

SBDA-GRID: implementación de un servicio de base de datos en una arquitectura grid

Nancy Ocotitla Rojas, Axel E. Moreno Cervantes, Rogelio Castro González

Escuela Superior de Cómputo-IPN, Av. Juan de Dios Batíz, esquina con Miguel Othón de Mendizábal, México, D.F., 07738. México
nanwen1@gmail.com

Resumen. En el presente trabajo se propone la implementación de un sistema llamado SBDA-GRID que consiste en la implementación de una arquitectura Grid para el desarrollo de un prototipo que permite brindar servicios de bases de datos a través de Internet. Así como el desarrollo de una aplicación web de prueba que brinda los servicios de control de acceso a usuarios y el servicio de creación, modificación y consulta a bases de datos (con estructura select atributo(s) from tabla(s) where condición(es)) con un máximo de dos condiciones. Usando además recursos de diferentes equipos para efectuar las operaciones que se soliciten.

Palabras Clave: GRID, Sistemas Distribuidos, Servicio Web, Bases de Datos.

1 Introducción

Durante mucho tiempo, se ha pretendido lograr que el procesamiento de la información no se realice en una sola computadora si no que se lleve a cabo en diferentes máquinas, incluso estando en lugares muy lejanos. Es por ello, que nace el concepto de computación distribuida ésta hace referencia a cualquier sistema en el que varios componentes autónomos cada uno con capacidad individual se comunican entre sí, afectando mutuamente su comportamiento [1].

Hace algunos años para poder realizar grandes cálculos o procesamiento de gran cantidad de información de cualquier proyecto de investigación, era necesario tener un supercomputador que fuera capaz de soportar trabajos complejos y de gran escala, el cual por sus características tienen un costo demasiado elevado. Pero en la actualidad gracias a los avances de la tecnología, las computadoras personales están llegando a tener mayor poder de procesamiento. Es por ello que se agrega el concepto de computación Grid, que se define como una tecnología nueva e innovadora, una nueva forma de computación distribuida. Los nodos se agrupan más allá del dominio de una red LAN (*Local Area Network*, Red de Área Local). La Grid es un concepto muy ambicioso, consiste en poder

procesar operaciones desde un nodo aunque éste no cuente con los recursos suficientes debido a que será compartida entre los recursos de toda la Grid [2] [3].

La tecnología Grid permite que diversas instituciones puedan unir sus recursos computacionales para obtener uno más poderoso. Con la ayuda de esta tecnología se pueden adelantar proyectos de investigación en países subdesarrollados los cuales no cuentan con el capital suficiente para adquirir supercomputadoras que almacenen y procesen inmensas cantidades de datos.

SBDA-GRID hace uso de la tecnología Grid, aprovechando las características mencionadas para la gestión de bases de datos relacionales, desarrollando herramientas de software que permite al usuario controlar la gestión de bases de datos en un servicio web, las operaciones que puede realizar el usuario son el control de acceso a usuarios y consultas básicas con estructura: select atributo(s) from tabla(s) where condición(es), con un máximo de dos condiciones. Se crean certificados X.509, que es un estándar para infraestructuras de claves públicas que garantizan la seguridad del sistema.

2 Estado del arte

Se analizaron los siguientes sistemas que involucran conceptos de Grid y Bases de Datos obteniendo los siguientes resultados:

1. **Oracle Solaris 11.** Permite a los clientes ejecutar sus aplicaciones en nubes privadas, híbridas o públicas. Permite a los usuarios de entornos corporativos colaborar con las personas adecuadas dentro de su empresa o del entorno de la misma. Se planea integrar perfiles de usuario, grupos, actividades, actualizaciones de estado, foros de discusión, intercambio de documentos, mensajería instantánea, email, notificaciones en tiempo real [4].
2. **Base de Datos Espaciales.** Es un sistema de hardware, software y procedimientos diseñados para realizar la captura, almacenamiento, manipulación, análisis, modelación y representación de datos referenciados espacialmente con el propósito de proporcionar el soporte para la resolución de problemas complejos de planificación y gestión [5].

Los sistemas no implementan específicamente la creación de bases de datos relacionales para el usuario final. SBDA-GRID pretende proporcionar al usuario la posibilidad de habilitar roles de usuario, crear bases de datos relacionales y permitir a cada usuario la gestión de las mismas mediante la arquitectura Grid.

