

Base de datos distribuidas

Prof. Alejandro Reyes Ortiz

“LOCALIZACIÓN DE DATOS”

Localización de datos distribuidos

- Hasta el momento no hemos tomado en cuenta la distribución de los datos (fragmentos).
- Este es el papel de la capa de localización: traduce una consulta hecha sobre relaciones globales a una consulta algebraica expresada en fragmentos.
- La localización utiliza información almacenada en el esquema de fragmentación.

Localización de datos distribuidos

- La fragmentación de una relación se define a través de las reglas de fragmentación.
- Una relación global puede ser reconstruida aplicando las reglas de reconstrucción.
- Un *programa de localización* será utilizado para identificar donde se encuentra los fragmentos y operar sobre ellos.

Programa de localización

- Sustituye la consulta sobre la relación global en una consulta distribuida y localizada.
- Esto puede ser visto como el reemplazo de las hojas del árbol del álgebra relacional (operadores) de la consulta con subárboles que corresponden a fragmentos y su localización.
- Al árbol de consulta obtenido por esta forma se le conoce como una *árbol de consulta distribuida o árbol de consulta localizada*.

Arbol de consulta localizada

- Pasos a seguir para obtener el árbol de consulta localizada:
 - Contar con un esquema de fragmentación (Horizontal, Vertical o Mixta).
 - Iniciar a partir de una consulta, de ahí el nombre de árbol de consulta localizada.
 - Primero es necesario contar, en su caso construir, el árbol de operadores (expresiones) obtenido en la fase de descomposición de la consulta.
 - Finalmente, considerando los fragmentos obtener el árbol de consulta localizada.
 - Reemplazando las hojas del árbol de operadores de la consulta, por subárboles que corresponden a fragmentos.

Arbol de consulta localizada

- Ejemplo: considere el siguiente esquema de fragmentación horizontal.

La relación EMP(ENO, ENAME, TITLE) puede ser dividida en tres fragmentos horizontales E_1 , E_2 y E_3 , definidos como sigue:

$$E_1 = \sigma_{ENO \leq "E3"}(EMP)$$

$$E_2 = \sigma_{"E3" < ENO \leq "E6"}(EMP)$$

$$E_3 = \sigma_{ENO > "E6"}(EMP)$$

El programa de localización conoce que:

$$EMP = E_1 \cup E_2 \cup E_3$$

Arbol de consulta localizada

La relación ASG puede ser dividida en dos fragmentos horizontales G_1 y G_2 definidos como sigue:

