

REALIDAD DE LA COMPUTACION GRID EN AMERICA LATINA REALITY OF COMPUTATION GRID IN LATIN AMERICA

Weimar Díaz Garzón

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS, Bogotá.

Estudiante Maestría en Ciencias de la Información y las Comunicaciones

[wdiazg@uniamazonia.edu.co](mailto:w Diazg@uniamazonia.edu.co), [wdiazg@gmail.com](mailto:w Diazg@gmail.com)

Resumen: Este artículo muestra la situación actual de la tecnología Grid en América latina, la cual está revolucionando el trabajo científico en Europa y Estados Unidos, puesto que apoya la colaboración entre investigadores, permite resolver problemas con gran demanda computacional y de datos.

Abstract: This Article presents the situation of the Grid technology in Latin America, which is revolutionizing the scientific work in Europe and the United States, since it supports the collaboration between investigators, allows to solve problems with great computational demand and of data.

Palabras Claves: GRID COMPUTING, CLARA, CONDOR, EELA

1 INTRODUCCIÓN

Hace algunos años para realizar grandes cálculos o procesar gran cantidad de información de cualquiera de los proyectos de investigación del mundo, era necesario tener a la mano un Supercomputador que fuera capaz de soportar trabajos complejos, el cual por sus características de hardware tienen un costo demasiado elevado. Hoy en día con el avance de la tecnología a pasos gigantescos, los computadores personales están llegando a tener mayor poder de procesamiento lo cual ratifica lo dicho por Moore, quien dice que la velocidad de la CPU se dobla cada 18 meses [Gordon E. Moore (1965)]. Si a lo anterior le sumamos la aparición y posterior masificación del Internet al cual se unen cada vez más personas en el mundo y que también su velocidad de interconexión aumenta cada año; podríamos imaginarnos un sistema el cual funcionara de forma transparente para el usuario, donde se pueda integrar recursos computacionales heterogéneos y así crear una infraestructura de hardware y software que pueda aumentar nuestras capacidades de cálculo y almacenamiento. Pues bien ya no es necesario imaginarlo por que en la actualidad existe una tecnología denominada GRID COMPUTING, que proporciona acceso dependiente, consistente, generalizado y económico a capacidades computacionales de altas

prestaciones [Foster, I. and Kesselman, C. (1999)], con 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.

2 PERO... ¿QUÉ ES EL GRID?

El termino "GRID" se a movido del obscuro ámbito académico al mas popular, leemos sobre Compute Grids, Data Grids, Science Grids, Access Grids, Knowledge Grids, Bio Grids, Sensor Grids, Cluster Grids, Campus Grids, etc. Pero en realidad ¿que es "GRID"? A mitad de los años 90 fue acuñado la palabra "GRID" para denotar una propuesta de infraestructura en computación distribuida para la ciencia y la ingeniera avanzada [Foster, I., C. Kesselmann, S. Tuecke (2001)]; una definición bastante aceptada en el mundo académico es la proporcionada por Ian Foster donde sugiere una lista de tres ítems que debe cumplir un sistema para ser llamado GRID [Foster, I. (2002)], que son:

- 1) *Coordinates resources that are not subject to centralized control.*
- 2) *Using standard, open, general-purpose protocols and interfaces.*
- 3) *To deliver nontrivial qualities of service.*

Que en español seria: sistema que coordina recursos, que no están sujetos a un control centralizado, usando interfaces y protocolos estándares, abiertos y de propósito general para proveer de servicios relevantes.

Hoy en día se puede definir como un conjunto de recursos hardware y software distribuidos por Internet que proporcionan servicios accesibles por medio de un conjunto de protocolos e interfaces abiertos (gestión de recursos, gestión remota de procesos, librerías de comunicación, seguridad, soporte a monitorización,..etc) [Fuentes, A., Vázquez, J. L., Huedo, E., Montero, R. S., Llorente I. M..(2005),].

3 ARQUITECTURA GRID

La arquitectura GRID esta compuesta por capas como lo muestra la figura 1, Los componentes dentro de cada parte de la capa poseen características comunes pero se pueden construir sobre capacidades y comportamientos proporcionados por cualquier capa mas baja. La especificación de varias capas de la arquitectura GRID siguen los principios del "hourglass model" [Realizing the Information Future (1994)]

Fabric: Interfaz a Control Local

Esta capa proporciona los recursos a los que el acceso compartido es conducido por los protocolos Grid (acceso local a recursos lógicos como CPU, software, ficheros). Un recurso puede ser una entidad lógica, un sistema de archivos distribuidos, computador cluster [Foster, I., C. Kesselmann, S. Tuecke (2001)].