3 Metodología

En la figura 1 se muestra la arquitectura de SBDA-GRID.

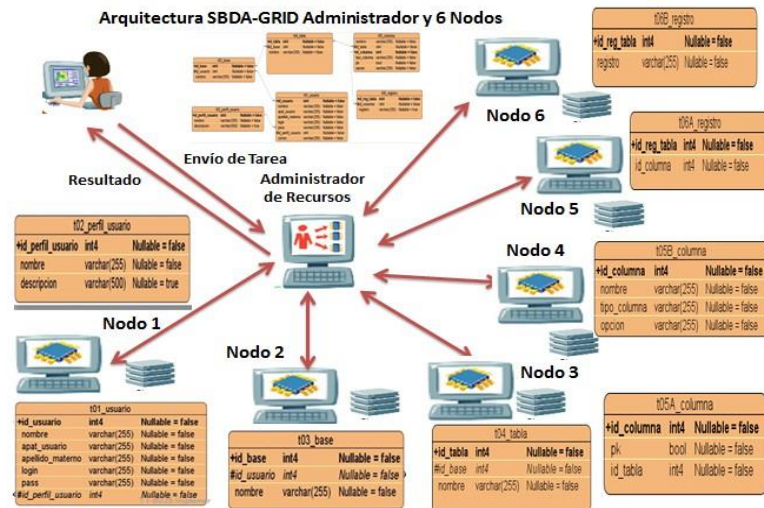


Fig. 1. Arquitectura de SBDA-GRID.

El funcionamiento de SBDA-GRID consiste en el envío de la tarea por parte del usuario a través de una aplicación web mediante el servicio web que es el encargado de establecer la comunicación con el administrador de recursos, éste se encarga de enviar las tareas a los nodos que se encuentren disponibles mediante los protocolos del GRID, la cual una vez procesada regresa al nodo administrador quien es el que envía el resultado final al usuario.

La arquitectura está compuesta por seis nodos clientes y un nodo servidor el cual contiene toda la base de datos que actúa como base de reserva en caso de que no se encuentre disponible alguno de los nodos para la realización del trabajo.

SBDA-GRID se compone de dos módulos principales los cuales son: el módulo GRID y el módulo Cliente.

A. El módulo Grid se compone de las conexiones a las bases de datos de cada uno de los nodos y la creación de los certificados para el manejo de la seguridad de la arquitectura Grid. El nodo administrador necesita crear una autoridad certificadora que es la que le permite implementar la seguridad en el Grid. Esto se logra haciendo uso de los certificados de seguridad X.509 [6]. Observe figura 2.



Fig. 2. Un certificado digital.

La arquitectura presenta una mayor carga de procesamiento en los nodos 1, 3, 4, y 5 (observe figura 1). Debido a que el almacenamiento de los nodos contiene los registros y las tablas de las bases de datos, por consecuente en ellos habrá un mayor número de conexiones realizadas a dichos nodos. Lo que se propone es minimizar dichas conexiones reuniendo los datos necesarios en una sola petición. Por esta razón se empleó el tipo de fragmentación vertical.

Se produjeron varios fragmentos agrupando un subconjunto de los atributos de una relación. A cada subconjunto se le asignó la llave primaria de la relación. El objetivo de la realización de la fragmentación vertical consiste en particionar la relación en un conjunto de relaciones más pequeñas de manera que las aplicaciones se ejecutarán sobre un fragmento, minimizando así el tiempo de ejecución de las consultas [6]. Además al tratar con relaciones más pequeñas, se ve reducido considerablemente el número de accesos.

SBDA-GRID empleó un enfoque de división, pues la fragmentación se basó en el comportamiento de acceso de las consultas que se realizaban sobre los atributos.

La fragmentación de las tablas columna y registro se realizó de la siguiente forma:

Nodo3: t04_tabla → id_tabla, id_base, nombre

T05A_columna → id_columna, pk, id_tabla

Nodo 4: t05B_columna → id_columna, nombre, tipo_columna, opción.

Nodo 5: t06A_registro → id_reg_tabla, id_columna.

Nodo 6: t06B_registro → id_reg_tabla, registro

Donde la tabla columna y registro son:

t05_columna → nombre, id_tabla, id_columna, tipo_columna, pk, opción

t06_registro → id_reg_tabla, id_columna, registro.

