



Universidad Técnica Federico Santa María
Departamento de industrias
ELO-322 Redes de computadores

Grid computing

Nombres:

- Roberto Etcheverria
- Eduardo Castillo
- Yanara Velásquez

Carrera: Ing. Civil Telemática

Fecha entrega :26 de junio de 2009

Resumen:

“La Grid es un sueño, grid computing es una realidad...”

“La Grid” toma su nombre de una analogía con el “poder de la red” eléctrica; en inglés la red eléctrica, el tramado de ella, se denomina Grid (la traducción al español de Grid es “Malla” o “Grilla”). La idea era que accediendo al poder de cómputo a través de una malla computacional sería tan simple como acceder al poder eléctrico desde una red eléctrica.

La computación grid es un servicio para compartir poder computacional y capacidad de almacenamiento de datos a través de Internet.

La diferencia entre la computación grid y la World Wide Web, es que la primera utiliza Internet para ayudar a compartir el poder de procesamiento, mientras la Web utiliza Internet para ayudarnos a compartir información.

La computación grid está haciendo grandes contribuciones a la investigación científica, ayudando a los científicos del mundo a analizar y almacenar cantidades masivas de datos.

EL SUEÑO

El sueño de la computación grid comenzó con una conversación acerca de la creación de una “Grid” todopoderosa: una grid compuesta por la agrupación de muchas grids pequeñas, formando una red global de computadoras que pudiesen operar como un único gran recurso computacional.

En la realidad de la computación grid ya existen cientos de grids alrededor del mundo, cada una de ellas creadas para ayudar a grupos específicos de investigadores o a ciertas agrupaciones de usuarios. Hoy, a través del mundo, investigadores e ingenieros de software, trabajan para acercar “la Grid” al logro del sueño.

Introducción

Hoy en día múltiples tecnologías nos han proporcionado múltiples ventajas a la hora de comunicarnos. Internet es una de las tecnologías que ha tenido más impactos a nivel mundial que ha permitido desarrollar aplicaciones, proporcionar aplicaciones entre varios, pero ahora nos cabe la duda, ¿existen aplicaciones que permitan mejorar las tecnologías de la información?, es aquí donde nace grid computing. Grid computing es una tecnología que permite que los ordenadores utilicen Internet para compartir con otras redes información, poder de cálculo y capacidad de procesamiento.

El término grid viene de la analogía eléctrica (Electric Powers grid): nos podemos enchufar a grid para obtener potencia de cálculo sin preocuparnos de donde viene al igual que hacemos cuando enchufamos un aparato eléctrico.

Grid computing tiene muchas ventajas, un caso típico que cabe destacar es que podemos ejecutar un programa de otra persona sin necesidad de haberlo instalado propiamente tal en nuestra máquina. En este caso grid computing interviene ejecutando remotamente en la computadora que almacena ese programa. Además cabe destacar que cuando la aplicación necesitada se encuentra ocupada, grid computing automáticamente buscará una copia de este sin importar donde se encuentren.

Grid computing es una herramienta útil y que necesita ser implementada en su totalidad, queremos que deje de ser ciertamente deje de ser un sueño.

¿Qué es Grid Computing?

Imagina muchos millones de computadoras en todo el mundo y pertenecientes a muchas personas distintas. Ahora imagina que conectas todos esos computadores en un único, enorme y súper potente computador. Ese único, inmenso y súper potente computador global es lo que muchas personas sueñan que será “La Grid”.

“La Grid” toma su nombre de una analogía con el “poder de la red” eléctrica; en inglés la red eléctrica, el tramado de ella, se denomina Grid (la traducción al español de Grid es “Malla” o “Grilla”). La idea era que accediendo al poder de cómputo a través de una malla computacional sería tan simple como acceder al poder eléctrico desde una red eléctrica.

¿CUÁN LEJOS PUEDE LLEGAR ESTA ANALOGÍA?

Comparemos una red de poder eléctrico con el sueño de la “Grid”:

Red de poder eléctrico

Nunca te has preocupado por saber desde dónde viene la electricidad que estás usando. Simplemente sabes que cuando enchufas la tostadora de pan al tomacorriente de la muralla, recibirá el poder eléctrico que necesitas para que haga el trabajo. La **infraestructura** que hace que esto sea posible se denomina “grid de poder” (malla de poder). Ella enlaza las plantas de poder de diversos tipos con tu hogar a través de estaciones de transmisión, estaciones de poder, transformadores, líneas de poder y otras.

La grid de poder es **omnipresente**: la electricidad está disponible esencialmente en todas partes y puedes acceder sencillamente a ella mediante un enchufe de pared estándar.

La “grid de poder” es un **servicio**: tú pides electricidad y la obtienes. Además, pagas por lo que recibes.

