



UNIVERSIDAD DE LOS LAGOS

FRAGMENTACIÓN VERTICAL

BASE DE DATOS AVANZADA

JUAN JOSÉ RAMÍREZ LAMA
JUAN.RAMIREZ@ULAGOS.CL

INGENIERÍA CIVIL EN INFORMÁTICA

CONTENIDO

- 1 Fragmentación Vertical
 - Matriz de Uso
 - Matriz de Afinidad
 - Ejemplo Matriz de Afinidad
- 2 Algoritmo de Fragmentación Vertical
 - Algoritmo de Agrupamiento
- 3 Fragmentación Híbrida
- 4 Conclusión
- 5 Ejercicio Fragmentación Vertical

CONTENIDO

- 1 Fragmentación Vertical
 - Matriz de Uso
 - Matriz de Afinidad
 - Ejemplo Matriz de Afinidad
- 2 Algoritmo de Fragmentación Vertical
 - Algoritmo de Agrupamiento
- 3 Fragmentación Híbrida
- 4 Conclusión
- 5 Ejercicio Fragmentación Vertical

FRAGMENTACIÓN VERTICAL

- ▶ El objetivo es dividir una relación R en un conjunto de relaciones más pequeñas (fragmentos verticales) de manera de que aplicaciones de usuarios corran sobre un único fragmento (idealmente)
- ▶ Una fragmentación óptima produce un esquema que minimiza el tiempo de ejecución de las aplicaciones

FRAGMENTACIÓN VERTICAL

- ▶ Dada una relación R una fragmentación vertical divide R en fragmentos R_1, \dots, R_n (usualmente 2).
- ▶ Cada fragmento contienen un subconjunto de los atributos de R
- ▶ Además cada fragmento contiene la clave primaria de R

FRAGMENTACIÓN VERTICAL

- ▶ Es más difícil que la fragmentación horizontal porque existen más pasos y condiciones.
- ▶ Existen dos enfoques:
 - Agrupamiento** asigna atributos a fragmentos.
 - División** parte desde una relación y decide cómo fragmentar sobre la base de mejorar el acceso de las aplicaciones, **genera fragmentos disjuntos**.

INFORMACIÓN DE APLICACIONES

Uso de atributos:

- Dado un conjunto de consultas $Q = (q_1, \dots, q_m)$ que corren sobre una relación $R[A_1, \dots, A_n]$, la **matriz de uso** se compone según:

$$uso(q_i, A_k) = \begin{cases} 1 & \text{si } A_k \text{ es utilizado por } q_i \\ 0 & \text{en caso contrario} \end{cases}$$

- Dicha matriz tiene las consultas como filas y los atributos como columnas.

uso	A_1	A_2
q_1	1	0
q_2	0	1
q_3	1	1

USO DE ATRIBUTOS

- Consideremos el siguiente esquema de la tabla “Jugador”:

Jugador (ID (PK), Nombre, Sueldo, Equipo)

Consideremos las siguientes consultas:

q₁

```
1 SELECT SUELDO
2 FROM JUGADOR
3 WHERE ID=valor
```

q₃

```
1 SELECT NOMBRE
2 FROM JUGADOR
3 WHERE EQUIPO='valor'
```

q₂

```
1 SELECT NOMBRE, SUELDO
2 FROM JUGADOR
```

q₄

```
1 SELECT SUM(SUELDO)
2 FROM JUGADOR
3 WHERE EQUIPO='Valor'
```


USO DE ATRIBUTOS

- Consideremos el siguiente esquema de la tabla “Jugador”:

Jugador (ID (PK), Nombre, Sueldo, Equipo)

Consideremos las siguientes consultas:

q₁

```
1 SELECT SUELDO
2 FROM JUGADOR
3 WHERE ID=valor
```

q₃:

```
1 SELECT NOMBRE
2 FROM JUGADOR
3 WHERE EQUIPO='valor'
```

q₂:

```
1 SELECT NOMBRE, SUELDO
2 FROM JUGADOR
```

q₄:

```
1 SELECT SUM(SUELDO)
2 FROM JUGADOR
3 WHERE EQUIPO='valor'
```

Donde: $A_1 = \text{ID}$, $A_2 = \text{NOMBRE}$, $A_3 = \text{SUELDO}$, $A_4 = \text{EQUIPO}$

USO DE ATRIBUTOS

Consideremos las siguientes consultas:

q_1

```
1 SELECT SUELDO
2 FROM JUGADOR
3 WHERE ID=valor
```

q_3 :

```
1 SELECT NOMBRE
2 FROM JUGADOR
3 WHERE EQUIPO='valor'
```

q_2 :

```
1 SELECT NOMBRE, SUELDO
2 FROM JUGADOR
```

q_4 :

```
1 SELECT SUM(SUELDO)
2 FROM JUGADOR
3 WHERE EQUIPO='valor'
```

Donde: $A_1 = \text{ID}$, $A_2 = \text{NOMBRE}$, $A_3 = \text{SUELDO}$, $A_4 = \text{EQUIPO}$

Sea entonces la siguiente tabla de uso de atributos por consulta:

uso	A_1	A_2	A_3	A_4
q_1	1	0	1	0
q_2	0	1	1	0
q_3	0	1	0	1
q_4	0	0	1	1

INFORMACIÓN DE APLICACIONES

Afinidad de atributos

- ▶ Indica cuan relacionados están los atributos de una relación.
- ▶ Se calcula por par de atributos analizando las aplicaciones que usan los atributos y las frecuencias de acceso de las aplicaciones.

MEDIDA DE AFINIDAD

La **medida de afinidad** entre 2 atributos $A_i, A_j \in R[A_1, \dots, A_n]$ con respecto a un conjunto de aplicaciones $Q = (q_1, \dots, q_n)$ se define como:

$$aff(A_i, A_j) = \sum_{\text{todas las consultas que acceden a } (A_i \wedge A_j)} \left(\sum_{\text{por cada sitio}} \text{frecuencia de la consulta} \cdot \frac{\text{Acceso}}{\text{Ejecución}} \right)$$

$\frac{\text{Acceso}}{\text{Ejecución}}$: número de veces que la consulta accede los atributos A_i, A_j por ejecución.

