

TEMA 10. BASES DE DATOS REPLICADAS

La replicación de bases de datos es un proceso consistente en la copia y el mantenimiento de objetos de la base de datos en múltiples bases de datos que forman un sistema de bases de datos distribuido. Permite a las organizaciones proporcionar a los usuarios acceso a los datos en todo lugar y en todo momento. Los cambios aplicados a un nodo se captura y almacenan localmente antes de reenviarlos y aplicarlos a cada una de las ubicaciones remotas.

La replicación utiliza tecnología de bases de datos distribuidas para compartir datos entre múltiples sitios. En una BD replicada los mismos datos están disponibles en múltiples ubicaciones. Los beneficios de la replicación de bases de datos son:

- Disponibilidad: Los datos están disponibles en múltiples ubicaciones, permitiendo el acceso a los usuarios y aplicaciones mediante la provisión de opciones alternativas de acceso a los datos.
- Fiabilidad: Al haber múltiples copias de los datos, se dispone de mecanismos excelentes de recuperación de los datos en caliente en caso de fallo.
- Rendimiento: Se incrementa el desempeño de las operaciones de consulta.
- Reducción de carga: El acceso se distribuye entre diferentes servidores y el coste disminuye (se puede acceder al servidor cuyo coste de acceso sea menor).
- Procesamiento offline: La replicación se puede implementar haciendo uso de instantáneas (Copias completas o parciales de la base de datos en un momento dado, conocido como snapshot: Los usuarios trabajan sobre un subconjunto de la base de datos mientras están desconectados y, cuando se restablece la conexión, se lleva a cabo la sincronización).
- Soporte multiusuario: La replicación permite generar muchas instantáneas personalizadas para las necesidades específicas de cada usuario.
- Soporta aplicaciones avanzadas: OLTP, OLAP, Data Mining, Mobile Computing.

Los inconvenientes de la replicación de bases de datos son:

- Mayor complejidad del SGBD que los sistemas centralizados.
- Sobrecarga debido a las operaciones de actualización.
- Mayor complejidad para la gestión del control de concurrencias y recuperación de la base de datos.

Las aplicaciones de la replicación son:

- Sincronización limitada: Equipos de ventas remotos que generalmente trabajan desconectados, con sincronización periódica de un gran número de nodos móviles.
- Sincronización continua: Aplicaciones financieras y gestiones bursátiles (sincronización continua e instantánea).

Los componentes básicos de la replicación son:

- Objeto de la replicación: Un objeto de la base de datos que existe en múltiples servidores de un sistema de bases de datos distribuido.
- Grupos de replicación: Colección de objetos de replicación que están lógicamente relacionados (facilita la administración de los objetos y pueden existir en múltiples sitios de replicación).
- Sitios de replicación con respecto a los grupos:
 - Maestros: Controla uno o más grupos de replicación (mantiene y propaga copias del grupo de replicación a los sitios esclavos), y se comunican entre ellos para propagar continuamente los cambios realizados en el grupo de replicación.
 - Esclavos: Pueden contener la copia del grupo de replicación o parte de ella, y contiene una instantánea del grupo de replicación.

Los entornos de la replicación son:

- Replicación síncrona: Los datos replicados se actualizan de forma inmediata al actualizar los datos de origen (protocolo de confirmación en 2 fases denominado 2PC, carga elevada de la red corporativa por elevado número de mensajes, y todos los nodos deben estar online).
- Replicación asíncrona: Hay un retardo entre la actualización de datos origen y destino, y organizaciones que trabajan con réplicas que no necesitan estar actualizadas.

Propiedad de los datos: Información acerca del sitio que mantiene el privilegio de actualización de los datos.

