EECS 484: Final Sample Exam #1

Please ***print*** your name and uniqname below.

Name:
Uniqname:

Instructions:

1. You have 80 minutes to complete this exam. The total number of points on this exam is **65**.
2. The questions vary widely in terms of difficulty. Watch your time carefully to avoid getting stuck on a hard question at the expense of an easier question.
3. Some of the questions on this exam require you to provide an explanation. You will be graded on the clarity of your explanation, in addition to the correctness of your answer.
4. At the end of the exam, sign the Honor Pledge given below.

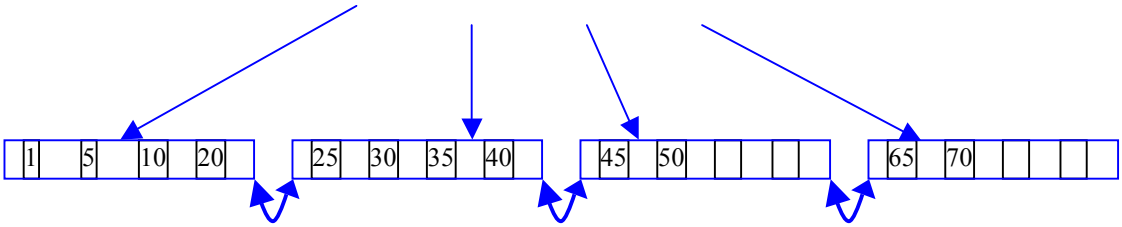
Honor Code Pledge: I have neither given nor received assistance in this exam, nor have I concealed any violation of the Honor Code.

Signature:

(For Instructor's Use)

Q1: Indexes (20 points)	
Q2: Query Evaluation. (25 points)	
Q3: Transactions & Recovery (20 points)	
TOTAL	

Question 1: (20 points) Indexes



(a) (4 points) Consider the B+-tree shown above. Draw the tree after inserting the key value **32***. Assume the convention that the left pointer points to values that are strictly less than the key value. Assume also that each node in this B+-tree can hold at most four data entries (i.e., the order of the tree is 2).

(b) (4 points) Using your answer from part (a), draw the tree after additionally inserting key value **15***.

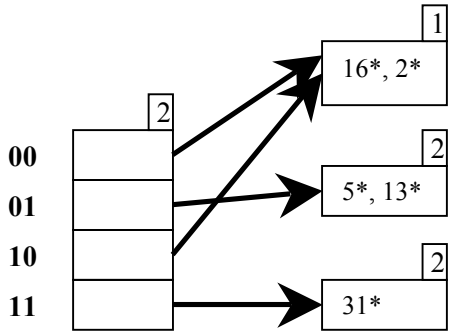
(c) (4 points) Consider a relation *R(a,b,c,d)* containing a million records. Suppose that *R* is stored in a heap file with unclustered indexes. The indexes use Alternative 2, where data entries are (value, rid) pairs. Suppose also that the records in *R* are randomly ordered. Assume that *a* is a candidate key for *R*, with values in the range 0 to 999,999. For each of the following queries, choose the approach that would likely require the fewest I/Os to complete the query. The approaches are as follows:

i) Scan through the entire heap file containing *R*.
ii) Use a B+-tree index on *R.a*.
iii) Use a hash index on *R.a*.

Find all tuples in R. _____
Find all tuples in R such that a < 50. _____
Find all tuples in R such that a = 50. _____
Find all tuples in R such that a > 50 and a < 100. _____

(d) (8 points) Consider the following extendible hash index. Assume that the hash function is h(K) = K mod 32, and that at most 2 data entries fit on a page.

Re-draw the index to show how it would appear after inserting the key values **8*** and **22*** (in that order).



Question 2: (25 points) Query Evaluation & Optimization

a) (15 points) Consider the join R $\bowtie_{R.a = S.b}$ S, and suppose you are given the following information about the relations to be joined. For all of the following questions, assume that the cost metric is the total number of page I/Os and that the cost of writing out the result should be uniformly ignored. For all questions, show your work to receive full credit.

