Database Management System – cs422 DE

Assignment 9 – Week 13

This assignment is based on lecture 11 (chapter 23 – Query Processing)

- Submit your *own work* on time. No credit will be given if the assignment is submitted after the due date.
- Note that the completed assignment should be submitted in .doc, .docx, .rtf or .pdf format only.
- In MCQs, if you think that your answer needs more explanation to get credit then please write it down.
- You are encouraged to discuss these questions in the Sakai forum.

---

(1) What are the objectives of query processing?

ANS:

i. To perform a query written in a high-level language i.e., SQL into an accurate and efficient execution strategy expressed in a low-level language.
ii. It is also used to execute the strategy to retrieve the data from the database.

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

ANS:

i. Query decomposition (consisting of parsing and validation)
ii. Query optimization
iii. Code generation
iv. Execution

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

ANS:

i. Perform selection operations as early as possible.
ii. Combine cartesian product with subsequent selection whose predicate represents join condition into a join operation.
iii. Use associativity of binary operations to rearrange leaf nodes so leaf nodes with most restrictive selection operations executed first.
iv. Perform projection as early as possible.
v. Compute common expression once.

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

ANS:

i. Success of estimation depends on amount and currency of statistical information DBMS holds.
ii. Keeping statistics current can be problematic.
iii. If statistics updated every time tuple is changed, this would impact the performance.
iv. DBMS could update statistics on a periodic basis.

For each base relation R

- nTuples(R) – the number of tuples (records) in relation R (that is, its cardinality).
- n bFactor(R) – the blocking factor of R (that is, the number of tuples of R that 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) = [nTuples(R)/bFactor(R)]

We use [x] to indicate that the result of the calculation is rounded to the smallest integer that is greater than or equal to x.

For each attribute A of base relation R

- nDistinctA(R) – the number of distinct values that appear for attribute A in relation R.
- minA(R), maxA(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. This is the average number of tuples that satisfy an equality condition on attribute A.

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

ANS:

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

Pipelining: Sometimes known as stream-based processing or on the fly processing. And is an alternative approach is to pipeline the results of one operation to another operation without creating a temporary relation to hold the intermediate result. Clearly, if we can use pipelining, we can save on the cost of creating temporary relations and reading the results back in again.