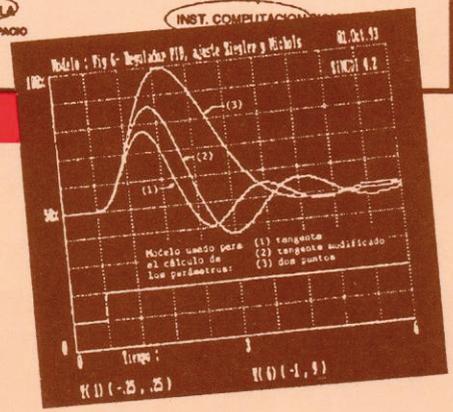
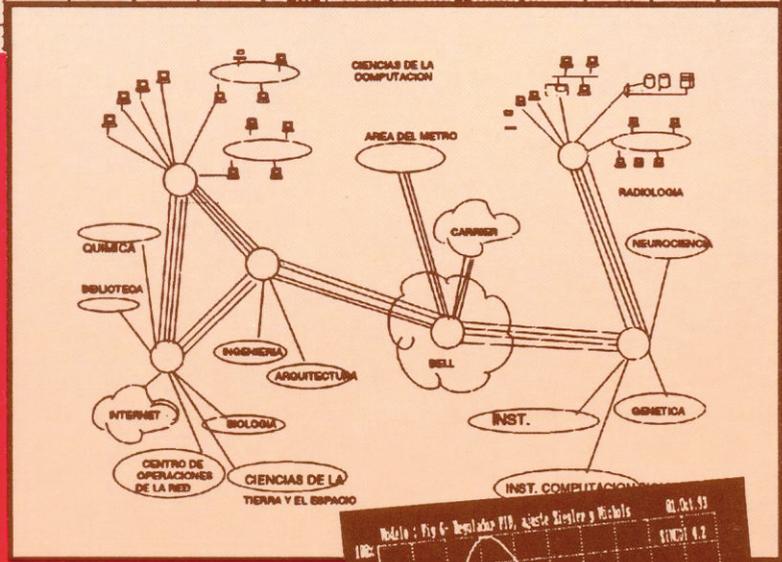
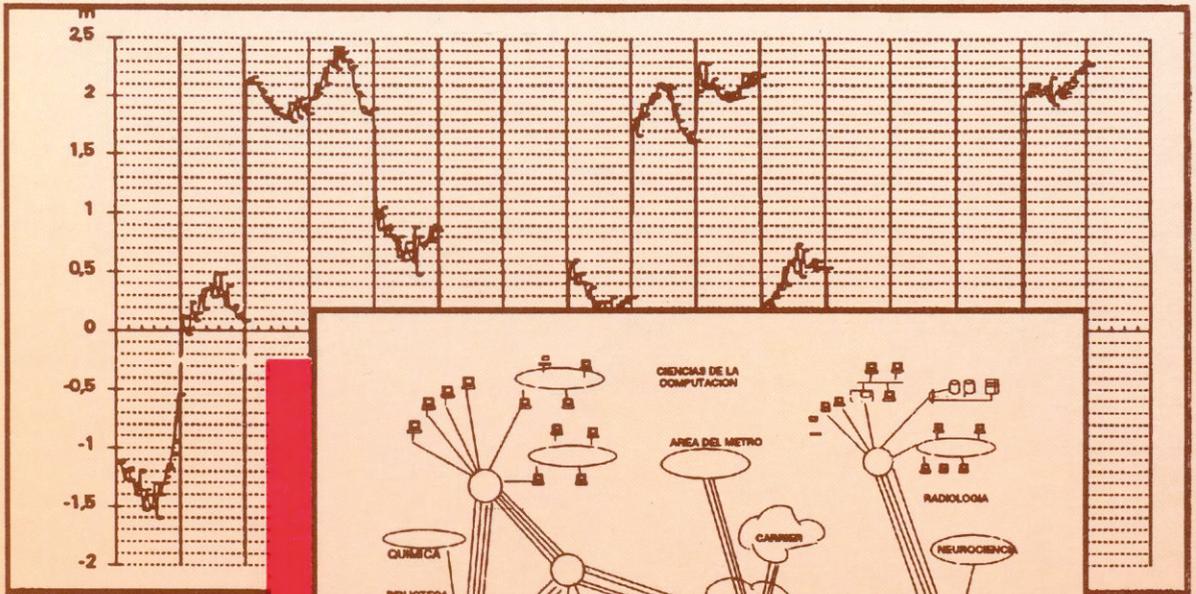


