Database Systems (CSF212)

BITS Pilani
Hyderabad Campus

Prof.R.Gururaj
CS&IS Dept.

SQL Query Optimization (T1-Ch.18)

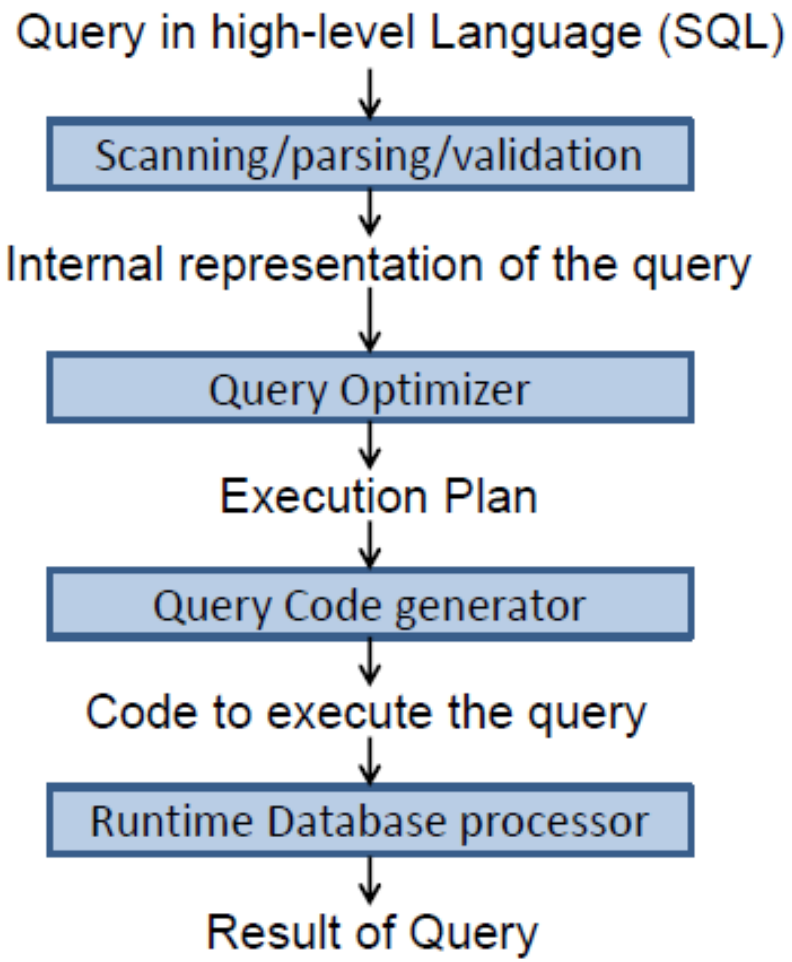
innovate achieve lead

- *Introduction to query execution*
- *SQL to Relational algebra*
- *Query tree Heuristics in Query optimization*

Steps in SQL Query execution

Here we discuss the techniques used by a DBMS to process, optimize, and execute high-level queries. The process involves following steps:

- Scanning - identifying the language tokens
- Parsing- Syntax checking
- Validation – checking the attributes and relations mentioned in the query valid or not.
- Generate query tree representation of the query
- Devise an execution strategy- to retrieve data from the files
- Query optimization- Choosing a suitable strategy from the available set of strategies for the query, and generate execution plan.
- Code generation – generating code for the plan
- Execution



Code can be executed
Directly (interpreted)
Or
Stored and executed Later

Translating SQL queries into Relational Algebra

An SQL query is first translated into an equivalent relational algebra expression- represented as a query tree data structure.

SQL queries are decomposed into query blocks.

A query block forms the basic unit that can be translated into algebraic operators and optimized.

A query block contains single SELECT-FROM-WHERE clause with optional GROUP BY and Having clauses.

Hence nested queries within a query are identified as separate query blocks.