La Grid

Nunca te preocuparás por saber desde dónde proviene el poder de cómputo que estás empleando. Simplemente sabes que cuando conectas tu computador a Internet, éste obtendrá el poder de cómputo que necesitas para realizar tu trabajo. La **infraestructura** que hace que esto sea posible se denomina “la Grid”. Ella enlaza recursos computacionales tales como PCs, estaciones de trabajo, servidores, elementos de almacenamiento, y provee los mecanismos necesarios para acceder a ellos.

La Grid será **omnipresente**: los recursos computacionales remotos serán accesibles desde distintas plataformas, incluyendo computadoras de escritorio, PDAs y teléfonos móviles, y accederás sencillamente a la Grid a través de tu navegador Web. La Grid es un **servicio**: tú pides poder de cómputo o capacidad de almacenaje y lo obtienes. También, pagas por lo que recibes.

Las grid del mundo

Los biólogos están empleando grids para simular miles de posibles drogas en sus computadores, con el objetivo de descubrir una molécula capaz de bloquear proteínas específicas de ciertas enfermedades.

Los científicos de la Tierra están empleando grids para registrar los niveles de ozono, usando satélites, descargando diariamente cientos de Gigabytes de datos (¡el equivalente a 150 CDs cada día!)

Los físicos de altas energías están aplicándolas grids en su búsqueda por una mejor comprensión del Universo, sobre la base de una grid de decenas de miles de computadoras de escritorio para almacenar y analizar los 10 Petabytes de datos (¡equivalentes a los datos que pueden contener alrededor de 20 millones de CDs!) producidos anualmente por el Gran Colisionador de Hadrones (Large Hadron Collider). Miles de físicos en docenas de universidades alrededor del mundo quieren analizar esos datos..

Los ingenieros están usando las grid para estudiar energías alternativas, tales como la fusión de energía.

Los artistas hoy usan las grid para crear complejas animaciones para las películas (puedes ver un buen ejemplo en Kung Fu Panda).

Los científicos sociales hoy estudian la vida de las abejas, el maquillaje que emplea nuestra sociedad y los secretos de la historia, mediante el uso de las grid.

Arquitectura de grid

Habitualmente se describe la arquitectura del grid en términos de “capas”, ejecutando cada una de ellas una determinada función. Las capas más altas son las más cercanas al usuario y las inferiores las más próximas a las redes de computación, distinguiendo entre:

Para poder hacer todo lo anterior, las aplicaciones que se desarrollen para ser ejecutadas en un ordenador concreto, tendrán que adaptarse para poder invocar los servicios adecuados y utilizar los protocolos correctos. Sin embargo, una vez adaptadas al grid, miles de usuarios podrán usar las mismas aplicaciones, utilizando las capas de middleware para adaptarse a los posibles cambios en el tejido del grid.

- **Capa de aplicación.** Formada por todas las aplicaciones de los usuarios, portales y herramientas de desarrollo que soportan esas aplicaciones. Es la capa que ve el usuario y que proporciona el llamado serviceware, que recoge las funciones generales de gestión tales como la contabilidad del uso del grid que hace cada usuario.
- **Capa de middleware.** Responsable de proporcionar herramientas que permiten que los distintos recursos participen de forma coordinada y segura en un entorno grid unificado.

- **Capa de recursos.** Constituida por los recursos que son parte del grid: ordenadores, supercomputadoras, sistemas de almacenamiento, catálogos electrónicos de datos, bases de datos, sensores, etc.
- **Capa de red.** Encargada de asegurar la conexión entre los recursos que forman el grid.

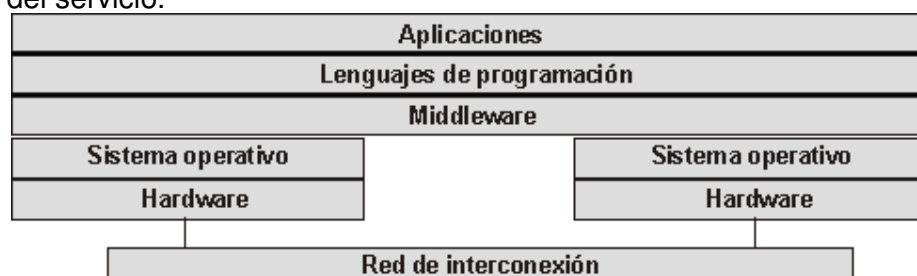
Para poder hacer todo lo anterior, las aplicaciones que se desarrollen para ser ejecutadas en un ordenador concreto, tendrán que adaptarse para poder invocar los servicios adecuados y utilizar los protocolos correctos. Sin embargo, una vez adaptadas al grid, miles de usuarios podrán usar las mismas aplicaciones, utilizando las capas de middleware para adaptarse a los posibles cambios en el tejido del grid.

Middleware

De todas estas capas, la más interesante es el **middleware**, el auténtico cerebro del grid, que se ocupa de las siguientes funciones:

- Encontrar el lugar conveniente para ejecutar la tarea solicitada por el usuario.
- Optimizar el uso de recursos que pueden estar muy dispersos.
- Organizar el acceso eficiente a los datos.
- Autenticar los diferentes elementos.
- Ejecutar las tareas.
- Monitorizar el progreso de los trabajos en ejecución.
- Gestionar automáticamente la recuperación frente a fallos.
- Avisar cuando se haya terminado la tarea y devolver los resultados.

