

▼ ¿Por qué sistemas de bases de datos distribuidas?

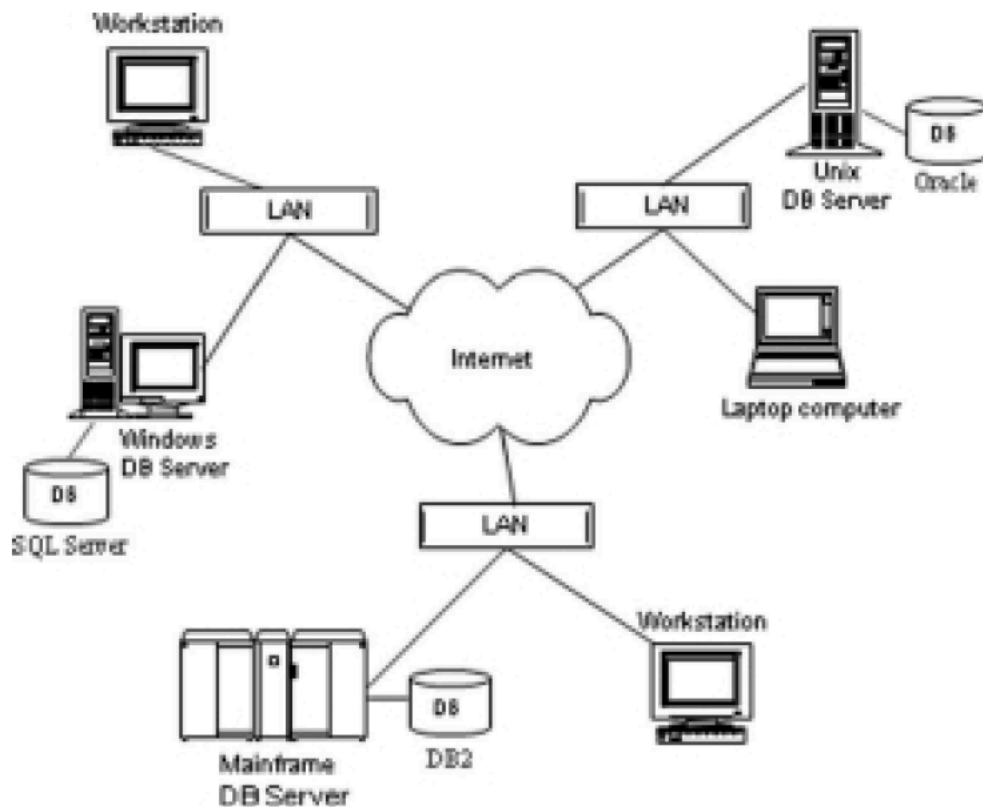
Con la amplia prevalencia de los sistemas de bases de datos centralizados, los usuarios y los programas de aplicación acceden a la base de datos desde sitios locales, así como desde localidades remotas. En contraste, una **base de datos distribuida** (*distributed database*) no se almacena completamente en una localidad central, sino que se distribuye en una red de localidades que pueden estar geográficamente separadas y conectadas por enlaces de comunicaciones. Cada localidad tendrá su propia base de datos y está capacitada para acceder a los datos de otras localidades.

Hay varias razones para desarrollar y usar un **sistema de bases de datos distribuida** entre las que se incluyen las siguientes:

1. A menudo, las organizaciones tienen ramales o divisiones en diferentes localidades. Para una localidad dada, L , pudiera haber un conjunto de datos que use frecuentemente y quizás exclusivamente. Además L puede usar datos que se usen con más frecuencia en otra localidad L' .
 Por ejemplo, en un negocio de venta al por menor cada tienda se puede beneficiar de su propia base de datos de inventarios, ventas, cuentas de clientes y empleados. Durante las operaciones de negocios, las transacciones se procesan convenientemente en la base de datos local. Al final de las operaciones del día se debe transmitir un resumen con los resultados a las oficinas centrales de la corporación en donde se mantiene una base de datos general. Periódicamente cada tienda puede beneficiarse de tener acceso a la base de datos general y poder comparar sus ventas y ganancias con las de las demás tiendas como medida de su rendimiento.
 En un sistema de base de datos centralizado cada sitio tendría que estar usando enlaces de comunicación con la base de datos para ambos tipos de información, por lo que las comunicaciones pueden convertirse en un cuello de botella importante.
2. Permitir que cada sitio almacene y mantenga su propia base de datos facilita el acceso inmediato y eficaz a los datos que se usan con más frecuencia. Tales datos podrían usarse en otros sitios, pero con menos frecuencia. Datos similares almacenados en otras localidades también podrían accederse si fuese necesario.
3. Las bases de datos distribuidas pueden mejorar la fiabilidad. Si los computadores de un sitio fallan, o si se queda fuera de servicio algún enlace de comunicación, el resto de la red puede seguir funcionando. Es más, si los datos están duplicados en más de un sitio, los datos necesarios pueden estar disponibles en otro lugar que aún se mantiene operativo.
4. Permitir el control local de los datos que se usan con más frecuencia en un sitio puede mejorar el grado de satisfacción de los usuarios con relación al sistema de bases de datos. Esto es, las bases de datos locales pueden reflejar mejor la estructura administrativa local y, por lo tanto, brindar un mejor servicio a las necesidades de la gestión.

