

GRID: la unión hace la fuerza

Cuando muchos pequeños se unen
y superan a los grandes



Por Juan Carlos Martínez
Ingeniero Superior de Telecomunicación

Para tener una visión actual de las tecnologías *GRID*, veámoslas como una solución a la creciente demanda de Computación y a la vez una aplicación natural de la enorme disponibilidad global de redes. Algunos ejemplos actuales e implementaciones nos van a servir para ver las ventajas e inconvenientes que acarrearán. Haremos una breve mención de tecnologías subyacentes, a modo de resumen, con referencias a sitios Web donde acceder a estas herramientas y sistemas en funcionamiento.

La aparición de la Web para el intercambio de contenidos e información fue el disparo de salida en la carrera del uso masivo de Internet. El siguiente paso en el desarrollo de Internet está ya viniendo de la mano de las tecnologías de "Malla" o "*Grid computing*". Al igual que la Web nos permitió acceder a la información que se necesitase, disponible desde cualquier sitio y en cualquier momento, la Grids o Mallas pondrán la capacidad de proceso que se necesite a disposición de usuarios, a lo largo y ancho de Internet. Por su

sencillez y concreción nos puede ayudar mencionar la definición de Ian Foster, co-director de la GriPhyN (*Grid Physics Networks*), que nos define o reconoce una Grid, con su "*Three Point Checklist*", como un sistema que: 1) ...coordina recursos no sujetos a control centralizado, 2) ...usan protocolos e interfaces estándar, abiertos y de propósito general, y 3) ...entregan una calidad de servicio importante.

La Malla o *Grid* conecta grupos de ordenadores, unidades de almacenamiento y redes, permitiendo a los cen-





tros de investigación y empresas, asignar dinámicamente los recursos de acuerdo a las necesidades del negocio. Al igual que la infraestructura física de caminos, puentes, suministro de agua, energía y telefónica que sustentan la sociedad moderna, la *Grid* es parte de una ciber-infraestructura que incluye estas tecnologías distribuidas de información y comunicación para potenciar el esfuerzo científico de investigación contemporáneo. Esta nueva tendencia del *Grid Computing* supone un cambio radical en la colaboración de sistemas conectados a Internet. Es un nuevo paradigma de computación, un modelo compartido que permite no solo la comunicación y almacenamiento sino el procesamiento de información por todo el mundo. Mediante la explotación de los rápidos avances en las tecnologías, las mallas empresariales y científicas permiten a las organizaciones de los sectores público y privado adaptar sus recursos de tecnología de la información a sus necesidades de negocio e investigación.

La evolución es continua. A finales de los 70 la conexión de los dos primeros ordenadores hace realidad el sueño del mundo académico de compartir información, nace Internet. A mediados de los 90, el hipertexto para enlazar millones de documentos construyó la *World Wide Web*. Con el milenio se usan las "farms" o granjas de nodos en *Intranets*, mientras en Internet ideas como Gnutella comparten datos entre personas a nivel mundial. El siguiente paso ya ha comenzado, la unión de ordenadores que comparten ciclos de procesamiento no usados, y cada vez aparece más claramente como una futura aplicación de Internet. Las *GRID* y *Peer-to-peer* (P2P) tienen mucho en común y especialmente la idea básica de compartición de recursos. Entre las características diferentes podemos ver la P2P como más anónima y generalizada en ordenadores de usuarios de Internet, mientras que las *GRIDs* nacen de una estructura de nodos más controlada y

jerarquizada en centros científicos. Una primera experiencia fue GriPhyN, para unir nodos en el proceso de Físicas de Altas Energías en EE.UU. En Europa con el apoyo de proyectos de CERN (Centro Europeo Investigación Nuclear) y el programa marco europeo se creó el *software* y red EDG (*Grid* de Datos Europea).

Podemos resaltar los puntos fundamentales de las Mallas o *Grid*:

1) su objetivo es la compartición de recursos en Internet de forma uniforme, transparente, segura, eficiente y fiable.

2) análoga a las redes de suministro eléctrico: ofrecen un único punto de acceso a un conjunto de recursos distribuidos geográficamente en diferentes dominios de administración (supercomputadores, *clusters*, almacenamiento, instrumentos, personal, bases de datos...), c) la tecnología *Grid* permite interconectar recursos en diferentes dominios de administración respetando sus políticas internas de seguridad y su *software* de gestión de recursos en la Intranet. Las *Grids* son apropiadas en particular en la computación de altas prestaciones debido a su enorme potencial respecto al intercambio y gestión de recursos. Hay tendencias a aprovechar los puntos fuertes de cada uno y a englobarlos dentro de una definición global de modelo *Grid*. Proyectos famosos que han introducido este tipo de conceptos comunes son *Emule* o *Limewire*, e igualmente en todo el mundo se usa el *Seti@Home*. Éste cuenta con miles de PCs repartidos por Internet que ceden tiempo de sus procesadores, ciclos de proceso desocupados, para analizar señales buscando patrones inteligentes extraterrestres. A nivel de centros de investigación se trabaja en estas tecnologías desde finales del 90, en proyectos desde Físicas de Partículas a Astrofísica o Biología, con una mejora exponencial en los últimos años.