Por simplicidad y al menos que se diga lo contrario $\frac{\text{Acceso}}{\text{Ejecución}} = 1$.

Se requiere conocer que aplicaciones utilizan ambos atributos al mismo tiempo para conocer su afinidad.

MEDIDA DE AFINIDAD

- Consideremos que las consultas acceden una vez a los atributos por ejecución.
- Las siguientes frecuencias de acceso de las aplicaciones, con $\frac{\text{Acceso}}{\text{Ejecución}} = 1$

uso	A_1	A_2	A_3	A_4
q_1	1	0	1	0
q_2	0	1	1	0
q_3	0	1	0	1
q_4	0	0	1	1
	S_1	S_2	S_3	
q_1	15	20	10	
q_2	5	0	0	
q_3	25	25	25	
q_4	3	0	0	

Afinidad entre los atributos A_1 y A_3 :

$$\text{aff}(A_1, A_3) = (S_1 \cdot \frac{\text{Acceso}}{\text{Ejecución}} + S_2 \cdot \frac{\text{Acceso}}{\text{Ejecución}} + S_3 \cdot \frac{\text{Acceso}}{\text{Ejecución}}) q_1$$

$$\text{aff}(A_1, A_3) = 15 \cdot 1 + 20 \cdot 1 + 10 \cdot 1 = 45$$

q_1 es la única consulta que accede a los 2 atributos al mismo tiempo.

MEDIDA DE AFINIDAD

- Consideremos que las consultas acceden una vez a los atributos por ejecución.
- Las siguientes frecuencias de acceso de las aplicaciones, con $\frac{\text{Acceso}}{\text{Ejecución}} = 1$
- Matriz de afinidad AA:

uso	A_1	A_2	A_3	A_4
q_1	1	0	1	0
q_2	0	1	1	0
q_3	0	1	0	1
q_4	0	0	1	1
	S_1	S_2	S_3	
q_1	15	20	10	
q_2	5	0	0	
q_3	25	25	25	
q_4	3	0	0	
	A_1	A_2	A_3	A_4
A_1	45	0	45	0
A_2	0	80	5	75
A_3	45	5	53	3
A_4	0	75	3	78

EJEMPLO 1

Considere el siguiente esquema relacional:

PROF (ID, NOMBRE, TITULO)

ASIG (IDA, TIPO, NASIG, DUR)

Asuma la siguiente definición de vista:

```
1 CREATE VIEW PROFVIEW(ID, NOMBRE, TIPO, NASIG) AS
2   SELECT P.ID, P.NOMBRE, A.TIPO, A.NASIG
3   FROM PROF P, ASIG A
4   WHERE P.ID=A.IDA AND A.DUR = 90
```

La cual es utilizada por la consulta q_1 en los sitios 1 y 2 con frecuencias 20 y 30 respectivamente.

Además, asuma la siguiente consulta q_2 :

```
1 SELECT IDA, DUR
2 FROM ASIG
```

La cual se ejecuta en los sitios 2 y 3 con frecuencias 10 y 30 respectivamente.

- Construir la matriz $uso(q_i, A_j)$ y Acceso.
- Generar la matriz de afinidad que contenga todos los atributos de PROF y ASIG.

EJEMPLO 1: SOLUCIÓN

$A_1 = ID, A_2 = NOMBRE, A_3 = TITULO, A_4 = IDA, A_5 = TIPO, A_6 = NASIG, A_7 = DUR.$

$S_1 = 20$ y $S_2 = 30$ acceden a q_1

$S_2 = 10$ y $S_3 = 30$ acceden a q_2

q_1

```
1 CREATE VIEW PROFVIEW(ID, NOMBRE, TIPO, NASIG)
2 AS
3 SELECT P.ID, P.NOMBRE, A.TIPO, A.NASIG
4 FROM PROF P, ASIG A
5 WHERE P.ID=A.IDA AND A.DUR = 90
```

Matriz de $uso(q_i, A_i)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
q_1							
q_2							

q_2

```
1 SELECT IDA, DUR
2 FROM ASIG
```

Matriz de Acceso

	S_1	S_2	S_3
q_1			
q_2			

EJEMPLO 1: SOLUCIÓN

$A_1 = ID, A_2 = NOMBRE, A_3 = TITULO, A_4 = IDA, A_5 = TIPO, A_6 = NASIG, A_7 = DUR.$

$S_1 = 20$ y $S_2 = 30$ acceden a q_1

$S_2 = 10$ y $S_3 = 30$ acceden a q_2

q_1

```
1 CREATE VIEW PROFVIEW(ID, NOMBRE, TIPO, NASIG)
2 AS
3 SELECT P.ID, P.NOMBRE, A.TIPO, A.NASIG
4 FROM PROF P, ASIG A
5 WHERE P.ID=A.IDA AND A.DUR = 90
```

q_2

```
1 SELECT IDA, DUR
2 FROM ASIG
```

Matriz de $uso(q_i, A_i)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
q_1	1	1	0	1	1	1	1
q_2	0	0	0	1	0	0	1

Matriz de Acceso

	S_1	S_2	S_3
q_1	20	30	0
q_2	0	10	30

EJEMPLO 1: SOLUCIÓN

► $S_1 = 20$ y $S_2 = 30$ acceden a q_1

► $S_2 = 10$ y $S_3 = 30$ acceden a q_2

Matriz de $uso(q_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
q_1	1	1	0	1	1	1	1
q_2	0	0	0	1	0	0	1

Matriz de Acceso

	S_1	S_2	S_3
q_1	20	30	0
q_2	0	10	30

Matriz de Afinidad $AA = aff(A_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1							
A_2							
A_3							
A_4							
A_5							
A_6							
A_7							

EJEMPLO 1: SOLUCIÓN

- $S_1 = 20$ y $S_2 = 30$ acceden a q_1

Matriz de $uso(q_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
q_1	1	1	0	1	1	1	1
q_2	0	0	0	1	0	0	1

