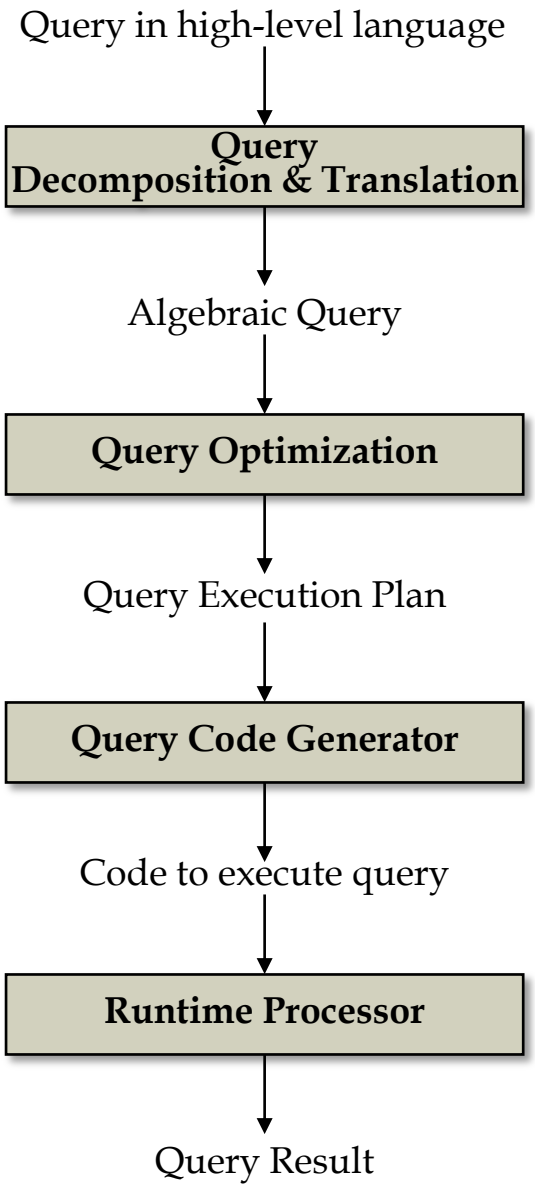
QUERY PROCESSING & OPTIMIZATION

CHAPTER 19 (6/E)

CHAPTER 15 (5/E)

QUERY PROCESSING METHODOLOGY



- generate alternative procedures for executing the query (query execution plans)
- select an efficient query execution plan

SQL

- check SQL syntax
- check existence of relations, attributes
- replace views by their definitions
- transform query into an internal form

QUERY OPTIMIZATION ISSUES

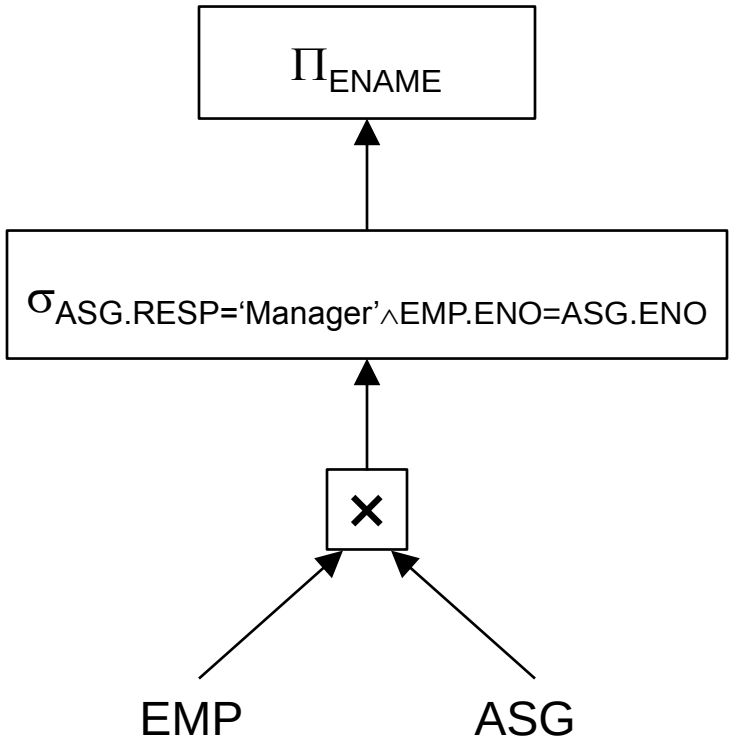
- Determining the “shape” of the query execution plan
  - Order of execution
  - Use transformation (equivalence) rules
    - e.g., $(R \bowtie S) \bowtie T \Leftrightarrow R \bowtie (S \bowtie T)$