Características

Llamando *GRID* al tipo de sistema de computación distribuido que permite

compartir recursos geográficamente distribuidos para resolver problemas a gran escala. Podemos utilizar recursos como ciclos de procesadores, disco, redes, dispositivos, instrumentación... Estos están distribuidos en la red de forma transparente, pero manteniendo un alto nivel de seguridad y una correcta política de gestión que tenga en cuenta parámetros tanto técnicos como económicos. Estas características son una aportación a las P2P. La infraestructura que facilita esta compartición se coordina en organizaciones virtuales, multi-institucionales, y dinámicas. Así, la *Grid* se puede ver como un prototipo o modelo que permite la creación de "las empresas virtuales" y "laboratorio virtuales", es una aplicación que trataremos más tarde.

Resumamos los recursos compartidos que pueden ser tan variados como:

- * Ordenadores, PCs, estaciones de trabajo, supercomputadoras, PDA, portátiles, móviles, etc;

- * Software - por ejemplo, los ASPs que alquilan aplicaciones especiales

- * Datos e Información, bases de datos catalogados - el acceso transparente a datos del genoma;

- * Instrumentos especiales - los radio telescopios cuyas señales van a SETI@Home para su estudio.

- * Personas/colaboradores.

La *GRID* aprovecha los recursos (de procesamiento, almacenamiento, periféricos, etc.) de todos los ordenadores conectados a una red, cuando estos no son utilizados. Esto permite que muchos ordenadores se alíen para conformar un enorme "ordenador virtual" capaz de procesar una cantidad de información inimaginable para un único ordenador, por muy potente que éste sea. El objetivo de la tecnología *Grid* es Integrar y optimizar, usando *middleware*, el uso de recursos distribuidos y de esta forma puede constituir la base para el desarrollo de los sistemas y las aplicaciones que guiarán el uso de Internet en pocos años. Las mallas permiten por tanto ejecu-





El objetivo de la tecnología Grid es Integrar y optimizar, usando middleware, el uso de recursos distribuidos y de esta forma puede constituir la base para el desarrollo de los sistemas y las aplicaciones que guiarán el uso de Internet en pocos años

tar el trabajo de forma transparente y para ello realizan una serie de funciones: encuentran los recursos, aseguran el acceso seguro y optimizado a los datos, monitorizan la ejecución y pueden "paralelizar" ciertos trabajos.

Negocio

Hay varias alianzas que veremos más adelante como la *Enterprise Grid Alliance* o *Globus*. Las empresas e instituciones que han participado en el desarrollo de estas tecnologías quieren entrar cuanto antes en un etapa de explotación comercial. Empresas como Microsoft y Sun Microsystems se han dado cuenta de la importancia que tendrá a medio plazo ofrecer GRID a sus clientes. Destacamos el papel de IBM invirtiendo en GRID como plataforma para ofrecer a sus clientes las ventajas como ahorro de tiempo y recursos económicos.

Estos proyectos, con vínculos y aportaciones mutuas entre instituciones y empresas, van generando una oferta de productos de calidad y servicio de alta disponibilidad, 24/7. Desde el punto de vista de usuarios de GRID, una de las claves del éxito empresarial está en el manejo de datos actualizados adecuadamente. Generar y mantener un flujo de información actualizado se convierte en una de las prioridades para que la toma de decisiones se haga con los datos necesarios y las últimas novedades sean integradas en el flujo de trabajo de la empresa. La GRID es la tecnología que permite compartir los recursos entre distintas empresas o entidades participantes en un proyecto.

Existen modelos de negocio desde la venta de *software* al alquiler de plataformas o servicios de computación. Unas empresas aportan experiencia de proveedor de servicios de red, y otras sistemas de apoyo al grupo de operaciones de centros de computación, herramientas que minimicen la carga de incidencias a operadores y faciliten las posibilidades de crecimiento de red.

Tendencias y Evolución de Computación en Red y Tecnología Grid

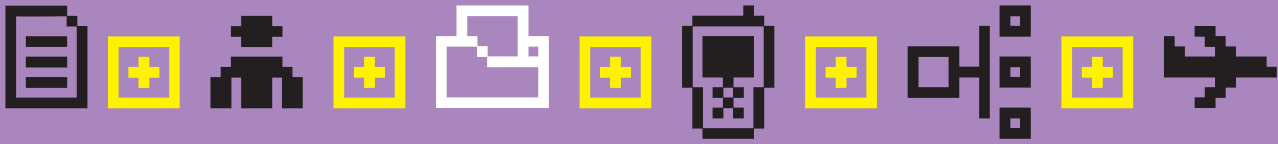
Paulatinamente se han desarrollado una variedad de tecnologías que permiten realizar cálculos masivos aprovechando la potencia computacional de sistemas interconectados, son diferentes modelos de computación en red que aportan mecanismos para aprovechar al máximo los recursos distribuidos normalmente poco utilizados. En la Evolución temporal de necesidades y tipos de aplicaciones que requieren potencia computacional, se trata en particular de dos grandes objetivos: Alto Rendimiento, o Ejecución de una aplicación en menos tiempo (ej. "paralelización" de tareas) y Alta Productividad, o Ejecución de un número mucho mayor de aplicaciones.

