PHYSICS PRACTICAL SHEETS

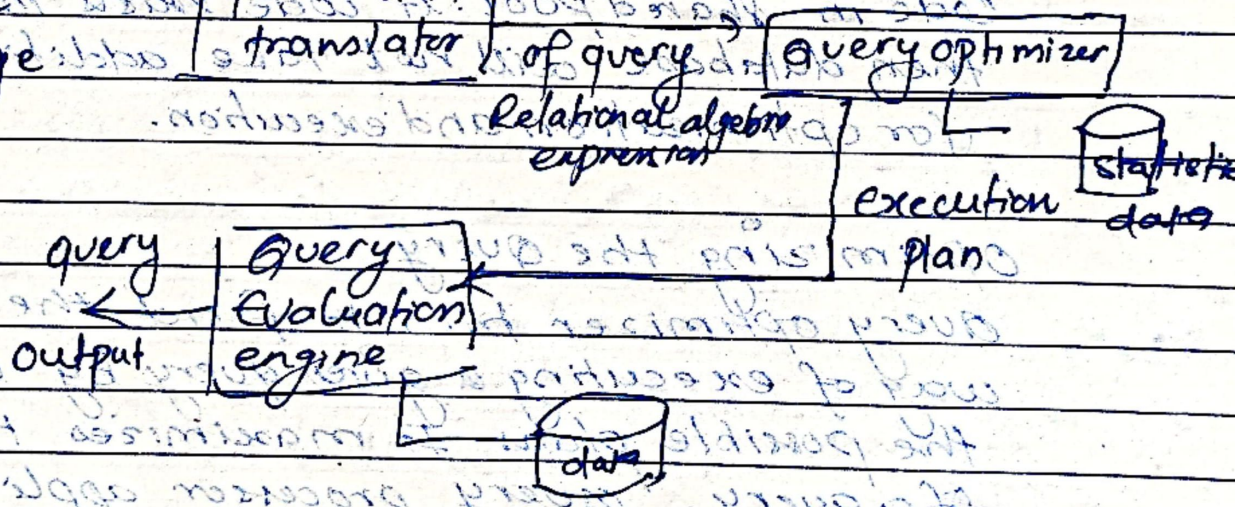
Date: ..... CAMPUS
Class: ..... Purusottam Adhikari
Roll No.: ..... chapter - 3
Shift: .....
Object of the Experiment (Block Letter)

Experiment No.: .....
Group: .....
Sub.: .....
Set: .....

1. What is query processing? explain its steps.

The aims of query processing are to transform a query written in high level language typically SQL, into a correct and efficient execution strategy expressed in a low level language and to execute the strategy to retrieve the required data.

Query processing is the activities involved in parsing, validating, optimizing and executing a query.



Parsing and Translating the query:
The main work of a query processor is to convert a query string into query objects ie. converting the query submitted by the user, into a form understood by the query processing engine. It converts the search string into definite instructions. During parse call, the database performs the following checks - syntax

semantic check and shared pool check, after converting the query into relational algebra.

eg: Syntax check - concludes SQL syntactic validity
eg: SECECT * FORM employee
Here error of spelling.

Semantic check: determines whether the statement is meaningful or not. eg: query contains a tablenam which doesnot exist it checked by this check.

Shared pool check:
Every query possess a hash code during its executio so this check determines existence of curiftten hash code in shared pool. If code exists in shared pool then database will not take additional steps for optimization and execution.

Optimizing the query
Query optimizer tries to find the most efficient way of executing a given query by considering the possible plans. It maximizes the performance of a query. Query processor applies rules to the internal data structures of the query to transfor these structures ~~of the query to transform these structure~~ into their equivalent but more effici representations. Rules may be based on various mathematical models and heuristics.
Database catelog stores the executionplans and then optimizer parses the lower cost plan for execution.

Evaluating the query

Final step in the query processing: An evaluation plan tells precisely the algorithm for each operation along with the coordination among the operations. The best evaluation plan that a user generates by optimization engine is selected and then executed. Evaluation plan comprises of a relational algebra tree, providing information at each node along with the implementation methods to be employed for each relational operator.

eg: SELECT Stu-name, stu-addren FROM STUDENT WHERE age≤ 25

This query can be translated into either of the following relational algebra expression.

- $\sigma_{age<26}$ ($\Pi_{stu\text{-}name, stu\text{-}address}$(student))
- $\Pi_{stu\text{-}name, stu\text{-}address}$ ($\sigma_{age<26}$ (student))

Once the evaluation/execution plan is selected, it is passed into the DBMS, query execution engine where the plan is executed and the results are returned.