Las principales capacidades que ha de proporcionar son las siguientes:

- Computational resources - Recursos computacionales.
- Storage resources - Recursos de almacenamiento.
- Network resources - Recursos de red.
- Code repositories - Repositorios de código.
- Catalogs - Catálogos.

Conectivity: Comunicación Fácil y Segura

Esta capa define los protocolos de comunicación y autenticación requeridos para las transacciones de red Grid. Los protocolos de comunicación permiten el intercambio de datos entre los recursos de la capa Fabric. Los protocolos de autenticación se construyen sobre servicios de comunicación para proveer mecanismos seguros criptográficos para verificar la identidad de usuarios y recursos. Son cubiertos por la pila de protocolos TCP/IP: Internet (IP e ICMP), transporte (TCP y UDP) y aplicación (DNS, OSPF) [Foster, I., C. Kesselmann, S. Tuecke (2001)].

Para la autenticación se tienen las siguientes características:

- Single sign on.
- Delegation - Delegación.
- Integration with various local security solutions - Integración con soluciones de seguridad locales.
- User-based trust relationships - Seguridad basada en usuario de confianza.

Resource: Compartición de Recursos individuales

Está constituida por protocolos que permitirán la negociación segura, iniciación, monitorización, control, pago y tarificación de las operaciones compartidas sobre recursos individuales, mediante la llamada a funciones de la capa Fabric para el acceso y control a los recursos individuales

[Foster, I., C. Kesselmann, S. Tuecke (2001)]. Se puede notar por tanto que están plenamente enfocados a recursos individuales.

Estos protocolos se pueden diferenciar en dos clases:

- Information protocols - Protocolos de información
- Management protocols - Protocolos de gestión

Collective: Coordinación de Recursos Múltiples

Está constituida por los protocolos y servicios no asociados a un recurso específico sino que son globales y capturan interacciones de diferentes recursos [Foster, I., C. Kesselmann, S. Tuecke (2001)]. Las principales funciones que lo caracterizan son:

- Directory services - Servicios de directorio
- Co-allocation, scheduling, and brokering services
- Monitoring and diagnostics services - Servicios de monitorización y diagnóstico
- Data replication services - Servicios de replicación de datos
- Grid-enabled programming systems
- Workload management systems and collaboration frameworks
- Software discovery services - Servicios de descubrimiento de software
- Community authorization servers
- Community accounting and payment services
- Collaboratory services - Servicios de colaboración

Aplicación

Mediante la utilización de APIs y SDKs (Software Development Kit) las aplicaciones intercambian mensajes de protocolo con el servicio/capa adecuado a fin de ejecutar las acciones adecuadas.

4 ¿POR QUE USAR GRID COMPUTING ?

Los acercamientos computacionales para resolver problemas han probado su valor en casi cada campo del esfuerzo humano, los computadores son usados para modelar y simular complejos problemas de la ciencia y la ingeniería, diagnosticar condiciones médicas, controlar equipo industrial, pronosticar el tiempo y muchos otros propósitos. Aunque todavía algunos problemas desafiantes pueden exceder nuestra habilidad para resolverlos, estamos usando los computadores mucho menos de lo que realmente debería ser. Un claro ejemplo sería, los investigadores de las universidades hacen un extenso uso de los computadores para estudiar el impacto de los cambios del suelo en la biodiversidad, pero los planificadores de la ciudad que seleccionan las rutas para las nuevas carreteras o planean las nuevas zonas de ordenamiento no lo hacen. Estas decisiones locales en últimas son las que forman nuestro futuro.

Hay varias razones para la falta de uso de métodos computacionales para la resolución de problemas, que incluyen la escasez de educación y de herramientas apropiadas. Pero un factor importante es el bajo promedio de ambientes computacionales para propósitos de cómputo sofisticados [Foster, I. and Kesselman, C. (1999)], por lo anterior se hace necesario un sistema donde la mayoría de instituciones tengan acceso y en el cual se pueda transmitir y compartir conocimiento de forma fácil y segura.