La solución clásica conocida es la Computación centralizada basada en servidor, sin embargo esto presenta una serie de problemas que podemos enumerar rápidamente por conocidos:

- * Falta de escalabilidad
- * Equipos muy caros
- * Mantenimiento muy caro
- * Una vez adquiridos pasan mucho tiempo desaprovechados
- * Las demandas de cálculo son puntuales
- * Problemas de fiabilidad

La siguiente opción que aparece son las soluciones de cluster, dedicado a computación paralela, como alternativa económica a la adquisición de un sistema multiprocesador. El *Cluster Computing* consiste en unir un conjunto homogéneo de computadores personales PCs, o estaciones, que suelen usar una red avanzada basada en router con *Fast Ethernet* (LAN) o *Myrinet* (SAN). Su principal ventaja es la mejor relación de coste a rendimiento y sus inconvenientes son dificultad de programación y mantenimiento. Los cluster ofrecen ciertas ventajas como la mejor relación coste/rendimiento, entre 3 y 10 veces mejor, sin embargo debemos resaltar sus principales inconvenientes:





* Alto coste de comunicaciones (Bus lento, Acceso secuencial bus , Sobrecarga TCP/IP)

* Mantenimiento

* Modelo de programación

Concluimos que los *cluster*, aun siendo una buena solución para aplicaciones medias o de Alta Productividad, no lo son en todos los casos. Aparece entonces la solución *Intranet Computing*, que nace de la unión de recursos *hardware* distribuidos en una red de área local. Dada la gran mayoría de PCs y estaciones que están frecuentemente desocupados, se trata de aprovechar esta potencia computacional. Son aquellas arquitecturas caracterizadas por:

- Explotación de potencia computacional distribuida
- Utilización de los equipos de una red departamental para ejecutar trabajos secuenciales o paralelos por medio de una herramienta de gestión de carga

La ventaja fundamental de *Intranet Computing* es que puede comportarse como los sistemas de alto rendimiento pero con un coste económico mínimo. Estas soluciones *Intranet* presentan un serie de ventajas:

- Aumentar el aprovechamiento de los recursos informáticos
- Ciclos de CPU a coste bajo
- Mejora de la escalabilidad
- Mejora de fiabilidad
- Facilidad de administración
- Facilidad de sustitución de equipos obsoletos

En *Intranet Computing* contamos ya con bastante software disponible para implementarlas:

SGE (Sun Grid Engine) de Sun Microsystems

www.sun.com/gridware

LSF de Platform Computing

www.platform.com

Condor de la Universidad de Wisconsin

www.cs.wisc.edu/condor

Ejemplos de soluciones y herramientas concretas son:

appLES de la Universidad de California

apples.ucsd.edu

Nimrod de la Universidad de Monash

www.csse.monash.edu.au/~rajkumar/ecogrid/

Otras herramientas de empresas como Avaki, United Devices, Entropia, Parabon...

Sin embargo tenemos también puntos negativos de estas soluciones:

* No pueden gestionar recursos fuera del dominio de administración. Aunque algunas herramientas (Condor, LSF o SGE) permiten colaboración inter-departamental asumiendo la misma estructura administrativa.

* No respetan las políticas de seguridad y de gestión de recursos de las organizaciones

* Protocolos e interfaces básicos no basados en estándares abiertos

* El único recurso que gestionan es la CPU , no ofrecen solución al manejo de los datos compartidos entre organizaciones.

Y esto a la larga se traduce en una serie de desventajas: a) tenemos escalabilidad limitada a la organización en picos de demanda, b) no podemos amortizar los recursos cuando están desaprovechados, c) ni compartir recursos con otras organizaciones.

Por último veamos dos tipos alternativos de sistemas:

* **Internet Computing:** Aparece en lógica evolución de red local a red global, buscando usar la potencia de los recursos distribuidos por Internet. Actualmente casi todas estas herramientas se limitan a ejecución paramétrica. Siguiendo el modelo. Su ventaja es la multiplicación de rendimiento y potencia. Pero los bajos ancho de banda y falta de seguridad, son inconvenientes de Internet que se reflejan aquí.

* **Computing Portal:** Viene de la aplicación y evolución paralela a la tecnología Web, frente a los modelos cliente/ servidor se desarrollan portales que proporcionen un único punto de acceso. Con un navegador web podemos invocar servicios como la ejecución en plataformas de alto rendimiento. Su ventaja es la facilidad en el acceso, al usar un método tan cotidiano como el web para acceder a pla-



La GRID aprovecha los recursos (de procesamiento, almacenamiento, periféricos, etc.) de todos los ordenadores conectados a una red, cuando estos no son utilizados. Esto permite que muchos ordenadores se alíen para conformar un enorme "ordenador virtual"



taformas de alto rendimiento.

Problemas y Soluciones Actuales

Siguiendo nuestra evolución de arquitecturas, y habiendo examinado las ventajas e inconvenientes de otras soluciones, enfoquémonos en los problemas que hay que resolver: 1) potencia, 2) coste de adquisición y mantenimiento, 3) desequilibrios de carga.

1. Potencia de cálculo

Es una necesidad básica y mientras un único sistema no será capaz de analizar los datos que almacenen sus discos, ni un único centro podrá analizar el volumen de información generado, la red sin embargo sí permitirá usar de forma eficiente recursos distribuidos. Según la ley de Moore y estadísticas, vemos que la potencia de los supercomputadores crece linealmente, la capacidad de almacenamiento se dobla cada 12 meses, y el ancho de banda de red se dobla cada 9 meses, y el rendimiento de un procesador se dobla cada 18 meses. La evolución que se espera entre 2001 y 2010 es en Ordenadores: x 60, y Redes: x 4000

2. Coste de equipos y dificultad de mantener

Para amortizar la inversión realizada se nos presenta la posibilidad de alquilar o compartir nuestros recursos durante la gran cantidad de tiempo que suelen estar desocupados.

