

INF 551 – Spring 2016

Quiz 4: Query Execution (10 points)

10 minutes

Consider natural-joining relations $R(A,B)$ and $S(B,C)$. Suppose that R has 1,000 blocks, S has 2,000 blocks, and there are 101 pages (blocks) of main memory available for the join. Ignore the cost of writing the output of join. Use all the available memory in each step of the join algorithm.

1. [5 points] Describe the steps of joining R and S using the sort-merge algorithm (be sure to indicate the size of runs and how many ways the merging is). How many block I/O's are needed?

Answer:

Step 1:

split R into runs of size 101 blocks (except for the last run which has 91 blocks). $\lceil 1000/101 \rceil = 10$ runs

split S into runs of size 101 blocks (except for the last run which has 81 blocks). $\lceil 2000/101 \rceil = 20$ runs

cost: $2B(R) + 2B(S)$

Step 2:

merge runs from B & S : 10 pages of buffer for R and 20 for S (so 30-way merging)

cost: $B(R) + B(S)$

Total cost: $3B(R) + 3B(S) = 9000$

2. [5 points] Describe the steps of joining R and S using the partitioned-hash algorithm (be sure to indicate the size of buckets). How many block I/O's are needed?

Answer:

Step 1: Hash S into 100 buckets, send all buckets to disk. 10 blocks/bucket.

Step 2: Hash R into 100 buckets, send all buckets to disk. 20 blocks/bucket.

Step 3: Join corresponding buckets from R and S (i.e., with the same bucket number).

Total cost: $3B(R) + 3B(S) = 9000$