- $aff(A_1, A_1) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_1, A_2) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_1, A_3) = 0$
- $aff(A_1, A_4) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_1, A_5) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_1, A_6) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_1, A_7) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_2, A_2) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_2, A_3) = 0$
- $aff(A_2, A_4) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_2, A_5) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_2, A_6) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_2, A_7) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$

- $S_2 = 10$ y $S_3 = 30$ acceden a q_2

Matriz de Acceso

	S_1	S_2	S_3
q_1	20	30	0
q_2	0	10	30

Matriz de Afinidad $AA = aff(A_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	50	50	0	50	50	50	50
A_2	50	50	0	50	50	50	50
A_3	0	0	0	0	0	0	0
A_4	50	50	0	90	50	50	90
A_5	50	50	0	50	50	50	50
A_6	50	50	0	50	50	50	50
A_7	50	50	0	90	50	50	90

EJEMPLO 1: SOLUCIÓN

- $S_1 = 20$ y $S_2 = 30$ acceden a q_1

Matriz de $uso(q_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
q_1	1	1	0	1	1	1	1
q_2	0	0	0	1	0	0	1

- $aff(A_3, A_3) = 0$
- $aff(A_3, A_4) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_3, A_5) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_3, A_6) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_3, A_7) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_4, A_4) =$
 $20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + 10 \cdot 1 + 30 \cdot 1 = 90_{q_1, q_2}$
- $aff(A_4, A_5) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_4, A_6) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_4, A_7) =$
 $20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + 10 \cdot 1 + 30 \cdot 1 = 90_{q_1, q_2}$
- $aff(A_5, A_5) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_5, A_6) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_5, A_7) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$

- $S_2 = 10$ y $S_3 = 30$ acceden a q_2

Matriz de Acceso

	S_1	S_2	S_3
q_1	20	30	0
q_2	0	10	30

Matriz de Afinidad $AA = aff(A_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	50	50	0	50	50	50	50
A_2	50	50	0	50	50	50	50
A_3	0	0	0	0	0	0	0
A_4	50	50	0	90	50	50	90
A_5	50	50	0	50	50	50	50
A_6	50	50	0	50	50	50	50
A_7	50	50	0	90	50	50	90

- $aff(A_6, A_6) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_6, A_7) = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50_{q_1}$
- $aff(A_7, A_7) =$
 $20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + 10 \cdot 1 + 30 \cdot 1 = 90_{q_1, q_2}$

EJEMPLO 1: SOLUCIÓN

- $S_1 = 20$ y $S_2 = 30$ acceden a q_1

Matriz de $uso(q_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
q_1	1	1	0	1	1	1	1
q_2	0	0	0	1	0	0	1

$$\sum \sum q_1 = 20 \cdot 1 + 30 \cdot 1 + 0 \cdot 1 = 50$$

$$\sum \sum q_2 = 0 \cdot 1 + 10 \cdot 1 + 30 \cdot 1 = 40$$

- $S_2 = 10$ y $S_3 = 30$ acceden a q_2

Matriz de Acceso

	S_1	S_2	S_3
q_1	20	30	0
q_2	0	10	30

Matriz de Afinidad $AA = aff(A_i, A_j)$

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	50	50	0	50	50	50	50
A_2	50	50	0	50	50	50	50
A_3	0	0	0	0	0	0	0
A_4	50	50	0	90	50	50	90
A_5	50	50	0	50	50	50	50
A_6	50	50	0	50	50	50	50
A_7	50	50	0	90	50	50	90

EJEMPLO 2

Considere el conjunto de consultas $Q = q_1, q_2, q_3, q_4$, atributos $A = A_1, A_2, A_3, A_4$, sitios $S = S_1, S_2, S_3$, y las siguientes matrices de uso de atributos por consultas, y frecuencias de acceso de aplicaciones:

	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

Asuma que la razón $\frac{\text{acceso}}{\text{ejecucion}} = 1$ para todas las consultas q_k y sitios S_i , y que A_1 es el atributo clave. Obtenga la Matriz de Afinidad correspondiente.

EJEMPLO 2: SOLUCIÓN

	A_1	A_2	A_3	A_4
A_1	70	25	25	45
A_2	25	50	50	0
A_3	25	50	80	0
A_4	45	0	0	45

$$\Sigma q_1 = 25$$

$$\Sigma q_2 = 25$$

$$\Sigma q_3 = 45$$

$$\Sigma q_4 = 30$$

CONTENIDO

- 1 Fragmentación Vertical
 - Matriz de Uso
 - Matriz de Afinidad
 - Ejemplo Matriz de Afinidad
- 2 Algoritmo de Fragmentación Vertical
 - Algoritmo de Agrupamiento
- 3 Fragmentación Híbrida
- 4 Conclusión
- 5 Ejercicio Fragmentación Vertical

ALGORITMO DE AGRUPAMIENTO

- ▶ El algoritmo de agrupamiento toma la matriz de afinidad y reorganiza los atributos para formar grupos de atributos que tienen una alta afinidad entre ellos
- ▶ **Bond Energy Algorithm (BEA)** encuentra un orden de atributos, tal que la medida global de **Afinidad** se **Máximiza** (AM) es la afinidad de A_i , A_j con sus vecinos.

$$AM = \sum_{i=1}^n \sum_{j=1}^n aff(A_i, A_j) [aff(A_i, A_{j-1}) + aff(A_i, A_{j+1})]$$

- ▶ **Entrada:** matriz de afinidad **AA**.
- ▶ **Salida:** Matriz de afinidad agrupada **CA**.
 - ① **Inicialización:** Elige una columna de **AA** que se fija en **CA**.
 - ② **Iteración:** Posiciona el resto de las $n - i$ columnas (i : número de columnas posicionadas en **CA**) en la matriz **CA**.
 - ③ **Mayor contribución:** Por cada columna elige el lugar que contribuye más a la medida de afinidad global.
 - ④ **Orden de columnas:** Finalmente ordena las filas de acuerdo al orden de las columnas