3. Desequilibrio de carga en el tiempo.

Por un lado la compartición vía Grid o Mallas nos permite asegurar una estadística de uso mucho más homogénea, especialmente si compartimos con centros localizados en otras franjas horarias y con diversos patrones de carga de proyecto en tiempo. Además, si tenemos un pico de demanda para nuestros proyectos podemos multiplicar nuestra capacidad reutilizando recursos que están fuera de nuestro centro o empresa, y que de otra forma serían impensables.

Como solución a estos puntos 1 a 3 las "Grid Computing" son la respuesta. Y si revisamos las características más importantes de estas formas de computación, podemos concluir que estas

tecnologías son casos especiales de un nuevo paradigma de computación, la GRID. Se trata de un modelo de proceso distribuido que en poco tiempo va a revolucionar la computación de altas prestaciones y el uso de Internet en general.

Ventajas e Inconvenientes de GRID

Además de la principal ventaja y razón de ser de las GRIDS, la creciente necesidad de potencia y compartición de recursos, podemos hacer una clasificación y resumen de los puntos tratados hasta ahora:

Beneficios de Grid

- * Permite alquilar recursos según necesidad
- * Amortización de recursos propios
- * Gran potencia de cálculo a precio bajo sin necesidad de adquirir equipamiento.
- * Mayor colaboración y compartición de recursos entre varios centros.
- * Creación de organizaciones virtuales.
- * Modelo de Negocios basados en proveer recursos.

Desafíos

- * Recursos heterogéneos
- * Descubrimiento, selección, reserva, asignación, gestión y monitorización de recursos
- * Desarrollo de aplicaciones.
- * Desarrollo de modelos eficientes de uso.
- * Comunicación lenta y no uniforme.
- * Organizativos: dominios de administración, modelo de explotación y costes, política de seguridad...
- * Económicos: precio de los recursos, oferta/demanda...

Proyectos y Tecnologías Grid

Una vez introducida de forma natural las GRIDS, así como su evolución desde tecnologías paralelas a clusters, veamos algunas tecnologías que hacen posible el Grid Computing. Aunque hay una evolución constante, nosotros presentamos una breve enumeración y resumen de los presentados en la red Española de Middleware[4], en la red Europea EGGE[13], y en las experiencias de

Interquanta [8].

Para tener una visión global, podemos distribuir los proyectos por su origen mundial, por continentes:

* Australia

Nimrod-G, Gridbus, GridSim, Virtual Lab, DISCWorld, GrangeNet,...

* Asia

India con I-Grid, Japón: Ninf DataFarm, y Corea: N*Grid

* Europa

EGEE, CrossGrid, UNICORE, Cactus, UKeScience, DataGrid, EuroGrid, MetaMPI, XtremeWeb,...

* EEUU

Globus, Legión, OGSA, Apples, NASA IPG, Cóndor-G, Jxta, NetSolve, AccessGrid...

Mallas existentes y de pruebas anteriores, con sistemas Grid para prototipos y producción:

- Grid Physics Network:

www.griphyn.org

- NASA's Information Power Grid:

www.nas.nasa.gov/About/IPG/ipg.html

- European Data Grid (ahora EGEE)

www.eu-datagrid.org

- Teragrid

www.teragrid.org

- PPARC PartPhysics & Astronomy:

www.pparc.ac.uk/Nw/GridPP2.asp

- Tidewater Research Grid:

www.tidewatergrp.org

Infraestructura Grid: Servicios y protocolos básicos para interconectar recursos:

- Globus:

www.globus.org

- MPI/G:

www.globus.org/mpi

- Legion:

www.cs.virginia.edu/~legion/

- Polder:

www.science.uva.nl/research/scs/PSCS4.html

- MOL:

www.uni-paderborn.de/pc2/projects/mol/

* Toolkits de aplicación: Módulos para construir aplicaciones Grid específicas

- DRMAA:

www.drmaa.org Distributed Resource Management Application API

- Nimrod/G:

www.csse.monash.edu.au/~raj कुमार/eco





Proyectos parecidos al de SETI están funcionando en la comunidad científica y académica, y son usados para estudiar enfermedades o para realizar grandes cálculos que antes eran impensables

grid/

- Condor:

www.cs.wisc.edu/condor/condorg

- Portal:

dast.nlanr.net/Projects/GridPortal

Proyectos, alianzas y frums:

- Globus

- GGF (Global Grid Forum)

- EGA (Enterprise Grid Alliance, 2004)

* IEEE TFCC,

- Grid Conference, and P2P Conference

Aplicaciones Grid

- CACTUS

(www.cactuscode.org)

- Virtual Laboratory

(www.csse.monash.edu.au/~rajkumar/vlab/)

b/)

Observatorios Virtuales:

(astro.u-strasbg.fr/~ccma/vo/virtual-observatory.html)

observatory.html)

- Virtual Observatories

(www.interquanta.com/vos, siguiente apartado)

apartado)

Componentes de Globus que hacen posible una GRID

Es de destacar el resultado de un grupo de universidades e instituciones que han unido fuerzas en el Proyecto Globus, homogeneizando protocolos y aplicaciones de Grid. En este proyecto encontramos potentes instituciones de EEUU, como NASA, y empresas como IBM, Microsoft, y Cisco Systems. El estado actual de Globus Toolkit, que permite compartir recursos localizados en diferentes dominios de administración, con diferentes políticas de seguridad y gestión de recursos. Ya tenemos disponible la versión GT4 que, como los propios socios la definen, es... "Un software middleware, Un conjunto de librerías, servicios y APIs", y para mayor claridad nos recuerdan que Globus no es... "una herramienta de usuario o planificador, ni una aplicación".

