“Computación **GRID**”

Ronald Barrera L., Héctor Peñafiel, María Verónica Serrano, Jesús Gaibor M,
Facultad de Ingeniería Eléctrica y Computación, Escuela Superior Politécnica del Litoral (ESPOL)
Profesor Director: Ing. Carlos Jordán. cjordan@espol.edu.ec
Materia: Organización y Arquitectura de Computadores
Correos: rbarrera@fiec.espol.edu.ec, mserrano@fiec.espol.edu.ec, hpenafi@fiec.espol.edu.ec,
jgaibor@fiec.espol.edu.ec

Definición y Funcionamiento de Grid Computing.

La Computación Grid ha surgido como un importante campo. Contrario a otras arquitecturas donde el enfoque es crear Computadores de gran rendimiento medido en número de cálculos de punto flotante por unidad de tiempo.

La importancia del Grid esta definida en términos de Cantidad de trabajo que puede despachar en una unidad de tiempo.

La tecnología GRID no es revolucionaria, más bien se puede decir que ha evolucionado de otras tecnologías como son: Sistemas Distribuidos, Vitalización, Web Services, Internet, Criptografía. Y así se puede observar que ninguna de estas tecnologías es realmente nueva, han existido por mucho tiempo y han estado sirviendo por ya muchos años.

Grid Computing toma importantes características de estas tecnologías para desarrollar un sistema que pueda proveer recursos computacionales para tareas específicas.

Las tareas pueden ser: Predecir tendencias del Mercado, Predecir Terremotos, Maremotos, Estudio del Espacio, Meteoritos, Servicios Empresariales Distribuidos en zonas Geográficas Extensas.

Por lo tanto Grid es una tecnología Evolucionaría que apalanca la Infraestructura de las Tecnologías de la Información.

Virtualización es un término importante, en grid se refiere a integración uniforme de sistemas heterogéneos geográficamente distribuidos. Y así permite que a los usuarios hacer uso de los servicios del Grid de forma transparente. Lo que se traduce en, que el usuario no necesita estar enterado donde esta localizado los recursos computacionales que lo van a satisfacer. Y así desde la perspectiva del usuario hay solo un punto de entrada al sistema Grid.

Ellos solo envían una petición de servicio desde este punto, esta petición es recibida por el Grid que se encarga de encontrar recursos computacionales disponibles que puedan servir al usuario.

“Organización Virtual”.- Dinámica colección de múltiples Organizaciones que comparten recursos de forma coordinada.

Para entender Sistema Grid es importante tener en cuenta tres valiosos conceptos: El primero es Virtualización que ya lo hemos explicado. El segundo es Heterogeneidad, que implica que al hablar de Instituciones diferentes con recursos diferentes en hardware, software, sistemas operativos y ancho de banda decimos que una Organización Virtual es una colección de recursos diferentes (heterogeneos).El tercer y ultimo término es dinamico, Las organizaciones pueden unirse o abandonar una Organización Virtual (VO) para su conveniencia. Pero tienen que recordar que una VO es dinámica en lo que a asignación de recursos computacionales.

Grid vs Otros Sistemas Distribuidos

Los sistemas distribuidos actuales por ejemplo RMI y CORBA sirven a propósitos de Organizaciones Simples y tienen un control Centralizado. Por el contrario Grid no tiene un control Centralizado y sirve al propósito de un largo número de Organizaciones. Para definir Grid Computing se necesita recursos heterogéneos, sistema dinámico, virtualizable. Los sistemas distribuidos tienen recursos heterogéneos pero limitados a una simple organización. Y más importante los Sistemas Distribuidos se enfocan en Compartir Información muy a menudo usando una arquitectura Cliente Servidor. En Grid el compartir no se limita solo a información, este se puede extender a hardware y software. Grid soporta descubrimiento de Recursos y Monitorización a una escala Global.