El middleware está formado por muchos programas software; algunos de estos programas actúan como **agentes** (agents) y otros como **intermediarios** (brokers), negociando entre sí, de forma automática, en representación de los usuarios del grid y de los proveedores de recursos. Un elemento fundamental del middleware son los **metadatos** (datos sobre los datos), que contienen, entre otras cosas, toda la información sobre el formato de los datos y dónde se almacenan (a veces en varios sitios distintos). Los agentes individuales presentan los metadatos referidos a los usuarios, datos y recursos. Por otro lado, los intermediarios se encargan de las negociaciones entre máquinas para la autenticación y autorización de los usuarios, de definir los acuerdos de acceso a los datos y recursos y, en caso de que corresponda, el pago por los mismos. Cuando queda establecido el acuerdo, un intermediario planifica y las tareas de cómputo y supervisa las transferencias de datos necesarias para acometer cada trabajo concreto. Al mismo tiempo, una serie de agentes supervisores especiales optimizan las rutas a través de la red y monitorizan la calidad del servicio.



JDL (Job Description Language)

es el lenguaje usado para enviar trabajos a la grid , usualmente se puede usar para re direccionar a un código en otro lenguaje pasado por parámetro, uno de los lenguajes más usados para re direccionar es python por su facilidad de programación. JDL en su implementación más común nos entrega dos archivos de texto uno con el resultado de la ejecución y otro con los errores que ocurrieron durante esta

GRID en CERN

El LHC es el experimento más grande jamás visto en la historia de la física, consiste en un gran acelerador de partículas de 25 kilómetros de perímetro a 100 metros bajo tierra, en este acelerador de partículas se generan alrededor de 40 millones de colisiones por segundo las cuales son filtradas por un proceso llamado trigger de las cuales quedan tan solo 100 colisiones “interesantes” cada colisión genera alrededor de 1 megabyte de información lo que se traduce en 100[MB/s] esta información debe ser almacenada distribuida y procesada manteniendo su integridad en todo el proceso. Pese a que CERN tiene una gran capacidad de cómputo esta no es suficiente para esta enorme cantidad de datos, por lo que se decidió implementar grid computing, pero esto represento un gran desafío puesto que al incorporar a todos los países involucrados resulto que ya varios contenían pequeñas grids con arquitecturas distintas (middleware), por esto se debió generar un supra middleware que permitiera coordinar todas estas “pequeñas” grid’s, este sistema está muy lejos de ser perfecto aun en la transferencia de estos datos se pierde información por incompatibilidad de versiones no detectadas entre otros factores.

UTFSM y la Grid

La UTFSM es la primera universidad en pertenecer a una grid en Chile, formando parte de EELA y recientemente de PIC (VO conformada principalmente por España y Portugal) con un pequeño cluster de 40 cores aspirando a potenciarlo sin embargo esta no es la mayor limitación ya que nuestro mayor problema son las bajas velocidades de transferencia

Siendo así los objetivos próximos para la universidad la adquisición de un cluster mas poderoso, la construcción de un data center, y por ultimo y probablemente las más difícil de alcanzar un enlace de 180 [Mb]

Conclusión

Grid computing favorece al trabajo de colaboración distribuida, y es y será una herramienta muy poderosa, capaz de suplir la necesidad de un único súper computador, disminuyendo los tiempos de espera en simulaciones y análisis de datos, y permitiendo así un gran desarrollo para con la ciencia

Sus principales inconvenientes provienen de la dificultad para sincronizar los procesos de todos estos equipos, monitorizando recursos, asignando cargas de trabajo y estableciendo políticas de seguridad fiables.

En definitiva, nos encontramos ante un paradigma de computación distribuida altamente versátil, escalable y que permite combinar la potencia de muchos equipos para lograr una capacidad global prácticamente ilimitada.

Se trata de un paradigma computacional en fase de desarrollo actualmente que ya ofrece servicios a muchos campos de investigación y que en el futuro tendrá una mayor influencia si cabe en dichos campos, al gozar de una estructura evolucionada respecto a la actual, con mayor robustez, mucho más rápida en cuanto a la comunicación entre sus equipos y, obviamente, con una mayor capacidad computacional de cada máquina debido a la evolución en términos de procesamiento y de rendimiento del hardware.

Bibliografía

- <http://www.gridcafe.org/>
- <https://twiki.cern.ch/twiki/bin/view/Atlas/ADCoS>
- <http://www.ing.unp.edu.ar/wicc2007/trabajos/PDP/131.pdf>
- <http://www.ramonmillan.com/tutorialeshtml/gridcomputing.htm>
- <http://www.textoscientificos.com/redes/computacion-grid/funcionamiento>