En una Grid, la información sobre recursos es crítica, y reunir esta información es una actividad esencial. El servicio de información de Globus incluye el MDS, para monitorizar y encontrar recursos de Grid, el Servicio (GRIS) de Información de Recursos, y

el Servicio (GIIS) de Información e índice. Tenemos en Globus :

* La seguridad (GSI); el Servicio de la Seguridad (autenticación) basada en PKI.

* La entrega de trabajos y la administración (GRAM); y distribución uniforme.

* Los servicios de información (MDS); el Servicio de Información basada en LDAP.

* La gestión remota del archivos (GASS); el Servicio de Acceso a Almacenamiento.

* (a) acceso de archivos GASS API: Reemplaza el call con globus_gass_call; las llamadas de lectura/escritura entonces pueden avanzar directamente.

* (b) las extensiones de RSL: los URLs utilizaron para denominar ejecutables, stdout, stderr.

* (c) utilidad Remota de administración de reserva.

* (d) APIs de bajo nivel para conductas especializadas.

* Los Catálogos Remotos e Instrumentos de Gestión; por Globus 2,0 (2002) a GT4.0 (2005).

En la topología de malla, podemos distinguir unos nodos servidores de recursos como los CE (Elemento de computación), con sus WN (Nodo de trabajo), y SE (Elemento de Almacenamiento), por otro lado para el uso de Grid únicamente necesitamos la aplicación de usuario UI..

Se enumeran herramientas, aplicaciones y protocolos, siendo las fuentes de web elementos de referencia, sirva así como un resumen o visión global de tecnologías subyacentes al paradigma de Grid. A continuación resumimos los protocolos, **software** y herramientas que se usan para soportar las GRID, y constituyen parte integrante de ellas. Enumeremos estas tecnologías clasificadas en funcionalidades:

Seguridad: Globus Security Infrastructure (GSI) se apoya en TLS/SSL (Transport Layer Security, Secure Socket Layer) y la implementación OpenSSL, usa PKI (Public Key Infrastructure) con certificados X.509, junto con GSS-



API(Generic Security Services) y CAS (Community Authorization Service).

Comunicaciones: Inicialmente se trabajaba con Nexus, que ahora está en desuso, y actualmente se usa Globus I/O sobre sistema de ficheros, sistema seguro GSI, y protocolos TCP (Transport Control Protocol) y UDP (User Datagram Protocol).

Detección de fallos: Mediante sistema Heart Beat Monitor (HBM), sobre Nexus o Globus.

Herramientas: alguna de alto nivel como: Órdenes Globus, MPICH-G2, Condor-G y Nimrod/G.

Gestión de recursos: Un sistema que reutiliza gestores locales (PBS, LSF, NQE,) es GRAM (Globus Resource Allocation Manager). Los recursos se definen con un RSL (Resource Specification Language) y se usan con LoadLeveler o UNIX fork. DUROC (Dynamic User Runtime On-line Co-allocator) usa GRAM y Globus para gestionar la distribución de recursos.

Descubrimiento y monitorización de recursos: Se usan los sistemas de Índice GIIS (Globus Index Information Service, usa GRIS), Recursos GRIS (Globus Resource Information Service, usa GRAM), apoyados en GSI y LDAP (Lightweight Directory Access Protocol), y por último Monitoreo MDS (Monitoring and Discovery Service) que usa los dos anteriores GIIS y GRIS.

Gestión de datos: Sobre el sistema GSI tenemos la evolución de FTP que da lugar al GridFTP (Grid File Transfer Protocol), sobre este y Globus funciona GASS (Global Access to Secondary Storage). Para la gestión de catálogos de datos Globus Replica Catalog funciona sobre LDAP y encima de ella tenemos Globus Replica Management (usa Globus Replica Catalog, GridFTP y GSI)

Ejemplos y Experiencias con GRIDS

Proyectos parecidos al de SETI están funcionando en la comunidad científica y académica, y son usados para estudiar enfermedades o para realizar grandes cálculos que antes eran impensables. Como vimos la Grid apa-

rece como el siguiente paso en el desarrollo de Internet, al igual que ya la web supuso el impulso con el intercambio de contenidos e información, la Grid abrirá el camino para el procesamiento de esa información.

Aparece así un nuevo motivo para la expansión de la red Internet, a la vez que una aplicación fundamental de la globalización que Internet ya posee en todo el mundo. Todos podrán acceder a los ciclos de procesamiento que necesiten, con la capacidad de conectar casi cualquier cosa: sensores, mp3, telescopios,... con su información disponible desde cualquier sitio y en cualquier momento.

Inicialmente la gran diversidad tecnológica y de programas existentes en el mundo hicieron pensar que la idea de compartir el procesamiento de información era una utopía. La progresiva expansión de sistemas de uso personal, de los usuarios de Windows o Linux ha allanado el camino. Inmediatamente vino la suma de potencia de grandes servidores o por lo menos de las estaciones de trabajo profesionales.

Los mayores ordenadores virtuales en Portugal y España, han sido puestos en marcha como parte de los proyectos predecesores a la red pan-europea EGEE. Actualmente estos dos países conforman la Federación Suroeste de Europa de EGEE, y en ella se integran centros de computación de Madrid, Galicia, Barcelona y Valencia. Estos centros, situándose a la cabeza de España en capacidad de cálculo disponible, permiten a sus investigadores avanzar en sus proyectos a una velocidad desconocida hasta el momento.

