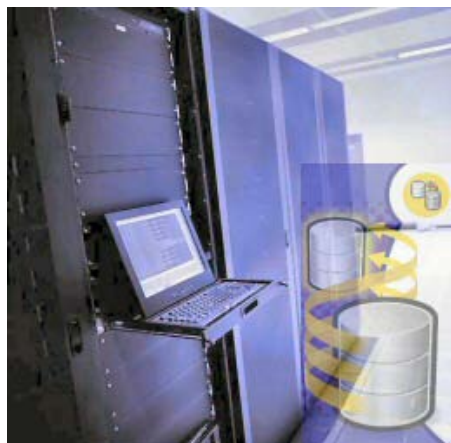


2

FRAGMENTACIÓN DE TABLAS



Contenido:

Fragmentación de tablas

- 2.1. Introducción
 - 2.1.1. Definición.
- 2.2. Tipos de Fragmentación
 - 2.2.1. Fragmentación Horizontal
 - 2.2.2. Fragmentación Vertical
 - 2.2.3. Fragmentación Mixta
 - 2.2.4. Grados de Fragmentación



CAPITULO II

1. FRAGMENTACIÓN DE TABLAS

1.1. INTRODUCCIÓN

Es preciso tomar decisiones respecto a los sitios en los que se almacenarán las porciones de la base de datos, debemos determinar las unidades lógicas de la base de datos que van a ser distribuidas, las unidades lógicas se llamarán: fragmentos.

El mayor reto es identificar la fragmentación ideal para los datos, además de resolver los grados de fragmentación.

Los datos pueden almacenarse en localidades donde son utilizados con mayor frecuencia, de tal manera que la mayor parte de las operaciones sean sólo locales lo cual reducirá el tráfico en la red.

1.1.1. DEFINICIÓN

FRAGMENTACIÓN

Es una relación que corresponde a una tabla, consiste en dividirla en fragmentos menores, cada fragmento se guarda en sitio diferente, tiene como objetivo buscar alternativas para dividir una las tablas o instancias en otras más pequeñas. Para ello hay las siguientes alternativas lógicas, como son: La fragmentación se



puede realizar por tuplas individuales (fragmentación horizontal), por atributos individuales (fragmentación vertical) o una combinación de ambas (fragmentación híbrida).

Fragmentos: Cada relación global puede ser dividida en porciones llamados fragmentos. El mapa resultante se denomina esquema de fragmentación. Una relación global puede dividirse en n fragmentos y un fragmento sólo puede pertenecer a una relación global. Los fragmentos se referencian por un nombre de relación global y un subíndice

RAZONES PARA FRAGMENTAR

- Encontrar unidad de distribución más adecuada.
- Disminuir cantidad de accesos remotos.
- Incrementar el nivel de concurrencia.

VENTAJAS DE LA FRAGMENTACIÓN DE TABLAS

- Mejorar el rendimiento de las aplicaciones al trabajar con subconjuntos de relaciones
- Da una respuesta eficiente a aplicaciones que trabajan con los mismos datos en diferentes nodos
- Los fragmentos permiten aumentar el número de ejecuciones concurrentes.



DESVENTAJAS DE LA FRAGMENTACIÓN DE TABLAS

- Disminuye la eficiencia en las aplicaciones que trabajan con varios fragmentos
- La comprobación de las restricciones de integridad puede ser más costosa.

CONDICIONES PARA DEFINIR FRAGMENTOS

Se deben cumplir tres reglas durante el proceso de fragmentación. Los cuales se explicara más detalladamente en los grados de fragmentación.

- ✓ Completitud
- ✓ Reconstrucción
- ✓ Disyunción

Completitud: La descomposición de una relación R en los fragmentos R_1, R_2, \dots, R_n es completa si y solamente si cada elemento de datos en R se encuentra en algún fragmento. [www.03]¹

Reconstrucción: Si la relación R se descompone en los fragmentos R_1, R_2, \dots, R_n , entonces debe existir algún operador que permita reconstruir la Relación Original R . [www.03]¹

¹ http://www.google.com.ec/url?sa=t&source=web&cd=3&ved=0CBwQFjAC&url=http%3A%2F%2Fwww.pascualbravo.edu.co%2Fbuzon%2FTRABAJO%2520DE%2520BASES%2520DE%2520DATOS.ppt&rct=j&q=Completitud+%2B+reglas+de+fragmentacion+de+tablas&ei=d-4jTK_KF4KB8gbwv62_BQ&usg=AFQjCNF-ZylyOaCw_MezqcebMaKRYG3QWQ



