COMP 3311: Database Management Systems

Tutorial 7
Query Processing

Exercise 1: Given relations: $R_1(A, B, C)$ and $R_2(\underline{C}, D, E)$

R_1 20,000 tuples	bf_{R1}: 25 tuples/page	# pages: 800
R_2 45,000 tuples	bf_{R2}: 30 tuples/page	# pages: 1500

Assume:
- 100 main memory pages.
- R_2 has a B$^+$-tree index with 3 levels on the join attribute C, the primary key of R_2.
- R_1 and R_2 are not initially sorted on the join attribute.

Estimate the number of page I/Os required, *in the worst case*, using each of the following join algorithms for $R_1 \bowtie R_2$.

a) Block nested-loop join (*worst case cost*)

    i.   using R_1 as the outer relation

    ii.   using R_2 as the outer relation

b) Indexed nested-loop join (*worst case cost*)

c) Sort-merge join (R_1 and R_2 are not initially sorted on the join attribute)

d) Hash join using 10 buckets

Name: (1) _____ Student#: (1) _____ Date: _____

Name: (2) _____ Student#: (2) _____

NOTE: You are highly encouraged to do this exercise with a partner.

COMP 3311: Database Management Systems

Tutorial 7
Query Processing

Exercise 2: Given relation R(A, B, C, D, E), organized as a sequential file on search key A, and the information below, answer the questions.

Tuple size: 200 bytes Attribute A: 16 bytes Page size: 2400 bytes
Number of tuples: 500,000 Pointer size: 4 bytes

a) How many pages are required to store R?

b) How many index pages are required if the search key A is organized using a static, multi-level index?

c) Consider the query: select * from R where A=xxx. For each of the query evaluation strategies given below, determine the cost in page I/Os of each strategy.

  i. linear scan

  ii. binary search

  iii. index search

d) Consider the query: select * from R where A>700000. What is the cost in page I/Os to answer this query using the index assuming that A is uniformly distributed on the interval [200,000; 800,000] and the leaf index pages are chained?