Un submódulo del sistema consiste en la ejecución de consultas de la forma:

```
select atributos(s), from tabla(s), where condición(es);
```

Por ejemplo: dada la siguiente consulta:

```
select nombre, boleta from alumno where id=5;
```

En donde la petición es llegada al administrador de la Grid. Éste se encarga de verificar si alguno de los nodos se encuentra disponible. Consultado a las bases primero, que es en donde se realizó la consulta. Luego las tablas (alumno) en otro nodo, columnas

en dos nodos diferentes (nombre, boleta) y finalmente los registros (resultado) que se encuentran en los nodos restantes.

Las funciones principales del módulo Grid son:

1. La actualización de los nodos conectados a la red.
2. La ejecución de las operaciones a las bases de datos.
3. La comunicación con el cliente.
4. El monitoreo de los nodos disponibles.
5. Ejecución de submódulo consulta en los nodos de la Grid.
6. Envío de las operaciones en cada nodo disponible.

En la figura 3 se muestra la distribución del contenido de la base en los nodos de la GRID sin considerar el nodo usuario.

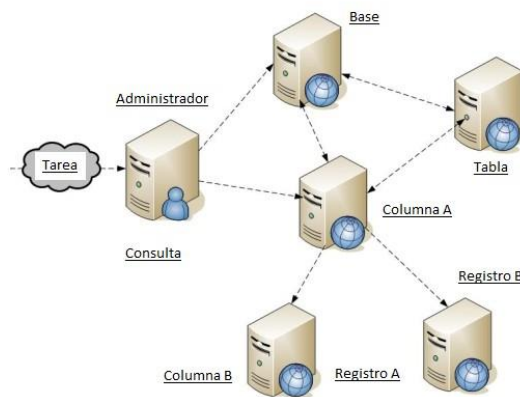


Fig. 3. Realización de operaciones de consulta.

B. El módulo cliente se conforma de las vistas del usuario final en el navegador web. Una aplicación web que hace uso de la conexión mediante el servicio web para extraer los datos que desee el usuario final.

En la figura 4 se muestra la relación de los componentes de SBDA-GRID: la vista de usuario, la WEB y la GRID.

La aplicación web puede estar en el servidor de Grid, se encargará de establecer la comunicación con el servicio web y mostrar los resultados al usuario, mediante archivos realizados en Java y lenguaje HTML. Además se realizó el manejo de sesión de los usuarios para controlar el acceso a las bases de datos específicas a un usuario y su tipo.

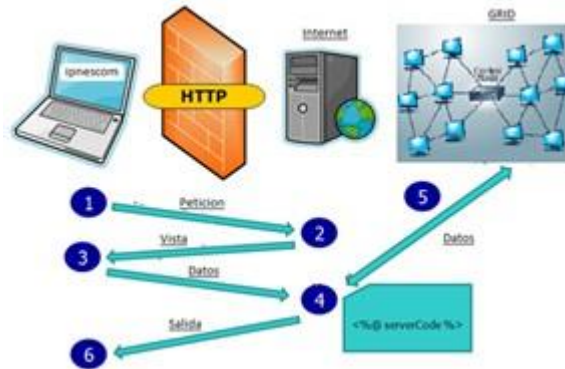


Fig. 4. Componentes SBDA-GRID.

La aplicación web hace uso de las clases stubs del servicio web para crear objetos que permiten la conexión con la arquitectura donde solamente crea las vistas para mostrarle los datos finales al usuario sin necesidad de realizar ninguna conexión a la Base de Datos.

Las funciones principales del módulo Cliente son:

1. Mostrar las vistas al usuario.
2. Extraer la información del usuario
3. La comunicación con el servidor grid.
4. Mostrar el resultado al usuario.

4 Resultados

Para verificar el correcto funcionamiento de SBDA-GRID se realizaron pruebas de consistencia de las bases repartidas en los nodos, dado que la consistencia de los datos fue programada ésta debe conservar los datos actuales al detener un nodo y volverlo a ejecutar. Es decir, debe mantener los datos de la base de datos de respaldo que se encuentran en el nodo principal. Se logró que la base de los nodos coincidiera con los registros de las demás bases en su totalidad.

Se tomó la velocidad que se tarda desde la petición del usuario en el cliente llevando a cabo operaciones en varios nodos y en el nodo principal arrojando el tiempo de respuesta de 800 y 900 milisegundos

Por parte del cliente se verificó la consistencia de los datos, tanto de la parte de Gestión de Usuario como de la de Gestión de Bases de Datos. Para verificar que la información mostrada en el cliente coincidiera con los datos que se encuentran en la base de datos. En la figura 5 se puede observar las bases de datos desde la vista del cliente.