Disyunción: Si la relación R se descompone en los fragmentos R_1 , R_2 , ..., R_n , y el dato di está en R_j , entonces, no debe estar en ningún otro fragmento. [www.03]⁴

Repetición y Fragmentación

Las técnicas de réplica y fragmentación se pueden aplicar sucesivamente a la misma relación de partida. Un fragmento se puede replicar y a su vez esa réplica ser fragmentada, para luego replicar alguno de esos fragmentos.

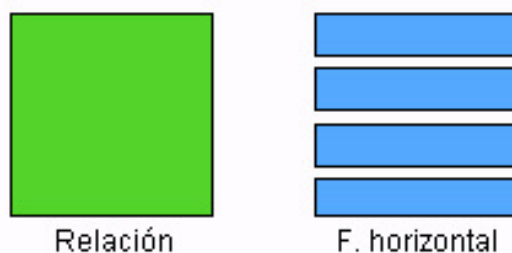
1.2. TIPOS DE FRAGMENTACIÓN.

La forma en cómo se pueden extraer los datos al ser consultados en un ambiente distribuido, se puede hacer una fragmentación de distintas tablas pertenecientes a diversas Bases de Datos localizadas en diversos servidores. Dado que una relación que corresponde esencialmente con una tabla esta se la puede dividir en fragmentos menores, inmediatamente surgen alternativas lógicas para llevar a cabo el proceso:

Existen tres tipos de fragmentación:

1. Fragmentación horizontal
2. Fragmentación vertical
3. Fragmentación híbrida

1.2.1. FRAGMENTACIÓN HORIZONTAL.



1. Fragmentación Horizontal [IMAG.01]²

Se realiza sobre las tuplas de la relación, es decir que cada fragmento será un subconjunto de las tuplas de la relación.

Una tabla T se divide en subconjuntos, T_1, T_2, \dots, T_n . Los fragmentos se definen mediante una operación de selección. Su reconstrucción se realizará mediante la unión de los fragmentos componentes. [www.04]³

Existen dos tipos de fragmentación como es: fragmentación horizontal primaria y la fragmentación horizontal derivada.

La fragmentación horizontal primaria. Consiste en particionar las tuplas (columnas) de una relación global en subconjuntos, donde cada subconjunto tenga propiedades comunes, se puede definir expresando cada fragmento como una operación de selección sobre la relación global.

² http://cablemodem.fibertel.com.ar/bucanero/facultad/Lab4_TP6.doc

³ <http://www.oei.eui.upm.es/Asignaturas/BD/DYOB/DISTRIBUIDAS.pdf>



Ejemplo: Considere la Relación J

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

J1 Escuela = CIME

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME

J2 Escuela = EISIC

Jno	NOMBRE	NOTA	ESCUELA
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

La fragmentación horizontal derivada Consiste en dividir una relación partiendo de los predicados definidos sobre alguna otra, debido a que la relación R depende de la relación Q, sobre cuyos atributos está definido el predicado de la fragmentación

Las tres entradas necesarias para desarrollar la fragmentación horizontal derivada son las siguientes: el conjunto de particiones de la relación propietaria, la relación miembro y el conjunto de predicados resultados de aplicar el semi-yunto entre la propietaria y la miembro.



Es decir La fragmentación de la tabla primaria, se aplica a tablas secundarias, o bien se debe partir de una fragmentación horizontal primaria.

Nota: La relación al final de la conexión es llamada *propietario* de la conexión y la relación que está en la cabeza de la conexión es llamada *miembro*.

Ejemplo: Considere las Relaciones

ID	MATERIA	ESCUELA	CRÉDITOS	NIVEL
1	Análisis Matemático	EISIC	4	1
2	Sistemas Operativos	EISIC	6	3
3	Programación II	CIME	6	2
4	Tecnología Eléctrica	EISIC	4	2
5	Técnicas de Aprendizaje	CIME	4	1
6	Dibujo Mecánico	CIME	6	3

Jno	NOMBRE	MATERIA	NOTA
J1	LUIS YANEZ	Sistemas Operativos	8
J2	ERIKA QUIROZ	Dibujo Mecánico	8
J3	DANIEL MURILLO	Técnicas de Aprendizaje	9
J4	MARIA JOSE MENDEZ	Análisis Matemático	10
J5	ADONIS PABON	Programación II	9

Esta relación puede ser fragmentada horizontalmente basada en la escuela en la cual el estudiante está matriculado. La escuela sin embargo no es un atributo del estudiante.