SQL Query tree

```
SELECT Lname, Fname
FROM EMPLOYEE
WHERE Salary > (SELECT MAX(Salary) FROM Employee WHERE Dno=5);
```

Inner block is:

```
(SELECT MAX(Salary)
FROM Employee
 WHERE Dno=5);   // Let  'c' represents the result returned from the inner block.
```

$£_{\text{MAX Salary}}(\sigma_{\text{Dno=5}}(\text{EMPLOYEE}))$

The outer block is:

```
SELECT Lname, Fname
FROM EMPLOYEE
WHERE Salary > c ;
```

$\pi_{\text{Lname, Fname}}(\sigma_{\text{Salary>c}}(\text{EMPLOYEE}))$

The query optimizer would then choose an execution plan for each block.

Query Optimization

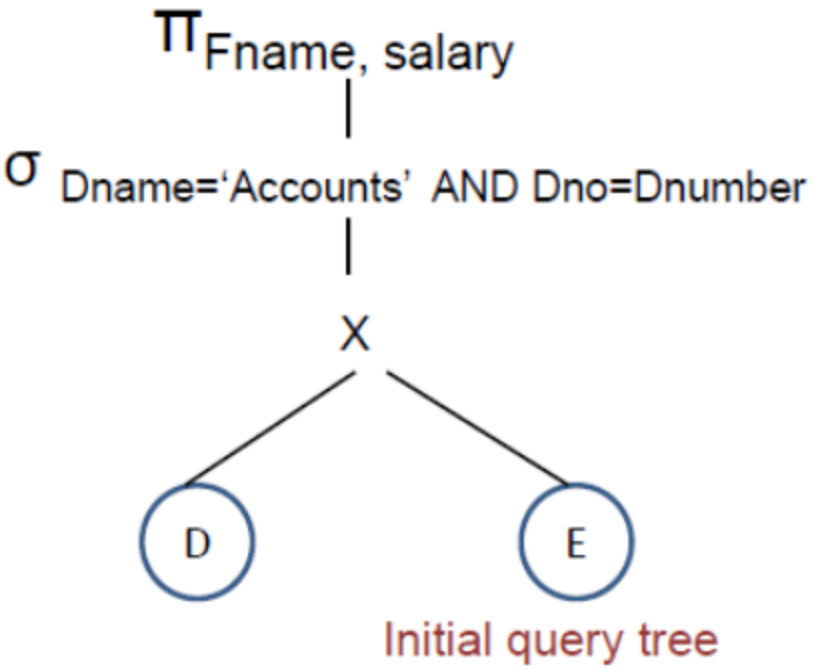
A *Query Tree* is a tree data structure that corresponds to a relational algebra expression.
It represents the input relations of the query as leaf nodes of the tree, and represents the relational algebra query as internal nodes.

The execution of the query tree consists of executing an internal node operation whenever its operands are available and then replacing the internal node by the relation that results from executing the operation.

The execution terminates when the root node is executed and produces the result relation for the query.

```
SELECT Fname , salary
FROM EMPLOYEE, DEPARTMENT
WHERE Dname='Accounts' and Dno=Dnumber;
```

$\pi_{Fname,\ salary}\ (EMPLOYEE \bowtie_{Dno=Dnumber} (\sigma_{Dname='Accounts'}\ (Department)))$



Initial query tree

A query parser will generate a standard initial query tree that corresponds to the SQL query, without doing any optimization.

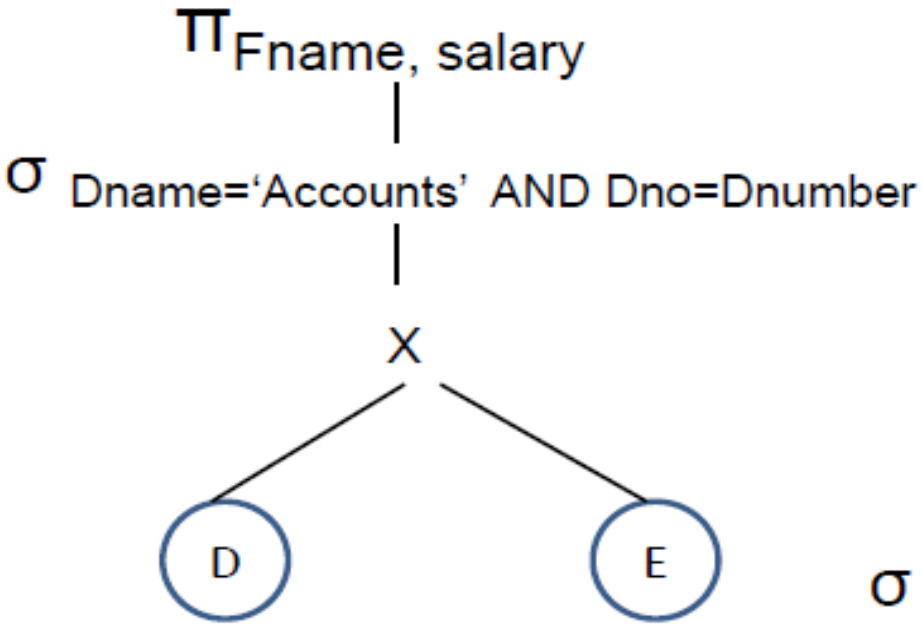
These initial trees are inefficient if executed directly.