$$G_1 = \sigma_{ENO \leq "E3"} (ASG)$$

$$G_2 = \sigma_{ENO > "E3"} (ASG)$$

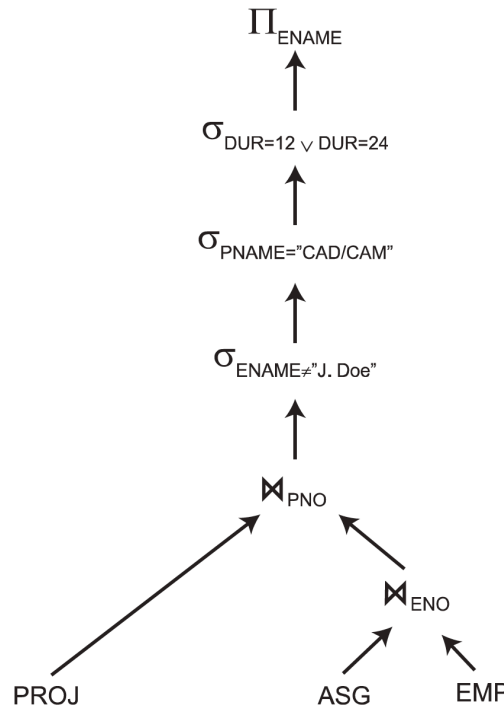
$$ASG = G_1 \cup G_2$$

Arbol de consulta localizada

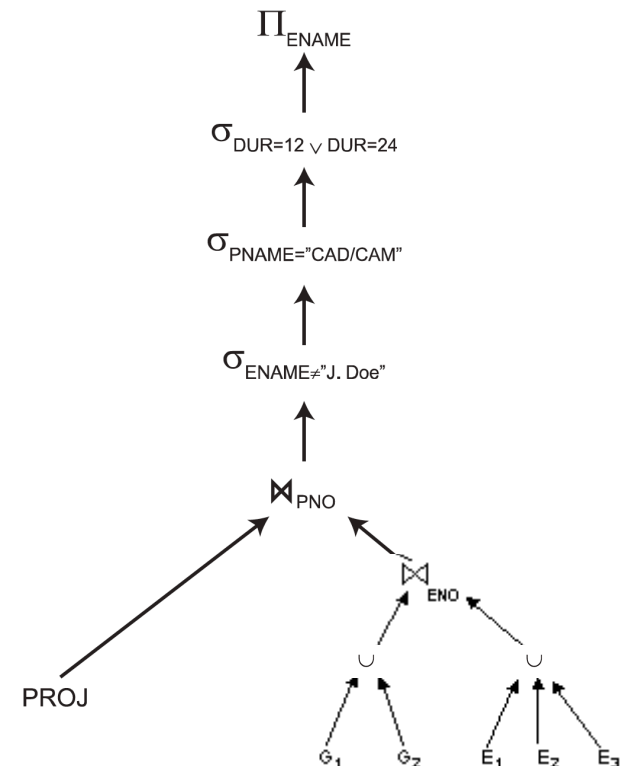
- Considere la siguiente consulta para obtener el *árbol de consulta localizada*.

```

SELECT ENAME
FROM   PROJ, ASG, EMP
WHERE  ASG.ENO = EMP.ENO
AND    ASG.PNO = PROJ.PNO
AND    ENAME != "J. Doe"
AND    PROJ.PNAME = "CAD/CAM"
AND    (DUR = 12 OR DUR = 24)
    
```



árbol de operadores



árbol de consulta localizada

Ejercicio

- Considere que la siguiente consulta ya ha sido analizada, normalizada y se ha eliminado redundancia.
- Obtener el árbol de operadores y el árbol de consulta localizada.

```
SELECT DUR
FROM PROJ, ASG
AND ASG.PNO = PROJ.PNO
AND (ENO = "E4" OR ENO = "E6")
AND PNAME = "CAD/CAM"
```

Programa de localización

- Este árbol de *consulta localizada* puede ser ser ineficiente, dado que en un escenario fragmento (distribuidos), *varias simplificaciones y reestructuraciones* de la consulta genérica aún pueden ser realizadas.
- Es necesario aplicar técnicas de reducción para cada tipo de fragmentación (H, V, Mixta) para generar consultas simples y optimizadas.

Programa de localización

- **Reducción para fragmentación horizontal.**
- **Reducción para fragmentación vertical.**

A partir de estas reducciones se generará el *árbol de consulta reducida*.

Reducción para fragmentación horizontal

- La fragmentación horizontal distribuye una relación basada en predicados de selección.
- En este escenario se aplica la **reduccion por selección**, la cual consiste en eliminar las hojas del árbol de *consulta localizada*, que no intervien en la consulta, **considerando la selección de tuplas**.
 - Esto significa que al quitar dichas hojas (frgamentos) no se verá afectada la **selección de tuplas** en la consulta.

Reducción para fragmentación horizontal

- El mismo esquema de fragmentación horizontal.

La relación EMP(ENO, ENAME, TITLE) puede ser dividida en tres fragmentos horizontales E_1 , E_2 y E_3 , definidos como sigue:

$$E_1 = \sigma_{ENO \leq "E3"} \text{ (EMP)}$$

$$E_2 = \sigma_{"E3" < ENO \leq "E6"} \text{ (EMP)}$$

$$E_3 = \sigma_{ENO > "E6"} \text{ (EMP)}$$

$$EMP = E_1 \cup E_2 \cup E_3$$

La relación ASG puede ser dividida en dos fragmentos horizontales G_1 y G_2 definidos como sigue:

$$G_1 = \sigma_{ENO \leq "E3"} \text{ (ASG)}$$

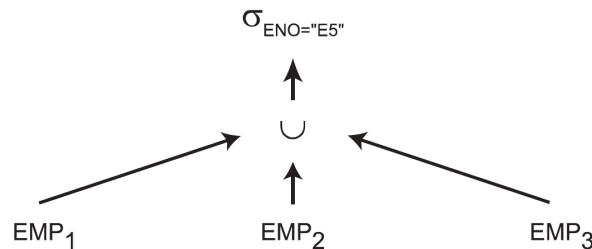
$$G_2 = \sigma_{ENO > "E3"} \text{ (ASG)}$$

