

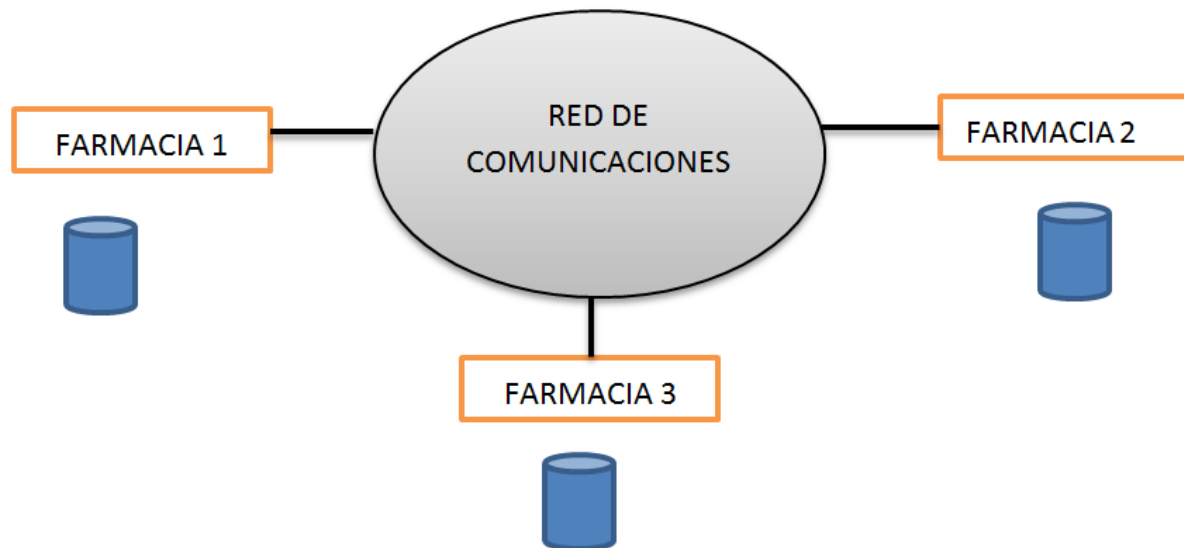
Base de Datos Distribuidas

Prof. José Alejandro Reyes Ortiz

“REPLICACIÓN”



Base de datos fragmentada de un hospital



Objetivos de la sesión

Al finalizar la sesión, deberán ser capaces de:

- Definir el concepto de replicación de base de datos distribuidas.
- Identificar y explicar los tipos de replicación.
- Identificar y explicar las ventajas y desventajas de la replicación.
- Explicar los modelos de replicación activa y pasiva.

A qué se refiere la replicación ...

- En base de datos distribuidas

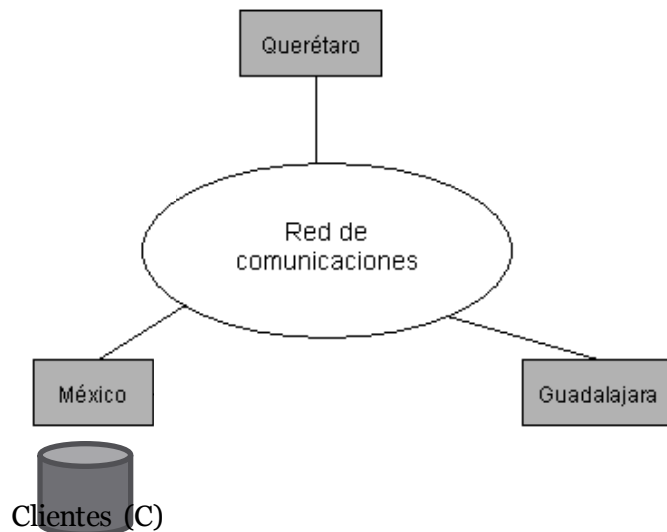
“se refiere al almacenamiento de copias en múltiples sitios (servidores) por una red de computadoras”

Pueden guardarse copias en varios sitios para satisfacer requerimientos de información específicos.