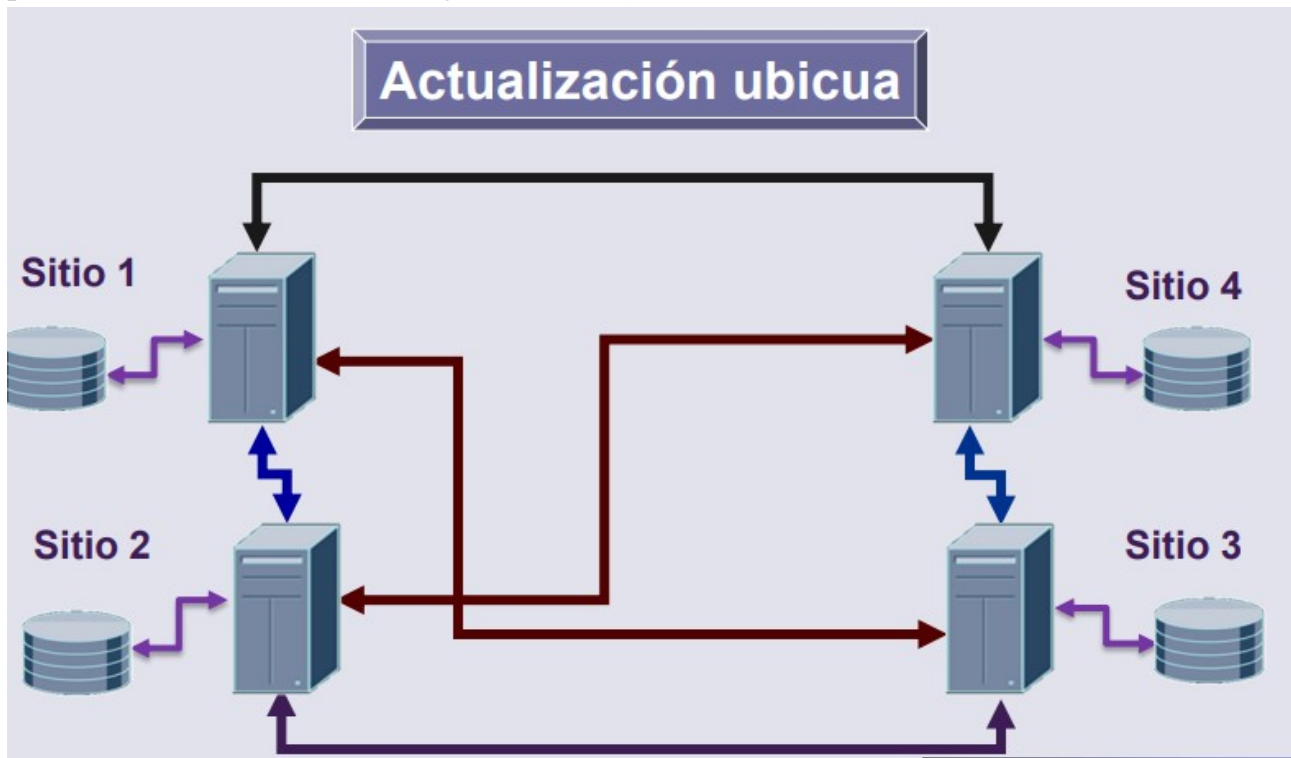
Propiedad Maestro/Esclavo: Los datos son asíncronamente replicados por el sitio maestro que tiene privilegios de actualización de los mismos. Mediante la publicación/suscripción el sitio maestro publica los datos y los sitios esclavos se suscriben, recibiendo copias de sólo lectura. Potencialmente cada sitio puede ser maestro de una serie de conjuntos de datos no solapados. No existen copias de datos iguales en diferentes sitios maestros: Un maestro contiene la extensión completa de una tabla. La extensión de una tabla es fragmentada entre distintos maestros (replicación asimétrica).



Propiedad Flujo de trabajo: Al igual que maestro/esclavo impide los conflictos de actualización aunque es más dinámico, permite que el privilegio de actualización se transfiera entre sitios, aunque en cada momento sólo un sitio tiene ese privilegio.

Propiedad de tipo de actualización ubicua (replicación simétrica): Entorno igualitario donde los múltiples nodos disponen de los mismos derechos para actualizar los datos replicados (permite que los sitios locales operen de forma autónoma, incluso cuando

otros sitios no están disponibles). Puede conducir a conflictos (incorporan procedimientos de detección y resolución).



El servidor de replicación es una técnica alternativa y potencialmente más eficiente que la distribución de los datos. Permite la actualización síncrona y asíncrona.

- Escalabilidad: Debe gestionar la replicación tanto de pequeños como de grandes volúmenes de datos.
- Mapeado y transformación: Debe ser capaz de gestionar la replicación entre distintos SGBD y plataformas.
- Replicación de objetos: Debe ser posible replicar objetos además de los datos.
- Especificación del esquema de replicación: Debe proporcionar mecanismos para permitir que un usuario con privilegios especifique los objetos y datos que se desea replicar.
- Mecanismo de suscripción: Debe proporcionar un mecanismo que permita la inicialización de una réplica de destino.
- Fácil administración: Simplicidad para el DBA de administrar el sistema, comprobar el estado, y monitorizar el rendimiento de los componentes del sistema de replicación.

El servidor de replicación presenta los siguientes problemas:

- Actualizaciones transaccionales: La estructura de la transacción en la base de datos origen se mantiene en la base de datos destino.
- Instantáneas: Permiten la replicación del estado de la base de datos origen de acuerdo a una planificación predefinida (calendario). Utilizan el registro de

operación de la base de datos para detectar las modificaciones en los datos origen. Implementan una gestión de colas para las actualizaciones de las réplicas.

- Disparadores: Es responsabilidad del usuario generar código dentro de un disparador que se ejecute cada vez que se produzca un suceso determinado. Es una técnica más flexible que las instantáneas pero presenta problemas (consumo de recursos adicionales, no puede planificarse y se ejecutan siempre que se produce el suceso, y es complejo cuando se actualizan muchas réplicas).
- Detección y resolución de conflictos: Envío de datos antiguos y datos nuevos (comprobación en el sitio esclavo). Se han propuesto los siguientes mecanismos:
 - Marcas temporales inferiores y superiores.
 - Prioridad de los sitios.
 - Actualizaciones aditivas y promediadas: Los datos satisfacen esta propiedad.
 - Valores mínimo y máximo.
 - Definido por el usuario.
 - Resolución manual.

Se han propuesto las siguientes soluciones:

- Nominación sitio + valor duplicado.
- Codificar (número de secuencia) el valor duplicado.
- Desestimar el valor duplicado.