$$ASG = G_1 \cup G_2$$

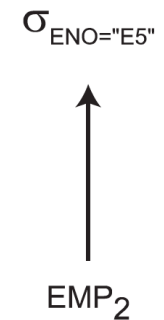
Reducción para fragmentación horizontal

- Ejemplo:
- Al árbol de consulta localizada se le puede aplicar **reducción por selección** para obtener el *árbol de consulta reducida*.

```
SELECT *  
FROM EMP  
WHERE ENO = "E5"
```



árbol de consulta localizada



árbol de consulta reducida

Ejemplo

- Considere el árbol de consulta localizada obtenido de la siguiente consulta.

```
SELECT DUR
FROM PROJ, ASG
AND ASG.PNO = PROJ.PNO
AND (ENO = "E4" OR ENO = "E6")
AND PNAME = "CAD/CAM"
```

¿ Se puede eliminar algún
fragmento con reducción por
selección ?

Reducción para fragmentación vertical

- La fragmentación vertical distribuye una relación de acuerdo a los atributos de proyección.
- En este escenario se aplica la **reduccion por proyección**, la cual consiste en eliminar las hojas del árbol de *consulta localizada*, que no intervien en la consulta, considerando la **proyección de atributos**.
 - Esto significa que al quitar dichas hojas (frgamentos) no se verá afectada la **proyección de atributos** en la consulta.

Reducción para fragmentación vertical

- Veamos un ejemplo:

Considere el siguiente esquema de fragmentación vertical:

$$EMP_1 = \Pi_{ENO,ENAME}(EMP)$$

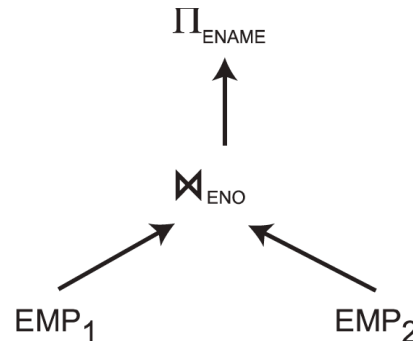
$$EMP_2 = \Pi_{ENO,TITLE}(EMP)$$

$$EMP = EMP_1 \bowtie_{ENO} EMP_2$$

Reducción para fragmentación vertical

- El árbol de consulta localizada se puede reducir considerando la técnica de **reducción por proyección**.

```
SELECT ENAME  
FROM EMP
```



