



## CAPÍTULO 14º: DISEÑO DE BASES DE DATOS RELACIONALES DISTRIBUIDAS

### 1 Introducción .-

Las propiedades y características de las Bases de Datos distribuidas requieren la adaptación de los diseños de las Bases de Datos centralizadas a las mismas. Diseño que puede abordarse de diferentes maneras en función de su definición, las propiedades que las caracterizan y las arquitecturas que poseen.

Una Base de Datos distribuida es una colección de varias Bases de Datos interrelacionadas lógicamente y distribuidas sobre una red de ordenadores.

Una Base de Datos Distribuida (BDD) se caracteriza por:

- ❖ La *integración* de los datos, esto es, los datos no son una colección de ficheros que están almacenados individualmente en los nodos de la red, sino que forman parte de una estructura global que los relaciona lógicamente.
- ❖ La *distribución* de los datos en distintos nodos de una red que para los usuarios de la BDD aparecerán como homogéneos y únicos.

En resumen, pueden caracterizarse las BDD de la siguiente forma:

- ◆ Los datos deben estar físicamente en mas de un ordenador, esto es, se encuentran almacenados en distintas *sedes*.
- ◆ Las *sedes* deben estar interconectadas mediante una red. Cada *sede* es un emplazamiento o nodo de la red. Para realizar el diseño de la BDD no se tendrá en cuenta la topología, tipo y rendimiento de la red que une estas *sedes*, aunque estas propiedades son relevantes para obtener buenos rendimientos y evitar atascos y retrasos en el tráfico de datos por la red. Se considera que éstos son aspectos muy concretos y específicos de cada red y que cada caso particular necesita su propio estudio.
- ◆ Los datos han de estar lógicamente integrados en una única estructura lógica o esquema global común.
- ◆ Los usuarios han de tener acceso (recuperación y actualización) a los datos pertenecientes a la BDD ya se encuentren éstos en la misma *sede* (*acceso local*) o en otra (*acceso remoto*).
- ◆ Cada emplazamiento o nodo debe proporcionar un entorno para la ejecución de transacciones, tanto locales como globales.



- ♦ En una única operación, tanto de actualización como de recuperación, puede accederse a los datos que se encuentran en mas de una sede sin que el usuario sepa la distribución de los mismos en las distintas sedes (*propiedad de transparencia de red*).

## **2 Sistema de Gestión de Bases de Datos Distribuidas (SGBDD) .-**

El SGBDD es un producto software que permite el manejo de bases de datos y hace la distribución transparente para los usuarios, cumpliendo con respecto a las BDD el mismo papel que cumple el Sistema Gestor de Bases de Datos (SGBD) respecto a las bases de datos centralizadas.

Sus componentes principales son:

- *Procesador de datos locales.*- Es el componente que se encarga de la gestión local de los datos de forma parecida al software de un SGBD centralizado, por lo que, además de ejecutar transacciones locales, se encarga de la concurrencia y la recuperación ante fallos a nivel local.
- *Diccionario o Directorio Global.*- Es aquel componente en el que se guardará la información acerca de dónde y cómo se almacenan los datos, el modo de acceso y otras características físicas. En resumen, es el componente que contiene las especificaciones necesarias para pasar de la representación externa o esquema externo de los datos a la representación interna de los mismos.
- *Procesador de aplicaciones distribuidas.*- Es el componente responsable de las funciones distribuidas. Accede a la información sobre la ubicación de los datos, que se encuentra en el diccionario, y se ocupa de procesar todas las peticiones que involucran mas de una sede para generar un plan de ejecución distribuido. Es el elemento diferenciador de los sistemas distribuidos ya que, dada una operación, es el que se encarga de repartir el trabajo a los distintos procesadores locales que intervienen en dicha operación.
- *Software y Red de comunicaciones.*- No forma parte estrictamente del SGBDD sino que provee al procesador de aplicaciones distribuidas de primitivas y servicios de comunicaciones para que éste lleve a cabo su labor.

## **3 Clases de los SGBDD .-**

Dependiendo de la homogeneidad de los SGBD locales, de la distribución de los datos y de su autonomía, habrá distintos tipos de SGBDD:

Atendiendo a la *homogeneidad*, si todos los SGBD locales son iguales se obtendrá un **Sistema Homogéneo** en el que se tiene un único producto y un único lenguaje de consultas , por lo que suelen ser sistemas muy integrados.