Ingeniería

Revista de la Universidad de Costa Rica
ENERO/JUNIO 1994 VOLUMEN 4 N° 1



SISTEMAS DE BASES DE DATOS MÚLTIPLES

*Elzbieta Malinowski Gajda**

RESUMEN

Los ambientes distribuidos donde se puede integrar y compartir los recursos son la parte principal de muchas investigaciones realizadas en todo el mundo. El concepto de Base de Datos Distribuidas es ampliado ahora a Bases de Datos Múltiples, lo cual abre la posibilidad de integrar diferentes plataformas en el nivel de "hardware", "software" y modelo de datos.

SUMMARY

The distributed environments where resources can be integrated and shared are the main components of many researches carried out throughout the world. The concept of Distributed Databases has now been broadened to Multidatabases, which permits the integration of different platforms at the hardware, software, and data model.

INTRODUCCION

Los Sistemas de Bases de Datos constituyen una de las partes más importantes en el desarrollo y funcionamiento de las instituciones. Estos sistemas deberían permitir manejar grandes volúmenes de información en forma eficiente, rápida y confiable. Las organizaciones en su mayoría tienen diferentes computadoras que manejan diferentes DBMS (DBMS-Database Management System; Sistema de Manejo de Bases de Datos). Mientras el usuario necesita compartir la información debe preservarse en muchos de los casos este ambiente debido al difícil, costoso y a veces innecesario proceso de unificar las DBMS existentes. Los Sistemas de Bases de Datos Múltiples (Multidatabase System -MDBS) sirven para integrar la información existente desde las bases de datos locales heterogéneas en un ambiente distribuido y ofrecen un uso global de sus recursos con los métodos transparentes para el usuario de todo el sistema.

El área de MDBS es un área importante de investigaciones actuales y esto se evidencia por la gran cantidad de proyectos académicos e industriales desarrollados en todo el mundo, como: General Motors Research (EEUU), CRAI (Italia),

Centro de Desarrollo en Procesamiento de Información (Japón), etc. [Bright, 1992].

DESARROLLO

Existe un gran rango de soluciones para compartir la información en un sistema distribuido y en la literatura se encuentran muchos términos para describirlos como: bases de datos distribuidas, bases de datos múltiples, bases de datos federativas y sistemas de interoperación. Todos estos sistemas describen un sistema distribuido, el cual incluye un componente global para acceder la información compartida y los múltiples componentes locales, los cuales administran la información solo en un sitio. La diferencia entre cada uno de los sistemas mencionados al principio se encuentra en la estructura del componente global y en la forma como éste interactúa con los componentes locales. Para distinguirlos más claramente se tomará en cuenta el aspecto de cuán estrecha es la integración del sistema global con las DBMS's locales.

En un sistema de unión estrecha entre el sistema global y una DBMS local, el sistema

global tiene el acceso a las funciones de bajo nivel de la DBMS local. Esto permite una buena sincronización entre los sitios y un proceso global muy eficiente.

Sin embargo esto también implica que las funciones globales pueden tener prioridad sobre las funciones locales y en consecuencia la DBMS local no tiene control sobre sus propios recursos. En un sistema menos estrecho las funciones globales accesan las DBMS's locales por medio de la interfase externa preparada por la DBMS local. La sincronización local y la eficiencia no es tan alta como en el caso anterior, aunque la DBMS local tiene el control total sobre sus datos y el procesamiento de ellos.

Según Bright, Hurson y Pakzad se puede comenzar a analizar las clases existentes de los sistemas distribuidos con aquellas que tienen la más estrecha unión entre el sistema global y la DBMS local [Bright, 1992]. Estas clases se presentan de la siguiente forma:

