

Fragmentación

La fragmentación es un conjunto de técnicas para dividir la BD en unidades lógicas, llamadas fragmentos, cuyo almacenamiento puede asignarse a los diversos sitios. Estas técnicas se utilizan durante el proceso de diseño de BDD. La información concerniente a la fragmentación de los datos se almacena en un catálogo global del sistema al que tiene acceso el software cliente cuando es necesario.

La fragmentación es un concepto que surge en los modelos que incluyen algún tipo de particionado y tiene implicaciones físicas y lógicas

- *Fragmento lógico: Bloque de datos localizado en algún SGBDD que puede ser descrito (en términos relacionales) como el resultado de una proyección y/o una selección de la información contenida en la BD, de modo que resulten subconjuntos disjuntos de ella. Por tanto, una vez que la BD ha sido dividida en un conjunto de fragmentos, todos los datos aparecen en uno y sólo uno de estos fragmentos.*
- *Fragmento físico: fragmento lógico que se almacena completo en un nodo. Cada uno de los fragmentos físicos se localizan en uno o más nodos dependiendo del modelo de distribución de los datos.*

El esquema de fragmentación es el conjunto de las distintas relaciones que se han obtenido del EG y las condiciones empleadas para esta división expresado en algebra relacional.

Existen distintas formas de fragmentar una relación

- Fragmentación vertical
- Fragmentación horizontal
- Fragmentación mixta

Razones para la fragmentación

Los esquemas de fragmentación se diseñan teniendo en cuenta el uso de los datos que se almacenan en cada una de las sedes, construyendo relaciones más pequeñas y más adaptadas a las operaciones de recuperación y actualización que utilizan las aplicaciones, tratando así de aligerar las comunicaciones entre los nodos debido al alto coste que éstas tienen.

Las cuatro razones para la fragmentación son las siguientes:

- Es útil ya que las aplicaciones de BD suelen funcionar con vistas, y por ello se pueden utilizar distintas relaciones en distintos nodos para formar la unidad distribuida.
- Se consigue una mayor eficiencia dado que los datos suelen estar almacenados cerca del nodo que más utiliza dichos datos.
- Permite aumentar el grado de concurrencia porque la fragmentación de las relaciones permite que una transacción pueda dividirse en subconsultas que operan sobre estos fragmentos.
- Aporta una mayor seguridad, dado que los datos no utilizados por un nodo local no se almacenan en él y por lo tanto no están al alcance para personas sin autorización.

Alternativas de fragmentación

Esencialmente las relaciones son las tablas, así que la dificultad reside en encontrar diferentes formas de dividir una tabla en otras más pequeñas. Es evidente que hay dos alternativas para esto: ***la fragmentación horizontal o la fragmentación vertical.***

La fragmentación puede ser anidada. Si las fragmentaciones son de tipos diferentes, se obtiene la fragmentación mixta o híbrida. Aun cuando la fragmentación híbrida no es considerada como un tipo primitivo de las estrategias de fragmentación es bastante obvio que en muchas ocasiones de la vida real las particiones pueden ser híbridas.

Grado de fragmentación

La medida en que la base de datos debe ser fragmentada es una decisión importante que afecta el rendimiento de ejecución de las consultas. El grado de fragmentación va de un extremo, es decir, no fragmentar nada, al otro extremo, a fragmentar el nivel de tuplas individuales (en el caso de la fragmentación horizontal) o al nivel de los atributos individuales (en el caso de la fragmentación vertical).

Lo que necesitamos es encontrar un nivel adecuado de fragmentación, que es un compromiso entre los dos extremos. Ese nivel sólo se puede definir con respecto a las aplicaciones que se ejecutan en la base de datos. La cuestión es ¿cómo? En general, las aplicaciones deben ser caracterizadas con respecto a una serie de parámetros. Los fragmentos individuales pueden ser identificados de acuerdo con los valores de estos parámetros.