CS304 Database System Concepts

Assignment 7

Due: April 6, 2012
(Please submit hard copies to class or to Zheng on due date.)

Name: Matric No:

Q1. (2 points)Let relations $r_1(A, B, C)$ and $r_2(C, D, E)$ have the following properties: r_1 has 50,000 tuples, $r_2$has 45,000 tuples, 25 tuples of r_1 fit on one block, and 30 tuples of r_2 fit on one block. Estimate the number of block transfers and seeks required, using each of the following join strategies for $r_1 \bowtie r_2$

a) Nested-loop join
b) Block nested-loop join
c) Merge join
d) Hash join

Q2. (2 points)Answer the following questions of the scenario: a file with 2,000,000 blocks and 17 available buffer blocks.

1. How many runs will you produce in the first pass?
2. How many passes will it take to sort the file completely?
3. What is the total I/O cost of sorting the file?
4. How many buffer blocks do you need to sort the file completely in just two passes?

Q3. (2 points)Pipelining is used to avoid writing intermediate results to disk. Suppose youneed to sort relation r using sort–merge and merge-join the result with analready sorted relation s.

a) Describe how the output of the sort of r can be pipelined to the merge joinwithout being written back to disk.

b) The same idea is applicable even if both inputs to the merge-join are theoutputs of sort–merge operations. However, the available memory has tobe shared between the two merge operations (the merge-join algorithmitself needs very little memory).What is the effect of having to share memoryon the cost of each sort–merge operation.

Q4. (2 points)Suppose that a B+-tree index on building is available on relation department,and that no other index is available.Whatwould be the bestway to handlethe following selections that involve negation?

a) $\sigma\neg(\text{building} < \text{“}Watson\text{”})(department)$

b) $\sigma\neg(\text{building} = \text{“Watson”})(\text{department})$

c) $\sigma\neg(\text{building} < \text{“}Watson\text{”} \vee \text{budget} < 50000)(department)$

Q5. (2 points)Suppose two relations r and s have histograms on attributes r.A and s.A, respectively,but with different ranges. Suggest how to use the histograms toestimate the size of $r \bowtie s$. Hint: Split the ranges of each histogram further.