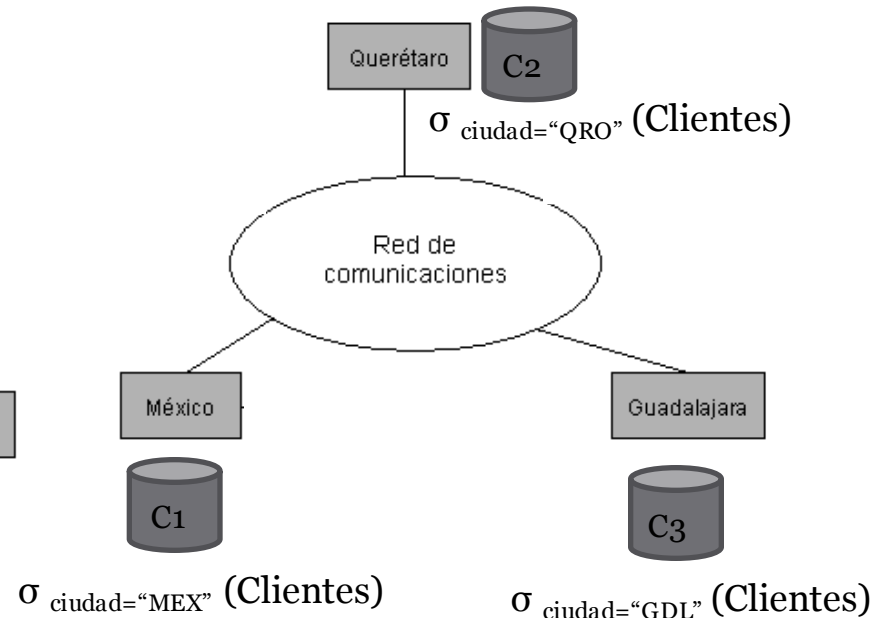
¿Cómo se lleva a cabo la replicación?

- Una base de datos distribuida, está dividida o fragmentada.
- Un tipo de fragmentación se presenta a nivel de los datos.

Base de datos centralizada

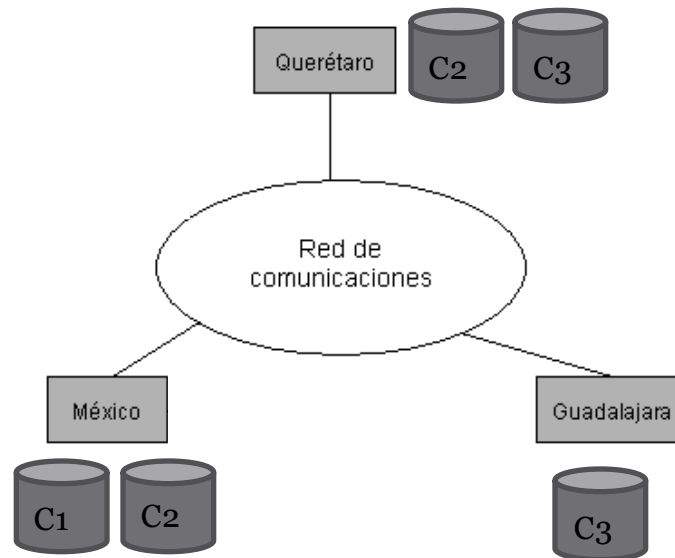


Base de datos distribuida



¿Cómo se lleva a cabo la replicación?

- La base de datos Clientes (C) está dividida en tres fragmentos: C1, C2, C3.
- La **replicación** en una base de datos distribuida implica que C2 y C1 se almacenan en el sitio “México”, mientras que en “Querétaro” se almacenan los fragmentos C2 y C3.



