

Fall 2020 – Quiz 8 (query execution)

10 points, 15 minutes

Monday afternoon section:

Consider joining tables R and S using the **partitioned hash join** algorithm. Suppose $B(R) = 3,000$, $B(S) = 20,000$, and $M = 101$ pages.

1. [8 points] Describe the steps (including input, output, and their sizes at each step) for the join algorithm.

step1: Hash R into 100 buckets, 30 blocks/bucket, and send all buckets back to disk [3pts]

step2: Hash S into 100 buckets, 200 blocks/bucket, and send all buckets back to disk [3pts]

step3: For each bucket R_i and S_i , join corresponding matched tuples [2pts]

2. [2 points] What is the total number of block I/O's needed for the algorithm?

Total Cost = $3B(R) + 3B(S) = 3 * 3000 + 3 * 20000 = 69000$

Makeup:

Consider joining tables R and S using the **sort-merge** algorithm. Suppose $B(R) = 3,000$, $B(S) = 20,000$, and $M = 101$ pages. Suppose 100 pages are used for sorting.

Note that when the number of runs of a relation is too large for merging, the runs will be further merged first. Select the relation with a larger number of runs for further merging if both have too many runs.

1. [8 points] Describe the steps (including input, output, and their sizes at each step) for the algorithm.

a. sort R using 100 pages into 30 runs. Input: R, output: sorted R

b. sort S using 100 pages into 200 runs. Input: S, output: sorted 200 runs of S

c. sort S again using 100 pages into 2 runs. Input: 200 runs of S, output: sorted 2 runs of S

d. merge R and S. Input: sorted R and S, output (doesn't count in cost): joined tuples

2. [2 points] What is the total number of block I/O's needed for the algorithm?

$3B(R) + 5 * B(S) = 9,000 + 100,000 = 109,000$

1. each step 2 points, if more or wrong steps provided, each wrong step -0.5

2. right equation 1.5 according to student's question 1 steps, calculation 0.5