- Determining how each “node” in the plan should be executed
  - Operator implementations

- These are interdependent and an optimizer would do both in generating the execution plan

ALTERNATIVE EXECUTION ORDERS

```
SELECT   ENAME
FROM     EMP,ASG
WHERE    EMP.ENO = ASG.ENO
AND      ASG.RESP = "Manager"
```

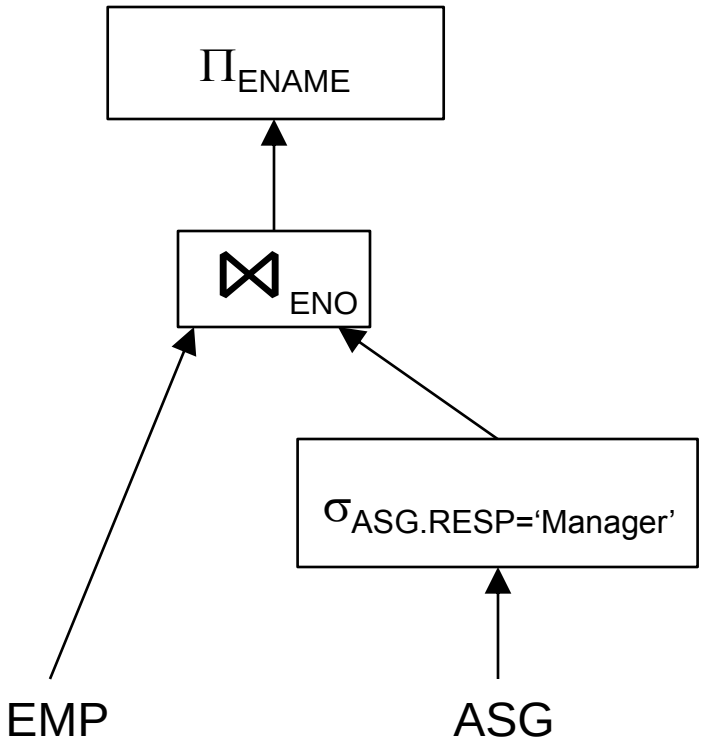


Strategy 1

$$\Pi_{ENAME}(\sigma_{RESP=\text{“Manager”} \wedge EMP.ENO=ASG.ENO}(EMP \times ASG))$$

ALTERNATIVE EXECUTION ORDERS

```
SELECT   ENAME
FROM     EMP,ASG
WHERE    EMP.ENO = ASG.ENO
AND      ASG.RESP = "Manager"
```



Strategy 1

$$\Pi_{ENAME}(\sigma_{RESP=\text{“Manager”}\wedge EMP.ENO=ASG.ENO}(EMP\times ASG))$$

Strategy 2

$$\Pi_{ENAME}(EMP \bowtie_{ENO} (\sigma_{RESP=\text{“Manager”}}(ASG))$$

Strategy 2 avoids Cartesian product, so may be “better”

EXAMPLE – JOIN ALGORITHMS

```
SELECT C.Cnum, A.Balance
FROM Customer C, Accounts A
WHERE C.Cnum = A.Cnum
```

- Nested loops join:

```
for each tuple c in Customer do
   for each tuple a in Accounts do
            if c.Cnum = a.Cnum then
               output c.Cnum,a.Balance
   end
end
```

- Sort-merge join:

  phase 1: sort Customer and Accounts on Cnum
  phase 2: merge the resulting sorted relations

COST OF PLANS

- Alternative execution plans may be compared according to cost
- Cost of a plan is the sum of the costs of its component operations
- Many possible cost metrics
  - However, most metrics reflect the amounts of system resources consumed by the plan
  - System resources may include:
    - disk block I/O's
    - processing time
    - network bandwidth
- Typically, exact cost cannot be determined beforehand
  - Estimation

LECTURE SUMMARY

- Query processing methodology
- Query optimization issues
  - Selecting an execution order for the operators
  - Selecting an algorithm for each operator
  - Choosing an efficient plan based on (estimated) costs