Escuela = CIME (Fragmentación Horizontal primaria)

ID	MATERIA	ESCUELA	CRÉDITOS	NIVEL
1	Análisis Matemático	EISIC	4	1
2	Sistemas Operativos	EISIC	6	3
4	Tecnología Eléctrica	EISIC	4	2

Fragmentación Horizontal derivada

Jno	NOMBRE	MATERIA	NOTA
J1	LUIS YANEZ	Sistemas Operativos	8
J4	MARIA JOSE MENDEZ	Análisis Matemático	10



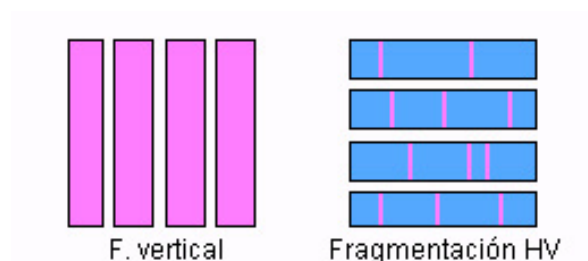
Escuela = EISIC (Fragmentación Horizontal primaria)

ID	MATERIA	ESCUELA	CRÉDITOS	NIVEL
3	Programación II	CIME	6	2
5	Técnicas de Aprendizaje	CIME	4	1
6	Dibujo Mecánico	CIME	6	3

Fragmentación Horizontal derivada

Jno	NOMBRE	MATERIA	NOTA
J2	ERIKA QUIROZ	Dibujo Mecánico	8
J3	DANIEL MURILLO	Técnicas de Aprendizaje	9
J5	ADONIS PABON	Programación II	9

1.2.2. FRAGMENTACIÓN VERTICAL



2. Fragmentación Vertical [IMAG.02]⁴

El objetivo de la fragmentación vertical consiste en dividir la relación en un conjunto de relaciones más pequeñas tal que algunas de las aplicaciones de usuario sólo hagan uso de un fragmento. Sobre este marco, una fragmentación óptima es aquella que produce un esquema de división que minimiza el tiempo de ejecución de las aplicaciones que emplean esos fragmentos.

⁴ http://cablemodem.fibertel.com.ar/bucanero/facultad/Lab4_TP6.doc



La fragmentación vertical se basa en los atributos de la relación para realizar la división, es decir: la subdivisión de atributos en grupos. La fragmentación es correcta si cada atributo se mapea en al menos un atributo del fragmento.

La partición vertical resulta más complicada que la horizontal. Esto se debe al aumento del número total de alternativas que tenemos disponibles.

Existen dos enfoques heurísticos para la fragmentación vertical de relaciones:

- ✓ **Agrupación:** Comienza asignando cada atributo a un fragmento, y en cada paso, junta algunos de los fragmentos hasta que satisface un determinado criterio. La agrupación surgió en principio para bases de datos centralizadas y se usó posteriormente para las bases de datos distribuidas.

- ✓ **Escisión:** A partir de la relación se deciden que fragmentos resultan mejores, basándose en las características de acceso de las aplicaciones a los atributos. Esta técnica se presentó, también, para bases de datos centralizadas. Posteriormente, se extendió al entorno distribuido.



La escisión genera fragmentos no solapados mientras que la agrupación normalmente produce fragmentos solapados. Dentro del contexto de los sistemas de bases de datos distribuidos, son preferibles los fragmentos no solapados por obvias razones. Los fragmentos no solapados se refieren únicamente a atributos clave no primarios.