- ¿Cómo determinamos el mejor lugar para posicionar una columna A_k ?

$$\text{cont}(A_i, A_k, A_j) = 2 \cdot \text{bond}(A_i, A_k) + 2 \cdot \text{bond}(A_k, A_j) - 2 \cdot \text{bond}(A_i, A_j)$$

donde:

$$\text{bond}(A_x, A_y) = \sum_{z=1}^n \text{aff}(A_z, A_x) \cdot \text{aff}(A_z, A_y)$$

EJEMPLO BEA

- Considere la siguiente matriz de afinidad AA y la matriz CA donde A_1 y A_2 han sido ya situadas. Debemos situar A_3 :

$$AA = \begin{array}{c|cccc} & A_1 & A_2 & A_3 & A_4 \\ \hline A_1 & 45 & 0 & 45 & 0 \\ A_2 & 0 & 80 & 5 & 75 \\ A_3 & 45 & 5 & 53 & 3 \\ A_4 & 0 & 75 & 3 & 78 \end{array}$$

$$CA = \begin{array}{c|cc} & A_1 & A_2 \\ \hline & 45 & 0 \\ & 0 & 80 \\ & 45 & 5 \\ & 0 & 75 \end{array}$$

Considerando el orden: $A_f - A_3 - A_1$:

$$\begin{aligned} cont(A_f, A_3, A_1) &= 2bond(A_f, A_3) + 2bond(A_3, A_1) - 2bond(A_f, A_1) \\ &= 2 * 0 \quad \quad \quad + 2 * 4410 \quad \quad \quad - 2 * 0 = \mathbf{8820} \end{aligned}$$

$$\begin{aligned} bond(A_f, A_3) &= aff(A_1, A_f) * aff(A_1, A_3) \\ &\quad + aff(A_2, A_f) * aff(A_2, A_3) \\ &\quad + aff(A_3, A_f) * aff(A_3, A_3) \\ &\quad + aff(A_4, A_f) * aff(A_4, A_3) \\ &= 0 * 45 + 0 * 5 + 0 * 53 + 0 * 3 \\ &= 0 \end{aligned}$$

$$\begin{aligned} bond(A_3, A_1) &= aff(A_1, A_3) * aff(A_1, A_1) \\ &\quad + aff(A_2, A_3) * aff(A_2, A_1) \\ &\quad + aff(A_3, A_3) * aff(A_3, A_1) \\ &\quad + aff(A_4, A_3) * aff(A_4, A_1) \\ &= 45 * 45 + 5 * 0 + 53 * 45 + 3 * 0 \\ &= 4410 \end{aligned}$$

EJEMPLO BEA

- Considere la siguiente matriz de afinidad AA y la matriz CA donde A_1 y A_2 han sido ya situadas. Debemos situar A_3 :

$$AA = \begin{array}{c|cccc} & A_1 & A_2 & A_3 & A_4 \\ \hline A_1 & 45 & 0 & 45 & 0 \\ A_2 & 0 & 80 & 5 & 75 \\ A_3 & 45 & 5 & 53 & 3 \\ A_4 & 0 & 75 & 3 & 78 \end{array}$$

$$CA = \begin{array}{c|cc} & A_1 & A_2 \\ \hline & 45 & 0 \\ & 0 & 80 \\ & 45 & 5 \\ & 0 & 75 \end{array}$$

Considerando el orden: $A_f - A_3 - A_1$:

$$\begin{aligned} cont(A_f, A_3, A_1) &= 2bond(A_f, A_3) + 2bond(A_3, A_1) - 2bond(A_f, A_1) \\ &= 2 * 0 + 2 * 4410 - 2 * 0 = \mathbf{8820} \end{aligned}$$

Considerando el orden: $A_1 - A_3 - A_2$:

$$\begin{aligned} cont(A_1, A_3, A_2) &= 2bond(A_1, A_3) + 2bond(A_3, A_2) - 2bond(A_1, A_2) \\ &= 2 * 4410 + 2 * 890 - 2 * 225 = \mathbf{10150} \end{aligned}$$

Considerando el orden: $A_2 - A_3 - A_4$:

$$cont(A_2, A_3, A_f) = 2bond(A_2, A_3) + 2bond(A_3, A_f) - 2bond(A_2, A_f) = \mathbf{1780}$$

EJEMPLO BEA

- Dado que el orden $A_1 - A_3 - A_2$ entrega la mayor contribución, la matriz CA nos queda de la siguiente manera:

CA =	A_1	A_3	A_2
	45	45	0
	0	5	80
	45	53	5
	0	3	75

EJEMPLO BEA

► Luego se procede de igual forma para el atributo A_4

$$\begin{aligned} cont(A_f, A_4, A_1) &= 2bond(A_f, A_4) + 2bond(A_4, A_1) - 2bond(A_f, A_1) = \\ &= 2 * 0 \quad \quad \quad + 2 * 135 \quad \quad \quad - 2 * 0 \quad \quad \quad = 270 \end{aligned}$$

$$\begin{aligned} cont(A_1, A_4, A_3) &= 2bond(A_1, A_4) + 2bond(A_4, A_3) - 2bond(A_1, A_3) = \\ &= 2 * 135 \quad \quad \quad + 2 * 768 \quad \quad \quad - 2 * 4410 \quad \quad \quad = -7014 \end{aligned}$$

$$\begin{aligned} cont(A_3, A_4, A_2) &= 2bond(A_3, A_4) + 2bond(A_4, A_2) - 2bond(A_3, A_2) = \\ &= 2 * 768 \quad \quad \quad + 2 * 11865 \quad \quad \quad - 2 * 890 \quad \quad \quad = 23486 \end{aligned}$$

$$\begin{aligned} cont(A_2, A_4, A_f) &= 2bond(A_2, A_4) + 2bond(A_4, A_f) - 2bond(A_2, A_f) = \\ &= 2 * 11865 \quad \quad \quad + 2 * 0 \quad \quad \quad - 2 * 0 \quad \quad \quad = 23730 \end{aligned}$$

EJEMPLO BEA

- Dado que el orden $A_2 - A_4 - A_f$ entrega la mayor contribución, la matriz CA nos queda de la siguiente manera:

$$CA = \begin{array}{c|cccc} & A_1 & A_3 & A_2 & A_4 \\ \hline & 45 & 45 & 0 & 0 \\ & 0 & 5 & 80 & 75 \\ & 45 & 53 & 5 & 3 \\ & 0 & 3 & 75 & 78 \end{array}$$

Finalmente queda reorganizar las filas tal como quedaron las columnas.