Now it is the job of the heuristic query optimizer to transform this initial query tree into a final query tree that is efficient to execute.

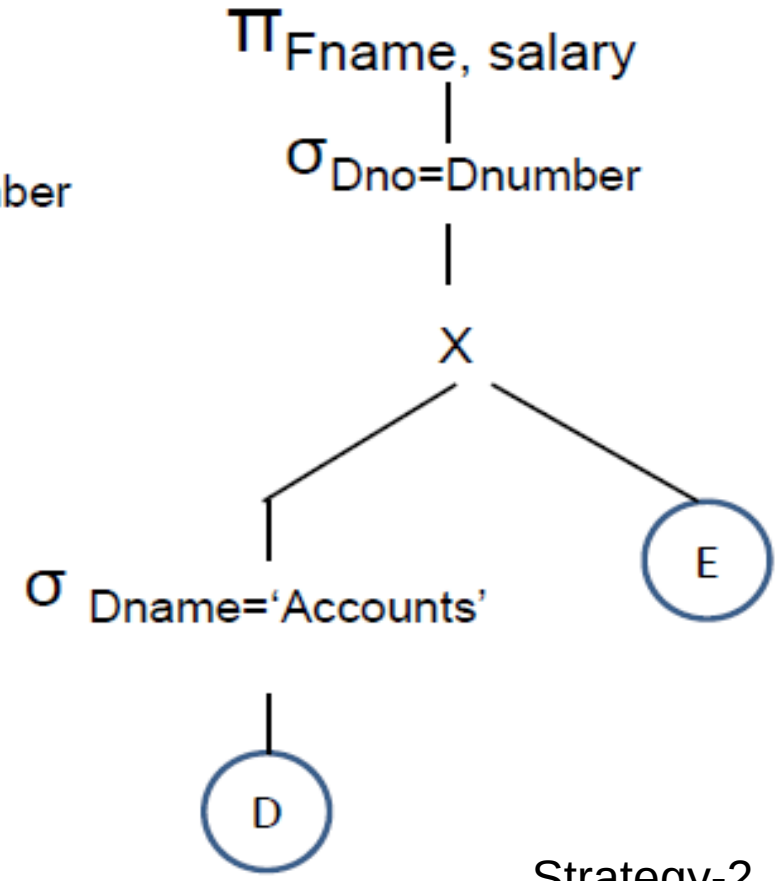
The optimizer must include rules for equivalence among relational algebra expressions that can be applied to initial tree.

Heuristic query optimization rules then utilize these equivalence expressions to transform the initial tree into the final, optimized query tree.

