

Capítulo 1.

Introducción a las Bases de Datos Distribuidas y Sistemas de Multibase de Datos (MDBMS)

La evolución de los sistemas de información y el crecimiento no planeado de la información dentro de las organizaciones, ha traído dispersión de los datos en sitios local o geográficamente dispersos. La necesidad de integrar y compartir dicha información, implica el nacimiento de una nueva tecnología capaz de conformar de manera consistente la información de las organizaciones. Una de las tecnologías que trabaja en el problema de integración de información, es la de Bases de Datos Distribuidas (BDD).

Los Sistemas de Multibase de Datos, como un tipo especial de BDD, trabajan en el proceso de integración de componentes distribuidos, con el fin de ofrecer un esquema integrado (esquema global); esquema capaz de ofrecer transparencia de acceso a las Bases de Datos Componentes (BDC's) y garantizar la autonomía de operación de cada una de ellas.

El presente capítulo ofrece, en primera instancia, un panorama global de las Bases de Datos Distribuidas, sus funciones de operación, tipos de BDD y algunos conceptos tales como autonomía, grado de homogeneidad y grado de heterogeneidad. La segunda parte de este capítulo describe de manera detallada la taxonomía de los Sistemas de Multibase de Datos (MDBMS) propuesta por [Sheth y Larson 1990], además de presentar el proceso de evolución de un MDBMS.

1.1 DEFINICIÓN DE BASES DE DATOS DISTRIBUIDAS

La heterogeneidad debido a diversos DBMS's es común en organizaciones que crecen sin una planeación en cuanto a sus sistemas de información. Dichos sistemas evolucionan paulatinamente en diferentes DBMS's o diferentes modelos de conceptualización, tales como: jerárquico, de red, relacional u orientado a objetos.

La heterogeneidad semántica ocurre cuando existe una diferencia en el significado, interpretación o uso del mismo modelo de datos [Larson & Sheth 1990]. Una descripción más detallada de dicho concepto se presenta en el siguiente capítulo.

Existe otro tipo de DDBMS denominado **MDBMS** (Multi database management system - **Sistema de Multibase de Datos**), también llamado sistema de base de datos federado. En un sistema así, cada servidor es un DBMS centralizado independiente y autónomo que tiene sus propios usuarios locales, transacciones locales y administrador de base de datos (DBA), y por ende posee un alto grado de autonomía local.. En éste tipo de DDBMS cada servidor puede autorizar el acceso a porciones específicas de la BD definiendo un esquema de exportación, el cual especifica la parte de la BD a la cual puede tener acceso una cierta parte de usuarios no locales [Elmasri y Navathe 1997]. Los MDBMS son catalogados como un sistema híbrido entre los sistemas distribuidos y los centralizados.

Un tercer aspecto a considerar para clasificar a las BDD es **el grado de transparencia** de la distribución. Si el usuario percibe un solo esquema integrado sin información alguna relativa a la fragmentación, replicación o distribución, se dice que el DDBMS tiene un alto grado de transparencia de distribución. Por otro lado, si el usuario puede ver toda la fragmentación, el reparto y la aplicación, el DDBMS no tiene transparencia de distribución.

1.1.3 Autonomía en BDD

Para lograr la independencia entre bases de datos componentes de un sistema distribuido, es necesario considerar los tipos de autonomía descritos por [Sheth y Larson 1990], los cuales se detallan a continuación:

- a. **Autonomía de diseño.** - La capacidad de que cada Base de Datos Componente (BDC) decida los aspectos concernientes con su diseño. Es decir, las personas involucradas son libres de decidir cualquier particularidad e incluso decidir que DBMS usar. Los aspectos a considerar en la autonomía de diseño son:

El universo de datos relevante para el sistema

La representación (modelo de datos, lenguaje de consultas) y el nombrado de los datos

La conceptualización o interpretación semántica de los datos (heterogeneidad semántica)

Restricciones usadas para administrar los datos

La funcionalidad del sistema

La asociación y compartición con otros sistemas y

La implementación (p.e. registros y estructuras de archivos).

- b. **Autonomía de comunicación.**- La habilidad de que una BDC decida

comunicarse o no con otro componente de una misma federación.

- c. **Autonomía de ejecución.**- Es la habilidad de una BDC para ejecutar operaciones locales sin la interferencia de operaciones externas, en el orden que la BDC lo decida.
- d. **Autonomía de asociación.**- Cada BDC decide cuanto y cuando puede compartir su funcionalidad y recursos con otros componentes, inclusive la capacidad de asociarse o retirarse de una o más federaciones.

