

## **Fragmentación**

### **3.3.1 Fragmentación Horizontal.**

La fragmentación horizontal divide una relación a lo largo de sus tuplas. Así cada fragmento tiene un subconjunto de las tuplas de la relación. Hay dos versiones de partición horizontal: primaria y derivada. La fragmentación horizontal primaria de una relación se realiza utilizando predicados que se definen en esa relación. La fragmentación horizontal derivada, por otro lado, es la partición de una relación que resulta de la definición de predicados en otra relación

#### **Requisitos de información de la fragmentación horizontal**

- **Información de la base de datos.**

Se refiere al esquema conceptual global. Las relaciones de la base de datos están conectadas entre sí, especialmente con uniones. En el modelo relacional, estas relaciones también se representan como relaciones. Sin embargo, en otros modelos de datos, como el modelo entidad-relación, estas relaciones entre los objetos de la base de datos se representan explícitamente

- **Información de la aplicación.**

Se requiere información tanto cualitativa como cuantitativa sobre las aplicaciones. La información cualitativa guía la actividad de fragmentación, mientras que la información cuantitativa se incorpora principalmente a los modelos de asignación. La información cualitativa fundamental consiste en los predicados utilizados en las consultas de los usuarios.

#### **Fragmentación Horizontal Primaria**

Una fragmentación horizontal primaria se define mediante una operación de selección en las relaciones de propietario de un esquema de base de datos.

### **Fragmentación horizontal derivada**

Una fragmentación horizontal derivada se define en una relación de miembro de un enlace de acuerdo con una operación de selección especificada en su propietario. Es importante recordar dos puntos. En primer lugar, el vínculo entre el propietario y las relaciones miembro se define como una unión equitativa. En segundo lugar, se puede implementar una unión equitativa por medio de semi uniones. Este segundo punto es especialmente importante para nuestros propósitos, ya que queremos particionar una relación miembro de acuerdo con la fragmentación de su propietario, pero también queremos que el fragmento resultante se defina solo en los atributos de la relación miembro. Para llevar a cabo una fragmentación horizontal derivada se necesitan tres entradas: el conjunto de particiones de la relación propietario, la relación miembro y el conjunto de predicados asemejan entre el propietario y el miembro.

### **Comprobación de la corrección**

La integridad de una fragmentación horizontal primaria se basa en los predicados de selección utilizados. Siempre que los predicados de selección estén completos, el resultado también se garantiza que la fragmentación sea completa. Dado que la base del algoritmo de fragmentación es un conjunto de predicados mínimos y completos,  $Pr0$ , la integridad está garantizada siempre que no se cometan errores al definir  $Pr0$ . La completitud de una fragmentación horizontal derivada es algo más difícil de definir. La dificultad se debe al hecho de que el predicado que determina la fragmentación implica dos relaciones.

### ***3.3.2 Fragmentación Vertical***

El objetivo de la fragmentación vertical es dividir una relación en un conjunto de relaciones más pequeñas para que muchas de las aplicaciones de usuario se ejecuten en un solo fragmento.

Una fragmentación “óptima” es aquella que produce un esquema de fragmentación que minimiza el tiempo de ejecución de las aplicaciones de usuario que se ejecutan en estos fragmentos. La fragmentación vertical se ha investigado tanto en el contexto de los sistemas de bases de datos centralizados como en los distribuidos. Su motivación dentro del contexto centralizado es como una herramienta de diseño, que permite que las consultas de los usuarios atiendan relaciones más pequeñas, provocando así un menor número de accesos a la página. La partición vertical es inherentemente más complicada que la partición horizontal. Esto se debe al número total de alternativas disponibles. Por ejemplo, en la partición horizontal, si el número total de predicados simples en  $Pr$  es  $n$ , hay  $2^n$  posibles predicados de mini término que se pueden definir en él.

