



TAREA #2

Bases de Datos Distribuidas

Descripción breve

Fragmentación horizontal primaria en bases de datos relacionales de Tamer Ozsu

- Urrutia González Brenda
 - Olea García Alan
- Juárez Anguiano Mario Alexis

Equipo 03
02/06/2025

Fragmentación Horizontal Primaria en Bases de Datos Relacionales

(Basado en Tamer Özsu, *Principles of Distributed Database Systems*, 3ª ed., 2011)

La fragmentación horizontal primaria consiste en dividir una relación (tabla) en subconjuntos de filas (tuplas), llamados fragmentos horizontales. Cada fragmento contiene un subconjunto de las filas de la relación original, de acuerdo a ciertos criterios, como condiciones de selección.

Ejemplo sencillo:

Si tenemos una tabla Empleados, podríamos fragmentarla en:

- Empleados_México: todos los empleados cuya ubicación sea México.
- Empleados_USA: todos los empleados cuya ubicación sea EE.UU.

Características clave:

- Basada en condiciones de selección: cada fragmento se obtiene aplicando un predicado de selección sobre la relación original.
- Propiedades para que sea correcta:
 1. Completitud: la unión de todos los fragmentos debe reconstruir la relación original.
 2. Reconstructibilidad: debe poder reconstruirse la relación original a partir de los fragmentos.
 3. No solapamiento: cada tupla debe aparecer en un solo fragmento (a menos que sea una fragmentación redundante).

Ejemplo en términos de predicados:

La fragmentación horizontal primaria se basa en predicados de selección que determinan qué tuplas van en cada fragmento. Cada fragmento se obtiene aplicando un predicado específico a la relación original.

Por ejemplo, supongamos que tenemos la tabla Empleados con los siguientes atributos:

Empleados(ID, Nombre, Departamento, País)

Queremos fragmentar esta tabla en función del país donde trabaja cada empleado. Podemos definir dos predicados de selección:

p1: País = 'México'

p2: País = 'EE.UU.'

Entonces, la fragmentación horizontal primaria se expresa así:

Empleados_México = $\sigma_{\{\text{País}=\text{'México'}\}}(\text{Empleados})$

Empleados_USA = $\sigma_{\{\text{País}=\text{'EE.UU.'}\}}(\text{Empleados})$

Esto significa:

- Empleados_México contendrá todas las filas (tuplas) donde el país sea 'México'.
- Empleados_USA contendrá todas las filas donde el país sea 'EE.UU.'.

Ventajas:

- Mejora la eficiencia: las consultas pueden acceder solo a los fragmentos relevantes.
- Facilita la distribución de la base de datos: cada fragmento puede almacenarse cerca de donde más se usa.

Bibliografías

Özsu, M. T., & Valduriez, P. (2011). *Principles of distributed database systems (3rd ed.)*. Springer.