Topics

Hash Based Duplicate Elimination

HQ_9_1

Suppose we want to join two clustered relations R and S with B(R)=250 and B(S)=400. Which of the following algorithm is applicable if the memory capacity M is 15?

✖ Sort-merge join

✔ Nested loop join

-

✖ Hash-based join

✖ None

HQ_9_2

Consider two relations R and S, with B(R) = 50 and B(S) = 80. If the cost of using a sort-based two-pass algorithm is 390 for a certain operation-- what operation is performed?

✖ Grouping

✔ Sort-based union

✖ Aggregation

✖ Sort-merge join

HQ_9_3

Given relation R with 35000 tuples with 700 tuples/block, and relation S with 40000 tuples with 500 tuples/block. Calculate M, the number of blocks in main memory, if the cost of joining R and S using the 'block-based nested loop join' is 150.

✔ 41

- B(S) + B(R) * B(S) / (M - 1) = 150

✖ 40

- See the correct answer for calculation.

✖ 57

- See the correct answer for calculation.

✖ None

- See the correct answer for calculation.

HQ_9_4

Suppose we have two relations, R(x, y) and S(y, z). We know that R occupies 400 blocks on disk and S 300 blocks. Neither R nor S are sorted on any of their attributes. The memory buffer fits 51 blocks (M=51). Suppose we want to join R and S using a hash-based join. Is this possible?

✖ Possible, since 400 > 51.

- See the correct answer for calculation.

✔ Possible, since 300/50 <= 50.

✖ Impossible, since 51 < 300.

- See the correct answer for calculation.

✖ Impossible, since 400/50 > 300/50.

- See the correct answer for calculation.

HQ_9_5

Assume a clustered relation R has 30,000 tuples with 100 tuples per block. What should be the minimum number of blocks in main memory that allows the two-pass merge sort operation on R?

✖ None

✖ 30

✖ 17

✔ 18

HQ_9_6

For two-pass algorithms based on hashing, when partitioning a relation R into buckets on disk, with memory of size M, each bucket has size approximately:

✖ B(R)

✔ B(R)/M

✖ M^2/B(R)

HQ_9_7

What is the cost of performing a table scan on relation R where B(R) = 20 and T(R) = 100?

✔ 20 or 100

✖ 20

✖ 100

✖ 5

HQ_9_8

How should we sort the data when performing a sort-merge join?

✖ Based on the whole tuple, from the first to the last attribute.

✖ Based on the key of each relation.

✔ Based on the join attribute.

HQ_9_9

What is the cost of performing a hash-based join on relations R and S where R fits in main memory, but S is much larger than main memory and both R and S are clustered?

✔ B(R) + B(S)

✖ B(R) * B(S)

✖ 3B(R)

HQ_9_10

What is the worst-case time complexity of a two-pass multi-way merge sort where B(R) < M^2?

✖ 3B(R)

✖ 4B(R)

✔ T(R) + 2B(R)

✖ 3T(R)