Relation R contains 10,000 tuples and has 10 tuples per page
Relation S contains 2,000 tuples and has 10 tuples per page
Attribute S.b is the primary key for S
Both relations are stored as unsorted heap files
Neither relation has any indexes
52 buffer pages are available

What is the cost of joining R and S using a page-oriented nested loops join?

What is the cost of joining R and S using Grace hash join?

Suppose you chose to join R and S using a block-oriented nested loops join, and that you want to do this using ≤ 1200 I/Os. What is the minimum number of buffer pages necessary? Explain your answer.

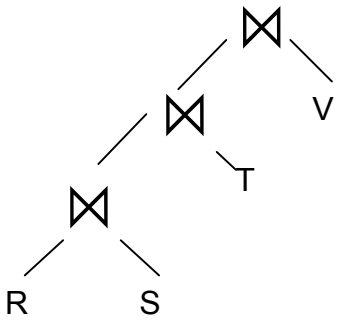
b) (3 points) Consider a relational schema including A(a1,a2,a3) and B(b1, b2, b3), and the following relational algebra expression over these relations. To which of the given expressions is it equivalent? (Select all that apply.)

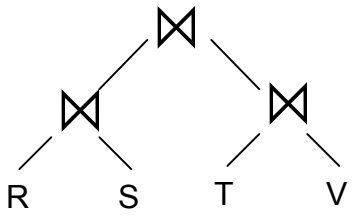
$\pi_{a3}\ (\sigma_{b2 = 5}\ (A \bowtie_{a1 = b1} B))$

i. $\pi_{a3}\ (\sigma_{b2 = 5}\ (\sigma_{a1 = b1}\ (\sigma_{b2 = 5\ V\ b2 = 10}\ (A \times B))))$
ii. $\pi_{a3}\ (\sigma_{b2 = 5}\ (\pi_{a3}\ (A \bowtie_{a1 = b1} B)))$
iii. $\pi_{a3}\ ((\sigma_{b2 = 5}\ B) \bowtie_{a1 = b1} A))$

Answer: _____

c) (4 points) Consider the four-way join $R \bowtie_{R.a = S.b} S \bowtie_{S.c = T.d} T \bowtie_{S.e = V.f} V$, where each relation (R,S,T,V) is the same size, and the result size for each two-way join is the same. Consider also the following two join plans (one left-deep and one bushy). Which plan would you choose? Explain your answer in terms of what the query engine must do for each.





d) (3 points) *Briefly* explain what role the system catalogs play in query optimization.

Question 3: (20 points) Transactions & Recovery

a) (4 points) A schedule produced by following the strict two-phase locking (Strict 2PL) protocol is guaranteed to have which of the following properties? (Select all that apply.)

i. The schedule must be serializable.
ii. The schedule must be recoverable.
iii. The schedule must avoid cascading aborts.
iv. The schedule must not result in a deadlock.

Answer: _____

b) (4 points) Consider the following (incomplete) schedule involving transactions T1, T2, and T3. Does this schedule avoid cascading aborts? Explain.

T1	**T2**	**T3**
R(X)		
W(X)		
		R(X)
		W(X)
	R(Y)	
	W(Y)	

☐ Yes ☐ No

Explanation:

c) (4 points) Consider the following (complete) schedule involving transactions T1, T2, and T3, and consider a system that implements ordinary 2-phase locking (2PL). Could this schedule possibly arise given this locking protocol? Explain.

T1	**T2**	**T3**
R(X)		
W(X)		
	R(Y)	
		R(X)
	W(Y)	
		W(X)
		Commit
Commit		
	Abort	

☐ Yes ☐ No

Explanation:

d) (3 points) *Briefly* explain the difference between a system crash and a media failure.

d) (5 points) Consider a system that implements the ARIES logging and recovery protocol. If the system fails repeatedly during recovery, what is the maximum number of log records that can be written (as a function of the number of update and other log records written before the crash) before restart completes successfully? Explain.