EJEMPLO BEA

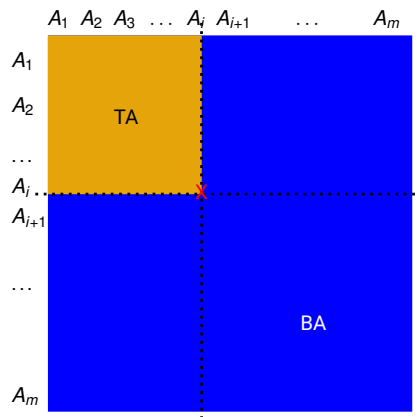
- Al reorganizar las filas obtenemos la matriz CA final:

	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_2	0	5	80	75
A_3	45	53	5	3
A_4	0	3	75	78

	A_1	A_3	A_2	A_4
$CA = A_1$	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

ALGORITMO DE AGRUPAMIENTO

¿Cómo se divide un conjunto de atributos agrupados (A_1, \dots, A_m) en dos (o más) conjuntos (A_1, \dots, A_i), (A_i, \dots, A_m) de tal forma que no existan aplicaciones que accedan a ambos (o a más de un) conjuntos?



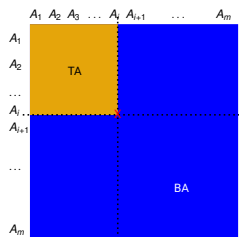
► **TA:** Atributos de Arriba

► **BA:** Atributos de Abajo

ALGORITMO DE AGRUPAMIENTO

- ▶ Uso de grupos de atributos por aplicaciones:
 - ▶ TQ = conjunto de aplicaciones que acceden solo a TA
 - ▶ BQ = conjunto de aplicaciones que acceden solo a BA
 - ▶ OQ = conjunto de aplicaciones que acceden a TA y BA
- ▶ Número de accesos
 - ▶ CTQ = número total de accesos a atributos por aplicaciones que acceden solo a TA
 - ▶ CBQ = número total de accesos a atributos por aplicaciones que acceden solo a BA
 - ▶ COQ = número total de accesos a atributos por aplicaciones que acceden a TA y BA
- ▶ Encontrar el punto en la diagonal que maximice:

$$MAX(CTQ \cdot CBQ - COQ^2)$$



ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_1, A_1]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: $A_1 = 0$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: A_3 o A_2 o $A_4 = q_1, q_4 = 25 + 30 = 55$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_2, q_3 = 25 + 45 = 70$

$$CTQ \cdot CBQ - (COQ^2) = 0 \cdot 55 - 70^2 = -4900$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_1, A_1]

- ▶ **CTQ**: Consultas que Acceden **solamente a arriba**: $A_1 = 0$
- ▶ **CBQ**: Consultas que Acceden **solamente abajo**: A_3 o A_2 o $A_4 = q_1, q_4 = 25 + 30 = 55$
- ▶ **COQ**: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_2, q_3 = 25 + 45 = 70$

$$CTQ \cdot CBQ - (COQ^2) = 0 \cdot 55 - 70^2 = -4900$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_1, A_1]

- ▶ **CTQ**: Consultas que Acceden **solamente a arriba**: $A_1 = 0$
- ▶ **CBQ**: Consultas que Acceden **solamente abajo**: A_3 o A_2 o $A_4 = q_1, q_4 = 25 + 30 = 55$
- ▶ **COQ**: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_2, q_3 = 25 + 45 = 70$

$$CTQ \cdot CBQ - (COQ^2) = 0 \cdot 55 - 70^2 = -4900$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_1, A_1]

- ▶ **CTQ**: Consultas que Acceden **solamente a arriba**: $A_1 = 0$
- ▶ **CBQ**: Consultas que Acceden **solamente abajo**: A_3 o A_2 o $A_4 = q_1, q_4 = 25 + 30 = 55$
- ▶ **COQ**: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_2, q_3 = 25 + 45 = 70$

$$CTQ \cdot CBQ - (COQ^2) = 0 \cdot 55 - 70^2 = -4900$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_1, A_1]

- ▶ **CTQ**: Consultas que Acceden **solamente a arriba**: $A_1 = 0$
- ▶ **CBQ**: Consultas que Acceden **solamente abajo**: A_3 o A_2 o $A_4 = q_1, q_4 = 25 + 30 = 55$
- ▶ **COQ**: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_2, q_3 = 25 + 45 = 70$

$$CTQ \cdot CBQ - (COQ^2) = 0 \cdot 55 - 70^2 = -4900$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_3, A_3]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o $A_3 = q_4 = 30$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: A_2 o $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_1, q_2, q_3 = 25 + 25 + 45 = 95$

$$CTQ \cdot CBQ - (COQ^2) = 30 \cdot 0 - 95^2 = -9025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_3, A_3]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o $A_3 = q_4 = 30$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: A_2 o $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_1, q_2, q_3 = 25 + 25 + 45 = 95$

$$CTQ \cdot CBQ - (COQ^2) = 30 \cdot 0 - 95^2 = -9025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_3, A_3]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o $A_3 = q_4 = 30$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: A_2 o $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_1, q_2, q_3 = 25 + 25 + 45 = 95$

$$CTQ \cdot CBQ - (COQ^2) = 30 \cdot 0 - 95^2 = -9025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_3, A_3]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o $A_3 = q_4 = 30$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: A_2 o $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_1, q_2, q_3 = 25 + 25 + 45 = 95$