El trabajo de GriPhyN (La Red Grid de Física)

En el año 2000 los investigadores principales I. Foster y P. Avery, de las universidades de Chicago y Florida, llevan adelante el proyecto GriPhyN. Su objetivo es proporcionar la gran cantidad de proceso necesaria para los grandes proyectos de Investigación Física, especialmente pensando en una serie



La Malla o Grid conecta grupos de ordenadores, unidades de almacenamiento y redes, permitiendo a los centros de investigación y empresas, asignar dinámicamente los recursos de acuerdo a las necesidades del negocio





de experimentos internacionales para el estudio de partículas de altas energías. El proyecto está terminando en este año 2005, con el nombre y fase actual, dejando tras de sí una enorme experiencia en la tecnología GRID.

El problema semilla que da razón de ser al inicio de *GriPhyN*, así como al EDG/EGEE que trataremos más tarde, es la enorme cantidad de información que va a generar un nuevo acelerador de Partículas a instalar en el CERN. Hablamos del LCH (*Large Hadron Collider*) que irá acompañado de una serie de instrumentos y experimentos (como CMS, ATLAS, ...) que entregarán del orden del petabyte (10^6 Gigabytes) de datos anuales. Para hacer frente a la cantidad de proceso y almacenamiento necesaria, el proyecto *GriPhyN* reúne un equipo sobresaliente de investigadores de informática y físicos experimentales. La gestión de datos y procesadores geográficamente dispersos sentará las bases de una forma de computación intensiva orientada a la ciencia en el siglo XXI. Es la creación de PVDG (*Petascale Virtuales Data Grids*) que respondan a las necesidades cálculo intensivo de una comunidad diversa de millares de científicos que se distribuyen alrededor del globo.

Además de los experimentos de Partículas LHC el sistema trata con otros dos campos y proyectos punteros en la ciencia actual: son el LIGO (*Laser Interferometer Gravitational-wave Observatory*) que pretende observar Ondas Gravimétricas producidas por pulsars y supernovas, y el proyecto SDSS (*Sloan Digital Sky Survey*) que llevará a cabo una inspección automatizada del cielo que permita un estudio sistemático de estrellas, galaxias y de las estructuras a gran escala.

Otros esfuerzos de colaboración, apoyados por la NSF (National Science Foundation) y otras agencias, generan una extensa actividad de base en la comunidad de investigación científica, han creado por ejemplo la Red de NASA PowerGrid, la PPDG (*Particle Physics Data Grid*), o el Observatorio

Virtual Nacional (*NVO National Virtual Observatory*).

Ejemplo de GRID Europea: EGEE 2004 y EDG

La actividad se empezó con el proyecto Europeo de DataGrid o EDG, formada por redes más pequeñas dedicadas a la física de Partículas y de nuevo enfocada al *Large Hadron Collider* (LHC), el potentísimo acelerador de partículas, del CERN que producirá una enorme cantidad de datos. Ahora continúa con el proyecto EGEE (*Enabling Grids for E-Science in Europe*, www.eu-egee.org) que se puso en marcha en abril de 2004, y permitirá compartir datos y procesadores equivalentes a 20.000 ordenadores personales, y entre investigadores de 27 países.

Dentro de la red se encuentran actividades de diseminación que pretenden difundir esta nueva tecnología en Europa y en las que empresas e instituciones pueden participar. En estas experiencias se trabaja con una instalación encima de *Globus* (GT3 Tool Kit) y del software EDG propio, se usa software específico LCG2.0 (*LHC Computing Grid*) desarrollado por el equipo europeo encabezado por CERN. El proyecto de LCG juega un papel central al proporcionar una infraestructura operacional de la etapa más temprana de la red. Los objetivos claves de la GRID europea, en Operación, Soporte y Gestión:

* **Los servicios de infraestructura core o núcleo de red:** un conjunto de servicios esenciales, como los servicios de información, gestores de recursos, de datos y la administración de las organizaciones virtuales que unen los recursos distribuidos de forma coherente

* **Control y monitorización de Grid:** del estado operacional y su capacidad de trabajo, incluyendo las acciones correctivas de posibles problemas, tanto en infraestructura como con los recursos.

* **El despliegue de Middleware:** validación de las entregas de *middleware*, últimas versiones, y desplegados a

centros de recurso a través de la *Grid*. Se proporciona ayuda a centros nuevos de recurso para unirse a la red, con instalación de *middleware* e introducción de procedimientos operacionales en centros de recurso. Un esfuerzo extra será ofrecido a centros de recurso que ofrecen los recursos paralelos y las supercomputadoras vectoriales, que juegan un papel estratégico para varias aplicaciones científicas.

* **Soporte de recursos y usuarios:** recibir, responder y coordinar la resolución de problemas en la operación de *Grid*, en ambos centros del recurso y usuarios; se filtran y agregan problemas, proporcionando soluciones tanto en infraestructura de centro o *middleware* o re-dirección a los expertos apropiados.

* **La gestión de Grid:** coordinación de la realización de objetivos por los centros regionales ROC (*Regional Operations Centres*) y CIC (*Core Infrastructure Centres*), relación con proveedores de recursos, negociación de acuerdos nivel de servicio y participación en organismos de estándares.

* **La colaboración Internacional:** con la colaboración con organizaciones paralelas en América y en Asia-Pacífico; para asegurar la interoperabilidad de infraestructuras y servicios en EGEE. Tanto los ROCs como CICs serán centros OMC (*Operations Management Centre*) de Administración de Operaciones.