Si, por el contrario, los SGBD son distintos, el sistema distribuido será un **Sistema Heterogéneo** donde, aunque se pueda utilizar el mismo modelo de datos, se tienen distintos productos y lenguajes de consultas que requieren su integración.

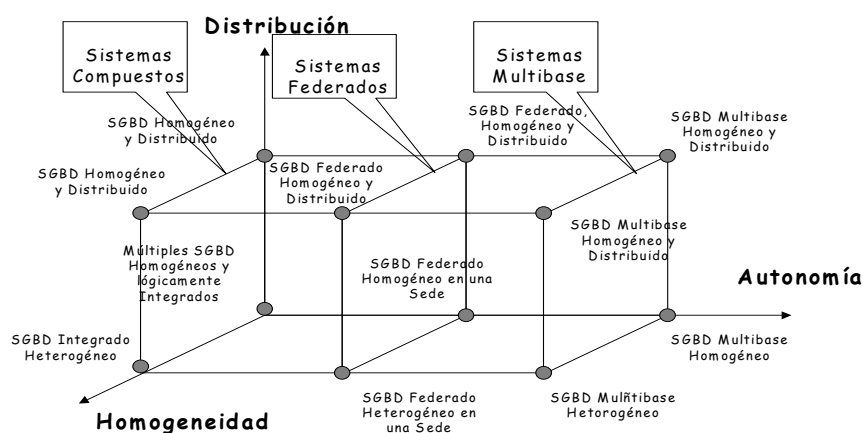
La *distribución* determina si los datos están distribuidos físicamente entre los múltiples sitios que se comunican entre sí (datos distribuidos) o si se mantienen en un único lugar (datos centralizados).

Por último, la *autonomía* se refiere a la distribución de control. Habrá *Sistemas Estrechamente Ligados* (Sistemas Compuestos) si todo el acceso a los datos se realiza siempre a través del procesador de datos distribuidos, con las sedes locales totalmente dependientes de éste, el procesador gestiona, por tanto, todos los accesos y funciones de administración.

Los *Sistemas Semiautónomos* (Sistemas Federados) son aquellos que dejan que cada procesador local actúe de forma autónoma e independiente. De hecho, cada uno de ellos tiene sus usuarios, transacciones, administradores y aplicaciones exclusivamente locales además de poseer porciones específicas de la Base de Datos Distribuida.

Por último, los *Sistemas con total autonomía* (Sistemas Multi-Base de Datos) son aquellos en los que los procesadores locales no conocen la existencia de otros SGBD ni cómo comunicarse con ellos. En los sistemas Multi-Base, al no haber un control global, el acceso a los datos es especialmente complicado.

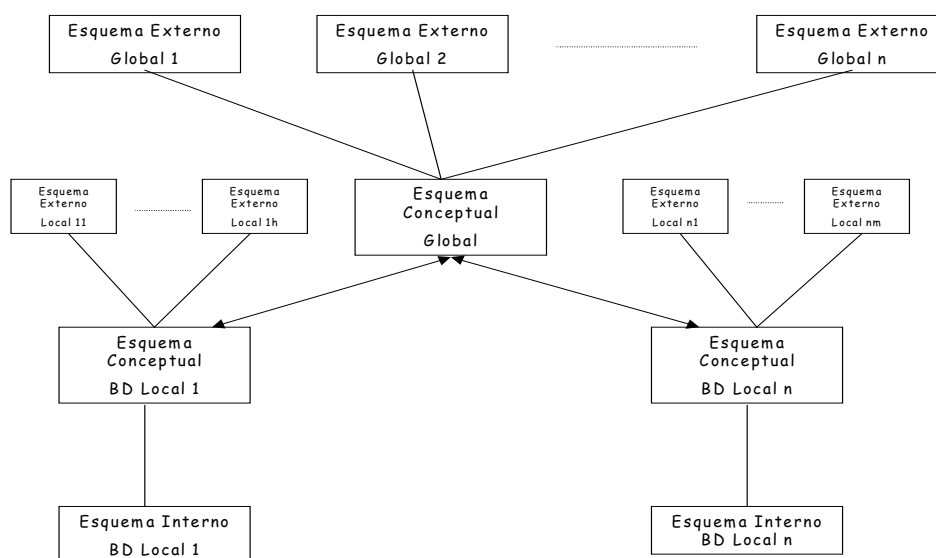
Gráficamente, pueden resumirse las distintas clases de SGBDD representando cada uno de los factores estudiados anteriormente en un sistema de coordenadas tridimensional, observando que distribución, autonomía y heterogeneidad son aspectos ortogonales.