$$CTQ \cdot CBQ - (COQ^2) = 30 \cdot 0 - 95^2 = -9025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_3, A_3]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o $A_3 = q_4 = 30$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: A_2 o $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_1, q_2, q_3 = 25 + 25 + 45 = 95$

$$CTQ \cdot CBQ - (COQ^2) = 30 \cdot 0 - 95^2 = -9025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_2, A_2]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o A_3 o $A_2 = q_1, q_2, q_4 = 25 + 25 + 30 = 80$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_3 = 45$

$$CTQ \cdot CBQ - (COQ^2) = 80 \cdot 0 - 45^2 = -2025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_2, A_2]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o A_3 o $A_2 = q_1, q_2, q_4 = 25 + 25 + 30 = 80$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_3 = 45$

$$CTQ \cdot CBQ - (COQ^2) = 80 \cdot 0 - 45^2 = -2025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_2, A_2]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o A_3 o $A_2 = q_1, q_2, q_4 = 25 + 25 + 30 = 80$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_3 = 45$

$$CTQ \cdot CBQ - (COQ^2) = 80 \cdot 0 - 45^2 = -2025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_2, A_2]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o A_3 o $A_2 = q_1, q_2, q_4 = 25 + 25 + 30 = 80$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_3 = 45$

$$CTQ \cdot CBQ - (COQ^2) = 80 \cdot 0 - 45^2 = -2025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Matriz[A_2, A_2]

- ▶ CTQ: Consultas que Acceden **solamente a arriba**: A_1 o A_3 o $A_2 = q_1, q_2, q_4 = 25 + 25 + 30 = 80$
- ▶ CBQ: Consultas que Acceden **solamente abajo**: $A_4 = 0$
- ▶ COQ: Consultas que Acceden a **ambas**: $A_1 \dots A_4 = q_3 = 45$

$$CTQ \cdot CBQ - (COQ^2) = 80 \cdot 0 - 45^2 = -2025$$

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

¿Debo continuar?

ALGORITMO DE AGRUPAMIENTO

uso	A_1	A_2	A_3	A_4
q_1	0	1	1	0
q_2	1	1	1	0
q_3	1	0	0	1
q_4	0	0	1	0

	S_1	S_2	S_3
q_1	20	5	0
q_2	10	0	15
q_3	0	40	5
q_4	0	20	10

CA	A_1	A_3	A_2	A_4
A_1	45	45	0	0
A_3	45	53	5	3
A_2	0	5	80	75
A_4	0	3	75	78

Asumiendo que A_1 es la clave primaria de $A = \{A_1, A_2, A_3, A_4\}$, usando BEA y el Algoritmo de Particionado podemos definir las siguientes tablas fragmentadas verticalmente.

- ▶ $A_1 = \{\underline{A_1}, A_3, A_2\}$
- ▶ $A_2 = \{\underline{A_1}, A_4\}$

CORRECCIÓN DE LA FRAGMENTACIÓN

- ▶ Una relación R definida sobre el conjunto de atributos A con clave K genera los fragmentos verticales $FV_R = (R_1, R_2, \dots, R_n)$

- ▶ **Compleitud:** lo siguiente debe ser verdad para A :

$$A = \bigcup A_{R_i}$$

- ▶ **Reconstrucción:** La reconstrucción debe ser posible a través de:

$$R = \bowtie_K R_i, \forall R_i \in F_R$$

- ▶ **Fragmentos disjuntos:**

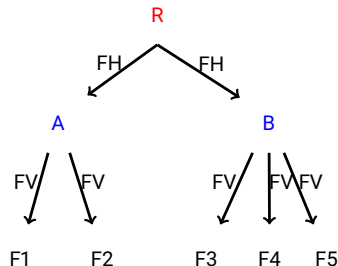
- ▶ Las claves primarias no se consideran como traslape

CONTENIDO

- 1 Fragmentación Vertical
 - Matriz de Uso
 - Matriz de Afinidad
 - Ejemplo Matriz de Afinidad
- 2 Algoritmo de Fragmentación Vertical
 - Algoritmo de Agrupamiento
- 3 Fragmentación Híbrida
- 4 Conclusión
- 5 Ejercicio Fragmentación Vertical

FRAGMENTACIÓN HÍBRIDA

- ▶ En la mayoría de los casos, una simple fragmentación horizontal o vertical de un esquema de base de datos no será suficiente para satisfacer los requisitos de las aplicaciones de usuario.
- ▶ En este caso, una fragmentación vertical puede ir seguida de una horizontal, o viceversa, produciendo una partición con estructura de árbol.
- ▶ Dado que los dos tipos de estrategias de partición se aplican uno después del otro, esta alternativa se denomina fragmentación híbrida.



CONTENIDO

- 1 Fragmentación Vertical
 - Matriz de Uso
 - Matriz de Afinidad
 - Ejemplo Matriz de Afinidad
- 2 Algoritmo de Fragmentación Vertical
 - Algoritmo de Agrupamiento
- 3 Fragmentación Híbrida
- 4 Conclusión
- 5 Ejercicio Fragmentación Vertical

CONCLUSIÓN

- ▶ Para fragmentar verticalmente, se debe elaborar las medidas de afinidad
- ▶ Luego, implementar el algoritmo BEA
- ▶ Y por último, aplicar el algoritmo de agrupamiento

CONTENIDO

- 1 Fragmentación Vertical
 - Matriz de Uso
 - Matriz de Afinidad
 - Ejemplo Matriz de Afinidad
- 2 Algoritmo de Fragmentación Vertical
 - Algoritmo de Agrupamiento
- 3 Fragmentación Híbrida
- 4 Conclusión
- 5 Ejercicio Fragmentación Vertical

EJERCICIO

Considere la relación JUG(ID, NOMBRE, SUELDO, EQUIPO) y las siguientes consultas:

► q_1 :

```
1 | SELECT NOMBRE, SUELDO FROM JUG WHERE ID = valor AND EQUIPO =  
  | "valor"
```