La distribución de LCG y software del proyecto está disponible en <http://grid-deployment.web.cern.ch/grid-deployment>. El producto resultante deberá ser fácilmente configurable, con una alta escalabilidad y arquitectura modular. El sistema deberá estar basado en estándares y se apoyará preferiblemente en software abierto.

El proyecto EGEE avanzará hacia la sistematización de los procedimientos de operación. Con este objetivo, implementará un sistema de monitorización y control automático de incidencias. El proyecto dará soluciones al producto en distintas fases, amplian-



do en cada fase las prestaciones de la proposición anterior. Se desarrollarán los interfaces necesarios para que un operador, sin conocimiento expreso del sistema, pueda monitorizar su estado y reconfigurar los parámetros que puedan ser necesarios, en la medida de lo posible.

Una de las redes menores integradas es la que va a preparar el *Particle Physics and Astronomy Research Council* británico. Se llamará *GridPP2* y su potencia combinada será equivalente a la del segundo superordenador más grande del planeta, tras el *Japan Earth Simulator* para simulaciones geofísicas. La *GridPP2* estará conectada a la *European Grid* y colaborará en el procesamiento de datos del LHC. La *GridPP2* ha estado funcionando en forma de prototipo dentro de una red de pocos ordenadores, situados en 10 localidades británicas.

Aplicaciones a E-CIENCIA, Laboratorio virtual, OBSERVATORIO VIRTUAL

La Investigación Virtual y Laboratorios Virtuales son conceptos que aparecen de desarrollar todas las posibilidades del paradigma de computación *Grid*. Podemos verlo cómo un sistema que permite compartir los datos de laboratorios distribuidos de forma transparente a través de redes, permitiendo al científico encontrar, acceder, analizar y procesar o combinar los datos de varios laboratorios y bases heterogéneas. Este concepto puede ser aplicado a una variedad de campos, como telecontrol, mediciones sísmicas, y otras áreas de investigación. Un caso de particular interés es en el trabajo con enormes bases de datos de Astronomía.

El Observatorio Virtual surge de la aplicación del concepto de Laboratorio Virtual a Astronomía, permitiendo a astrónomos encontrar, acceder, analizar y combinar los datos astronómicos. Tres principales recursos: Datos, Catálogos y Procesadores en línea accesibles vía Internet. Además del observatorio virtual se presenta tam-

bién como una plataforma ideal para enseñar astronomía. El observatorio astronómico virtual es un concepto que engloba todo un entorno informático que hace accesibles datos astronómicos en una red distribuida de archivos como Internet. Además se pretende que sea un sistema que permita describir las características de los datos astronómicos en profundidad.

Podemos ver su utilidad para la comunidad científica como el futuro de los archivos astronómicos, que con el elevado ritmo de adquisición de datos necesita de la últimas tecnologías y modelos *Grid*. Especialmente, dado que las líneas de investigación más avanzadas requieren el análisis de datos a lo largo de todo el espectro electromagnético, lo que ha multiplicado el número de instrumentos necesarios y de archivos astronómicos.

Siguiendo esta evolución, la comunidad internacional construye Observatorios virtuales, un tipo de organización de los datos, aplicaciones e instrumentos mundiales en una forma coherente que mejorará dramáticamente la habilidad de hacer los estudios de la astronomía moderna. Las actividades internacionales como las AVO/NVO, descritas a continuación, y las plataformas de *Grid*, como *AstroGrid*, se han iniciado para permitir al mundo astronómico la explotación científica de archivo lejanos. Un ejemplo de archivo astronómico que tenemos en España es el sistema INES. Pero veamos brevemente los trabajos realizados en este área:

* NVO - Observatorio Virtual Nacional - www.us-vo.org

El NVO desarrolla herramientas que hacen fácil localizar y recuperar catálogos mundiales de observaciones astronómicas con múltiples instrumentos, analizar sus datos y comparar modelos y simulaciones teóricas con observaciones. <http://nvo.gsfc.nasa.gov/nvo-index.html>

* AVO - Observatorio Virtual Astrofísico - www.euro-vo.org

El Proyecto Observatorio Virtual es un estudio para el diseño e implemen-

tación de un observatorio virtual para la astronomía europea. Un observatorio virtual es una colección de archivo de datos, de software y de instrumentos que usan Internet para formar un entorno científico de investigación astronómica. Al igual que un observatorio verdadero consiste en catalejos, cada colección de instrumentos astronómicos extraordinarios, el AVO consiste en una colección de datos central, cada uno con colecciones extraordinarias de datos astronómicos, sistemas de software y procesamiento de las capacidades.

*Esfuerzo Conjunto: www.ivoa.net

El NVO en EEUU, AVO en Europa y otros esfuerzos internacionales se unen para crear estándares internacionales desarrollados en la colaboración IVOA (*International Virtual Observatory Alliance*).

* UK-AstroGrid

El proyecto *AstroGrid* en el Reino Unido se centra en resultados a corto plazo, aplicaciones y herramientas para los centros de datos que los manejan: (a) SuperCOSMOS, Sloan, INT-WFC, UKIRT WFCAM, el XMM-NEWTON, Chandra, el ESMEREJON y datasets relacionado de VLA; (b) SOHO y Yohkoh; y (c) el Grupo y EISCAT. Las diferencias entre el tipo de datos y medios de diversas federaciones presenta un desafío claro, cuyas soluciones pasan por desarrollar un "Observatorio Virtual" global.