Un ejemplo de DDBE

Supongamos que nuestra organización tiene tres DBEs centralizados implementados como se muestra en la Figura. Tenemos una instancia de DB2 DBMS de IBM ejecutándose en nuestro mainframe de IBM, que contiene información de misión crítica sobre toda la organización y es compatible con nuestras aplicaciones de nómina, recursos humanos y cumplimiento de clientes. También tenemos dos grupos departamentales diferentes dentro de nuestra organización: un grupo está usando el DBMS de Oracle para mantener la información del proyecto y el otro grupo está utilizando el DBMS de SQL Server de Microsoft para realizar un seguimiento de la información de producción.



Ejemplo del posible despliegue de un CDBE de una organización.

En este ejemplo, aunque la información del proyecto en el DBE de Oracle y la información del cliente en el DBE de DB2 se acceden de manera independiente la mayor parte del tiempo, hay ocasiones en que necesitamos hacer una referencia cruzada de la información en estos dos sistemas. Supongamos que un gerente desea generar un informe para cada cliente, con una lista de detalles sobre el cliente, así como detalles sobre todos los proyectos de la compañía que involucran a ese cliente.

Para satisfacer los requisitos de este administrador, necesitamos combinar la información que se almacena en dos DBEs diferentes, controlados por dos tipos diferentes de DBMS. Podríamos elegir entre dos enfoques cuando intentamos satisfacer los requisitos de este gerente: un enfoque manual y un enfoque automatizado. En el

enfoque manual, iniciamos sesión en cada servidor individual, escribimos algunas consultas para obtener la información de cada base de datos, transferimos los resultados de cada base de datos a nuestra estación de trabajo personal y luego escribimos un programa que combina toda la información en el formato del informe final. Este es un proceso que requiere mucho tiempo y requiere que alguien esté familiarizado con cada plataforma de hardware, cada DBE, las instalaciones de comunicaciones necesarias para iniciar sesión en los sistemas y transferir los archivos, y las utilidades o habilidades de programación necesarias para obtener la información de cada sistema y fusionar los archivos en el resultado final. Incluso cuando tenemos a alguien que puede realizar todas estas tareas, no tenemos una forma fácil de garantizar que el informe final sea válido.

Para el enfoque automatizado, nuestra empresa necesita combinar los tres CDBE mencionados anteriormente en un nuevo DDBE único que debemos implementar. Podemos utilizar los servicios de DBMS locales, cada uno de los cuales se ejecuta en una computadora separada, y los servicios del subsistema de comunicación (Ethernet, Token Ring e Internet) para coordinar las operaciones de lectura o consulta necesarias. Idealmente, los usuarios de nuestro nuevo DDBE desconocerán por completo que el contenido de los datos está disperso en diferentes computadoras y está controlado por diferentes productos DBMS; tienen la ilusión de que todo el contenido de nuestros datos combinados está almacenado y controlado localmente por un sistema (el DDBE). Nuestros usuarios de DDBE no necesitan saber nada sobre los DBMS que tenemos. No necesitan conocer los ML requeridos para cada DBMS, nada sobre el hardware de la computadora que ejecuta las DBEs, o cualquier detalle sobre cómo estos sistemas de hardware y software están interconectados. Nuestros usuarios pueden enviar una solicitud de consulta al procesador de consultas DDBE que maneja toda la coordinación, ejecución y fusión necesarias automáticamente. Aunque el uso de este nuevo sistema DDBE es muy fácil, su implementación puede ser bastante difícil.