Sri Lanka Institute of Information Technology

B.Sc. Special Honours Degree/Diploma
in
Information Technology

Final Examination
Year 2, Semester II (2016)
June Intake

IT202 - Database Management Systems II

Duration: 3 Hours

October 2016

Instructions to Candidates:

- This paper has 5 questions. Answer all questions.
- Write answers in the booklet given.
- Total marks 100.
- This paper contains 7 pages including the cover page.
- Electronic devices capable of storing and retrieving text, including calculators and mobile phones are not allowed.

Question 1 (20 marks)

Consider the following schema of a database developed for a university.

Lecturer (*lecturerId*: integer, *lname*: varchar (30), *grade*: varchar (20), *salary*: real, type: *varchar (10), department*: varchar (30))
Qualifications (*lecturerId*: integer, *qualification*: varchar (30))
Course (*cid*: integer, *cname;* varchar (25), *credits:* integer, *program:* varchar (30))
Teach (*lecturerId*: integer, *cid*: integer, *year*: integer, *semester:* integer)

Lecturer table stores the id (*lecturerId*), name (*lname*), *grade* ('Assistant Lecturer', 'Lecturer' or 'Senior Lecturer'), *salary*, *type* of the lecturer ('Permanent' or 'Visiting') and the attached *department* (such as 'IT' and 'IS') for each lecturer in the university. **Qualifications** table stores information related to the qualifications of lecturers. The table stores the IDs of the lecturers (*lecturerId*) and the *qualifications* they have received (such as 'BSc in IT', 'MSc in GIS', 'MBA' and 'Phd in Computer Science'). **Course** table stores the ID of the course (*cid*), name (*cname*), number of credits (*credits*) and the *program* the course belongs to (such as 'BSc in IT', and 'BEng in Mechanical Engineering') for each course offered. The **Teach** table stores information on teaching. For each lecturer teaching a course, the Teach table stores the ID of the lecturer (*lecturerId*), ID of the course he/she has taught (*cid*), the *year* and the *semester* he/she has taught the course. *LecturerId* attribute in Qualifications and Teach tables are foreign keys that references the Lecturer table. *Cid* attribute in Teach table is a foreign key that references the Course table. All primary keys of the tables are underlined.

a) Use SQL queries to answer following questions.
   i. Find the names of the courses in the 'MBA' program which are taught by the lecturers attached to the 'IT' department. (3 marks)

   ii. Find the names of the courses that are not offered in semester 2 of 2016. (3 marks)

   iii. Find the names and the IDs of the lecturers who teach a course for the same qualification he has obtained.
   Hint: When a student completes a program he obtains that qualification (4 marks)

b) Create a function that calculates and returns the workload for a given lecturer in a given year and a semester. Workload indicate the number of hours a teacher teach weekly.
Assume that a teacher spend 2 hours weekly in teaching every course he teach (5 marks)

c) Suppose that the university pays 5000 rupees for visiting lecturers for one hour of lectures. With the assumption that a lecturer teach 20 hours in each course he teach in one semester, write a procedure that calculate the payment to be made for all visiting lecturers for a given year and a semester and store it in the salary column. (5 marks)

Question 2 (20 marks)

a) Briefly describe two main techniques that could be used to improve the reliability of the disk. (3 marks)

b) What is meant by 'data striping'? What is the purpose of striping data? (2 marks)

c) Consider a disk with a sector size of 512 bytes, block size of 512 bytes, 1000 tracks per surface, 100 sectors per track and 5 single-sided platters. Seek time is 10 msec. The disk is spinning at a speed of 2400 rpm. You need to store a student file which contains 10000 records of 100 bytes which are in fixed-length format. No record spans multiple blocks.

i. How many student records can be stored in a block in the above disk? (1 mark)

ii. How many student records can be stored in a single track? (1 mark)

iii. How many cylinders are needed to store the entire student file? (2 marks)

iv. What is the time required to transfer a track of data from the above disk? (2 marks)

v. Assuming that the data in the above disk are stored using the Next Block concept and what would be the time taken to read the entire student file from the disk. (3 marks)

d) Explain with an aid of a diagram, the alternatives available for storing variable length fields in a record? (2 marks)

e) Consider the following buffer pool with 5 frames. Each frame has a frame number (Frame No), page id (PageID) of page, pin count (PinCount) and a dirty bit (Dirty).