Varios prototipos y redes dignas de mención relacionadas con Observatorios Virtuales, brevemente:

- Australia VO
<http://www.aus-vo.org/twiki/bin/view/TWiki/TWikiRegistration>
- IVO
http://www.atnf.csiro.au/_Mail_Archives/ivo
- NASA SkyView
<http://skyview.gsfc.nasa.gov/>
- Galex (Evolución)
<http://galex.stsci.edu/>,
- INES (Extracción Espectral)
<http://iuearc.vilspa.esa.es>
- NGASt





<http://archive.eso.org/NGAST>.
- MAST (Archivo xml STScI)
<http://archive.stsci.edu/>

España: Red Temática para la Coordinación de Actividades Middleware en Grid

En Marzo de 2004 se crea en España la Red Temática para la Coordinación de Actividades Middleware en Grids, financiada mediante una acción especial al Ministerio de Ciencia y Tecnología. La persona de contacto de esta red es Ignacio Martín del Centro de Astrobiología de CSIC-INTA y su página web es asds.dacya.ucm.es/GridMiddleware.

Entre los objetivos principales de la red está coordinar a todos los grupos de investigación middleware y empresas interesadas para que puedan participar en las decisiones que se tomarán en relación con el banco de pruebas IRISGrid. Y a su vez constituir un interfaz a las actividades dentro del proyecto EGEE que permita tener en cuenta las necesidades e intereses de los grupos de investigación middleware y empresas en España.

En la Primera Reunión de la Acción Especial en Noviembre en Toledo se presentaron trabajos de varios centros de Investigación y universidades, y ya podemos encontrar que junto al gran número de instituciones y universitarios encontramos tres empresas españolas en la web de la red. Y dedicadas a temas complementarios: infraestructura red, middleware y aplicaciones. El papel de GridSystems es pionero en España y probablemente en Europa.


Enterprise Grid Alliance 2004

La Enterprise Grid Alliance (EGA) se crea en California, en abril de 2004, por un grupo de empresas líderes en tecnología para desarrollar soluciones empresariales de informática distribuida y para acelerar el despliegue de esta tecnología en las empresas. Es un consorcio abierto enfocado en el desarrollo y promoción de soluciones de mallas empresariales. La participación en

esta alianza está abierta a todas las organizaciones a través de múltiples niveles y para dar el siguiente paso se encuentran en su portal www.gridalliance.org.

La EGA es una comunidad abierta, independiente y neutral dirigida al despliegue de las aplicaciones comerciales en un entorno distribuido a corto plazo. Las áreas de enfoque iniciales incluyen modelos de referencia, aprovisionamiento, seguridad y finanzas. La Alianza se centrará en resolver los obstáculos que las organizaciones se encuentran cuando utilizan mallas empresariales, con soluciones abiertas e interoperables. Al concentrarse exclusivamente en las necesidades de los usuarios empresariales, la EGA permitirá a las empresas beneficiarse de las múltiples ventajas del Grid Computing, tales como una reacción más rápida a los cambios en las necesidades comerciales, una mejor utilización y rendimiento a nivel de servicio y una reducción en los costes operativos de TI. La Junta directiva inicial de EGA está formada por EMC, Fujitsu Siemens Computers, HP, Intel, NEC, Network Appliance, Oracle y Sun Microsystems.

Proyecto Crossgrid

Actualmente en su fase final, es un estudio de requerimientos e implantación de aplicaciones con necesidades de tiempo de respuesta corto o tiempo real en entornos de computación distribuida según el modelo de Mallas o Grid, con especial énfasis en la visualización y minería o extracción de datos. En ella participan 27 instituciones de 11 países. Sus aplicaciones son: a) Simulación de circulación sanguínea, b) Simulación de inundaciones y avalanchas de agua, c) Altas energías, d) Predicción atmosférica y dispersión de contaminantes. 

Bibliografía y Referencias

- [1] Grid FAQ. M. Baker, R. Buyya, and D. Laforenza, **Grids and Grid Technologies for Wide Area Distributed Computing International Journal of Software: Practice and Experience** (SPE), Volume 32, Issue 15, Pages: 1437-1466, Wiley Press, USA, December 2002.
- [2] K. Krauter, R. Buyya, and M. Maheswaran, **A Taxonomy and Survey of Grid Resource Management Systems for Distributed Computing International Journal of Software: Practice and Experience** (SPE), ISSN: 0038-0644, Vol 32, Issue 2, Pag 135-164, Wiley Press, February 2002.
- [3] CESGA (Centro de Supercomputación de Galicia) www.cesga.es
- [4] **Red Middleware Presentaciones** Reuniones Octubre 2004 - Ignacio Martín Llorente -
- [5] Empresa de Grid y Middleware GridSystems.com
- [6] Tendencias-
www.gridalliance.org
www.pr21.com
- [7] Nuevas Tendencias Científicas <http://www.tendencias21.net/>
- [8] Interquanta, Resumen de Grids, y OV's
www.interquanta.com/grid
- [9] **Moore's Law vs. storage improvements vs. optical improvements.** Scientific American (Jan-2001)
- [10] **GriPhyN**, Comunicaciones de Paul Avery, University of Florida, - www.GriPhyN.org
- [11] Actualidad Grid www.gridcomputing.com
- [12] NSF Future Cyberinfrastructure Report
<http://www.nsf.gov/od/lpa/news/03/pr0318.htm>
- [13] EGEE project site www.egee.org
- [14] Ranking de Procesadores www.top500.org