Uso de Grid

Habilita la formación de una Organización Virtual.-Grid habilita la colaboración entre múltiples organizaciones para compartir sus recursos informáticos de infraestructura. Esta colaboración no se vera limitada solo a compartir archivos e implica acceso directo a recursos computacionales como hardware y software. Todos los miembros del GRID pueden ser organizados de manera dinámica en múltiples Organizaciones Virtuales, y cada una de estas Organizaciones Virtuales tendrá sus propias políticas Administrativas y de control. Todas serán parte del GRID y podrán compartir sus recursos computacionales. Los recursos computacionales que se repartirán podrían ser: datos, hardware especial, capacidad de procesamiento, información de los recursos existentes para el funcionamiento del GRID. Es muy importante saber que la Organización Virtual esconde la complejidad del grid del usuario. También se puede definir políticas de seguridad y de priorización de recursos para diferentes usuarios.

Tolerancia a Fallos.- Suponga que usuario del Grid ejecuta su trabajo y el Grid lo desarrolla en un nodo particular. El trabajo asigna apropiados recursos basado en disponibilidad y en la política programada del grid. Ahora suponga que el nodo que esta ejecutando el trabajo se cae por alguna razón. El grid esta en capacidad de hacer una reasignación automática del trabajo a otros recursos disponibles cuando una falla se a detectado y así se conjuga en concepto de virtualizacion. El usuario no se entero de la falla del grid y además el usuario realizo el trabajo.

Balanceo de Carga y Compartir una amplia gama de recursos .-Balancear las Cargas y compartir recursos son importantes aspectos del grid. Que proveen las principales características en la gestión de los recursos computacionales. Supón que un nodo del grid esta sobre saturado. Entonces el Algoritmo de Programación del Grid puede reorganizar las tareas a otros sistemas menos congestionados u ociosos.

Procesamiento en Paralelo .- Algunas tareas pueden ser divididas en múltiples subtareas, cada una de estas puede ser ejecuta o procesada en una maquina diferente. Como por ejemplo el modelo matemático de renderizar una imagen en 3d. Luego los resultados de estas subtareas procesadas en paralelo se combinan para producir una salida deseada. Hay sin embargo, condicionamientos como puede ser el tipo de tarea y si esta puede ser particionada en subtareas. También existe un límite en el número de subtareas que puede ser dividido, limitando el rendimiento. Si dos o más de estas subtareas están operando en la misma estructura de datos debe existir un mecanismo de bloqueo similar al control de concurrencia de una base de datos o las interrupciones de los Sistemas Operativos para lograr que los datos accedidos no sean inconsistentes.

Quality of Services.- Un grid puede usarse como un escenario donde los usuarios enviar sus trabajos y obtienen una salida, y esto ser medido basado en alguna métrica como puede ser tiempo de completar el trabajo. El usuario espere calidad en los servicios entregados o disponibles a su uso.

Esto es especificado en el "Service Level Agreement" (SLA). SLA especifica lo mínimo en calidad de servicio, disponibilidad, etc. Esperado por el usuario. Para ser más específicos SLA puede especificar el mínimo tiempo esperado de actualización del sistema. Así también como el grid debe especificar una tolerancia a fallos, fiabilidad, capacidad de procesamiento en paralelo para ciertas tareas. Basados en los requerimientos del usuario una tarea puede tener una mayor prioridad sobre otras tareas de usuario en el algoritmo de programación del grid. Por ejemplo un usuario puede requerir los servicios de una aplicación en tiempo real y con ello una elevada calidad de Servicio comparado a los servicios diferentes de otros usuarios. QoS puede ser entregado basado en reservación de recursos para ciertos trabajos. Si un recurso reservado para un usuario especifico esta libre aun. Este puede reportar su status como un nodo en el grid. El recurso puede entonces ser usado por el grid hasta que este se necesite.