árbol de consulta localizada



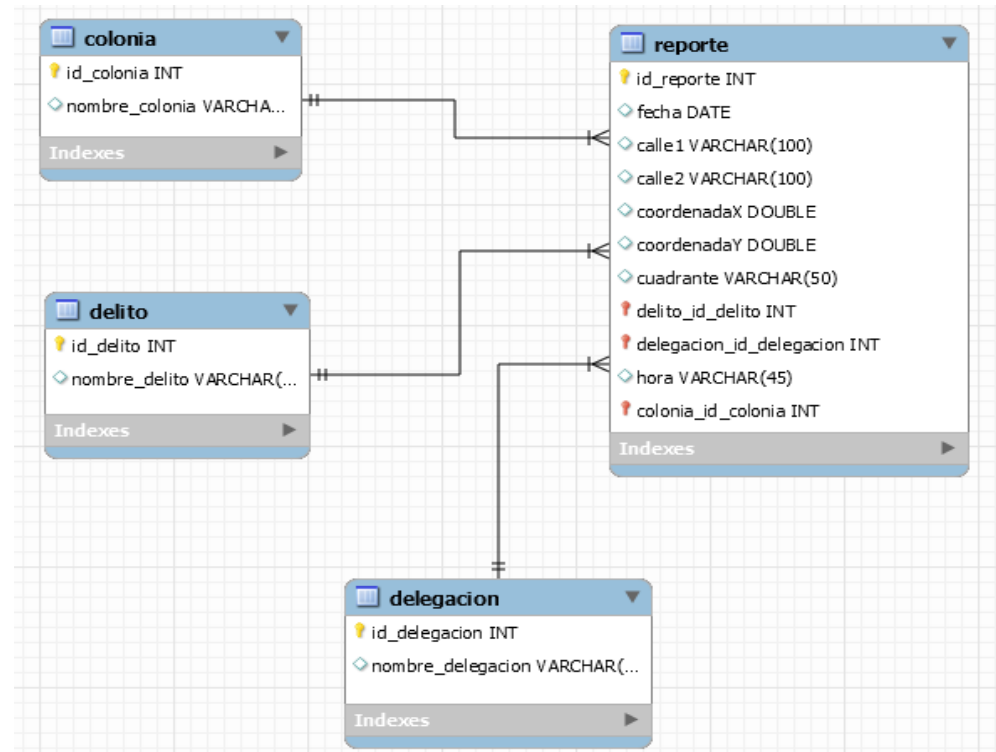
árbol de consulta reducida

- Dado que en el fragmento EMP1 se encuentra la proyección de atributos de la consulta (ENAME), se puede eliminar EMP2 sin afectar dicha consulta.

Ejercicios

```
SELECT fecha
```

```
FROM reporte
```



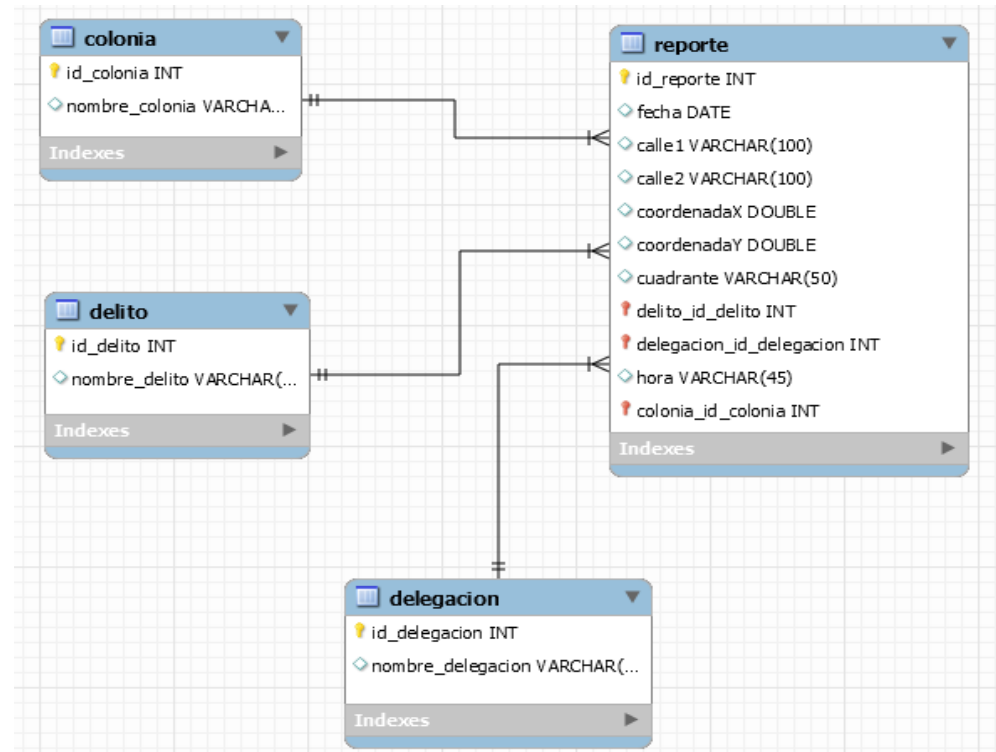
Considere la siguiente fragmentación vertical de la relación “reporte”:

- R1 = Contiene información del tiempo del reporte: fecha y hora.
- R2= Contiene información del lugar: coordenadas, calles, cuadrante, delegación y colonia.
- R3= tipos de delito.

Ejercicios

```
SELECT fecha
```

```
FROM reporte
```



Considere la siguiente fragmentación vertical de la relación “reporte”:

- $R_1 = \Pi_{id_reporte, fecha, hora} (reporte)$
- $R_2 = \Pi_{id_reporte, coordenadaX, coordenadaY, calle1, calle2, cuadrante, id_delegación, id_colonia} (reporte)$
- $R_3 = \Pi_{id_reporte, id_delito} (reporte)$