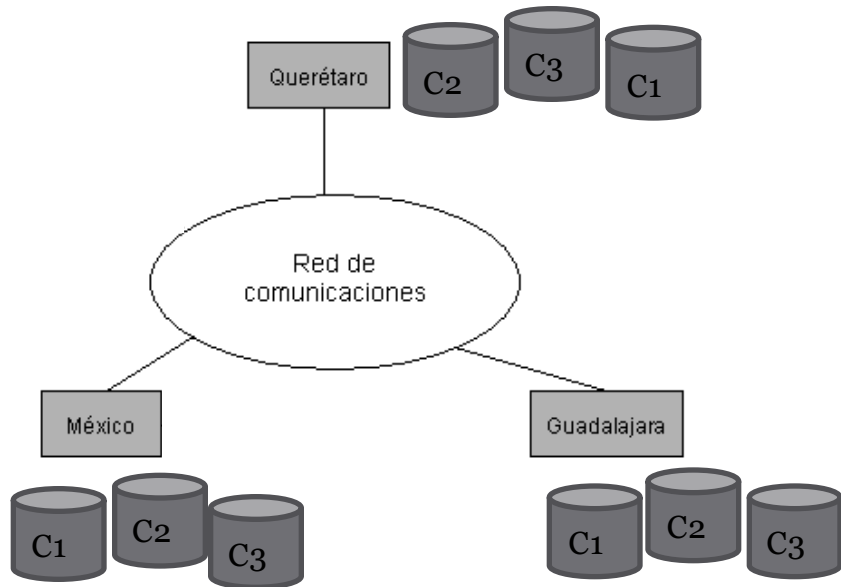
Base de datos distribuida replicada

Regla en la replicación...

- Los datos replicados se deben someter a la **regla de consistencia mutua**, la cual requiere que todas las copias de fragmentos de datos sean idénticas.
- Para mantener la **consistencia de los datos** entre replicas, se debe garantizar la actualización de la base de datos en todos los sitios donde existen replicas.
- Una operación **read (lectura)** selecciona cualquier copia (de preferencia la más cercana) para satisfacer la transacción.
- Una operación **write (escritura)** requiere que todas las copias se seleccionen y actualicen para satisfacer la regla de consistencia mutua.

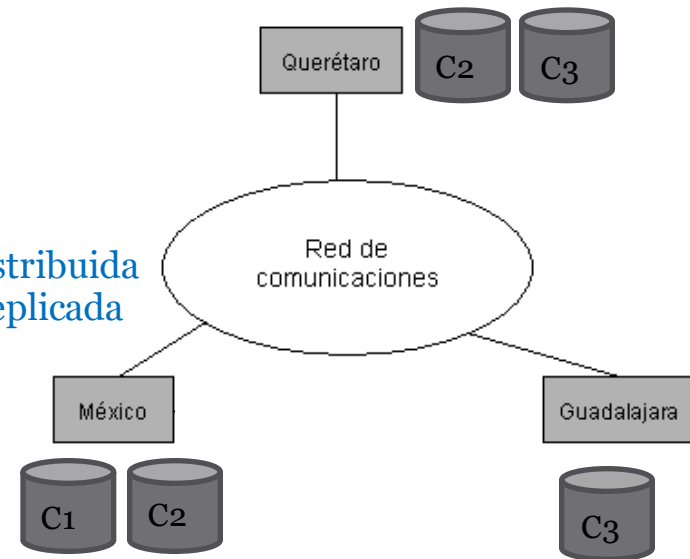
Tipos de replicación

- Una base de datos distribuida **totalmente replicada** guarda copias de todos los fragmento de la base de datos en múltiples sitios.
- Una base de datos distribuida **parcialmente replicada** guarda copias de algunos fragmentos de la base de datos en múltiples sitios.
- Una base de datos distribuida **no replicada** guarda cada fragmento de base de datos en un solo sitio.



Base de datos distribuida totalmente replicada

Base de datos distribuida parcialmente replicada



Base de datos distribuida no replicada



Comparación de las características en las estrategias de replicación

	Replicación total	Replicación parcial	No replicada
Paralelismo	Muy alto	Alto	Bajo
Disponibilidad	Muy alta	Alta	Baja
Costo	Elevado	Moderado	Bajo
Riesgo de pérdida de datos	Bajo	Medio	Alta
Realidad	Aplicación posible	Realista	Aplicación posible

Factores para la replicación

Varios factores influyen en la decisión de utilizar replicación de datos:

- Tamaño de la base de datos
- Frecuencia de uso.
- Costo—desempeño. Considerar el costo del software de coordinación, transacciones, replicación y software de administración de las base de datos.

Ventajas de la replicación

- **Mayor disponibilidad.** El fragmento F_i puede encontrarse en varios sitios.
- **Confiabilidad.** Seguir respondiendo a pesar del fallo de un sitio.
- **Paralelismo incrementado.** En caso de que el acceso al fragmento F_i sea de sólo lectura, varios sitios pueden procesar, en paralelo, las lecturas que impliquen a F_i .
- Cuantas más replicas de F_i haya mayor será la disponibilidad de que los datos se encuentren en el sitio que los solicitó.