Ejemplo: Considere la Relación J

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

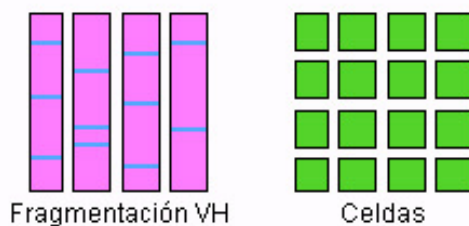
J1 Información de Notas

Jno	NOTA
J1	8
J2	8
J3	9
J4	10
J5	9

J2 Información de Nombres y Escuelas

Jno	NOMBRE	ESCUELA
J1	LUIS YANEZ	CIME
J2	ERIKA QUIROZ	CIME
J3	DANIEL MURILLO	EISIC
J4	MARIA JOSE MENDEZ	EISIC
J5	ADONIS PABON	EISIC

1.2.3. FRAGMENTACIÓN MIXTA



3. Fragmentación Mixta [IMAG.03]⁵

La fragmentación mixta puede llevarse a cabo de tres formas diferentes: desarrollando primero la fragmentación vertical y, posteriormente, aplicando la fragmentación horizontal sobre los fragmentos verticales (denominada partición VH), o aplicando primero una división horizontal para luego, sobre los fragmentos generados, desarrollar una fragmentación vertical (llamada partición HV), o bien, de forma directa considerando la semántica de las transacciones.

Ejemplo: Considere la Relación J

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

⁵ http://cablemodem.fibertel.com.ar/bucanero/facultad/Lab4_TP6.doc



Fragmentación Horizontal Previa

Escuela = EISIC

Jno	NOMBRE	NOTA	ESCUELA
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

Información de Notas (Fragmentación Vertical)

Jno	NOTA
J3	9
J4	10
J5	9

1.2.4. GRADOS DE FRAGMENTACIÓN

El grado de fragmentación puede considerarse en el que cada tupla o atributo forme un fragmento, o una ausencia de la división, considerando a las relaciones como unidades de fragmentación

Se debe buscar un compromiso intermedio, en el que se debería establecerse sobre las características de las aplicaciones que hacen uso de la base de datos, sobre estas una serie de parámetros, se podrá establecer el grado de fragmentación del banco de datos de acuerdo con sus valores.



REGLAS DE CORRECCIÓN DE LA FRAGMENTACIÓN.

Asegurarán la ausencia de cambios semánticos en la base de datos, estas son las tres reglas que se debe cumplir en el proceso de fragmentación.

- 1. Compleción.** Si una relación R se descompone en una serie de fragmentos R_1, R_2, \dots, R_n , cada elemento de datos que pueda encontrarse en R deberá poder encontrarse en uno o varios fragmentos R_i .

$$1 \leq i \leq n \text{ (No hay pérdida de información)}$$

Asegura que los datos de la relación global se proyectan sobre los fragmentos sin pérdida alguna.

En el caso de fragmentación horizontal el elemento de datos, normalmente, es una tupla, mientras que en el caso de fragmentación vertical es un atributo. [www.05]⁶

- 2. Reconstrucción.** Si una relación R se descompone en una serie de fragmentos R_1, R_2, \dots, R_n , puede definirse un operador relacional ∇ tal que el operador será diferente dependiendo de las diferentes formas de fragmentación.

$$R = \nabla R_i \quad \forall R_i \in F_R$$

⁶http://html.rincondelvago.com/bases-de-datos-distribuidas_1.html



La reconstrucción de la relación a partir de sus fragmentos asegura la preservación de las restricciones definidas sobre los datos en forma de dependencias. [www.06]º

3. Disyunción. Si una relación R se descompone horizontalmente en una serie de fragmentos R1, R2, ..., Rn, y un elemento de datos d_i se encuentra en algún fragmento Rj, entonces no se encuentra en otro fragmento Rk ($k \neq j$). Esta regla asegura que los fragmentos horizontales sean disjuntos. Si una relación R se descompone verticalmente, sus atributos primarios, clave normalmente, se repiten en todos sus fragmentos. [www.07]º

Ejemplo:

Considere la Relación J

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

J1 Escuela = CIME

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME

J2 Escuela = EISIC

Jno	NOMBRE	NOTA	ESCUELA
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC



Compleción: No hay pérdida de datos ya que todas las tuplas se encuentran fragmentadas en J1 y J2

Reconstrucción: Se puede reconstruir mediante la unión de todos los fragmentos, es decir $J1 \cup J2$

Disyunción: En la fragmentación $J1$ Escuela = CIME no se va a encontrar EISIC, y en $J2$ Escuela = EISIC no se va a encontrar CIME, entonces J_k ($k \neq j$).

ALTERNATIVAS DE ASIGNACIÓN.

Una vez que se ha realizado la fragmentación correctamente se debe decidir donde serán asignados cada uno de los fragmentos.