#### **Requisitos de información de la fragmentación vertical**

Se centra exclusivamente en lo que debe determinarse acerca de las aplicaciones que se ejecutarán en la base de datos distribuida. Dado que la partición vertical coloca en un fragmento esos atributos a los que generalmente se accede juntos, existe la necesidad de alguna medida que defina con mayor precisión la noción de "unión". Esta medida es la afinidad de los atributos, que indica qué tan cerca relacionados están los atributos.

#### **Algoritmo de agrupamiento**

La tarea fundamental en el diseño de un algoritmo de fragmentación vertical es encontrar alguna forma de agrupar los atributos de una relación basada en los valores de afinidad de atributos en  $AA$ . Se considera apropiado por las siguientes razones

- Está diseñado específicamente para determinar grupos de elementos similares en lugar de, por ejemplo, una ordenación lineal de los elementos (es decir, agrupa los atributos con valores de afinidad más grandes juntos y los que tienen valores más pequeños juntos).
- Las agrupaciones finales son insensibles al orden en que se presentan los elementos al algoritmo.
- El tiempo de cálculo del algoritmo es razonable:  $O(n^2)$ , donde  $n$  es el número de atributos.
- Las interrelaciones secundarias entre los grupos de atributos agrupados son identificables.

### **Algoritmo de partición**

El objetivo de la actividad de división es encontrar conjuntos de atributos a los que acceden únicamente, o en su mayor parte, distintos conjuntos de aplicaciones.

### **3.3.3 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 los usuarios. En este caso, una fragmentación vertical puede ser seguida por una horizontal, o viceversa, produciendo una partición estructurada en árbol. Dado que los dos tipos de estrategias de partición se aplican uno tras otro, esta alternativa se denomina fragmentación híbrida. También se le ha denominado fragmentación mixta o fragmentación anidada.

### **3.4 Asignación**

La asignación de recursos a través de los nodos de una red informática es un viejo problema que se ha estudiado ampliamente. Sin embargo, la mayor parte de este trabajo no aborda el problema del diseño de bases de datos distribuidas, sino el de colocar archivos individuales en una red informática.

#### **Requisitos de información**

Es en la etapa de asignación que necesitamos los datos cuantitativos sobre la base de datos, las aplicaciones que se ejecutan en ella, la red de comunicación, las capacidades de procesamiento y las limitaciones de almacenamiento de cada sitio en la red.

- **Información de la base de datos**

Para realizar la fragmentación horizontal, definimos la selectividad de los miniterminos. Ahora necesitamos extender esa definición a los fragmentos y definir la selectividad de un fragmento  $F_j$  con respecto a la consulta  $q_i$ . Este es el número de tuplas de  $F_j$  a las que se debe acceder para procesar  $q_i$ . Este valor se denotará como  $s_{li}(F_j)$ .

- **Información de la aplicación**

La mayor parte de la información relacionada con la aplicación ya se compila durante la actividad de fragmentación, pero el modelo de asignación requiere

algunos más. Las dos medidas importantes son el número de accesos de lectura que hace una consulta  $q_i$  a un fragmento  $F_j$  durante su ejecución (denotado como  $RR_{ij}$ ), y su contraparte para los accesos de actualización ( $UR_{ij}$ ).

- ***Información del sitio***

Para cada sitio de computadora, necesitamos saber su capacidad de almacenamiento y procesamiento. Obviamente, estos valores se pueden calcular mediante funciones elaboradas o mediante estimaciones simples. El costo unitario de almacenar datos en el sitio  $S_k$  se indicará como  $USC_k$ . También es necesario especificar una medida de costo  $LPC_k$  como el costo de procesar una unidad de trabajo en el sitio  $S_k$ . La unidad de trabajo debe ser idéntica a la de las medidas  $RR$  y  $UR$ .

- ***Información de red***

Para habilitar el cálculo de la cantidad de mensajes, usamos tamaño  $f$  como el tamaño (en bytes) de un cuadro.