Desventajas de la replicación

- **Sobrecarga** durante la actualización. Se debe asegurar la consistencia mutua, en caso contrario, pueden producirse cálculos erróneos. Para ello, se debe propagar la actualización a todos los sitios que contienen réplicas.
- **Mayor tiempo de procesamiento:** el intercambio de mensajes hacia las réplicas y los cálculos adicionales suponen una forma de tiempo extra que debe ser considerada.
- **Costos**
 - Costos del desarrollo de software: es necesario un software para la administración de la replicación.
 - Costos de almacenamiento de las réplicas.
 - Costos de procesamiento.

Disponibilidad de un Sistema de Bases de Datos Distribuidas

- La proporción de tiempo que un sistema está accesible con tiempos de respuesta razonables se le conoce como **disponibilidad**.
- La disponibilidad debe ser **cercana al 100%**.
- **Factores de pérdida de disponibilidad**
 - Fallos en el servidor
 - Particiones de red o desconexiones
 - Desconexión intencionada o no intencionada.

Disponibilidad de un Sistema de Bases de Datos Distribuidas

Fallos en el servidor de base de datos distribuidas

- Si un servidor tiene una posibilidad de fallo p del 5%, tendrá disponibilidad del 95%
- Si replicamos n veces el servidor, la disponibilidad será
$$1-p^n$$
- Con $n=2$ servidores: $1 - 0.05^2 = 99.75\%$

Disponibilidad de un Sistema de Bases de Datos Distribuidas

Considere una posibilidad de fallo del servidor de bases de datos del 25%. Calcular la disponibilidad de las aplicaciones en dos casos: si el número de réplicas es dos y tres respectivamente.

Disponibilidad de un Sistema de Bases de Datos Distribuidas

Calcular la disponibilidad del Sistema de Bases de Datos Distribuidas de la empresa “SeguridadCDMX S.A. De C.V”, considerando el siguiente escenario:

La empresa tiene su base de datos particionada en dos fragmentos (incidentes en proceso de investigación e casos resueltos) y con una replicación total, los cuales son asignados a los dos servidores de bases de datos con que cuenta la empresa:

- a) Uno creado por sus desarrolladores en las oficinas centrales. Este servidor tiene una disponibilidad del 85 %
- b) Un segundo servidor contratado en un proveedor en la nube, con una disponibilidad de 95 %

Tolerancia a fallos y replicación

- Una alta disponibilidad no implica necesariamente datos correctos.
 - Puede haber datos no actualizados, o inconsistentes.
- Podemos utilizar replicación para ganar tolerancia a fallos

Requisitos de la replicación

La replicación debe llevarse a cabo considerando

- **Transparencia:** los clientes no son conscientes de que hay múltiples copias del recurso al que acceden
- **Consistencia:** las operaciones sobre un conjunto de objetos replicados deben dar resultados que sigan la especificación de corrección definida para dichos objetos.

Reglas en la replicación...

- Los datos replicados se deben someter a la **regla de consistencia mutua**, la cual requiere que todas las copias de fragmentos de datos sean idénticas.
- Para mantener la **consistencia de los datos** entre replicas, se debe garantizar la actualización de la base de datos en todos los sitios donde existen replicas.
- Una operación **read (lectura)** selecciona cualquier copia (de preferencia la más cercana) para satisfacer la transacción.
- Una operación **write (escritura)** requiere que todas las copias se seleccionen y actualicen para satisfacer la regla de consistencia mutua.

Escenarios...

- BD distribuida (fragmentada) y sin réplica
 - Ejemplo: La base de datos de incidentes delictivos de la CDMX se encuentran fragmentados por la zona en que fueron cometidos (norte, sur, oriente y poniente), y cada fragmento se encuentra en **solo un servidor**.
- BD distribuida (fragmentada) y con réplica
 - Ejemplo: La base de datos de incidentes delictivos de la CDMX se encuentran fragmentados por la zona en que fueron cometidos (norte, sur, oriente y poniente), y, **al menos, un fragmento se encuentra en dos servidores**.