

## Query processing worksheet

For the following questions, express your answers in terms of the following statistics:

$nTuples(R)$  – cardinality of  $R$

$bFactor(R)$  – blocking factor of  $R$

$nBlocks(R)$  – number of blocks in  $R$

$nDistinct_A(R)$  – number of distinct values for  $A$  in  $R$

$min_A(R), max_A(R)$ , – range endpoints for  $A$  in  $R$

$SC_A(R)$  – selection cardinality of  $A$  in  $R$

1. Estimate the size of the result returned by a selection under the following conditions:
  - a. Inequality conditional ( $A > c$ )
  - b. Equal to any element in set  $\{c_1, \dots, c_n\}$
2. Estimate the cost in disk accesses for a selection under the following conditions:
  - a. Linear search on unordered file with no index
    - i. Equality condition on unique attribute
    - ii. Equality condition on non-unique attribute or inequality condition
  - b. Binary search on ordered file with no index
    - i. Equality condition on unique attribute
    - ii. Equality condition on non-unique attribute
  - c. Hash lookup on hashed attribute
  - d. Index lookup on indexed attribute
3. Estimate the size of the result returned by a join under the following conditions:
  - a.  $R$  equijoin  $S$  on a key in  $R$
  - b.  $R$  equijoin  $S$  on some non-unique attribute

4. Estimate the cost in disk accesses for a join under the following conditions:
  - a. Basic nested loop join
    - i. When all blocks for R and S can be loaded into memory
    - ii. Loading one block of R (outer loop) and loading each block of S (inner loop) one at a time
    - iii. Loading ( $\text{maxBlocks} - 2$ ) blocks of R at a time and joining with one block of S (inner loop) at a time
  - b. Sort-merge join
    - i. When the relations are already sorted on the join attribute
    - ii. When the relations are not sorted on the join attribute
5. Give upper and lower bounds for the size of the results of a:
  - a. Set union
  - b. Set intersection
  - c. Set difference