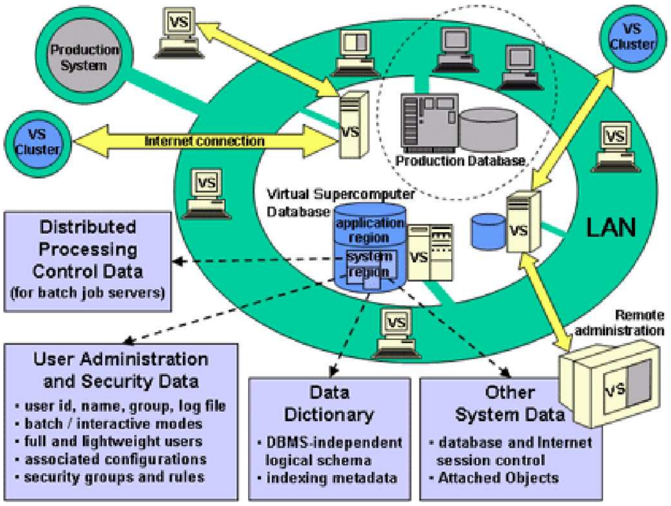
Conceptos Básicos de las Arquitecturas GRID

Seguridad.- Así como en cualquier otro sistema en el mundo, la seguridad es un aspecto vital de los sistemas Grid. Las tres características más importantes de seguridad que el Grid debe proveer son: Login, Autenticación, Autorización, La primera significa que un usuario es capaz de registrarse haciendo uso de sus credenciales de seguridad [usuario y contraseña] y tener acceso a los servicios del grid por un cierto periodo de tiempo. Autenticación significa a ser capaz de crear una identidad, como por ejemplo cuando login en tu cuenta de mail te autenticas con el servidor dando tu usuario y tu contraseña. Autorización es el proceso de gestionar y controlar los privilegios de los usuarios.

Gestión de Recursos.- El grid debe optimizar los recursos bajo su disposición para lograr el máximo posible de ancho de banda. La gestión de recursos incluye envió de trabajo remotamente, chequear el status cuando esta en progreso y obtener la salida al final de la ejecución. Cuando un trabajo es enviado, los recursos disponibles son descubiertos a través del directorio de servicios. Entonces los recursos son seleccionados para procesar el trabajo. Esta decisión es hecha por otro componente del Grid nombrada como GRID PLANIFICADOR. La decisión puede basarse en muchos factores. Por ejemplo, si una aplicación que consiste de algunos trabajos, el planificador puede planificar trabajos secuenciales para un recurso. La decisión puede también basarse en las prioridades de la tarea de un Usuario específico. Siempre que corresponda con SLA.

Gestión de Datos.- La gestión de datos en el GRID cubre una amplia variedad de aspectos necesarios para gestionar largas cantidades de datos. Esto incluye seguridad de acceso a los datos, replicación y migración de datos, manejo de la metadata, indexado, caching.

Monitoreo de Información y Descubrimiento de Información.-Se ha mencionado que el Planificador del Grid debe ser consiente de los recursos disponibles que este en posibilidad de procesar un trabajo. Bueno esta información se obtiene del servicio de descubrimiento. Este servicio provee una lista de recursos disponibles por el grid y su status. Cuando el Planificador realiza una petición al servicio de descubrimiento por lo generar envía condicionamientos como “hallar los recursos mas relevantes a solucionar un trabajo”. Hablando de Capacidad computacional entonces diríamos que se seleccionara solo la maquina que sea capaz de procesar la tarea en el tiempo estimado. Este servicio puede funcionar de dos formas distintas uno como Web services, donde tiene una interfaz de acceso a los datos y por otro lado como una lista de recursos disponibles.



Estándares de GRID

Web Services.- Así como nosotros podremos ver Grid Services, definidos por OGSA como una extensión de web services. Así también grid services re potenciará lo aprovechable de la especificación de Web services. Los cuatro servicios web básicos son:

1. XML
2. SOAP
3. WSDL
4. UDDI

Open Grid Services Architecture (OGSA).- Define el framework base para la implementación de un Grid. Buscando estandarizar los servicios del Grid como descubrimiento de recursos, gestión de recursos, seguridad, etc. A través de una interfaz de web services. Esto también incluye aquellas características que no son necesarias para la implementación de un Grid. Esta basado en la existente "Web Services Especification" y le añade características para hacerla la suit base para los ambientes grid.