Frame No: 0	Frame No: 1	Frame No: 2	Frame No: 3	Frame No: 4
PageID: 4	PageID: 17	PageID: 10	PageID: 21	PageID: 15
PinCount: 1	PinCount: 0	PinCount: 0	PinCount: 1	PinCount: 1
Dirty: 1	Dirty: 1	Dirty: 1	Dirty: 0	Dirty: 1

Assume that the following times states the last time the particular frame was accessed (on the same day):

Frame number 0 at 12:15 pm
Frame number 1 at 12:07 pm
Frame number 2 at 2:05 pm
Frame number 3 at 12:02 pm
Frame number 4 at 1:26 pm

Describe the steps followed by the buffer manager when a request for the page with PageID 50 occurs if MRU (Most-recently-used) replacement policy is used.

(4 marks)

Question 3 (20 marks)

a) Briefly explain the Sequential file organization. What are the advantages of this file organization compared to the Heap file organization? (3 marks)

b) "Range search in sequential file organization is not always efficient than in heap file organization". Accept or refute the above statement giving reasons. (3 marks)

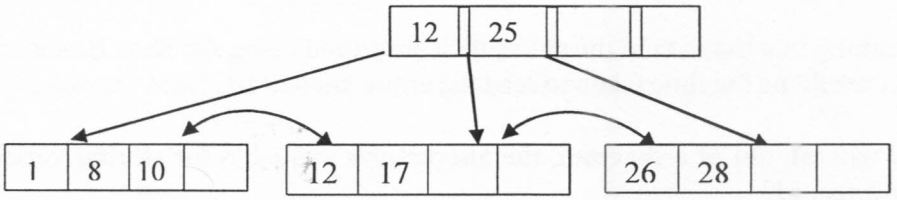
c) Suppose you have a student table with following fields

Student (studentId, name, address, gender, phone, birthdate, NIC, GPA)

Assume that you want to retrieve the IDs and the names of students with GPA higher than 3.0. The student table contains a clustered index on the GPA column of the table.

Explain the approach you will be using to retrieve the records you require. Why would you use the selected approach? (3 marks)

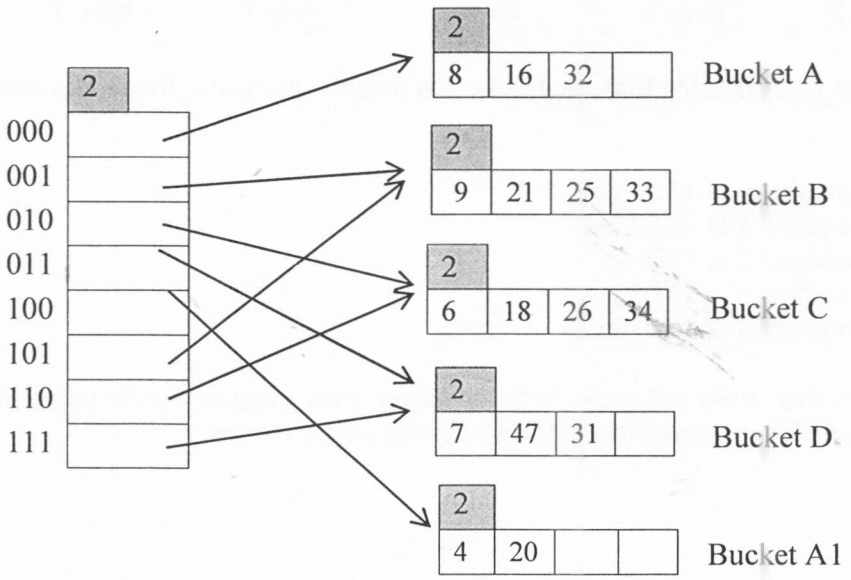
d) Consider the following B+ tree of order 2.



i. Illustrate the B+ tree after inserting 4 and 7 to the tree above. (3 marks)

ii. Illustrate the B+ tree after deleting 17 from the **original tree**. (3 marks)

e) Consider the following extendible hash index. The hashing function considers the last 2 digits of the binary representation.



