

UNIVERSIDAD DE CASTILLA-LA MANCHA
ESCUELA SUPERIOR DE INFORMÁTICA



Modelos Avanzados de Base de Datos

LISTADO DE PREGUNTAS DEL TRABAJO RENDIMIENTO 1:
BASES DE DATOS PARALELAS Y GRID

AUTORES: Funcionalidad 3

Juan Andrada Romero
Juan José Antequera Flores
Jose Domingo López López

9 de mayo de 2010

1. PREGUNTAS SELECCIONADAS

Pregunta En el trabajo se habla acerca de sistemas paralelos y se comenta que existen arquitecturas paralelas de *disco compartido* y *sin compartimiento*. Pero, ¿cuáles son los cuatro tipos de arquitecturas utilizadas en sistemas paralelos?.

Respuesta

- **Arquitectura de memoria compartida:** los procesadores y discos tienen un acceso común a memoria, normalmente a través de un bus o de una red de interconexión. Esta arquitectura es muy eficiente en la comunicación entre procesadores pero no se pueden tener más de 32 o 64 procesadores porque el bus o la red de interconexión se convertiría en un cuello de botella. En las arquitecturas de memoria compartida, cada procesador suele tener una memoria caché muy grande para evitar acceder muy a menudo a la memoria compartida, aunque se debe mantener la coherencia de los datos en todas las cachés.
- **Arquitectura de disco compartido** Todos los procesadores podrán acceder a todos los discos y cada procesador tiene memoria privada. De este modo, el bus deja de ser un cuello de botella y existe la tolerancia a fallos, pues se duplican los datos en cada disco y si un procesador falla, otros procesadores podrán ejecutar las sentencias de acceso a los datos.
- **Arquitectura sin compartimiento** Cada nodo de la máquina consta de procesador, memoria y uno o más discos. Un nodo se comunica con otro a través de una red de interconexión de alta velocidad, por la que viajan las peticiones a los datos, los accesos a los discos remotos y los resultados obtenidos. De este modo, cada nodo actúa como un servidor de datos, de manera similar a cómo sucede en las bases de datos distribuidas. Siguiendo esta arquitectura, la escalabilidad es muy alta, pero tiene como inconveniente el coste, ya que el software es más complejo para evitar bloqueos y mantener la coherencia en cada nodo.
- **Arquitectura jerárquica** Combina las características de los modelos anteriores. El sistema está formado por una arquitectura sin compartimiento donde cada nodo es en realidad un sistema de disco compartido o un sistema de memoria compartida.

Fuentes: [6], [1]

Pregunta En el trabajo se comenta que una base de datos *Grid* debe tener como requisito la seguridad. Pero, ¿cuáles son los mecanismos que se llevan a cabo para garantizar la seguridad en este tipo de bases de datos?

Respuesta Para la seguridad en bases de datos Grid, hay que tener en cuenta el control de acceso, autenticación y gestión de cuentas de usuarios. Esto se conoce como la *Triple A* (AAA), por sus siglas en inglés: Access Control, Authentication y Accounting.

Para poder acceder a los datos del sistema Grid, se debe comprobar la identidad del usuario (autenticación), sus permisos y privilegios de lectura/escritura (control de acceso) y registrar todas las operaciones que dicho usuario realice (tarea de *accounting*). Para ello, hoy en día se utiliza GSI (*Grid Security Infrastructure*), que permite la encriptación, confidencialidad e integridad de los datos, protección contra ataques de denegación de servicio y detección de paquetes fuera de orden.

GSI está basado en el uso de claves públicas encriptadas, certificados X.509 y otros mecanismos de seguridad como SSL. Para realizar la autenticación y otros mecanismos de seguridad, GSI se apoya en algunos *middleware*, como el *middleware* **Spitfire**, que permite el mapeo de certificados de usuarios a los distintos roles permitidos en la base de datos. También utiliza otros mecanismos, como el Gestor de Recursos de Almacenamiento (SRB - *Storage Resource Broker*), que permite la autenticación de los clientes ante los servidores.

Fuentes: [3], [4], [5]

Pregunta Como se indica en el trabajo, un sistema *Grid* permite la virtualización de recursos y el acceso de manera transparente para el usuario. ¿Cuáles son los diferentes niveles de transparencia que deben darse en los servicios proporcionados en un sistema de tipo *Grid*?