Open Grid Services Infraestructure (OGSI).- Recordemos que OGSA describe las características que son necesarias para implementar servicios en un Grid como un web services. Este sin embargo no nos provee los detalles de la implementación. OGSI provee una especificación formal y técnica para la implementación de servicios GRID. Este provee una descripción de WSDL, y como define a un GRID. Y además de los mecanismo para la creación, gestión e interacción entre servicios.

Web Services Resource Framework (WSRF).- Al desarrollar el WSRF definimos una framework abierto y genérico para modelar y acezar a los estados de los recursos usando web services. Se definen convenciones para gestión de estados, habilitación de aplicaciones para descubrir u obtener e interactuar con el estado de los web services de una manera estándar. Las aplicación del Grid necesitan sabes de los estado para poder desarrollar las tareas de los usuarios

Algunos Ejemplos de Proyectos en GRID

TeraGrid

TeraGrid es una infraestructura abierta la investigación científica que combina los recursos de clase de liderazgo a las once sitios asociados para crear un sistema integrado, los recursos computacionales persistente.

Uso de las conexiones de red de alto rendimiento, el TeraGrid integra ordenadores de alto rendimiento, los datos de los recursos y herramientas, y de gama alta las instalaciones experimentales de todo el país. Actualmente, los recursos TeraGrid incluir más de un petaflop de capacidad de procesamiento y más de 30 petabytes de almacenamiento en línea y el archivo de datos, con acceso rápido y recuperación a través de redes de alto rendimiento. Los investigadores también pueden acceder a más de 100 bases de datos específicas de la disciplina. Con esta combinación de recursos, el TeraGrid es la mayor del mundo, la ciberinfraestructura distribuido más completo para la investigación científica abierta.

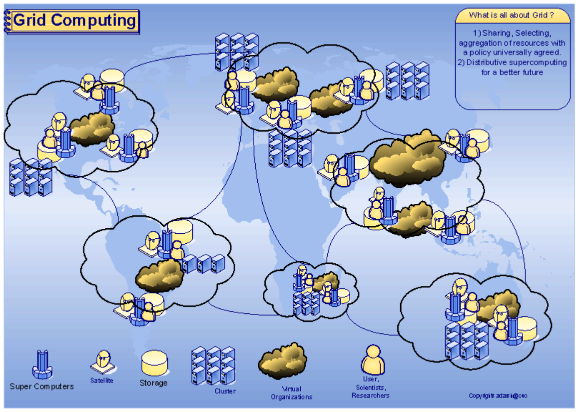
TeraGrid es coordinado por el Grupo de Infraestructura de Grid (GIG) de la Universidad de Chicago, trabajando en asociación con los sitios de proveedor de recursos: la Universidad de Indiana, Louisiana Optical Network Initiative, el Centro Nacional para Aplicaciones de Supercomputación, el Instituto Nacional de Ciencias Computacionales, Oak Ridge National Laboratory, Pittsburgh Supercomputing Center, Purdue University, San Diego Supercomputer Center, Texas Advanced Computing Center, y la Universidad de Chicago / Laboratorio Nacional de Argonne, y el Centro Nacional de Investigaciones Atmosféricas.

Grid5000

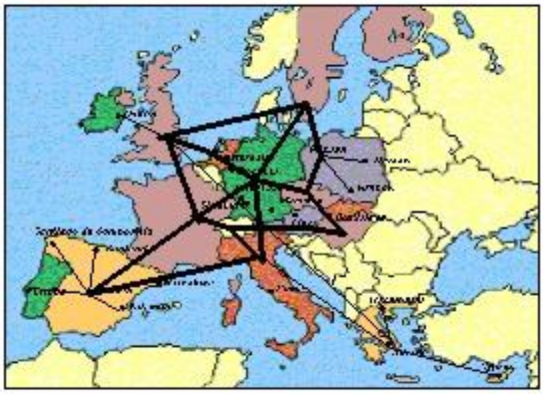
Este proyecto tiene por objeto la construcción de una plataforma experimental de recogida de cuadrícula 8 sitios distribuidos geográficamente en Francia. El principal objetivo de esta plataforma es la de servir como banco de pruebas experimentales para la investigación en Grid Computing.

