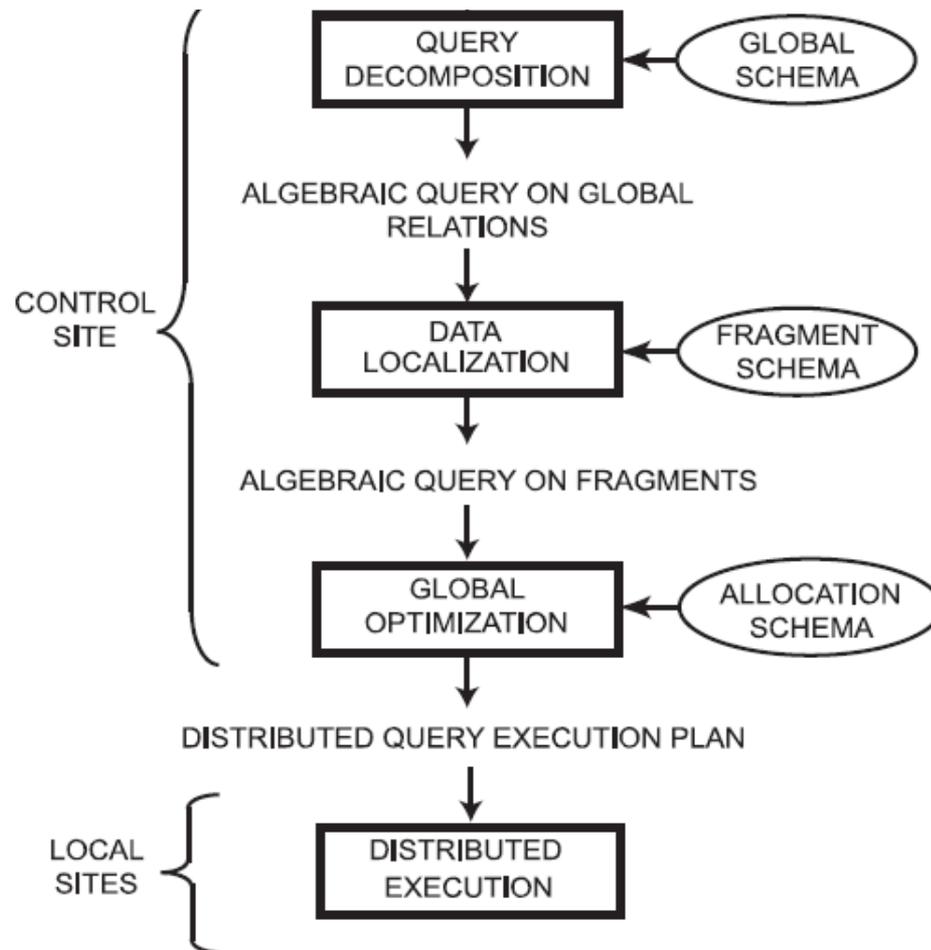


Base de datos distribuidas

Prof. Alejandro Reyes Ortiz

“DESCOMPOSICIÓN DE CONSULTAS”

Capas del procesamiento de consultas distribuidas



Descomposición de la consulta

- Primer fase del proceso del proceso de consultas que obtiene una consulta en algebra relacional.
- La salida refiere o hace uso de las relaciones globales en el escenario distribuido.
- Sus etapas son: normalización, análisis, eliminación de redundancia y reescritura

Normalización

- Su meta es la transformación de una consulta en su forma normalizada para facilitar el futuro procesamiento.
- En el lenguaje relacional SQL, la transformación mas importante está en la cláusula WHERE que puede ser compleja.
- Esta cláusula puede tener dos formas normales posibles:
 - Precedencia AND (\wedge)
 - Precedencia OR (\vee)

Normalización

- La forma normal conjuntiva es una conjunción (predicados \wedge) de disyunciones (predicados \vee).

$$(p_{11} \vee p_{12} \vee \cdots \vee p_{1n}) \wedge \cdots \wedge (p_{m1} \vee p_{m2} \vee \cdots \vee p_{mn})$$

- **La forma normal disyuntiva es:**

$$(p_{11} \wedge p_{12} \wedge \cdots \wedge p_{1n}) \vee \cdots \vee (p_{m1} \wedge p_{m2} \wedge \cdots \wedge p_{mn})$$

Normalización

- Su transformación es directa usando las reglas de equivalencia bien conocidas para las operaciones lógicas (\wedge , \vee , \neg):