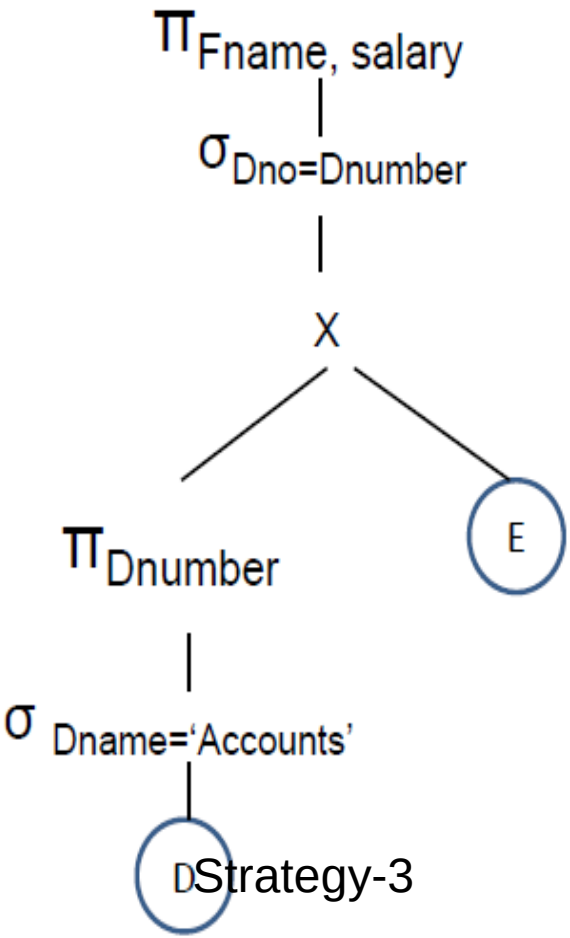
Query Graph



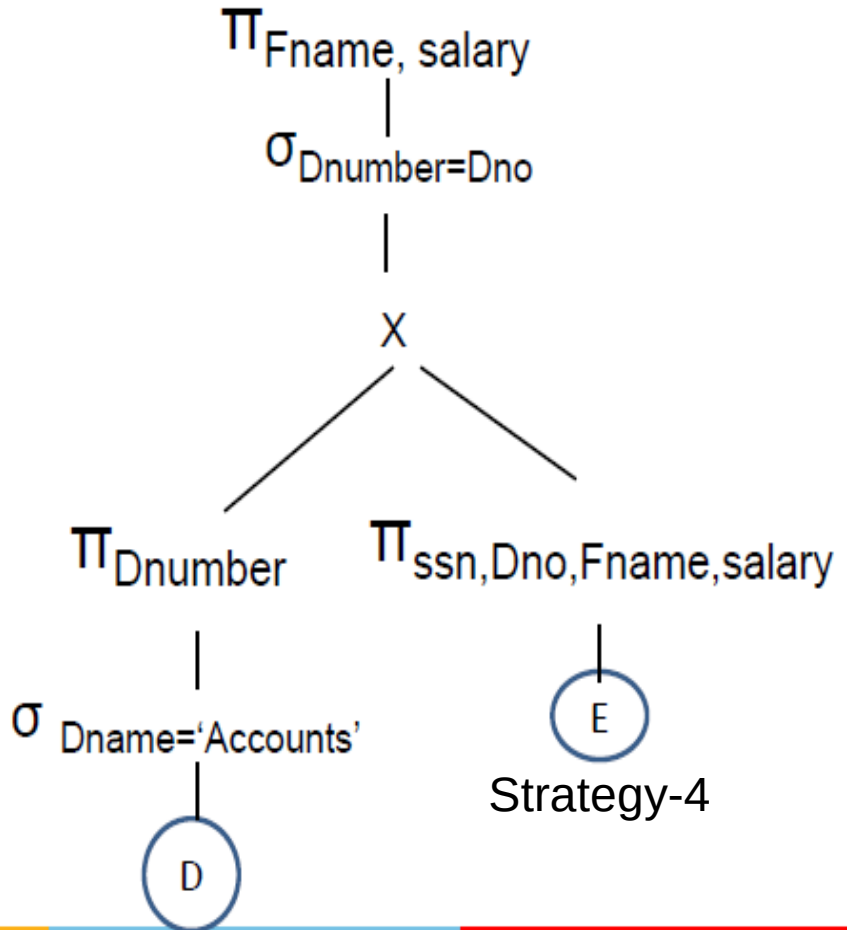
Strategy-1



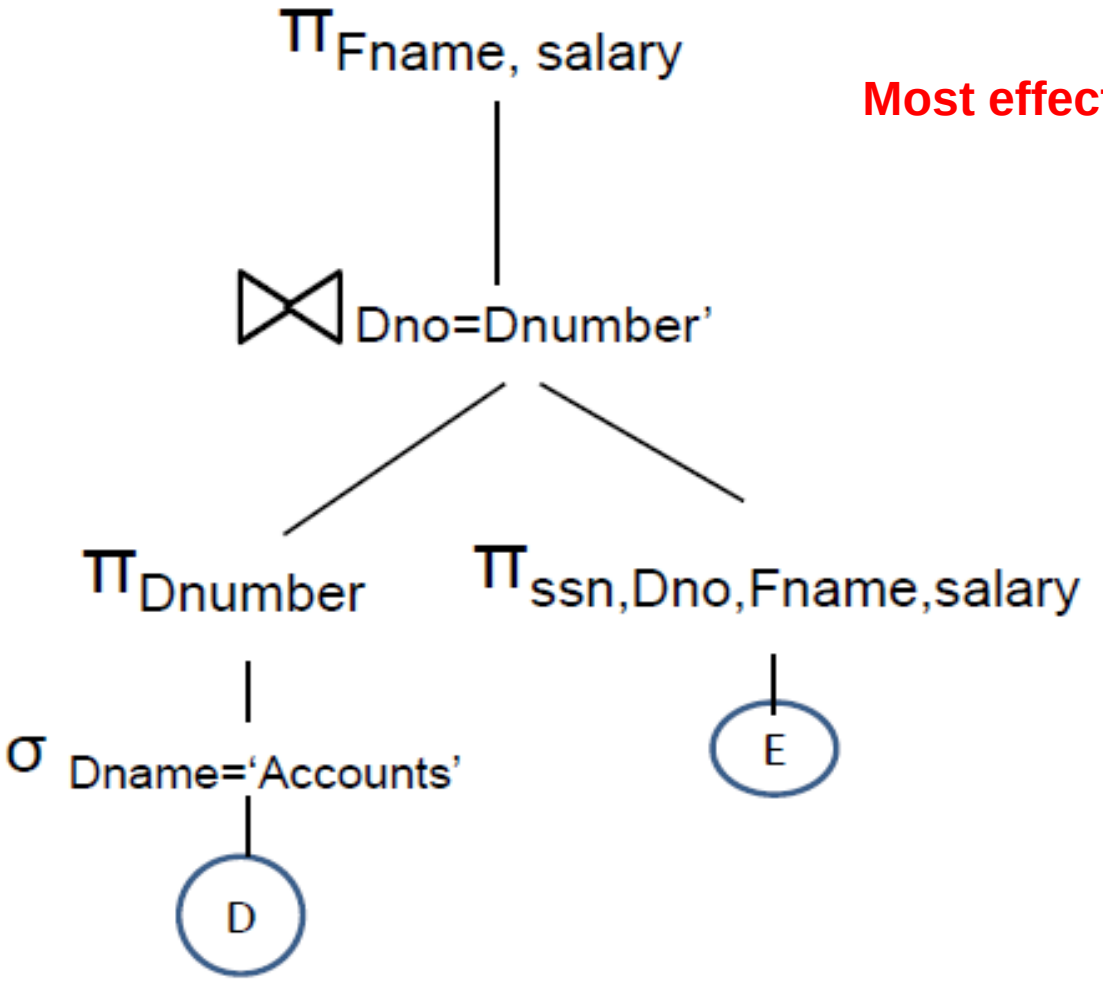
Strategy-2



Strategy-3



Strategy-4



Most effective strategy

Strategy-5

Outlines of Heuristic Optimization Algorithm

1) Break up conjunctive SELECT operations (helps in moving SELECT operations down in different branches).
2) Move each SELECT as far down the query tree as possible
3) More restrictive SELECT condition need to be executed first
4) Combine a CARTESIAN PRODUCT with a subsequent SELECT in the tree into a JOIN , if the condition represents a JOIN.
5) Break up PROJECT operations and move them as far down as possible in different branches.

Example

Consider the following SQL query on COMPANY database, and transform the query into equivalent relational algebraic expression. Draw the initial query tree for the relational algebraic expression and also derive the optimized query tree after applying standard heuristics on it.

```
SELECT E.eid, E.name, D.dname
FROM EMP as E, DEPT as  D
WHERE E.dno=D.dnum  AND  E.age>40   AND
D.location='Mumbai';
```

Summary

- ✓ Introduction to query execution
- ✓ SQL to Relational algebra
- ✓ Query tree Heuristics in Query optimization