Name: ______________________________ Student#: _________________________ Date: __________

COMP 3311: Database Management Systems

Final Review

1. A. A page can hold either 3 records or 10 (key, search-pointer) index entries. If a database contains n records, then how many pages are needed to store both the data file and a single-level dense index?

   a) n/30

   b) 3n/10

   c) 10n/3

   d) 13n/30

   B. In a B^+-tree, if the search-key value is 12 bytes, the page size is 1024 bytes and a pointer is 6 bytes, then the maximum number of search-key values that can be stored in each non-leaf node of the B^+-tree is:

   a) 54

   b) 56

   c) 57

   d) 58

   C. What is the minimum number of search-keys in any non-root node for a B^+-tree in which the maximum number of search-keys in a node is 4?

   a) 1

   b) 2

   c) 3

   d) 4

2. The relation Sailor(<u>sailorId</u>, name, rating, age) is not sorted. Assume each attribute is 25 bytes, the page size is 1,000 bytes and there are 11,000 tuples. For the following questions, apply external sorting using a buffer of 11 pages.

   A. How many sorted runs will be produced in <u>pass 0</u>?

   a) 10

   b) 11

   c) 100

   d) 110

   B. What is the <u>total number of passes</u> required to sort the relation completely (including pass 0)?

   a) 2

   b) 3

   c) 4

   d) 5

   C. What is the total page I/O cost of sorting the relation?

   a) 4,400

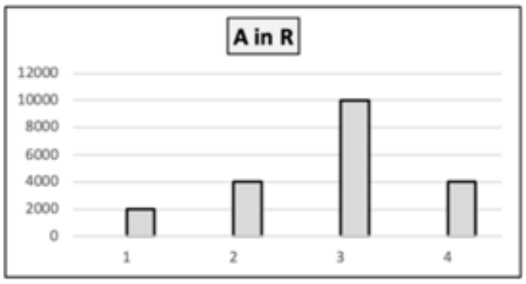
   b) 5,500

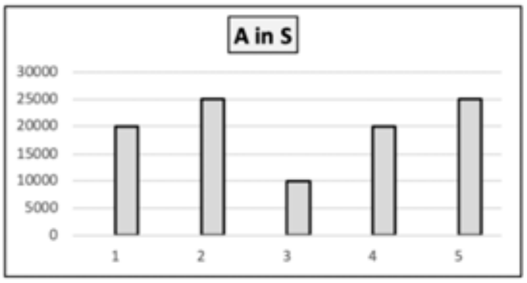
   c) 6,600

   d) 7,770

Name: ________________________________ Student#: ___________________________ Date: ___________

3. Consider the two relations R(<u>*B*, C</u>, A) and S(<u>B</u>, A, Y). R contains 20,000 tuples and S contains 100,000 tuples. Assume that for both relations 10 tuples fit per page (i.e., the size of R is 2,000 pages and that of S is 10,000 pages). The possible values of attribute A in R are {1, 2, 3, 4}, whereas the possible values of A in S are {1, 2, 3, 4, 5}. The following histograms present statistical information about the occurrences of values for A in R and S (e.g., there are 2,000 tuples with A=1 in R and 20,000 tuples with A=1 in S).





A. How many tuples are there in the query result of (R JOIN$_{R.A=S.A}$ S) (i.e., what is the cardinality of the output)?

a) 20,000
b) 100,000
c) 320,000,000
d) 400,000,000

B. What is the <u>minimum page I/O cost</u> to compute (R JOIN$_{R.A=S.A}$ S) using block nested-loop join; how many buffer pages are needed?

a) Minimum page I/O cost is 12,000; 2,000 buffer pages are needed.
b) Minimum page I/O cost is 12,000; 2,002 buffer pages are needed.
c) Minimum page I/O cost is 320,000; 2,002 buffer pages are needed.
d) Minimum page I/O cost is 320,000; 12,000 buffer pages are needed.

C. We want to compute (R JOIN$_{R.A=S.A}$ S) using block nested-loop join with R *as the outer relation*. What is the minimum number of buffer pages needed in order to achieve a page I/O cost of 42,000?

a) 502
b) 2,000
c) 2,002
d) 10,002

D. We want to compute (R JOIN$_{R.A=S.A}$ S) using hash join with R *as the build input*. How many buckets should be used for partitioning and what is the minimum buffer requirement?

a) 4 buckets; 202 buffer pages.
b) 4 buckets; 1,002 buffer pages.
c) 5 buckets; 202 buffer pages.
d) 5 buckets; 102 buffer pages.