1. Bases de Datos Distribuidas (Distributed Databases - DDB):

Las funciones globales y locales comparten la interfase local de cada sitio y están tan estrechamente integradas que la diferencia entre ellas es mínima. Las DBMS's locales son típicamente homogéneas (con respecto al modelo de datos implementado: solo relacional o solo jerárquico, etc.) y representan las mismas interfases funcionales en todos los niveles, aunque pueden ser implementadas en diferentes computadoras. El sistema global tiene el control sobre los datos locales y su procesamiento. El sistema mantiene típicamente un esquema global creado por la integración de los esquemas locales de cada sitio. Por esquema se entiende la descripción estructurada de la información disponible en la base de datos. Los usuarios globales atienden sus consultas dejando la decisión al esquema global del sistema. En este caso se tiene la sincronización global del procesamiento. Además, como las funciones globales tienen el control sobre las funciones locales, el procesamiento de las consultas puede ser optimizado tomando en cuenta los requerimientos globales. Como resultado, las

DDB tienen buen funcionamiento global a costo de la preparación previa por medio de las modificaciones locales; además pierden el control local. Esto está presentado muy claro en Ceri [Ceri, 1984] y Korth [Korth, 1988]. Este sistema en literatura se nombra también como un sistema lógicamente centralizado (refiriéndose al control) y físicamente descentralizado (refiriéndose a la ubicación física de los datos) [Larson, 1987].

2. Esquema Global de Bases de Datos Múltiples (Global Schema Multidatabases):

Este esquema no tiene una unión tan estrecha como el anterior, porque las funciones globales de acceso se realizan a través de una interfase externa de la DBMS local. Sin embargo el sistema global siempre mantiene un esquema global, y esto obliga a las DBMS's locales a tener una estrecha cooperación para mantener este esquema global.

Para formar este sistema se puede integrar las DBMS's locales ya existentes sin ninguna modificación. Aquí normalmente se puede integrar las DBMS's heterogéneas. Por heterogéneas se entiende los modelos de datos diferentes o las diferentes implementaciones de los mismos modelos de datos. Así pues la creación de un esquema global es un problema más difícil que en el caso de DDB. Este problema no se puede resolver en forma eficiente usando las herramientas ya existentes como por ejemplo, los Diccionarios de Base de Datos [Faulkner, 1992]. Larson considera este sistema igual al anterior solo que con diferentes niveles de centralización lógica [Larson, 1987].

3. Bases de Datos Federativas (Federated Databases):

A diferencia con el Esquema Global de Bases de Datos Múltiples aquí no se dispone de un esquema global. Cada sitio mantiene su esquema local de importación y exportación. El esquema de exportación es una descripción de la información del nodo local la cual puede ser compartida con los demás nodos del sistema global. El esquema de importación es una descripción de la información (representación

de datos y su origen) de los nodos remotos, los cuales se pueden acceder localmente. Cada esquema de importación es en verdad una porción del esquema global. Por lo tanto cada nodo debe cooperar estrecha y solamente con los nodos a los cuales tiene el acceso autorizado. Las consultas de los usuarios son limitadas y pueden ejecutarse solo sobre los datos locales y los datos representados en el esquema de importación.

Según Larsen [Larsen, 1987] este sistema corresponde a una base de datos lógica y físicamente descentralizada. Aquí además de la descripción de los datos para exportación e importación, claramente se define la porción de datos privados, a los cuales tienen acceso solo los dueños del nodo.

4. Sistema con Lenguaje de Bases de Datos Múltiples (Multidatabase Language Systems):

Este sistema no mantiene ningún esquema global (ni en forma parcial). El sistema global soporta todas las funciones globales de las bases de datos suministrando una herramienta de lenguaje de consultas para integrar la información desde diferentes bases de datos. Usando este lenguaje se puede especificar el acceso a los datos desde cualquier esquema local de cualquier nodo participante en el sistema. La herramienta del lenguaje incluye espacio para el nombre global y las funciones especiales para proyectar la información desde diferentes modelos y representaciones hacia un modelo y representación significativo para los usuarios.

Aquí se puede integrar las DBMS's locales heterogéneas sin ninguna modificación. Este es una de los sistemas más fuertemente desarrollados en la actualidad, según la necesidad de "conectar" las bases de datos existentes. Así, por ejemplo, desarrollaron sus sistemas SQL Access Group [Faullmer, 1992] o CORBA del Object Management Group [Mowbray, 1994].

En los proyectos que se están desarrollando, también se toma en cuenta si la migración del "software" para la ejecución de las consultas

debe realizarse en forma de programa objeto (compilado en el nodo, el cual desarrolla la consulta) o como programa fuente (compilado en otra computadora) [Booth, 1990].

