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: Following are the objectives of query processing:

1. Query optimization
2. Minimize resource usage by reducing total execution time of the query.
3. Maximizing the number of parallel operations.

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

ANS: The typical phases of query processing:

1. Query Decomposition
2. Query Optimization
3. Code Generation
4. Execution

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

ANS: The heuristics that should be applied to improve the processing of query are:

1. Perform projection operation.
2. Use associativity of binary operations to arrange leaf node nodes.
3. Combine the cartesian product with subsequent selection with join operation.
4. Selection process makes possible.
5. Managing common expression

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

ANS: The types of statistics should a DBMS is expected to hold.

- nTuples(R) -> The no of tuples in relation.
- bFactor(R) -> The blocking factor of Relation that fit into one block.
- nBlocks(R) -> The number of blocks required to store Relation.
- $nDistinct_A(R)$ -> The number of distinct values that appear for attribute A in Relation.
- $min_A(R)$,$max_A(R)$ -> The minimum and maximum possible values for the attribute A in R.

- $SC_A(R)$ -> 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: The result of intermediate relational algebra operations is written temporarily to disk. This process is known as materialization. The output of one operation is stored in a temporary relation for a processing by the next operation. Further aspect that is sometimes used to improve the performance of queries that is called pipelining. Already known as stream-based processing or on the fly processing. An alternative approach is to pipeline the results of one operation to another operation without creating a temporary relation to hold the intermediate result. By using pipelining, we can save on the cost of creating temporary relations and reading the results back in again.

When code allows input tuples to be processed as they received, pipelining is used, if this code process the same input tuples more than once, materialization is used.