Database Systems

25th April 2018

End Semester Exam

Duration: 3 hrs, Marks: 90

1. No clarifications during the exam.
2. Make reasonable assumptions and clearly state them to answer ambiguous questions.
3. Show your steps. Be concise and organized.
4. Calculators allowed. Sharing of calculators not allowed.

1) The Megatron 747 disk has the following characteristics:

1. There are 4 platters providing 8 surfaces, with 8192 tracks per surface.
2. Tracks hold an average of 256 sectors of 512 bytes each.
3. 10% of each track is used for gaps.
4. The disk rotates at 3840 rpm.
5. The time it takes the head to move n tracks is 1+0.002n milliseconds.

Suppose the Megatron 747 disk head is at track 1024, i.e. 1/8 of the way across the tracks. Suppose that the next request is for a block on a random track. Calculate the average time to read this block.

[10]

2) Suppose that blocks can hold either ten records or 99 keys and 100 pointers. Also assume that the average B-tree node is 70% full; i.e., it will have 69 keys and 70 pointers. The data file is a sequential file, sorted on the search key, with 10 records per block. The B-tree is a dense index. Determine: (i) the total number of blocks needed for a 1,000,000-record file, and (ii) the average number of disk I/O's to retrieve a record given its search key. You may assume nothing is in memory initially, and the search key is the primary key for the records.

[10]

3) Suppose we store a relation R(x,y) in a grid file. Both attributes have a range of values from 0 to 1000. The partitions of this grid file happen to be uniformly spaced; for x there are partitions every 20 units, at 20, 40, 60, and so on, while for y the partitions are every 50 units.

We wish to perform a nearest-neighbour query for the point (110, 205). We begin by searching the bucket with lower-left corner at (100, 200) and upper-right corner at (120, 250), and we find that the closest point in this bucket is (115, 220). What other buckets must be searched to verify that this point is the closest?

[10]

4) Suppose B(R) = B(S) = 10,000, and ~~M = 1000~~. What value of M would we need to compute R ⋈ S using the nested-loop algorithm with no more than 1,00,000 disk I/O's?

[10]

5) Below are the vital statistics for four relations, W, X, Y, and Z:

W(a, b): T(W) = 100; V(W, a) = 20; V(W, b) = 60
X(b, c): T(X) = 200; V(X, b) = 50; V(X, c) = 100
Y(c, d): T(Y) = 300; V(Y, c) = 50; V(Y, d) = 50
Z(d, e): T(Z) = 400; V(Z, d) = 40; V(Z, e) = 100

0.22

Estimate the sizes of relations that are the results of the following expression:

(a) W ⋈ X ⋈ Y ⋈ Z
(b) $\sigma_{a=10}(W)$

[1

Scanned by CamScanner

6) With respect to transaction serializability, consider these protocols:

(A) 2PL

(B) Validation test

(C) *Timestamp ordering*

Is the following schedule valid in each of the above protocols? Describe why or why not, for each one.

$R_1(a), R_2(b), R_3(c), R_1(d), R_2(c), R_3(d), W_3(c), W_1(d), W_3(d)$

[15]

7) The following is a sequence of undo-log records written by two transactions T and U: <START T>; <T, A, 10>; <START U>; <U, B, 20>; <T, C, 30>; <U, D, 40>; <COMMIT U>; <T, E, 50>; <COMMIT T>. If there is a crash and the last log record to appear on disk is:

a. <START U>

b. <COMMIT U>

c. <T, E, 50>

d. <COMMIT T>

For each of these situations, what values written by T and U *must* appear on disk?

[10]

8) Consider the following sequence of undo/redo log records: <START S>; <S, A, 60, 61>; <COMMIT S>; <START T>; <T, A, 61, 62>; <START U>; <U, B, 20, 21>; <T, C, 30, 31>; <START V>; <U, D, 40, 41>; <V, F, 70, 71>; <COMMIT U>; <T, E, 50, 51>; <COMMIT T>; <V, B. 21, 22>; <COMMIT V>. Suppose that we begin a non-quiescent checkpoint immediately after one of the following log records has been written (in memory):

a. <S, A, 60, 61>

b. <T, A, 61, 62>

c. <U, B, 20, 21>

d. <T, E, 50, 51>

For each, tell:

i. At what points could the <END CKPT> record be written, and

ii. For each possible point at which a crash could occur, how far back in the log we must look to *find all possible* incomplete transactions. Consider both the case that the <END CKPT> record was or *was not* written prior to the crash.

[15]

Scanned by CamScanner