Respuesta Según la arquitectura OGSA (*Open Grid Services Architecture*), se definen varios niveles de transparencia para los datos:

- **Transparencia de heterogeneidad:** se refiere a que los mecanismos de acceso a los datos deben ser independientes de las distintas implementaciones de los SGBD (DB2, Oracle, etc.) y de la propia estructura de los datos.
- **Transparencia de localización:** una aplicación debe ser capaz de acceder a los datos, independientemente de donde estén localizados.
- **Transparencia de nombres:** una aplicación debe ser capaz de acceder a los datos sin conocer el nombre del recurso que los contiene (como la URL, por ejemplo). Ésto debe ser proporcionado por algún tipo de mapeo o de DNS.
- **Transparencia de distribución:** una aplicación debe poder consultar y manipular datos sin tener que ser consciente de que los datos están distribuidos. Es decir, la aplicación

debe poder manejar los datos como si utilizase una base de datos centralizada y única, sin preocuparse de como están distribuidos los datos en los diferentes recursos del sistema *Grid*.

- **Transparencia de replicación:** la aplicación no tiene que ser consciente de que los datos están replicados en diferentes recursos, sino que las operaciones de consulta y manipulación de los datos se deben tratar como si existiese una única bases de datos. Por tanto, aunque los datos estén distribuidos y replicados, se debe proporcionar la sensación de tener una única base de datos lógica.

Fuentes: [4], [5]

2. PREGUNTAS DESCARTADAS

Pregunta ¿Qué SGBD comerciales implementan arquitecturas paralelas y en qué tipos de aplicaciones se utilizan las bases de datos paralelas?

Respuesta Las bases de datos paralelas se utilizan en SGBD comerciales como Oracle, Informix y Sybase.

En cuanto a las aplicaciones donde se utilizan este tipo de bases de datos, éstas se suelen utilizar en sistemas de **muy** altas prestaciones, pues aunque sea posible comprar sistemas cada día más grandes y rápidos, a veces no es económico sustituir el hardware cada pocos años (o incluso meses), periodo en el que se suele duplicar la información de una organización. Por ello, no es muy útil utilizar estas bases de datos en aplicaciones que no sean muy complejas ni en empresas que no necesiten una carga de datos y velocidad muy elevadas.

De este modo, los sistemas donde se suelen utilizar las bases de datos paralelas son aquellos sistemas que trabajan con enormes cantidades de datos y se necesita una velocidad de recuperación muy elevada. Se pueden destacar estos dos tipos de sistemas:

- **Sistemas de procesamiento analítico en línea:** son sistemas donde se trabajan con cargas de datos muy elevadas y donde hay que realizar consultas y cálculos a una gran velocidad, para disminuir el tiempo de espera del usuario. Por ello, las consultas y los cálculos se suelen paralelizar.
- **Sistemas de minería de datos;** al igual que en el caso anterior, se trabajan con cantidades de datos muy elevadas, siendo necesario paralelizar las consultas para intentar descubrir reglas o patrones en los datos e ir realizando en paralelo diferentes árboles de decisión.

Fuente: [6], [2]

REFERENCIAS

- [1] Arquitecturas de los Sistemas de Bases de Datos. http://www.google.es/url?sa=t&source=web&ct=res&cd=22&ved=OCBsQFjABOBQ&url=http%3A%2F%2Ffantares.itmorelia.edu.mx%2F~antolino%2FCurso_Tit2006%2FCurso_Tit_BDD_ArqIII.ppt&ei=IabmS5uCJ43w0gSH-anHBg&usg=AFQjCNE72YlYk9p8WcdPEwBKxR5G0KH9JA&sig2=fIcs-f10XKhQ2cEk9VFs_A.
- [2] La Disciplina de los Sistemas de Bases de Datos. Historia, Situación Actual y Perspectivas. <http://users.dsic.upv.es/~jorallo/docent/BDA/DisciplinaBD.pdf>.
- [3] Uso de bases de datos en GRID. http://ccoello.ucsur.tripod.com/docs/TE02_II2.pdf.
- [4] Grid Database Access and Integration: Requirements and Functionalities. <http://www.ogf.org/documents/GFD.13.pdf>.
- [5] Grid database management: issues, requirements and future directions. <http://www.scs-europe.net/conf/ecms2008/ecms2008%20CD/hpcs2008%20pdf/hpcs08-DBMS-Fiore.pdf>.
- [6] SGBDs. <http://www.utomde.com/asigna/sghd/MisResumenes.pdf>.