Los datos pueden replicarse para mantener una copia al momento que son asignados. La seguridad y a la eficiencia de las consultas de lectura son una de las razones para la réplica. En caso de fallo en el sistema se podría acceder a esos datos ubicados en sitios distintos, además las consultas que acceden a los mismos datos pueden ejecutarse en paralelo, ya que habrá copias en diferentes sitios.

La ejecución de actualizaciones, de escritura, implicará la actualización de todas las copias que existan, este proceso puede ser complicado y problemático, esto se puede resolver calculando la cantidad de consultas de lectura y el número de consultas de



escritura que se llevarán a cabo. Donde las consultas que se llevan a cabo sean la mayor parte de lectura, se podría alcanzar un alto grado de réplica, no así en el caso contrario. Una base de datos fragmentada es aquella donde no existe réplica alguna.

A lo largo de la red los fragmentos se alojan en sitios donde únicamente existe una copia de cada uno de ellos.

Una base de datos se puede considerarla totalmente replicada, cuando existe una copia de todo el banco de datos en cada sitio, o parcialmente replicada cuando existen copias de los fragmentos ubicados en diferentes sitios.

INFORMACIÓN NECESARIA PARA LA FRAGMENTACIÓN

En el diseño de distribución, un aspecto importante es la cantidad de factores, como es: las características de acceso de las aplicaciones a la base de datos, las características del sistema en cada sitio, la organización lógica de la base de datos, la localización de las aplicaciones, éstas tienen una influencia sobre la distribución.

Se divide en las siguientes categorías la información necesaria para el diseño de la distribución.

- La información del banco de datos.
- La información de la aplicación.



- La información sobre la red de ordenadores
- La información sobre los ordenadores en sí.

La información sobre la red de ordenadores y la información sobre los ordenadores en sí, son de carácter cuantitativo y servirán, principalmente, para desarrollar el proceso de asignación.

INFORMACIÓN NECESARIA PARA LA FRAGMENTACIÓN HORIZONTAL

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

Es importante saber cómo están conectadas las relaciones de la base de datos una con otra, es decir se refiere a la información del esquema conceptual de la base de datos.

La relación al final de la conexión (relación de origen) es llamada propietario de la conexión y la relación que está en la cabeza de la conexión (relaciones destino) es llamada miembro.

- **Información sobre la aplicación.**

Es Necesario información cualitativa y cuantitativa.

La información cualitativa guiará la fragmentación, mientras que la cuantitativa es requerida en los modelos de asignación.

Referente a la información cuantitativa, se necesita definir dos conjuntos de datos que son:



- ✓ Selectividad mintérmino. Es el número de tuplas de una relación a las que accede una consulta de acuerdo a un predicado mintérmino dado.
- ✓ Frecuencia de acceso. Es la frecuencia con la que un usuario accede a los datos.

INFORMACIÓN NECESARIA PARA LA FRAGMENTACIÓN VERTICAL

La principal información para la fragmentación vertical, es la relacionada a las aplicaciones. Debido a que en el particionamiento vertical coloca en un fragmento aquellos atributos los cuales son accesados juntos.

Los principales datos requeridos relacionados a las aplicaciones son sus frecuencias de acceso.

INFORMACIÓN NECESARIA PARA LA FRAGMENTACIÓN MIXTA

En la etapa de asignación necesitamos los datos cuantitativos acerca de la base de datos, las comunicaciones de la red, las aplicaciones que funcionan sobre ella, el limitantes de almacenamiento para cada sitio en la red y las capacidades de procesamiento.

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

Para llevar a cabo una fragmentación horizontal definimos la selección de mintérminos, ahora necesitamos extender la



definición de fragmentos, y definir la selección de un fragmento **F_j** con respecto a una consulta query **q_i**. Este es el número de tuplas de **F_j** que se necesitan ser accesadas para procesar **q_i**.

- **Información de la aplicación.**

Durante la actividad de la fragmentación, la mayoría de la información relacionada a la aplicación ya ha sido compilada, pero en algunos de los casos es requerido por el modelo de asignación

- **Información del sitio.**

Necesitamos conocer la capacidad de almacenamiento y procesamiento de cada sitio, estos valores pueden calcularse por medio de funciones o por estimaciones.

- **Información de la red.**

En nuestro modelo asumimos la existencia de una simple red donde el costo de comunicación es definido en términos de un marco de datos.