

- This is a closed book, **60 minutes** test.
- **No questions are allowed during the test. If in doubt, write down your doubts and assumptions and proceed with your answer.**

LAST NAME	SOLUTIONS
FIRST NAME	SOLUTIONS
YORK ID#	SOLUTIONS
CS LOGIN	SOLUTIONS

Exercise 1 [10 points]

Consider two relations R1(A, B, C) and R2(C, D, E) that have the following properties: R1 has 20,000 tuples, R2 has 45,000 tuples. 25 tuples of R1 fit in one block and 30 tuples of R2 fit in one block. Estimate the number of block transfers required, using each of the join strategies for R1 join R2.

- a. Nested-loop join
- b. Block nested-loop join

Assume that there are M blocks of available memory to perform the join. Join here means natural join, i.e., equijoin on the common attribute(s).

Answer:**Nested-loop join:**

Using R1 as the outer relation we need $20000 * 1500 + 800 = 30,000,800$ disk accesses;

if R2 is the outer relation we need $45000 * 800 + 1500 = 36,001,500$ disk accesses.

Block nested-loop join

If R1 is the outer relation, we need $\text{ceiling}(800 / (M-1)) * 1500 + 800$ disk accesses;

if R2 is the outer relation we need $\text{ceiling}(1500 / (M-1)) * 800 + 1500$ disk accesses.

Note, if $M > 800$, the join can easily be done in $1500 + 800$ disk accesses, using even plain nested-loop join

Exercise 2 [10 points]

Consider the relations R1 (A, B, C), R2(C, D, E) and R3(E, F), with primary keys A, C, and E, respectively. Assume that R1 has 1000 tuples, R2 has 1500 tuples, and R3 has 750 tuples.

- a. Estimate the size of $R1 \text{ join } R2 \text{ join } R3$. (*join* means natural join, i.e., equijoin on the common attribute(s)).
- b. Would any indices available on attributes of R1, R2, R3 could help the above join? If so, on which attributes and how would they help?

Answer:

- a. 1000 tuples at the most
- b. indices on C of R2 and E of R3

Exercise 3 [10 points]

Use the replacement method to produce runs (in the context of sort merge) from the following data stream.

23, 29, 5, 17, 9, 55, 41, 3, 51, 33, 18, 24, 11, 47.

- Assume that the stream is processed left to right.
- Assume that the main memory has a capacity of 4 (four) pages that can be used for the current set and also has additional pages – as needed, for the input and output sets.
- Assume that one entry occupies one page.
- Assume that an entry is brought in the current set and is available to be processed as soon as space in it becomes available.

Show the runs that are produced and also show the contents of the current set upon the start of each run.

Answer:

Run 1: **5, 9, 17, 23, 29, 41, 51, 55**

Run 2: **3, 11, 18, 24, 33, 47.**