Student No: _____

HAND IN

Queen's University

Faculty of Arts and Science

School of Computing

CISC 432* / 836*

Advanced Database Systems

Final Examination

December 14, 2002

Instructor: Pat Martin

Instructions:

1. This examination is 3 hours in length.

2. Please answer all questions on the exam paper in the space provided. Put your name and student number on every page. **HAND IN** the exam paper.

3. The exam has 3 sections. There are 11 questions in total. Marks are allocated to the sections as follows:

Disks and files	25
Query processing and optimization	35
Transaction management	40

4. There are 16 pages in the exam. Pages 6, 10, 14 and 16 of the exam are intentionally blank and may be used for your answers if required.

Please Note:

The candidate is urged to submit with the answer paper a clear statement of any assumptions made if doubt exists as to the interpretation of any question that requires a written answer.

GOOD LUCK!!!

Student No: _____

Section 1: Disks and files (25 marks)

1. [5 marks] Consider a RAID storage system
   (a) What is data striping?
   (b) What are the implications of fine-grained data striping to the performance of a RAID system?
   (c) What are the implications of coarse-grained data striping to the performance of a RAID system?

Student No: _____

2. [5 marks] Explain how the GCLOCK buffer page replacement algorithm works. Please provide a diagram.

Student No: _____

3. [5 marks] Suppose that you have a student records file with the following characteristics: the file size is 1000 blocks; the block size is 4000 bytes; the record size is 100 bytes; the data entry size for an index is 10 bytes; the **studentNumber** field is the primary key; the **subjectMajor** field has 50 possible values and the values are uniformly distributed in the file.

   (a) Suppose that there is a clustered B+ tree index on **studentNumber**. What is the average cost (in block accesses) of a search on **studentNumber**? Show your work.

   (b) Suppose that there is an unclustered hash index on **subjectMajor**. What is the average cost (in block accesses) of a search on **subjectMajor**? Show your work.

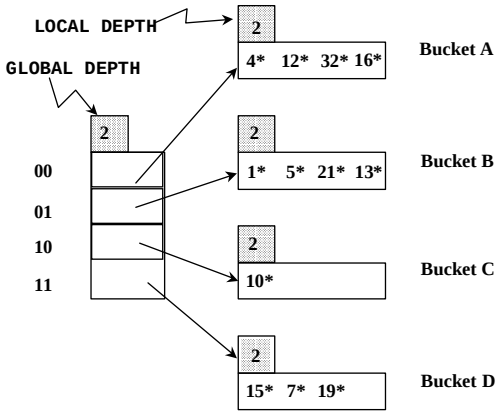
Student No: _____

4. [10 marks] Consider the extendible hash index structure shown below for the field **numberOfCredits** in the student records file of Question 3. Assume that the least significant bits in the **numberOfCredits** field identify the directory entries and each bucket holds up to 4 entries. Show what happens to the index structure and the local and global depths after each of the following operations. Assume that the effects of the operations are cumulative.

   (a) Insert 27.

   (b) Insert 9.

   (c) Insert 31.



Student No: _____

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Student No: _____

Section 2: Query processing and optimization (35 marks)

5. [10 marks] A national retail chain maintains a store file that contains a record for every store in the chain. Each record contains data about a store such as its address, its phone number, the name of its manager, and so on. The file is 1000 blocks in size.

   (a) Suppose that the file is originally unsorted and that you have 10 buffer pages available in memory. What is the cost, in block I/Os, to sort the file on store address? Show your work.

   (b) Suppose that you have a clustered B+ tree index on the store address field, that there are 50 leaf pages, that index pages have a maximum fanout of 40 and that you have 10 buffer pages available in memory. What is the cost, in block I/Os, to sort the file on store address using the index? Show your work.

Student No: _____

6. [10 marks] Suppose that you have the two relations **R** and **S** and wish to perform the join **R**$\triangleright\triangleleft_{\mathbf{R.A=S.B}}$ **S** where **R** contains 5000 tuples and has 10 tuples per page, **S** contains 2000 tuples and has 20 tuples per page, **A** is the primary key of **R**, **R** and **S** are stored as heap files and 102 buffer pages are available.

   (a) If no indexes are available then what is the cost (in block I/Os) of joining **R** and **S** using the blocked nested loops algorithm (ignoring the cost of writing the result)?

   (b) If a hash index exists on **R.A** (assume data entries are <key value, record id> pairs) then what is the cost (in block I/Os) of the joining **R** and **S** using the index nested loops algorithm (ignoring the cost of writing the result)?

Student No: _____

7. [15 marks] Assume the following schema

   Employee (<u>empNo</u>: <u>integer</u>, **salary**: real, **name**: char(10), **deptNo**: integer**)**
   Department (<u>deptNo</u>: <u>integer</u>, **dName**: char(15), **budget**: real**)**

   where **Employee** has 20,000 tuples, **Department** has 5000 tuples, the primary keys are underlined, all integer and real fields are 5 bytes, pages are 4000 bytes, department names are unique and **Employee.salary** is uniformly distributed in the range 50,000 to 150,000 and values are in $1000 increments.