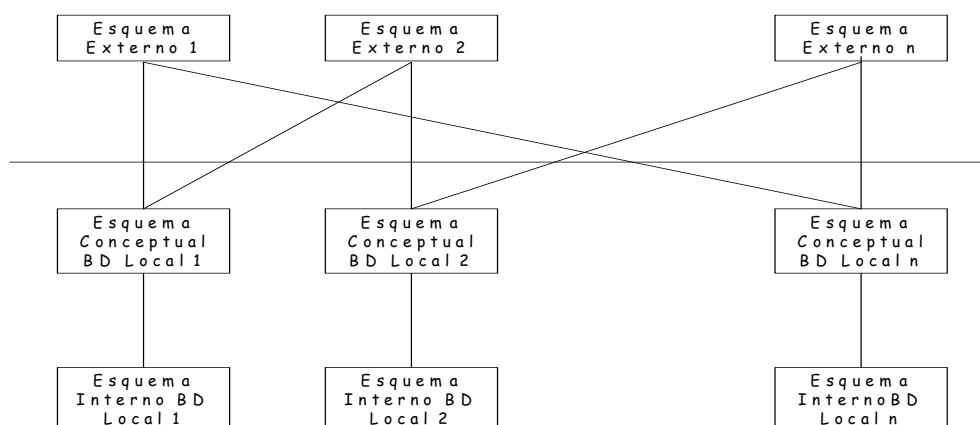
#### 4 Arquitecturas de los SGBDD .-

La arquitectura de los SGBDD en dos tipos: la arquitectura ANSI/X3/SPARC para sistemas gestores distribuidos



**Arquitectura ANSI/X3/SPARC para SGBDD**

y la arquitectura de los sistemas SGBD Multibase:

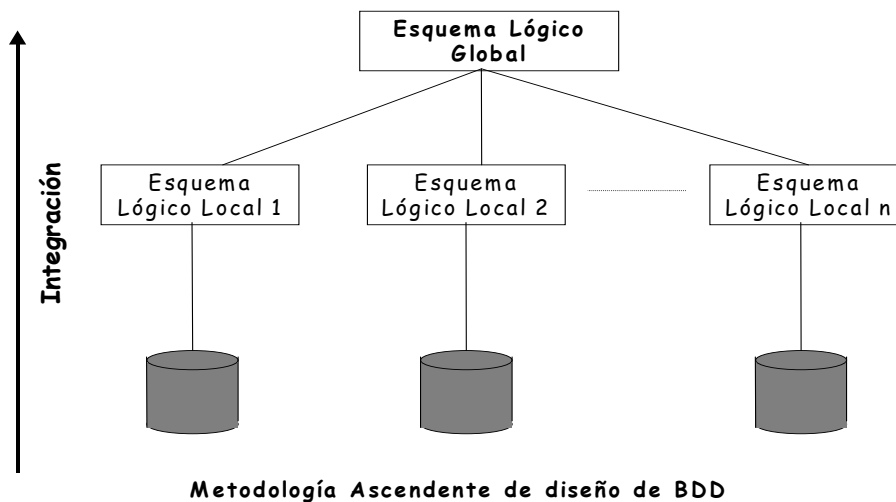


**Arquitectura de los SGBD MultiBase**

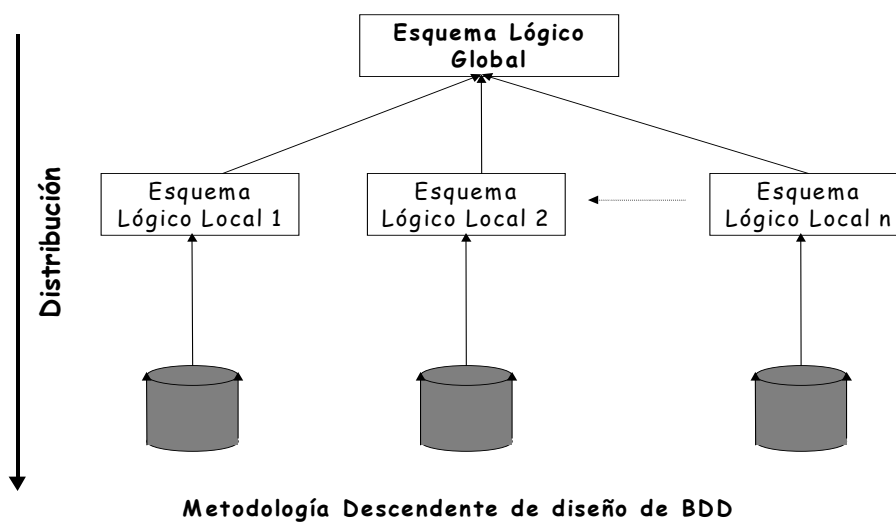


## 5 Diseño de Bases de Datos Distribuidas .-

En cualquier proceso de diseño hay dos aproximaciones básicas: la *Ascendente* o Top-Down y la *Descendente* o Botton-Up



La *metodología ascendente* se utiliza, en general, cuando existen varias bases de datos locales y quiere construirse una Base de Datos distribuida. Para ello se parte de Esquemas Lógicos Locales (ELL) que se corresponden con las distintas Bases de Datos ubicadas en los diferentes nodos de la red y se integran, parte de los ELL o todos, en un único Esquema Lógico Global (ELG)





Por su parte, la metodología descendente parte de un Esquema Lógico Global (ELG) que, mediante la fragmentación, replicación y asignación de los objetos pertenecientes a éste, se construyen los Esquemas Lógicos Locales (ELL).

En general, la metodología descendente es la más utilizada. Partiendo de un esquema lógico global de la base de datos (Esquema relacional) que se quiere distribuir, y, a partir de él se construyen los esquemas de fragmentación, de asignación y replicación de los datos para distribuirlos en los diferentes nodos de la red.

### 5.1 Esquema de fragmentación .-

En general los esquemas de fragmentación se basan en el análisis de los datos utilizados por las distintas aplicaciones que acceden a la BD para crear relaciones más pequeñas y más adaptadas a las operaciones de recuperación y actualización que las aplicaciones utilizan, esto es, tener los datos divididos en función de la utilización que se hace de ellos.

El esquema de fragmentación se compone de las distintas relaciones en las que se ha dividido una relación procedente del Esquema Lógico Global, junto con la condición empleada para esta división, expresado todo ello en álgebra relacional.

Para asegurar que la BD no sufrirá cambios semánticos durante la fragmentación de los datos, se definen tres normas que determinan la calidad de la fragmentación de una relación:

- *Compleitud*.- Todos los datos han de encontrarse en, al menos, un fragmento.
- *Disyunción*.- Los fragmentos deberán ser disjuntos.
- *Reconstrucción*.- Siempre se ha de poder reconstruir la BD global a partir de los fragmentos.

### 5.2 Esquemas de asignación .-

En los esquemas de asignación y replicación se fija desde qué nodo se demandan los datos y el tipo de operación que se realiza (si es de consulta o actualización), para que las operaciones puedan llevarse a cabo de forma local y minimizar, de esta forma, el tráfico por la red.

El esquema de asignación y replicación consiste en la realización de la correspondencia entre los fragmentos y los nodos que constituyen la red de comunicaciones de la BDD. Dicha correspondencia debe hacerse de forma óptima teniendo en cuenta los parámetros siguientes:

- ♦ *Mínimo coste*.- Haciendo mínimo el coste de mantenimiento de cada fragmento en el nodo correspondiente, el coste de modificar un fragmento que está replicado en distintos nodos y el coste de la transferencia de datos por la red.



- ◆ *Rendimiento.*- Minimizando los tiempos de respuesta y maximizando la capacidad de procesamiento del sistema en cada nodo.