2. How can you measure the cost of query?

Cost of query is the time taken by the query to hit the database and return the result. It involves query processing time ie. time taken to parse and translate the query, optimize it, evaluate and return the result to the user is called cost of query. Executing the optimized query involves hitting the primary and secondary memory based on the file organization and the indexes used, time taken to

retrieve the data may vary. query cost considered the no of different resources that are listed below:

- No of disk accesses / no of block transfer / size of table
- Time taken by cpu for executing the query.

The time taken by cpu is negligible in most system when compared with the no of disk access. If we consider the no of blocks transfer as main component in calculating the cost of query, it would include more subcomponents. these are:

- Rotational latency: time taken to bring and spin the required data under r/w head of disk
- Seek time: time taken to position the read-write head over the required track or cylinder
- Sequential I/O: reading data that are stored in contiguous block of disk
- Random I/O: reading data that are stored in different blocks that are not contiguous.

Query cost = $b \times t_T + S \times t_S$

where b - block transfer
S - seeks
t_T - time to transfer one block
t_S - time for one seek.

if $t_T = 0.1$ ms, $t_S = 4$ ms, the block size is 4kB and transfer rate is 40 MB per second.

3. What is query optimization? Discuss its optimization strategies.

The function of query optimization engine is to find an evaluation plan that reduces the overall execution cost of a query.

Query optimization strategies for lowering the execution time of queries include-

- Cost-based optimization
- Heuristic-based optimization
- Semantic-based optimization

Cost-based optimization

Cost-based optimization selects a lower cost plan / mechanism. It is based on cost of query. This method mainly depends on the statistics like Record size, no of records, no of records per block, no of blocks, table size, size of columns etc.

Some features are:-

- Based on cost of query that to be optimized
- Query can use a lot of paths based on the value of indexes available sorting methods, constraints etc
- The aim of query optimization is to choose the most efficient path of implementation the query at the possible lowest minimum cost in the form of an algorithm.
- cost of an algo also depends upon the cardinality of input

Heuristics based optimization

Heuristic optimization transforms the query-tree by using a set of rules that typically improves execution performances.

Some common rules are
- perform selection early (reduces the no. of tuples)
- perform projection early (reduces the number of attributes)
- perform most restrictive selection and join operations (i.e. smallest result size) before other similar operations.

~~Semantic~~ Semantic based optimization
- uses constraints specified on the database schema such as unique attributes and other more complex constraints - in order to modify one query into another query that is more efficient to execute

eg. SELECT e.lname, m.lname
FROM employee e, employee m
WHERE e.superssn = m.ssn and
e.salary > m.salary;

This query retrives the name of the employees who more than their supervisors. Semantic query optimizer checks for the existence of this constraint, it doesnot need to execute the query at all.

PHYSICS PRACTICAL SHEETS

..... CAMPUS

Date: .....
Class: .....
Roll No.: .....
Shift: .....
Object of the Experiment (Block Letter)

Experiment No.: .....
Group: .....
Sub.: .....
Set: .....

VARIETY PRODUCT

4. Differentiate heuristic and semantic based optimization.

heuristic	semantic
i) based on set of rules	i) based on constraints specified on database schema
ii) execute each and every rules	ii) doesnot need execute all query
iii) It performs selection early and projection early	iii) It checks the existence of constraint
iv) defined as transform the query tree by using set of rules to improve execution performance	iv) It uses unique attribute and other complex constraints inorder to modify one query to another that is more efficient to execute
v) It uses query tree and single query tree at last	v) It uses semantic constraint check and meta data.

5. Differentiate cost and semantic based optimization

cost based	semantic based
i) It is based on cost of query	i) It is based on semantic constraints
ii) physical approach	ii) semantic check approach
iii) based uses lots of paths based on value of indexes available existing methods.	iii) It checks whether the query results null or not
iv) cost of an algo depends upon cardinality of input	iv) searching through many constraint to find most applicable to given query.
v) aim is to find lowest minimum cost	v) aim is to find the existence of given query.
vi) based minimize cost of algo	vi) minimizes execution time

6. Differentiate cost and heuristic based optimization

cost	heuristic
i) optimization compares different strategies based on relative cost and selects and execute one that minimizes costs	i) transform the query tree by using set of rules that typically improves performance
ii) cost of algo depends upon cardinality of input	ii) heuristics rules include perform selection early- projection early & restrictive selection and join operations
iii) used to determine the best physical query plan	iii) pick best logical query plan

iv) No rule estimate is made	iv) No cost estimate is made
v) Expensive than heuristics	v) less expensive than cost

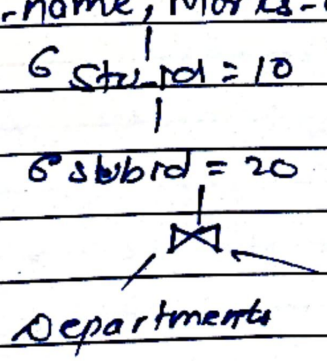
7. How can you express the query to optimized tree transformation.

