



INSTITUTO POLITECNICO NACIONAL
UNIDAD PROFESIONAL INTERDISCIPLINARIA EN
INGENIERIA Y TECNOLOGIAS AVANZADAS

**Tarea 2: Fragmentación horizontal primaria en
base de datos relacionales**

**INGENIERIA TELEMÁTICA
BASE DE DATOS DISTRIBUIDAS**

PROFESOR: De la Cruz Sosa Carlos

ALUMNOS:

Legorreta Rodriguez Maria Fernanda
Macías Galván Arturo Daniel
Palacios Reyes Leslie Noemi

GRUPO: 3TM3
SEMESTRE 2025/2

1. Introducción

En los sistemas de bases de datos distribuidas, uno de los objetivos principales es optimizar el acceso a los datos minimizando el tiempo de respuesta y reduciendo los costos de comunicación entre sitios. Para ello, se emplean técnicas como la **fragmentación de datos**, las cuales permiten dividir lógicamente la base de datos y asignar sus partes a diferentes nodos o servidores. En este documento, se aborda el concepto de **fragmentación horizontal primaria**, su importancia en entornos distribuidos, y se describe su diseño con base en las recomendaciones del autor **Tamer Özsu**, un referente en sistemas distribuidos.

2. Fragmentación Horizontal Primaria según Tamer Özsu

La **fragmentación horizontal primaria** consiste en dividir una tabla relacional en múltiples fragmentos, cada uno conteniendo un subconjunto de filas de la tabla original. Esta división se basa en condiciones de predicado (por ejemplo, departamento = 'Ventas' o país = 'México'). Todos los fragmentos resultantes mantienen el mismo esquema de columnas, pero contienen distintas tuplas según los criterios definidos.

Cada fragmento se almacena y administra de forma independiente, lo cual facilita su asignación a diferentes sitios físicos dentro de un sistema distribuido. Esta técnica no solo mejora el rendimiento al distribuir la carga, sino que también reduce la congestión y el tiempo de acceso cuando los datos son consultados localmente.

Ejemplo práctico:

Supongamos una tabla Empleados que contiene información de personal en distintas regiones. Si se realiza una fragmentación horizontal basada en la ubicación geográfica (país), podríamos obtener:

- Empleados_Mexico (donde país = 'México')
- Empleados_EEUU (donde país = 'Estados Unidos')
- Empleados_Canada (donde país = 'Canadá')

Esto permite que cada sede de la empresa tenga acceso eficiente solo a los datos que necesita.

3. Diseño de Fragmentos Primarios

Para diseñar fragmentos primarios de manera efectiva, se deben seguir los siguientes pasos:

3.1 Análisis de Patrones de Acceso

Se debe realizar un análisis exhaustivo sobre cómo se utilizan los datos:

- ¿Qué consultas son las más comunes?
- ¿Qué atributos se usan frecuentemente como filtros?
- ¿Desde qué ubicaciones o departamentos se accede a los datos?

3.2 Selección del Tipo de Fragmentación

- **Fragmentación Horizontal Primaria:**
Se crean fragmentos según predicados que aplican a las filas. Cada fragmento contiene la totalidad de columnas pero un subconjunto de tuplas.
- **Fragmentación Vertical Primaria (para referencia):**
Se crean fragmentos según subconjuntos de columnas. Cada fragmento incluye la clave primaria para asegurar la reconstrucción de la tabla original.

3.3 Validación de los Fragmentos

Para garantizar la calidad del diseño, cada fragmento debe cumplir con las siguientes propiedades:

- **Compleitud:**
La unión de todos los fragmentos debe reconstruir la tabla original sin pérdida de datos.
- **Reconstruibilidad:**
Se debe poder recomponer la tabla original utilizando operaciones relacionales estándar, como UNION (en horizontal) o JOIN (en vertical).
- **No Redundancia** (opcional):
Evitar la duplicación innecesaria de datos entre fragmentos, salvo que se requiera por cuestiones de disponibilidad.

3.4 Asignación de Fragmentos a Sitios

Una vez diseñados, los fragmentos se deben asignar estratégicamente a los nodos donde serán consultados con mayor frecuencia. Esto reduce la latencia y mejora el rendimiento gracias a la **localidad de acceso**.

4. Costo Mínimo de Comunicación (COM_MIN)

En el contexto de la fragmentación en bases de datos distribuidas, el parámetro **COM_MIN** se refiere al **costo mínimo de comunicación** necesario para ejecutar operaciones distribuidas (como consultas o actualizaciones) entre los distintos sitios que almacenan fragmentos.

Este parámetro se considera en el diseño para:

- Reducir la transferencia de datos entre nodos.
- Optimizar la ejecución de consultas distribuidas.
- Seleccionar las mejores estrategias de fragmentación y replicación.

Objetivo: minimizar el tráfico de red y garantizar un rendimiento óptimo del sistema distribuido.

5. Conclusión

La fragmentación horizontal primaria representa una técnica fundamental en el diseño de bases de datos distribuidas. Siguiendo una estrategia bien definida —como la propuesta por Tamer Özsu—, se puede lograr una asignación eficiente de datos, mejorar el rendimiento de las consultas y reducir costos de comunicación. Un diseño correcto permite a las organizaciones escalar sus sistemas de forma efectiva, manteniendo la integridad y disponibilidad de la información a través de múltiples nodos.