E. Given that R.B is a <u>not null</u> foreign key referencing S.B, how many tuples are in the query result of (R JOIN$_{R.B=S.B}$ S)?

a) 20,000
b) 100,000
c) 120,000
d) 320,000

Name: ______________________________ Student#: __________________________ Date: __________

F. We want to compute (R JOIN$_{R.B=S.B}$ S) using indexed nested-loop join with R *as the outer relation*. Assume that there is a hash index on S.B with no overflow buckets (i.e., finding an index entry has page I/O cost 1). What is the join page I/O cost?

a) 6,000
b) 12,000
c) 22,000
d) 42,000

G. How many tuples are there in the result of (($\sigma_{A=1}$R) JOIN$_{R.B=S.B}$ S) and what is the minimum query processing page I/O cost using indexed nested-loop join with R *as the outer relation*. Assume that the only index is a hash index on S.B with no overflow buckets.

a) The result has 2,000 tuples; the page I/O cost is 6,000.
b) The result has 2,000 tuples; the page I/O cost is 22,000.
c) The result has 20,000 tuples; the page I/O cost is 40,000.
d) The result has 20,000 tuples; the page I/O cost is 42,000.

H. How many tuples are there in the result of (($\sigma_{A=1}$R) JOIN$_{R.B=S.B}$ ($\sigma_{A=3}$S)) and what is the minimum query processing page I/O cost using indexed nested-loop join with R *as the outer relation*. Assume that the only index is a hash index on S.B with no overflow buckets.

a) The result has 200 tuples; the page I/O cost is 600.
b) The result has 200 tuples; the page I/O cost is 6,000.
c) The result has 2,000 tuples; the page I/O cost is 6,000.
d) The result has 2,000 tuples; the page I/O cost is 42,000.

4. Consider the following schedules of two transactions T_1 and T_2. Indicate for each whether it is serial, (conflict) serializable or not serializable. r denotes a read and w a write operation.

a) Schedule: $r_1(A)$ $w_1(A)$ $r_2(A)$ $w_2(B)$

T_1	T_2
read(A)	
write(A)	
	read(A)
	write(B)

b) Schedule: $r_1(A)$ $r_2(A)$ $w_1(A)$ $w_2(B)$

T_1	T_2
read(A)	
	read(A)
write(A)	
	write(B)

c) Schedule: $r_1(A)$ $r_2(A)$ $w_2(A)$ $w_1(A)$

T_1	T_2
read(A)	
	read(A)
	write(A)
write(A)	

Name: ______________________________ Student#: _________________________ Date: __________

5. Consider the schedule $r_1(A)$ $w_1(A)$ $r_2(A)$ $w_2(B)$ c_1 c_2 (where c_1 and c_2 indicate the commit statements).

   a) Is the schedule recoverable and cascadeless? Why?

   b) Change the time of the commits (c_1, c_2) in the schedule in a) so that it becomes a cascadeless schedule.

   c) Is the schedule $r_2(A)$ $r_1(A)$ $w_1(A)$ $w_2(B)$ c_2 c_1 recoverable and cascadeless?

6. Modify the schedule $r_2(A)$ $r_1(A)$ $w_1(A)$ $w_2(B)$ according to the 2PL protocol by adding lock-s, lock-x, unlock statements. Explain briefly whether the schedule is allowed by 2PL.

*T*₁	*T*₂
	read(A)
read(A)	
write(A)	
	write(B)

7. Modify the schedule $r_2(A)$ $r_1(A)$ $w_1(A)$ $w_2(B)$ according to timestamp-ordering protocol by adding RTS (read timestamp) and WTS (write timestamp) statements. Assume that the timestamps of T_1 and T_2 are 2 and 1, respectively. The initial read and write timestamps of A and B are both 0.

T_1 [TS=2]	T_2 [TS=1]
	read(A)
read(A)	
write(A)	
	write(B)

8. Modify the schedule $r_2(A)$ $r_1(A)$ $w_1(A)$ $w_2(B)$ according to the multi-version timestamp-ordering protocol by adding RTS (read timestamp) and WTS (write timestamp) statements and specify the versions of the items. Assume that the timestamps of T_1 and T_2 are 1 and 2, respectively and that the initial versions of items are A_0 and B_0 with values 0 for their read and write timestamps. Complete the correct version numbers (e.g., read(A_0) instead of read(A)).

T_1 [TS=1]	T_2 [TS=2]
	read(A)
read(A)	
write(A)	
	write(B)

You must upload this completed exercise sheet to Canvas by **11 p.m. today.**