After a high-level query has been parsed into an equivalent relational algebra expression, the query optimizer can perform heuristic rules on the expression and tree to transform the expression and tree into equivalent but optimized forms.

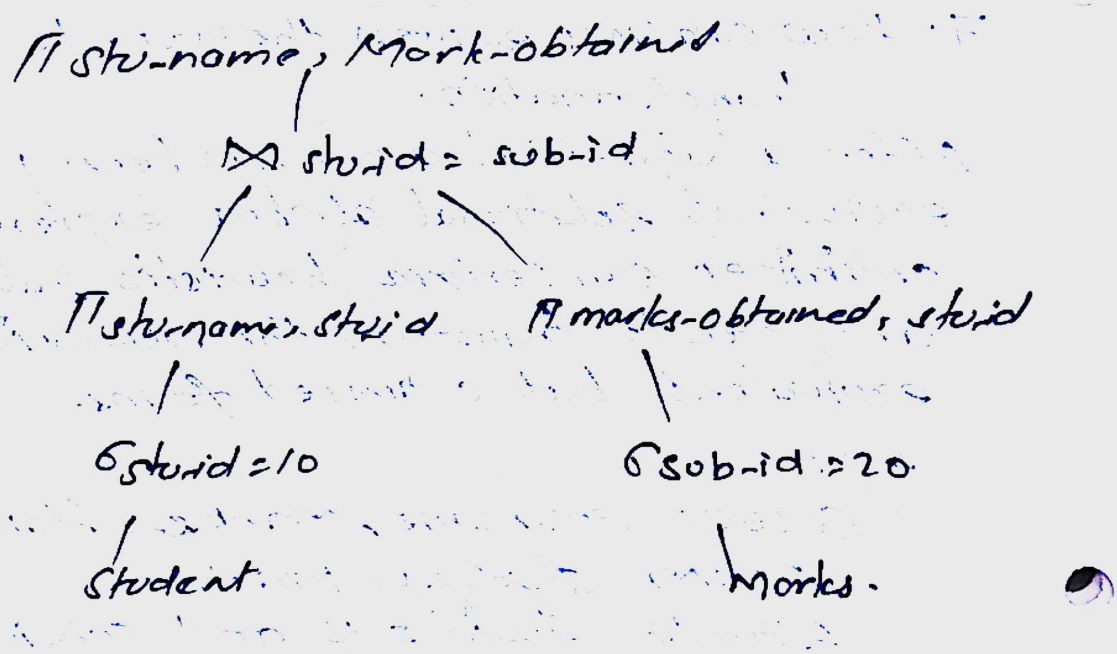
eg:

```
SELECT stu_name, Marks_obtained
FROM student, Marks
WHERE Stu_id = 10 and sub_id = 20;
```

Π stu_name, Marks_obtained (σ stu_id = 10 (σ sub_id = 20 (student $\bowtie$ Marks))



Suppose the student and marks relations both have 100 records each and the no of stu-id = 10 is 50. Note that cartesian product resulting in 10,000 records can be reduced by 80% if $\sigma_{stu\text{-}id = 10}$ operation is performed first. We can also combine the $\sigma_{stu\text{-}id = 10}$ and cartesian product operations into a more efficient join operation as well as eliminating any unneeded columns before the expensive join is performed



8. How can you choose the evaluation plans.

The important phase while evaluating a query is deciding which evaluation plan has to be selected so that it can be tranversed efficiently. It collects all the statistics, costs, evaluation paths, relational tree etc. It then analyses them and chooses the evaluation plan/path.

If some query is written in different forms of relational algebra corresponding trees for them too is drawn using heuristic.
statistics for them based on cost based evaluation and heuristic methods are collected. It checks the cost based on different methods. It checks for the operator, joining type, indexes, no of records, from data dictionary. once all these information are collected, it picks the best evaluation plan.


support me at
eSewa™
9810867824