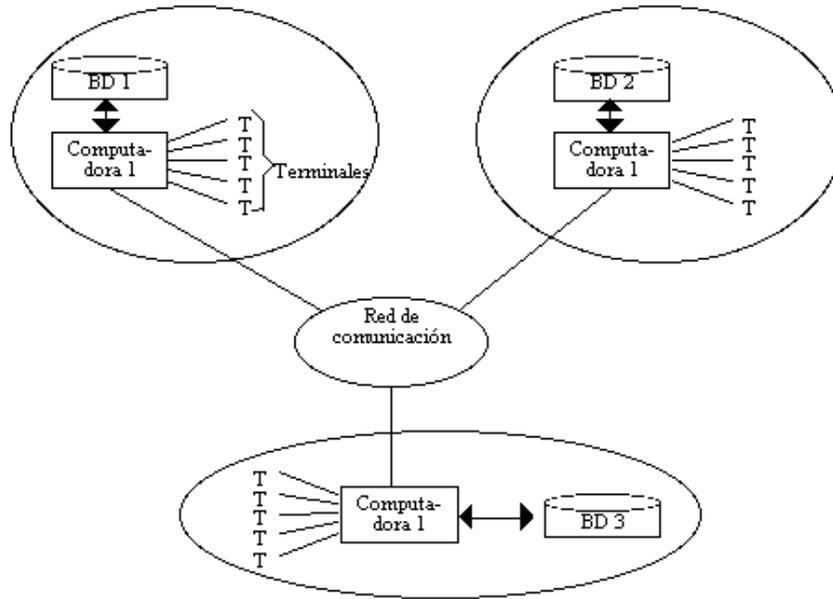


Figura 1.1 Una BDD en una red geográficamente dispersa

| Sistemas de Base de Datos | |
|--|--|
| <ul style="list-style-type: none"> ● Diferencias en DBMS's ● Heterogeneidad semántica | |
| <p style="text-align: center;">Sistema Operativo</p> <ul style="list-style-type: none"> ● Sistema de archivo ● Nombrado, tipos de archivo, operaciones ● Soporte a transacciones ● Comunicación interprocesos | C o m u n i c a c i ó n |
| <p style="text-align: center;">Sistema de HW</p> <ul style="list-style-type: none"> ● Conjunto de instrucciones ● Formatos y representación de datos configuración | |

1.2 SISTEMA DE MULTIBASE DE DATOS

Base de datos no federado.- Un sistema de base de datos no federado es una integración de Bases de Datos Componentes que no son autónomas [Sheth y Larson 1990]. Un sistema de base de datos no federado no distingue a usuarios locales y no locales. Una particularidad extra de estos sistemas, es que todas las bases de datos se integran para proveer un solo esquema global, llamado esquema unificado de MDBMS. Las operaciones de manipulación a una sistema de base de datos no federado, son a través de un mecanismo único de interacción con el sistema.

Base de datos federada.- Un sistema de base de datos federado (FDBS) consiste de componentes que son autónomos aún siendo participantes de una federación. Estos sistemas permiten la contribución parcial y controlada de sus datos. En la arquitectura federada no hay un control centralizado, debido a que sus componentes controlan el acceso a sus datos.

El FDBS soporta operaciones globales y locales. Las operaciones globales involucran el acceso a datos haciendo uso del FDBS y puede involucrar el acceso a datos de múltiples BDC's, dichos componentes deben conceder permisos de acceso a los datos solicitados. Las operaciones locales se realizan directamente sobre la BDC. Una BDC no necesita distinguir entre operaciones locales o globales.

Federación débil y fuertemente acoplada.- La forma en que los componentes se integran y la responsabilidad de manejar la federación, son los criterios a considerar para categorizar la federación como débil o fuertemente acopladas. Una federación es débilmente acoplada si es responsabilidad del usuario crear y mantener la federación. Una federación es fuertemente acoplada si su administrador tienen la responsabilidad de crear y mantener la federación y activar el control de acceso a las BDC's [Sheth y Larson 1990].

1.2.2 Proceso de evolución de un MDBMS

- a. **Preintegración.**- Cuando existen datos manejados en archivos planos, es necesario migrarlos a un DBMS a través de las siguientes actividades:
 1. Desarrollar un esquema componente que describa los datos en el archivo.
 2. Cargar el DBMS con datos desde el archivo.

3. Modificar programas de aplicación para acceder el DBMS en lugar de archivos.
- b. **Creación del MDBS** definiendo el mapa entre varios esquemas e implementando los procesos asociados.
 - c. **Operación y manipulación del MDBS.**

Con los conceptos planteados en este capítulo y la identificación de los diferentes tipos de BDD existentes, es posible identificar la naturaleza del proyecto. Además de crear expectativas en cuanto a las funcionalidades o capacidades que un MDBMS debe proporcionar. Este capítulo ofrece el marco de referencia para discutir y clasificar los sistemas de Multibase de Datos presentados en el capítulo 2.

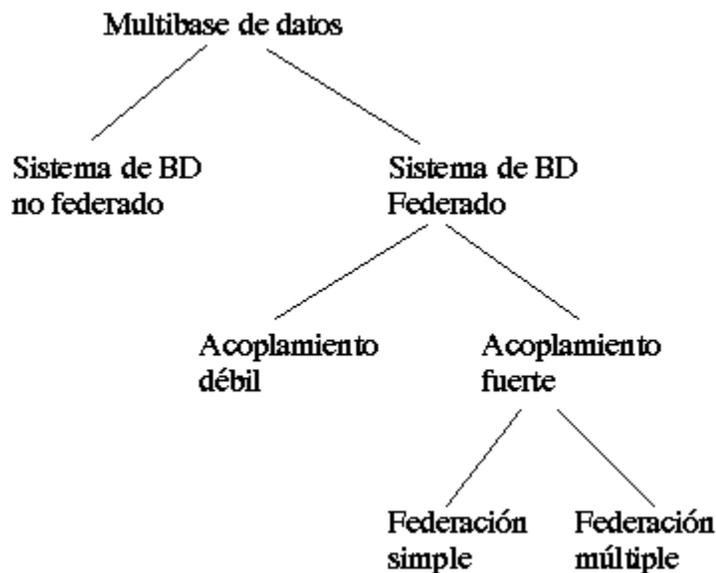


Figura 1.3 Taxonomía de sistemas de Multibase de Datos

Alvarez Carrión, G. 1999. **Integración de esquemas en bases de datos heterogéneas fuertemente acopladas.** Tesis Maestría. Ciencias con Especialidad en Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas Puebla. Mayo. Derechos Reservados © 1999.