Se realizaron pruebas de procesamiento, para las cuales, se utilizó el comando “top”, que muestra un listado en tiempo real de los procesos que se están ejecutando en el sistema, especificando además el porcentaje de CPU y el porcentaje memoria (KB) que se está utilizando.

En el servidor el uso de CPU es de 0 %, debido a que todo el procesamiento lo lleva a cabo el nodo1. En el nodo 1 (prueba 2), se obtuvo un promedio de 8% de uso del CPU y de uso de memoria un promedio de 8.66%.



Fig. 5. Visualización de bases de datos.

El procesamiento varía dependiendo del tipo de operación que se envía, tanto al Sistema Gestor de Base de Datos como al contenedor, aunque estos valores siempre tienden a un valor semejante. Sin embargo, el procesamiento se lleva a cabo en los nodos a los que se envíe la tarea, bajando la carga de trabajo al nodo administrador. La tabla 1 muestra la comparación de procesamiento de SBDA-GRID.

Tabla 1. Procesamiento de CPU.

Contenedor de Servicios Web					
	Memoria Virtual	Memoria Física	Memoria Compartida	%CPU	%MEM
Servidor	1265m	193m	9548	0	8.6%
Nodo 1 (prueba 1)	727m	171m	9496	7	8.6%
Nodo1 (prueba 2)	730m	171m	9499	8	8.6%
Nodo1 (prueba 3)	729m	172m	9500	9	8.8%
			PROMEDIO	8	8.66%

Los resultados obtenidos de las pruebas realizadas fueron satisfactorios cumpliendo con los requerimientos en su totalidad.

5 Conclusiones y trabajos a futuro

Se ha podido comprobar las ventajas que brindan las herramientas que emplean procesamiento en diferentes equipos, logrando disminuir la carga de las tareas de una sola máquina. Se elaboró un cliente web funcional, eficiente y principalmente transparente para el usuario. La tecnología Grid además del procesamiento que ofrece, una de las grandes ventajas es que permite el almacenamiento de información en los nodos; parte que se utiliza en la aplicación desarrollada. Se solventó la creación de bases de datos en la aplicación desarrollada, manipulando los metadatos de las mismas. Entonces podemos decir que se logró la gestión de BD y además minimizar los costos de procesamiento y almacenamiento.

Los trabajos a futuro para SBDA-GRID consisten en: Realizar consultas más elaboradas que impliquen el uso de operadores de MySQL y añadiendo más condiciones. Agregar nodos dinámicamente a la Grid para aumentar el aprovechamiento de recursos. Realizar respaldo de las bases de datos creadas por el usuario en la aplicación web.

Agradecimientos. Nosotros agradecemos a la ESCOM-IPN por las facilidades otorgadas para el desarrollo del SBDA-GRID: Implementación de un Servicio de Base de Datos en una Arquitectura Grid.

Referencias

- 1 IEEE (2005). Grid Architecture Storage Utilising Grid Computing for Dynamic Data Storage. 2da. Edición, IEEE Computer Society Washington, DC, USA, (2005)
- 2 N. Soriano. Computación distribuida en Grid y en la nube. [online] Retrieve from: <http://www.suite101.net/article/computacion-distribuida-en-grid-y-en-nube-a1937>. [Accessed: August (2012)].
- 3 IEEE COMPUTING. Grid Service Architecture plan gaining momentum. [online] Retrieve from: <http://cs.cinvestav.mx/jorge/bib/webservices/gridserplan.pdf>. [Accessed: August (2012)].
- 4 Oracle Solaris 11. The First Cloud OS. [online] Retrieve from: <http://www.oracle.com/us/products/servers-storage/solaris/solaris11/overview/index.htm>. [Accessed: August (2012)].
- 5 Uso de Grid Computer en la creación de BD espaciales. [online] Retrieve from: <http://es.scribd.com/doc/59521334/Grid-Computer-y-Bd-Espaciales>. [Accessed: March (2012)].
- 6 Ramez A. Elmasri, Shamkant B. Navathe . Fundamentos de Sistemas de Bases de Datos. 3era Edición. Editorial Pearson, (2007)