Extra exercises 7: Query operators II

Question 1. True or false: Whenever an index is available, index nested loops always perform better than block nested loops.

Question 2. True or false: Hash indexes can be used for index nested loops implementation.

Question 3. True or false: Hybrid hash joins never have a higher I/O cost than two-pass hash joins.

Question 4. True or false: Hash table overflows must always be solved with recursive partitioning.

Question 5:
Suppose that we want to join tables R (100000 records) and S (50000 records), stored on M = 1000 and N = 500 pages respectively.

Assume that all hash functions we use are perfectly uniform (partitioning N pages into K partitions will result in partitions of size at most ceil(N / K), for all N and K). In other words, we assume that the fudge factor is f = 1.

When answering the "total I/O cost" questions below, do **not** consider the cost of writing out the join results, which can be as large as M * N.

Answer the following questions about two-pass hash joins:

1. What is the smallest value of B, the number of buffer pages, needed to perform a two-pass hash join?
2. What is the total I/O cost for two-pass hash joins?

Question 6:
Answer the following questions about hybrid hash joins, using B = 50 buffer pages.

1. Assuming that k = B = 50, what is the largest possible value of t, the number of hash values to retain per phase, such that we can perform a hybrid hash join? With such t and k, the total I/O cost for hybrid hash join?

Question 7:
Answer the following questions about sort merge joins.

1. What is the number of pages in memory needed to perform a one-pass join?

2. Assuming we have 10 buffer pages, what is the total I/O cost to perform a multi-pass sort merge join (without refinement)?

Question 8:
Match the advantages to the corresponding join algorithm:

1. Good when M is large but N is small.
2. Good when table records are already sorted on the join key
3. Good when multiple cores are available for parallelization
4. Good when join condition is an inequality

Question 9:
Which of the indexes below can be used as an index-only scan for the following query? (select multiple)

SELECT AVG(score) FROM students WHERE class = ‘A’ GROUP BY year

A. B+ tree index on (score, id)
B. Hash index on (class, score, year)
C. B+ tree index on (score, class, year, id)
D. Hash index on (score, class)

Extra exercises 7: Query operators II
Solutions

Answer 1: false
Answer 2: true
Answer 3: true
Answer 4: false
Answer 5:

1. 23 or 24 (per the sqrt formula or the actual (B - 1)(B - 2) >= N formula)
2. 4500 = 3 * (M + N)

Answer 6:

1. 5, by solving t/k <= k/N.
2. 4200 = (3 - 2 * t/k)(M + N)

Answer 7:

1. 1500, to hold all pages of the tables in memory.
2. 12500 = 2 * 1000 * (1 + ceil(log9(1000 / 10))) + 2 * 500 * (1 + ceil(log(9(500/10)))) + 1000 + 500

Answer 8:

1. hash join
2. sort merge join
3. hash join
4. sort merge join

Answer 9:
B+C