COP 4720 - Information Systems and Databases II, Spring 2007, Final

(This test consists of 5 questions in 2 pages.) Good luck

Remarks:

- This test is out of 100 points. The value of each question/sub-question is written in square brackets, next to question number. You can get partial credits for your answers.
- Answer each question on a separate paper.
- Write your name on every paper you return.

Question 1 [24] Let \( R \) and \( S \) be two clustered relations where \( B(R) \leq B(S) \). For each of the following queries, briefly describe the smallest memory requirement for executing it in a single pass. Also write a brief pseudocode of this single pass algorithm.

A [12] \( S - set R \) (set difference)

B [12] \( R - bag S \) (bag difference)

Question 2 [30] Let \( R_1(X, Y_1) \), \( R_2(X, Y_2) \) and \( R_3(X, Y_3) \) be three clustered relations where \( B(R_1) = 4,000 \), \( B(R_2) = 6,000 \), \( B(R_3) = 8,000 \). Assume that the available memory size is \( M = 100 \) blocks. Answer the following questions:

A [10] Find the smallest number of disk I/Os for joining \( R_1 \) and \( R_2 \) using BNLJ (block nested loop join). In other words, \( R_1 \bowtie_\theta R_2 \), with \( \theta : R_1.X = R_2.X \)

B [20] Assume that the \( X \) attribute of each of the three relations takes integer value uniformly in the \([1 : 1,000]\) interval. Briefly describe how you would adapt the BNLJ to join the three relations \( R_1, R_2 \) and \( R_3 \) (i.e., \( R_1 \bowtie_\theta R_2 \bowtie_\theta R_3 \), with \( \theta : R_1.X = R_2.X = R_3.X \)). Write the pseudocode of your algorithm. Aim to minimize the number of disk I/Os in your algorithm.

What is the total number of disk I/Os for the algorithm you presented in B?

Question 3 [15] List at least one advantage of


Multi-version time stamps over time stamps.

**Question 4 [19]** Answer the questions below for the following schedule:

\[
S = r_2(C)r_1(A)w_3(C)w_3(D)r_2(B)w_2(A)r_1(D)w_3(B)w_4(C)
\]

A [7] Draw the precedence graph of \( S \).


C [10] Show how \( S \) will behave if 2PL is used.

**Question 5 [12]**

Assume that concurrency is maintained using time stamps. Find out what will happen for each of the schedules below. Note that \( st_i \) denotes the start of transaction \( T_i \).

A [6] \( st_1st_2w_2(B)r_1(A)w_2(A)w_1(B) \)

B [6] \( st_1r_1(B)st_2w_2(A)st_3r_3(A)r_1(A) \)