   Consider the query

   SELECT **E.empNo**, **E.name**
   FROM **Employee E**, **Department D**
   WHERE **E.deptNo** = **D. deptNo**
   AND **E.salary** > 100,000
   AND **D.dName** = "sales"

   (a) Give an optimized query tree for the above query.

   (b) Show the 1-relation plans that would be generated by the System R optimizer algorithms for the above query.

   (c) Show the 2-relation plans that would be generated by the System R optimizer algorithms for the above query.

Student No: _____

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Student No: _____

Section 3: Transaction management (40 marks)

8. [5 marks] Consider the different definitions of serializability we discussed in class.
   (a) Define conflict equivalence.
   (b) Define view equivalence.
   (c) Give an example of a schedule that is view serializable but not conflict serializable.

Student No: _____

9. [5 marks] Consider the following set of transactions:

   T1: R(Y), R(Z), W(Y), C
   T2: R(Z), W(Z), R(X), W(X), C
   T3: R(X), R(Y), W(Y), W(X), C

   where "R(X)" is a read operation on the object "X", "W(X)" is a write operation on the object "X" and "C" is a commit operation.

   Decide if the following schedule for the transactions T1, T2 and T3 is conflict serializable or not. If it is conflict serializable then give the equivalent serial schedule. A term "T1:R(X)" in the schedule represents a read of "X" by "T1". Show your work.

   T1:R(Y), T3:R(X), T2:R(Z), T2:W(Z), T3:R(Y), T3:W(Y), T2:R(X), T1:R(Z), T2:W(X), T2:C, T1:W(Y), T1:C, T3:W(X), T3:C.

Student No: _____

10. [15 marks] Consider the following schedule of operations for the 3 transactions in Question 9:

    T1:R(Y), T2:R(Z), T2:W(Z), T3:R(X), T3:R(Y), T3:W(Y), T1:R(Z), T1:W(Y), T2:R(X), T2:W(X), T3:W(X), T1:C, T2:C, T3:C

    Describe how each of the following concurrency control mechanisms handles the schedule.

    (a) Strict 2PL with deadlock detection. Add lock and unlock requests to the schedule as per the protocol. If a transaction is blocked then assume all of its actions are delayed and the DBMS continues with the next unblocked action in the schedule. Show the waits-for graph if a deadlock occurs.

    (b) Timestamp concurrency control with the Thomas write rule. Assume that the timestamp for transaction Ti is i. Assume that the initial read and write timestamps of all objects are 0.

    (c) Multiversion concurrency control. Assume there is initially 1 version of each object and that the read and write timestamps of that version are both 0.

Student No: _____

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Student No: _____

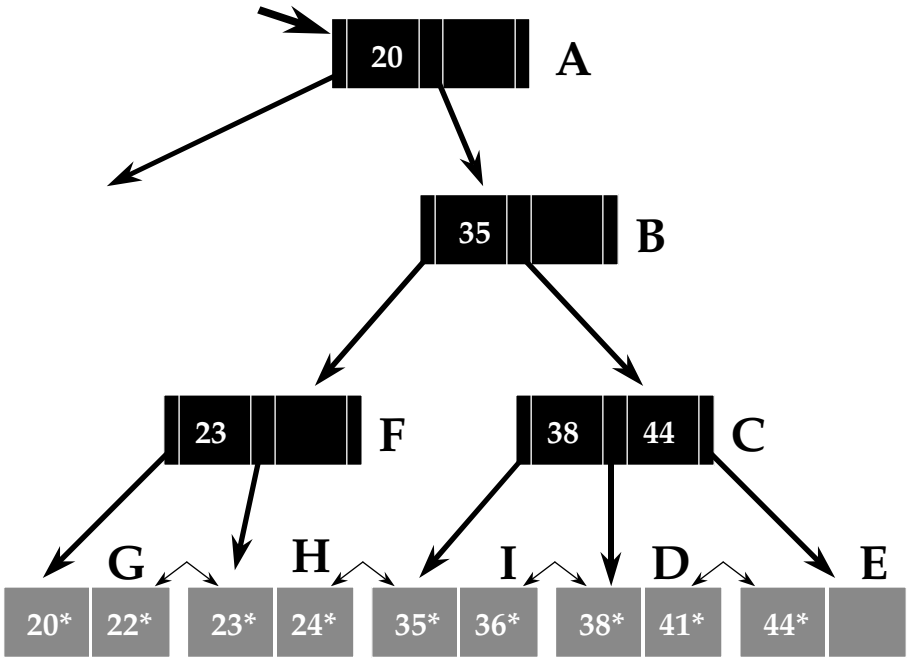
11. [15 marks] Consider the B+ tree of order 1 given below. Suppose that we wish to insert an entry for the search key 42*. Describe the process of acquiring and releasing locks on the B+ tree when each of the following index locking protocols are used.

    (a) Strict 2PL

    (b) Simple tree locking protocol (first tree locking protocol presented in class).

    (c) Tree locking protocol – variation 1 (second tree locking protocol presented in class).

    (d) Tree locking protocol – variation 2 (third tree locking protocol presented in class).



Student No: _____

THIS PAGE IS INTENTIONALLY LEFT BLANK.