Reminders

Due: Nov. 13, 3pm (in class)
Please, TYPE your answers - drawings may be handwritten (but NEAT!)
Total weight: 3% of the course grade.
Estimated time, for someone who knows the material: 30 minutes

1 Syntactic Query optimization [15 pts]


2 Equivalence of expressions [15 pts]

Consider the relations $r(A, B, C)$ and $s(A, B, C)$. Disprove the 'rule':

$$\pi_A( r \cap s ) \neq ? \pi_A(r) \cap \pi_A(s)$$

3 Selectivities [30 pts]

Consider relation $r(A, B, C)$ and $s(A, D, F)$, with the following statistics:

- \( n_r = 10,000 \) tuples in \( r \)
- \( n_s = 20,000 \) tuples in \( s \)
- \( b_r = 1,500 \) blocks (= pages) that \( r \) occupies
- \( b_s = 3,000 \) blocks (=pages) that \( s \) occupies
- \( V(A, r) = 1,000 \) distinct values of \( A \) in \( r \)
- \( V(B, r) = 5,000 \) distinct values of \( B \) in \( r \)
- \( V(A, s) = 3,000 \) distinct values of \( A \) in \( s \)

Give your estimates for the following questions:
1. How many tuples qualify for the query ([5 pts])

   ```
   select *
   from r
   where A = 12
   ```

2. How many tuples qualify for the query ([10 pts])

   ```
   select *
   from r
   where A = 12 or B = 13
   ```

3. How many tuples qualify for the join ([15 pts])

   ```
   r \bowtie s
   ```

4. **Joins [40 pts]**

   For the same two relations $r$ and $s$ as before, with the same statistics, estimate the number of disk (=block = page) accesses, for the join $r \bowtie s$, for the following settings:

   1. Block nested loop join, with $r$ as the outer relation, and $m=2$ buffers only (each buffer can hold one page) ([8 pts])
   2. Repeat, with $s$ as the outer relation, and $m=2$ buffers only ([8 pts])
   3. Block nested loop join, with $r$ as the outer relation, and $m=100$ buffers. Specify how many buffers are allocated to each relation, to minimize the number of disk accesses. ([10 pts])
   4. Consider the optimal setting for block nested loop join

      - What is the smallest number of disk accesses $D_{min}$ we could hope for, with block nested loop join, and with enough buffers $m_{best}$? ([4 pts])
      - What is the smallest number of buffers $m_{best}$ in this case? ([4 pts])
      - Which relation should be the outer relation in this optimal case? ([3 pts])
      - How many buffers should we allocate to the outer relation? ([3 pts])

   End of questions