1. $p_1 \wedge p_2 \Leftrightarrow p_2 \wedge p_1$
2. $p_1 \vee p_2 \Leftrightarrow p_2 \vee p_1$
3. $p_1 \wedge (p_2 \wedge p_3) \Leftrightarrow (p_1 \wedge p_2) \wedge p_3$
4. $p_1 \vee (p_2 \vee p_3) \Leftrightarrow (p_1 \vee p_2) \vee p_3$
5. $p_1 \wedge (p_2 \vee p_3) \Leftrightarrow (p_1 \wedge p_2) \vee (p_1 \wedge p_3)$
6. $p_1 \vee (p_2 \wedge p_3) \Leftrightarrow (p_1 \vee p_2) \wedge (p_1 \vee p_3)$
7. $\neg(p_1 \wedge p_2) \Leftrightarrow \neg p_1 \vee \neg p_2$
8. $\neg(p_1 \vee p_2) \Leftrightarrow \neg p_1 \wedge \neg p_2$
9. $\neg(\neg p) \Leftrightarrow p$

Normalización

- Ejemplo

“Encontrar los nombres de los empleados que han estado trabajando con el proyecto P1 por 12 o 24 meses”

```
SELECT ENAME
FROM EMP, ASG
WHERE EMP.ENO = ASG.ENO
AND ASG.PNO = "P1"
AND DUR = 12 OR DUR = 24
```

forma normal disyuntiva

$(EMP.ENO = ASG.ENO \wedge ASG.PNO = "P1" \wedge DUR = 12) \vee$
 $(EMP.ENO = ASG.ENO \wedge ASG.PNO = "P1" \wedge DUR = 24)$

forma normal conjuntiva

$EMP.ENO = ASG.ENO \wedge ASG.PNO = "P1" \wedge (DUR = 12 \vee DUR = 24)$

Análisis

- El análisis de consultas permite el rechazo de consultas normalizadas para las cuales el procesamiento adicional es **imposible** o **innecesario**.
- Las razones principales para el rechazo son que la consulta es **incorrecta en los tipos** o **semánticamente incorrecta**.
- Cuando se detecta uno de estos casos, la consulta **se devuelve** simplemente al usuario con una explicación.

Análisis

- Consulta de tipos de datos incorrectos.
 - Una consulta es de tipo incorrecto si alguno de sus nombres de atributos o relaciones no están definidos en el esquema global o si las operaciones están siendo aplicadas sobre atributos del tipo incorrecto.
 - La técnica utilizada para detectar consultas incorrectas de tipo es similar a la comprobación de tipos: para lenguajes de programación.

```
SELECT E#  
FROM EMP  
WHERE ENAME > 200
```

La siguiente consulta SQL es de tipo incorrecto por dos razones:

- El atributo E# no se declara en el esquema.
- La operación "> 200" es incompatible con la cadena de tipo de ENAME.

Análisis

- Consulta semánticamente incorrecta.
 - Una consulta es semánticamente incorrecta si sus componentes no contribuyen en modo alguno a la generación del resultado.
 - El cálculo relacional, no es suficiente para determinar la correctitud semántica de las consultas. Sin embargo, se utiliza la representación de la consulta como un grafo, llamado **grafo de consulta** o **grafo de conexión**.
- Una consulta será semánticamente incorrecta si su grafo de consulta no está conectado. Es decir, uno o más subgrafos (correspondientes a subconsultas) están desconectados del grafo que contiene el resultados.

Análisis

- ¿Cómo construir el grafo de la consulta?
 - Se define para los tipos de consultas más utilizados que implican: **proyección (select)**, **selección (where)** y **join**.
 - Un nodo representa el resultado y los demás nodos representan las relaciones involucradas.
 - Una arista entre dos nodos (no del nodo de resultado) corresponde al **join**.
 - Una arista cuyo nodo destino es el resultado, representa una **proyección (select)**.
 - Una arista cuyo origen y destino es el mismo nodo, representa una **selección (where)**