► q_2 :

```
1 | SELECT EQUIPO, SUELDO FROM JUG WHERE NOMBRE = "valor"
```

► q_3 :

```
1 | SELECT COUNT(NOMBRE) FROM JUG WHERE EQUIPO = "valor"
```

► q_4 :

```
1 | SELECT SUM(SUELDO) FROM JUG WHERE EQUIPO = "valor"
```

EJERCICIO

Considere la siguiente matriz de frecuencia de acceso (FA) de cada consulta en cada sitio y la fórmula para calcular la afinidad entre atributos:

FA	S_1	S_2	S_3
Q_1	15	20	10
Q_2	15	10	25
Q_3	30	20	10
Q_4	3	6	5

$$aff(A_i, A_j) = \sum_{\text{Consultas que acceden a } (A_i \wedge A_j)} (AC)$$

$$S_1 : \frac{\text{acceso}}{\text{ejecución}} = 3$$

$$S_2 : \frac{\text{acceso}}{\text{ejecución}} = 1$$

$$S_3 : \frac{\text{acceso}}{\text{ejecución}} = 2$$

EJERCICIO

a) Genere la matriz de uso de atributos, considere:
 A_1 =ID, A_2 =NOMBRE, A_3 =SUELDO, A_4 =EQUIPO.

<i>uso</i>	A_1	A_2	A_3	A_4
Q_1				
Q_2				
Q_3				
Q_4				

EJERCICIO

b) Genere la matriz de afinidad para la relación JUG.

<i>aff</i>	A_1	A_2	A_3	A_4
A_1				
A_2				
A_3				
A_4				

c) Asuma que el atributo A_1 se fija en la matriz CA, ubique el resto de las columnas y obtenga el correcto agrupamiento.

d) Determine los fragmentos de la tabla.

RESPUESTAS

a) Genere la matriz de uso de atributos, considere: A_1 =ID, A_2 =NOMBRE, A_3 =SUELDO, A_4 =EQUIPO

<i>uso</i>	A_1	A_2	A_3	A_4
Q_1	1	1	1	1
Q_2	0	1	1	1
Q_3	0	1	0	1
Q_4	0	0	1	1

RESPUESTAS

b) Genere la matriz de afinidad para la relación JUG.

$$\sum(Q_1) = 15*3 + 20*1 + 10*2 = 85$$

$$\sum(Q_2) = 15*3 + 10*1 + 25*2 = 105$$

$$\sum(Q_3) = 30*3 + 20*1 + 10*2 = 130$$

$$\sum(Q_4) = 3*3 + 6*1 + 5*2 = 25$$

RESPUESTAS

b) Genere la matriz de afinidad para la relación JUG.

$$\sum(Q_1) = 15*3 + 20*1 + 10*2 = 85$$

$$\sum(Q_2) = 15*3 + 10*1 + 25*2 = 105$$

$$\sum(Q_3) = 30*3 + 20*1 + 10*2 = 130$$

$$\sum(Q_4) = 3*3 + 6*1 + 5*2 = 25$$

$$\text{Aff}(A_1, A_1) = 85*1 + 105*0 + 130*0 + 25*0 = 85$$

$$\text{Aff}(A_1, A_2) = 85*1 + 105*0 + 130*0 + 25*0 = 85$$

$$\text{Aff}(A_1, A_3) = 85*1 + 105*0 + 130*0 + 25*0 = 85$$

$$\text{Aff}(A_1, A_4) = 85*1 + 105*0 + 130*0 + 25*0 = 85$$

$$\text{Aff}(A_2, A_2) = 85*1 + 105*1 + 130*1 + 25*0 = 320$$

$$\text{Aff}(A_2, A_3) = 85*1 + 105*1 + 130*0 + 25*0 = 190$$

$$\text{Aff}(A_2, A_4) = 85*1 + 105*1 + 130*1 + 25*0 = 320$$

$$\text{Aff}(A_3, A_3) = 85*1 + 105*1 + 130*0 + 25*1 = 215$$

$$\text{Aff}(A_3, A_4) = 85*1 + 105*1 + 130*0 + 25*1 = 215$$

$$\text{Aff}(A_4, A_4) = 85*1 + 105*1 + 130*1 + 25*1 = 345$$

RESPUESTAS

b) Genere la matriz de afinidad para la relación JUG.

<i>aff</i>	A_1	A_2	A_3	A_4
A_1	85	85	85	85
A_2	85	320	190	320
A_3	85	190	215	215
A_4	85	320	215	345

RESPUESTAS

c) Asuma que el atributo A_1 se fija en la matriz CA, ubique el resto de las columnas.

bond(A_1, A_2) =

$$\begin{aligned} & \text{Aff}(A_1, A_1) * \text{Aff}(A_1, A_2) + \\ & \text{Aff}(A_2, A_1) * \text{Aff}(A_2, A_2) + \\ & \text{Aff}(A_3, A_1) * \text{Aff}(A_3, A_2) + \\ & \text{Aff}(A_4, A_1) * \text{Aff}(A_4, A_2) = 77.775 \end{aligned}$$

bond(A_1, A_3) =

$$\begin{aligned} & \text{Aff}(A_1, A_1) * \text{Aff}(A_1, A_3) + \\ & \text{Aff}(A_2, A_1) * \text{Aff}(A_2, A_3) + \\ & \text{Aff}(A_3, A_1) * \text{Aff}(A_3, A_3) + \\ & \text{Aff}(A_4, A_1) * \text{Aff}(A_4, A_3) = 59.925 \end{aligned}$$

RESPUESTAS

$$\begin{aligned} \text{bond}(A_1, A_4) = & \\ & \text{Aff}(A_1, A_1) * \text{Aff}(A_1, A_4) + \\ & \text{Aff}(A_2, A_1) * \text{Aff}(A_2, A_4) + \\ & \text{Aff}(A_3, A_1) * \text{Aff}(A_3, A_4) + \\ & \text{Aff}(A_4, A_1) * \text{Aff}(A_4, A_4) = 82.025 \end{aligned}$$

$$\begin{aligned} \text{bond}(A_2, A_3) = & \\ & \text{Aff}(A_1, A_2) * \text{Aff}(A_1, A_3) + \\ & \text{Aff}(A_2, A_2) * \text{Aff}(A_2, A_3) + \\ & \text{Aff}(A_3, A_2) * \text{Aff}(A_3, A_3) + \\ & \text{Aff}(A_4, A_2) * \text{Aff}(A_4, A_3) = 177.675 \end{aligned}$$