S. Sistemas Interoperables (Interoperable Systems):

La función global está limitada por un simple intercambio de datos y no tiene la funcionalidad de una base de datos global. Se definen los protocolos estándar para la comunicación entre los nodos. Como el sistema global no es un sistema orientado a bases de datos, los sistemas locales pueden incluir otro tipo de sistemas de información, como sistemas expertos, bases de conocimiento, etc.

¿Cuál es entonces la definición de las Bases de Datos Múltiples (MDB)?

Este término genérico se usa para incluir en él las siguientes clases: Esquema Global de Bases de Datos Múltiples, Bases de Datos Federativas, Sistema con Lenguaje de Bases de Datos Múltiples.

La MDB es un sistema distribuido con un sistema global que provee total funcionalidad del sistema de base de datos e interactúa con las DBMS's locales por medio de su interfase externa. Aunque cada nodo local debe mantener algunas funciones globales para comunicarse con el sistema global, estos nodos locales son autónomos. El sistema global provee algunos medios (el esquema global o lenguaje de bases de datos múltiples) para resolver la diferencia en la representación de los datos y funciones entre las DBMS's locales. Esta resolución es indispensable, porque la misma información podría encontrarse en diferentes nodos y en diferentes formas. El usuario global del sistema puede acceder cualquier información desde los diferentes nodos con una consulta relativamente simple.

Es obvio que la clave principal de la MDB en contraste con la DDB es que en la MDB cada DBMS local retiene el completo control sobre los datos locales y su procesamiento. Esto permite asegurarle la autonomía local.

Pero este sistema no está libre de problemas, por ejemplo:

- El establecimiento de las reglas globales para asegurar la integridad del sistema.
- El proceso de consultas globales: su optimización, descomposición.
- El control de concurrencia (en muchos de los casos se pasa ejecutar las operaciones en serie).
- Problemas de seguridad.
- Etc.

CONCLUSIONES

Las tendencias existentes actualmente nos llevan inevitablemente a unir los recursos existentes en una organización. El costo de los equipos y "software" desarrollado es tan elevado, que no se debe omitir su utilización y se necesita integrar no solo "hardware" diferentes, sino también diferentes modelos de bases de datos. Esto permite ampliar la perspectiva y planificar un sistema de información apoyado en una eficiente base de datos. Para ello se necesita incorporar los nuevos conceptos desarrollados fuertemente por la ciencia, y además algunos ya implementados en forma de prototipos o sistemas comerciales, como es el caso de una base de datos federativa llamada IngresStar2 de la Relational Technology, Inc.

Cabe mencionar que en la literatura existe gran confusión sobre los términos usados para especificar los ambientes distribuidos. Se establecieron doce reglas para unificar los conceptos sobre las bases de datos en ambientes distribuidos [Krasowski, 1991], imitando los mismos pasos que realizó Codd en los años 70, estableciendo doce reglas para las bases de datos relacionales. Los ambientes distribuidos están en

proceso de investigación y los resultados de estas investigaciones marcan la pauta de seguimiento en el futuro, no así las reglas establecidas y poco flexibles.

Las Bases de Datos Múltiples son el camino de crecimiento natural de las DDB. En consecuencia algunas investigaciones realizadas en DDB se pueden aplicar para las MDB, lo cual permitirá su más rápida implementación.

BIBLIOGRAFIA

1. Booth Grace. "Distributed Data Bases on Unlike Computers". "Data Bases Management". New York, Auerbach Publishers, 1990.
2. Bright M. W., Hurson A.R., Pakzad S.H. •A Taxonomy and Current Issues in Multidatabase Systems". "Computer". Marzo 1992.
3. Ceri S., Pelagatti G. "Distributed Databases: Principles and Systems". McGraw-Hill, New York; 1984.
4. Faulkner Editorial Staff. "Data Base Management Systems Strategies and Trends". "Faulkner Technical Reports". Marzo 1992.
5. Korth Henry F., Silberschatz Abraham. "Fundamentos de Bases de Datos". McGraw-Hill, México, 1988.
6. Krasowski Michel D. "Integrating Distributed Data Bases into the Information Architecture". "Data Base Management". New York, Auerbach Publishers, 1991.
7. Larson I.A. "Alternative Architectures for Distributed DBMS's". "Data Base Management". New York, Auerbach Publishers, 1987.
8. Mowbray Thomas I "Distributed Objects Everywhere". "Object Magazine". Enero 1994.