Análisis

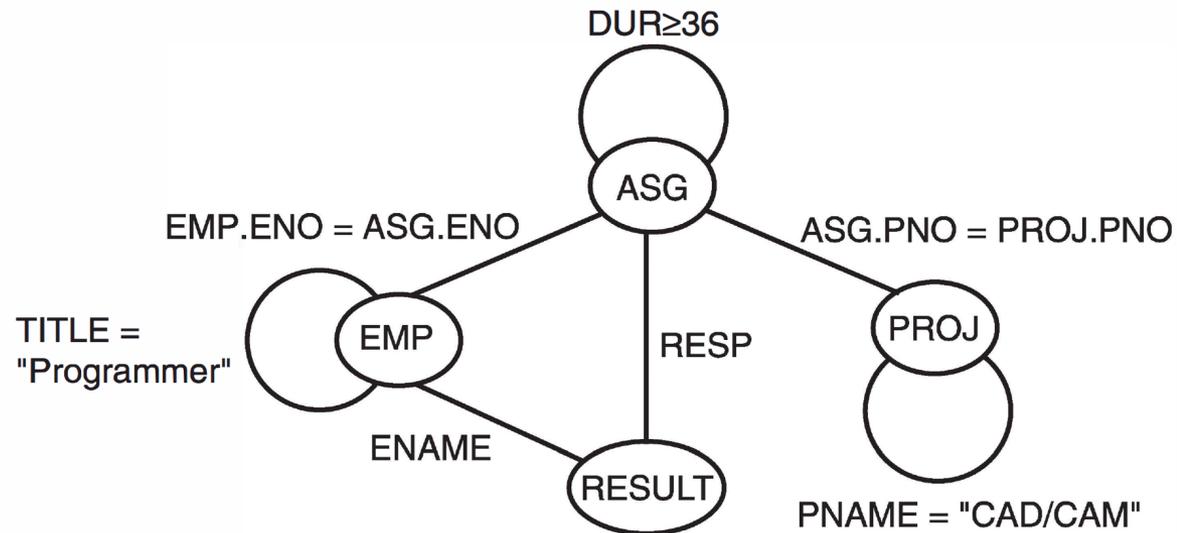
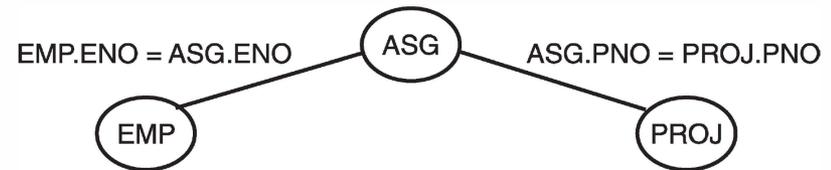
“Find the names and responsibilities of programmers who have been working on the CAD/CAM project for more than 3 years.”

“Encontrar los nombres y responsabilidades de los programadores que han trabajado sobre proyectos CAD/DAM por más de tres años”

```
SELECT ENAME, RESP
FROM EMP, ASG, PROJ
WHERE EMP.ENO = ASG.ENO
AND ASG.PNO = PROJ.PNO
AND PNAME = "CAD/CAM"
AND DUR ≥ 36
AND TITLE = "Programmer"
```

Análisis

```
SELECT ENAME, RESP
FROM   EMP, ASG, PROJ
WHERE  EMP.ENO = ASG.ENO
AND    ASG.PNO = PROJ.PNO
AND    PNAME = "CAD/CAM"
AND    DUR ≥ 36
AND    TITLE = "Programmer"
```

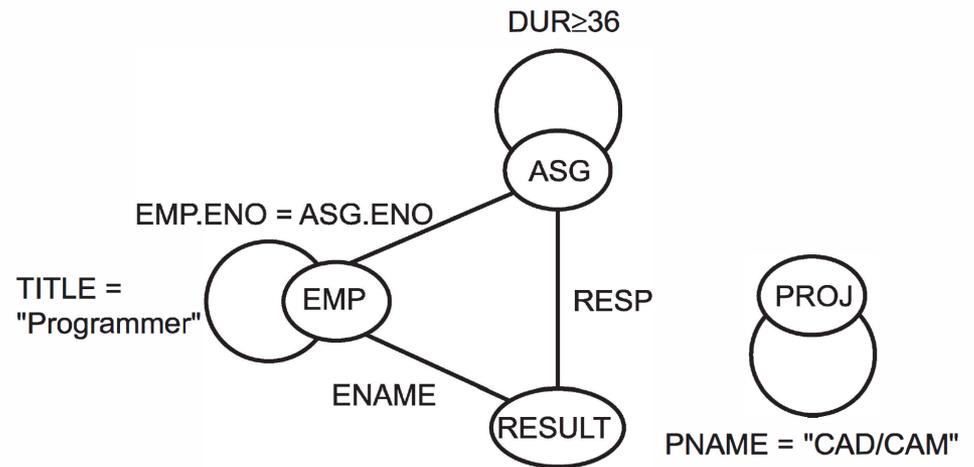


- Como es un grafo conectado, por lo tanto es una consulta semánticamente correcta.

Análisis

“Encontrar los nombres y responsabilidades de los programadores que han trabajado sobre proyectos CAD/DAM por más de tres años”

```
SELECT ENAME, RESP
FROM   EMP, ASG, PROJ
WHERE  EMP.ENO = ASG.ENO
AND    PNAME = "CAD/CAM"
AND    DUR ≥ 36
AND    TITLE = "Programmer"
```



- Como es un grafo desconectado, es una consulta semánticamente incorrecta.