RESPUESTAS

$$\begin{aligned} \text{bond}(A_2, A_4) = & \\ & \text{Aff}(A_1, A_2) * \text{Aff}(A_1, A_4) + \\ & \text{Aff}(A_2, A_2) * \text{Aff}(A_2, A_4) + \\ & \text{Aff}(A_3, A_2) * \text{Aff}(A_3, A_4) + \\ & \text{Aff}(A_4, A_2) * \text{Aff}(A_4, A_4) = 260.875 \end{aligned}$$

$$\begin{aligned} \text{bond}(A_3, A_4) = & \\ & \text{Aff}(A_1, A_3) * \text{Aff}(A_1, A_4) + \\ & \text{Aff}(A_2, A_3) * \text{Aff}(A_2, A_4) + \\ & \text{Aff}(A_3, A_3) * \text{Aff}(A_3, A_4) + \\ & \text{Aff}(A_4, A_3) * \text{Aff}(A_4, A_4) = 188.425 \end{aligned}$$

RESPUESTAS

UBICANDO A_2 :

$$\text{cont}(A_0, A_2, A_1) = 2 * \text{bond}(A_2, A_1) = 2 * 77.775 = \mathbf{155.550}$$

$$\text{cont}(A_1, A_2, A_3) = 2 * \text{bond}(A_1, A_2) = 2 * 77.775 = \mathbf{155.550}$$

\therefore El mejor orden puede ser A_1, A_2 o A_2, A_1

Escogeremos A_1, A_2 .

RESPUESTAS

UBICANDO A_3 :

$$\text{cont}(A_0, A_3, A_1) = 2 * \text{bond}(A_3, A_1) = 2 * 59.925 = 119.850$$

$$\begin{aligned} \text{cont}(A_1, A_3, A_2) &= 2 * \text{bond}(A_1, A_3) + 2 * \text{bond}(A_3, A_2) - 2 * \text{bond}(A_1, A_2) = \\ &= 2 * (59.925 + 177.675 - 77.775) = 319.650 \end{aligned}$$

$$\text{cont}(A_2, A_3, A_4) = 2 * \text{bond}(A_2, A_3) = 2 * 177.675 = \mathbf{355.350}$$

∴ El mejor orden es A_1, A_2, A_3

RESPUESTAS

UBICANDO A_4 :

$$\text{cont}(A_0, A_4, A_1) = 2 * \text{bond}(A_4, A_1) = 2 * 82.025 = 164.050$$

$$\begin{aligned} \text{cont}(A_1, A_4, A_2) &= 2 * \text{bond}(A_1, A_4) + 2 * \text{bond}(A_4, A_2) - 2 * \text{bond}(A_1, A_2) = \\ &= 2 * (82.025 + 260.875 - 77.775) = 530.250 \end{aligned}$$

$$\begin{aligned} \text{cont}(A_2, A_4, A_3) &= 2 * \text{bond}(A_2, A_4) + 2 * \text{bond}(A_4, A_3) - 2 * \text{bond}(A_2, A_3) = \\ &= 2 * (260.875 + 188.425 - 177.675) = \mathbf{543.250} \end{aligned}$$

$$\text{cont}(A_3, A_4, A_5) = 2 * \text{bond}(A_3, A_4) = 2 * 188.425 = 376.850$$

∴ El mejor orden es A_1, A_2, A_4, A_3

RESPUESTAS

Matriz CA sin ordenar.

CA	A_1	A_2	A_4	A_3
A_1	85	85	85	85
A_2	85	320	320	190
A_3	85	190	215	215
A_4	85	320	345	215

RESPUESTAS

Matriz CA ordenada.

CA	A_1	A_2	A_4	A_3
A_1	85	85	85	85
A_2	85	320	320	190
A_4	85	320	345	215
A_3	85	190	215	215

RESPUESTAS

uso	A ₁	A ₂	A ₃	A ₄
Q ₁	1	1	1	1
Q ₂	0	1	1	1
Q ₃	0	1	0	1
Q ₄	0	0	1	1

FA	S ₁	S ₂	S ₃	
Q ₁	15	20	10	$\sum Q_1 : 85$
Q ₂	15	10	25	$\sum Q_2 : 105$
Q ₃	30	20	10	$\sum Q_3 : 130$
Q ₄	3	6	5	$\sum Q_4 : 25$

CA	A ₁	A ₂	A ₄	A ₃
A ₁	85	85	85	85
A ₂	85	320	320	190
A ₄	85	320	345	215
A ₃	85	190	215	215

RESULTADO

- ▶ Matriz $[A_1, A_1]$: $CTQ \cdot CBQ - COQ^2 =$
 $0 \cdot (105_{Q_2} + 130_{Q_3} + 25_{Q_4}) - 85_{Q_1}^2 = 0 \cdot 260 - 7225 = -7225$
- ▶ Matriz $[A_2, A_2]$: $CTQ \cdot CBQ - COQ^2 =$
 $0 \cdot 25_{Q_4} - (85_{Q_1} + 105_{Q_2} + 130_{Q_3})^2 = 0 \cdot 25 - 102400 = -102400$
- ▶ Matriz $[A_4, A_4]$: $CTQ \cdot CBQ - COQ^2 =$
 $130_{Q_3} \cdot 0 - (85_{Q_1} + 105_{Q_2} + 25_{Q_4})^2 = 130 \cdot 0 - 46225 = -46225$

La mejor fragmentación es:

$$F_1 = \{A_1\}$$

$$F_2 = \{A_1, A_2, A_3, A_4\}$$

\therefore no es recomendable fragmentar.