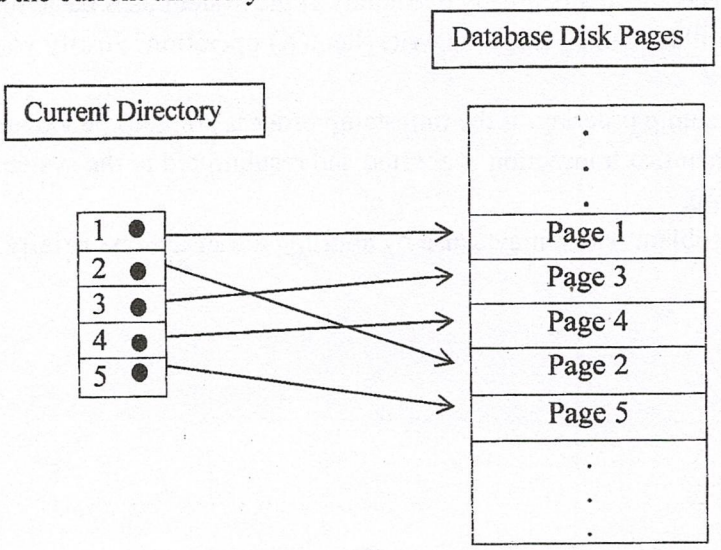
b. Show the typical states and operations of the transaction execution process using a state transition diagram.

c. Consider the schedule S given below to answer the questions in c (ii) & c (iii).

$S : r_1(Y); r_2(Y); w_1(Y); r_1(X); w_2(Y); w_1(X);$

i. Two transaction operations are called schedule conflict if they satisfy three important conditions. What are they?
ii. Name three pair of conflict operations of schedule S given above.
iii. Name one pair of non-conflict operation of the schedule S given above giving the reason for it to be non-conflict.

d. Shadow paging is one of the useful techniques used for the database recovery management process. It uses two directories called current directory and shadow directory. The diagram given below shows the current state of the database disk pages and the current directory.



i. Draw a diagram to display the possible states of the current directory, shadow directory and the database disk pages after updating the disk pages 1,2 and 5.
ii. Assume that the system has failed after updating the disk pages mentioned in d(i) above before committing the transaction. Suggest a possible method that the system can use to recover from that failure.

Continued…

4.

a.

i. Name the three phases of validation concurrency control technique.

ii. Write a major difference of validation concurrency control technique, with compared to other concurrency control techniques?

b. Isolation level is an attribute of a SQL transaction. Different isolation levels may cause different types of violations.

Name two types of such violations. Give suitable examples for each of them.

c.

i. What is a timestamp? State two ways of generating timestamp values.

ii. Given below is a part of a schedule which has already executed,

T_1 : read(x), write(x)

T_2 : read(y), write(y)

T_3 : read(x), write(x)

The timestamps of the transactions are,

$\mathbf{TS(T_1) < TS(T_2) < TS(T_3)}$

Suppose T_2 needs to issue the write_item(X) operation (assume that all the operations given above are already executed). If the system uses basic timestamp ordering algorithm, can T_2 issue the write_item(X) operation? Justify your answer.

iii. In basic time stamp ordering, if the timestamp order is going to be violated by a transaction, then that transaction is aborted and resubmitted to the system with a new time stamp.

What is the problem which may cause by aborting a transaction? Briefly explain it.

Continued...

d. Prove that the schedule given below is non-serializable by considering the initial values as x=8 and y=12 (Hint: use the different answers which x and y can have for the two situations which T1 followed by T2 and T2 followed by T1).
Modify the given schedule to ensure serializability by using the two-phase locking mechanism

T_1	T_2
read_lock(y)	
read_item(y)	
unlock(y)	
	read_lock(x)
	read_item(x)
	unlock(x)
write_lock(x)	
read_item(x)	
x=x-y	
write_item(x)	
unlock(x)	
	write_lock(y)
	read_item(y)
	y=y-x
	write_item(y)
	unlock(y)

----------*End*----------