Show the extendible hash index after inserting 12 and 22. (5 marks)

Question 4 (15 marks)

a) Briefly explain the query optimization phase of the query processing. (3 marks)

b) From where does the optimizer obtain the required information for query processing? Name three different types of information available for the optimizer to use. (2 marks)

c) What is an index only plan? (1 mark)

d) Consider the following relation:
Employee (eid:char(5), ename: varchar(50), job: varchar(30), salary: float)

The relation has 8000 pages with 100 tuples per page. Assume that 20 buffer fames available in buffer pool.

Consider the following query:

Query 1: SELECT * FROM Employee WHERE eid = 'E0001'

Estimate the cost (in Disk I/Os) of executing *Query 1* with the following Hash indexes. The data entries in the indexes have <k, rid> format. Hash indexes require 1.2 Disk I/Os on average to retrieve a data entry.

i. Unclustered hash index on Employee<eid> (2 marks)

ii. Clustered hash index on Employee<eid> (2 marks)

e) Consider the following query

Query 2: SELECT *ename* FROM *doctor* WHERE specialization = 'Surgeon'

You are given following information:
The *doctor* relation has four attributes namely, *name*, *designation* and *specialization* and *address*; all are string fields of the same length. The *ename* attribute is a candidate key. The relation contains 3000 pages. There are 10 buffer pages. Only 0.02 percent of tuples qualify the where clause condition. DBMS support only B+ tree indexes and require 3 disk I/Os to retrieve leaf page.

i. Suppose this query is one of the most frequently running queries. Therefore, recommend the most suitable index to get the best performance of the query. Briefly explain the query plan you have considered. (3 marks)

ii. Estimate the cost of executing the query using the plan and the index proposed above. (2 marks)

Question 5 (25 marks)

a) Explain the four properties of a transaction. (4 marks)

b) Briefly explain why the database executes *Serializable Schedules* rather than *Serial Schedules*? (3 marks)

c) Briefly explain the rules in Strict 2 Phase Locking Protocol. (3 marks)

d) Briefly explain what a Cascading Abort is using a sample schedule. What are the causes for a Cascading Abort? (4 marks)

e) Consider the transaction schedule below.

T_1	T_2	T_3	T_4
S(A)			
R(A)			
		S(C)	
		R(C)	
S(C)			
	X(B)		
	W(B)		
		X(B)	
			X(D)
			W(D)
	X(D)		
			X(A)

Assume that Transaction T_i is higher priority than transaction T_{i+1}(i.e. transaction T_1 has higher priority than T_2; and T_2 has higher priority than T_3).

i. Draw a *Wait-For-Graph* for the schedule given above. (2 marks)

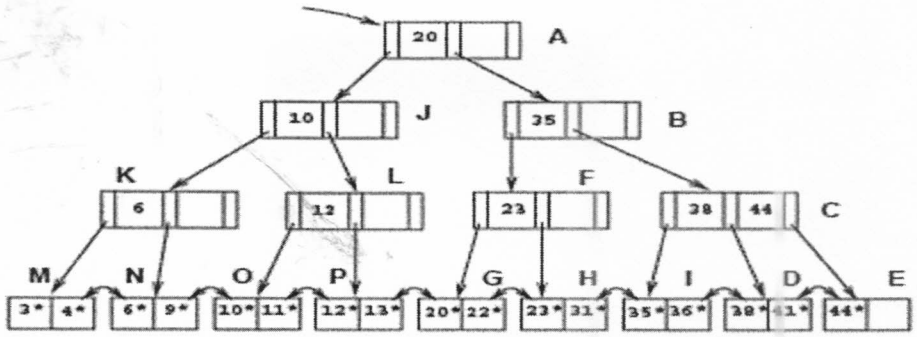
ii. Draw the above schedule considering deadlock prevention algorithm: Wait-die approach. (3 marks)

f) Consider the tree shown below. Describe the steps involved in executing each of the following operations according to the Simple Tree Locking Algorithm, in term of the order in which nodes are locked, unlocked, read and write.

i. Search 26 (2 marks)

ii. Insert 2 (2 marks)

iii. Delete 44 (2 marks)



-- End of the Question Paper --