

Assignment 9 – Week 13

---

(1) What are the objectives of query processing?

ANS:

The objectives of query processing are

- to transform the query which are written in the high-level language (e.g., SQL) into the correct and efficient execution strategy expressed in the low-level language (e.g., implementing RA) and
- to execute that strategy to retrieve required data.

(2) What are the typical phases of query processing?

ANS:

The typical phases of query processing are

1. decomposition (parsing and validating)
2. optimization
3. code generation
4. execution

(3) State the heuristics that should be applied to improve the processing of a query.

ANS:

- performing selection as early as possible
- combining cartesian product with the subsequent selection whose predicate represents join condition into a join operation
- using associativity of binary operations to rearrange leaf nodes so that leaf nodes with most restrictive selection operations can execute first
- performing projection as early as possible
- computing common expressions once

(4) What types of statistics should a DBMS hold to be able to derive estimates of relational algebra operations?

ANS:

A DBMS hold the following types of statistics to be able to derive estimates of relational algebra operations

for each base relation R:

- $nTuples(R)$  – the number of tuples or records or cardinality of relation R
- $bFactor(R)$  – the blocking factor of R or number of tuples of R which fit into one block
- $nBlocks(R)$  – the number of blocks required to store R. If the tuples of R are stored physically together, then  $nBlocks(R) = \lceil nTuples(R)/bFactor(R) \rceil$

for each attribute A of base relation R:

- $nDistinctA(R)$  – the number of distinct values that appear for attribute A in relation R

- $\min A(R)$ ,  $\max A(R)$  – the minimum and maximum possible values for the attribute A in relation R
- $SCA(R)$  – the selection cardinality of attribute A in relation R

(5) What are the differences between materialization and pipelining?

ANS:

Materialization is the process where the results of intermediate relational algebra operations are written temporarily to disk. The output of one operation is stored in a temporary relation for processing by the next operation.

Pipelining or stream-based processing or on-the-fly processing is to pipeline the results of one operation to another operation without creating a temporary relation to hold the intermediate result so that we can save on the cost of creating temporary relations and reading the results back in again.

MUM-DBMS