INFS2200/7903 – Relational Database Systems

School of Information Technology and Electrical Engineering (ITEE), UQ

Tutorial 7

Question 1 Consider a relation R(A,B,C) containing 5,000,000 records, where each data page (i.e., block) of the relation holds 10 records. R is organized as an ordered file that is sorted on R.A. Assume that R.A is a unique key for R, with values lying in the range 0 to 4,999,999. For each of the following relational algebra queries, state which of the following two approaches is most likely to be more efficient (i.e., reads fewer number of blocks) and justify your answer.

Approaches:
1. Access the sorted file for R directly.
2. Use a B+ tree index on attribute R.A.

Relational algebra queries:

A. $\sigma_{A \leq 50,000}(R)$

B. $\sigma_{A \geq 50,000 \text{ and } A < 50,010}(R)$

C. $\sigma_{A \neq 50,000}(R)$

Question 2 Consider the join $R \bowtie_{R.A = S.B} S$, given the following information about the relations to be joined. The cost metric is the number of block read operations.
- Relation R contains 10,000 tuples and has 10 tuples per block.
- Relation S contains 2,000 tuples and also has 10 tuples per block.
- Attribute B of relation S is the primary key for S.
- Both relations are stored as simple heap (unsorted) files.
- Neither relation has any index built on it.
- 52 buffer blocks are available.

A. What is the cost of joining R and S using a tuple-based nested loop join?

B. What is the cost of joining R and S using a page-oriented nested loop join?

C. What is the cost of joining R and S using a block nested loop join?

D. How many tuples does the join of R and S produce at most, and how many blocks are required to store the result of the join back on disk?

E. Would your answer to Q2(A) change, if you were told that R.A is a foreign key that refers to S.B?