Grid'5000 permitirá experimentos Grid Francia amplia en todas estas capas de software. Los planes actuales son para montar una plataforma física con 9 grupos, cada uno con 100 a mil ordenadores, conectados por la Educación y la Red de Investigación Renater. Todos los grupos estarán conectados a Renater a 1 Gb / s (10 Gb / s se espera en un futuro próximo).

Este gran esfuerzo de investigación en colaboración es un proyecto financiado por el Ministerio francés de Educación e Investigación, INRIA, CNRS, las Universidades de todos los sitios y algunos consejos regionales.



CERN

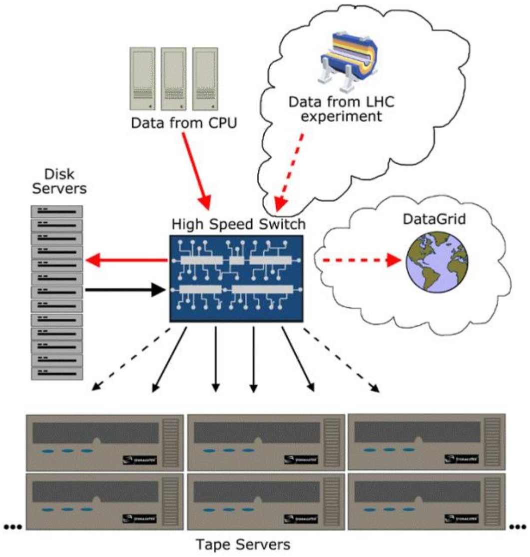


El Centro Europeo para la Investigación Nuclear (CERN) desarrollo del proyecto “DataGrid”, que tiene como objetivo unir a grandes bases de datos y usuarios en una futura red informática de alta velocidad.
El acuerdo, firmado, prevé que la Unión Europea aportará en un periodo de tres años, 9,8 millones de euros para el desarrollo de la nueva red. Los especialistas en informática consideran que el proyecto, que ha recibido el nombre de “DataGrid” es “el banco de pruebas ideal de un nuevo modelo de tratamiento informático a escala mundial y la evolución natural de la WWW”.

Según el CERN, que es uno de los mayores centros de investigación física del mundo, DataGrid unirá a superordenadores, procesadores, sistemas informáticos y usuarios en una red de alta velocidad. DataGrid se define pues como una alternativa para superar los límites de capacidad de tratamiento de datos que padece la WWW y su lentitud debido a la multiplicación del número de usuarios. Según el comunicado del CERN, DataGrid ofrecerá a los científicos de todo el mundo y a todos los usuarios, un acceso rápido a los recursos informáticos. Su primer uso, y al mismo tiempo su primer desafío, es el tratamiento de la gran cantidad de datos que generará la puesta en marcha del Gran Colisionador de Hadrones, LHC.

Todos los datos proporcionados por el LHC, el nuevo acelerador de partículas que se construye en el laboratorio del CERN, en la frontera franco-ginebrina, y que está previsto para el 2005, necesitará una potencia de cálculo equivalente a 100.000 ordenadores individuales de los que ahora existen.

DataGrid será útil, en particular, para los biólogos, ya que el uso de la genómica y la proteómica necesitan una gran potencia de cálculo, almacenamiento de múltiples datos y rápido acceso a la informática.



Referencias Bibliograficas e Internet

- Ebook – CRC Introduction To Grid Computing
  Marzo 2009

- Ebook - From P2p to Web Services and Grid
  Marzo 2009

- http://press.web.cern.ch/Press/PressReleases/Releases2003/PR06.03EStoragetek.html - Proyecto CERN - DATA GRID Architecture

- CRC – Introduction To Grid Computing Mar 2009

- Grid5000 – pagina Oficial - www.grid5000.fr

- TERAGrid – Pagina Oficial: www.teragrid.org