

# OPTIMIZACIÓN BASADA EN COSTOS

# COSTOS DE EJECUCIÓN

- Es la cantidad de procesamiento que requiere un DBMS para resolver una consulta.
- Generalmente medida en cantidad de accesos a memoria secundaria (disco) por el alto costo.
- El costo se puede ver afectado por:
  - Estructuras de almacenamiento: tamaño de páginas, tamaño de tuplas, ordenamientos.
  - Disponibilidad (o no) de índices.
  - Algoritmos de implementación de operaciones.

# PLAN FÍSICO

- Es la asociación de una implementación a cada operación que aparece en un plan lógico.
- Un plan lógico puede originar diferentes planes físicos dado que se pueden considerar diferentes implementaciones para cada operador.
- Es necesario estimar el costo de los diferentes planes que se generan y elegir el de costo mínimo
- La calidad de la estimación del costo y tamaño depende de la calidad de la información estadística almacenada en el diccionario de datos del DBMS.

# ESTIMACIÓN – PLAN FÍSICO

- A cada operador del plan lógico se le asigna una implementación
- Luego hay que estimar el costo de todo el plan basándose en los costos de cada algoritmo.
- Es importante conocer la estrategia de evaluación, esto es conocer como fluyen los datos de una operación a la otra

# ESTRATEGIAS DE EVALUACIÓN

- PIPELINED (en secuencia)
  - Operadores que se ejecutan en secuencia y pueden pasarse los resultados a medida que los generan
  - No necesita grabar los resultados intermedios
- No PIPELINED (materialización)
  - Los operadores se ejecutan de forma secuencial y es necesario grabar los resultados intermedios
  - La selección y los Join no se implementan por pipelining

# IMPLEMENT. DE OPERADORES

- Búsqueda lineal (full scan): lee cada registro y si cumple la condición lo agrega al resultado
- Búsqueda binaria: Costo balanceado se aplica únicamente a registros ordenados.
- Loops anidados (neested loop): es para implementar los joins; se aplica a registros o a bloques. Combina cada elemento de un conjunto con todos los pares de este en el segundo conjunto
  - Nota: no es la lista completa!

# ESTIMACIÓN DE TAMAÑO

- Se trata de estimar al realizar una determinada operación:
  - La cantidad de tuplas que se obtendrán
  - El tamaño (en bytes) de la tupla resultante
  - La cant. de bloques necesarios para almacenar el resultado
  - Para el producto cartesiano:

- Cantidad de tuplas

$$n_{R \times S} = n_R \times n_S$$

- Cada tupla ocupa

$$t_{R \times S} = t_R + t_S \text{ bytes}$$

# COSTO EN SUBCONSULTAS

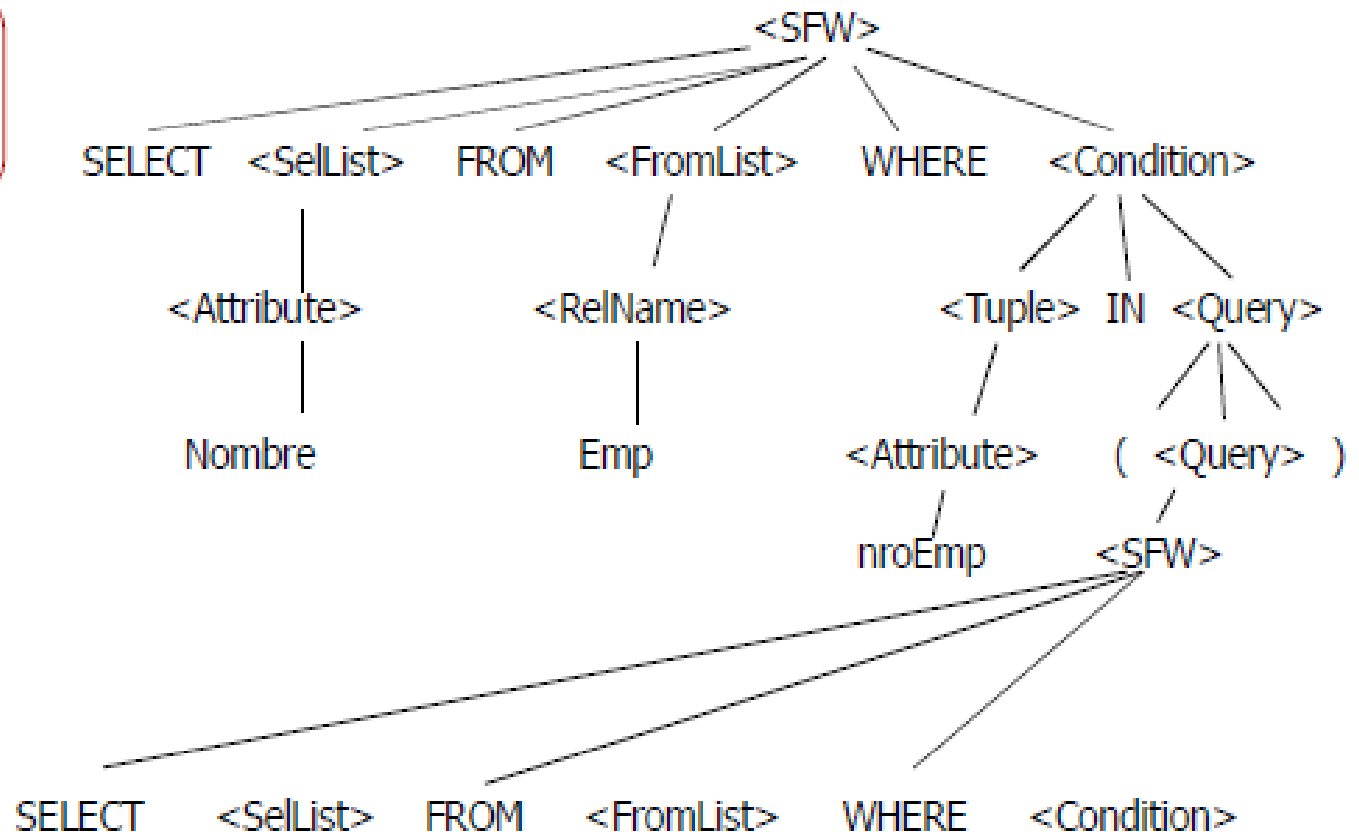
- El costo depende del tipo de la sub-query
- Correlacionada: Es evaluada para cada tupla de la consulta maestra. No se almacenan a disco los resultados temporales de la sub-query, habiendo solamente costos de lectura.
- No correlacionada: El resultado de la sub-query se computa una única vez. Pueden estar asociados a costos de lectura si el resultado no puede almacenarse en memoria. Si el resultado intermedio esta en disco se lee para cada registro de la consulta maestra.



# NO CORRELACIONADA

## Consulta NO Correlacionada

Select E.Nombre  
From Emp E  
Where E.nro IN  
(  
Select nroEmp  
From Trabaja  
Where nroPrj=5  
)



# CORRELACIONADA

Select E.Nombre  
From Emp E  
Where Exists  
(Select nroEmp  
From Trabaja T  
Where T.nroEmp=E.nro  
And nroPrj=5 )

**Consulta  
Correlacionada**