El cambio está siendo progresivo, lo que está conduciendo a la mayoría de los grupos de investigación a buscar soluciones a medida ya que los métodos en uso se iban quedando cortos. La primera fase de contención ha consistido en la instauración de sistemas de clusters y adaptación de algoritmos y procedimientos para su ejecución paralela a pequeña escala. Estas soluciones, si bien están permitiendo desarrollar y probar nuevos algoritmos y soluciones, resultan insuficientes para el tratamiento de datos experimentales masivos [Área temática de bioinformática (2003)].

Por ello se puede expresar las necesidades básicas como una mayor demanda de capacidad de almacenamiento y tratamiento de información, y un crecimiento desmesurado de la capacidad de cálculo. Los costes asociados al tratamiento de información derivada de las nuevas técnicas experimentales superan con creces la capacidad de prácticamente cualquier grupo o entidad aislada, aún recurriendo a soluciones paralelas de bajo coste como las granjas de computadores personales. En estas condiciones, la única solución viable consiste en compartir recursos entre grupos de forma solidaria hasta reunir recursos suficientes para abordar los problemas experimentales, en otras palabras utilizar GRID COMPUTING.

5 CAMPOS DE APLICACION DE LA COMPUTACION GRID

El Grid es una tecnología que ha nacido y todavía se sigue desarrollando mayoritariamente en el mundo científico, por eso se dice que es el Instrumento fundamental de la e-Ciencia, la cual se puede definir como el "conjunto de actividades científicas desarrolladas mediante el uso de recursos distribuidos accesibles a través de Internet" [Fundación Española para la Ciencia y la Tecnología (2004)]. Algunos de los campos representativos de aplicaciones científicas dónde el uso de la tecnología Grid está siendo utilizada son [Doallo, R., Touriño, Heymann, J., E. (2006)]:

- Simulación molecular
- Física de partículas
- Modelado del clima
- Observación de la Tierra
- Estudio del genoma humano
- Física de altas energías (HEP), biomedicina y química computacional son áreas científicas dónde sobresale el grado de explotación de la tecnología Grid.

6 PROYECTOS DE GRID COMPUTING

Los proyectos GRID tienen diferentes finalidades por ejemplo en algunos se concentran en el desarrollo de herramientas software (middleware), otros en el desarrollo de servicios, y otros en la puesta a punto para aplicaciones científicas específicas o en la optimización de la red de interconexión.

En Europa, Estados Unidos y Japón se vienen desarrollando gran cantidad de proyectos GRID con múltiples propósitos, algunos de los más importantes son:

- **CONDOR (Condor High Troughput Computing):** es un potente gestor de tareas distribuídas operativo desde 1988, desarrollado bajo el modelo Open Source desde la Universidad de Wisconsin EEUU, este proyecto esta dentro de la categoría de Middleware. Actualmente se encuentra en la versión 6.8.4 y es uno de los proyectos Open Source más saludables, liderado por el prestigioso Miron Livny y por Todd Tanenbaum [Condor Project], es muy robusto, ofrece alta disponibilidad, y gran escalabilidad. Soporta múltiples estándares (Globus, OSGA) y es seguro (soporta cifrado, autenticado, y

certificado). Condor se utiliza principalmente en centros y redes de investigación, pero también en la industria y en finanzas, como los bancos UBS y JP Morgan y la aseguradora Hartford Life.

- **CROSSGRID**: el proyecto CrossGrid apunta a desarrollar servicios Grid y herramientas de programación para simulación y visualización en tiempo real en los campos de la medicina, la física y la geología. Algunos de los países involucrados son: Austria, Alemania, Grecia, Italia, Portugal, España. Las categorías en las cuales se enfoca el proyecto son: Middleware, Infrastructure, Programming Environment, Applications [CrossGrid Exploitation (2004)].
- El proyecto **Biogrid** se inicio con la construcción de una red de supercomputación, como parte del programa de IT del ministerio de Educación, cultura, deporte, ciencia y tecnología. Bajo el proyecto, la universidad de Osaka y otras instituciones de importancia están en el proceso de desarrollar tecnología grid computing para abarcar áreas de biología y ciencia medica, el proyecto abarca dos categorías Middleware y networking [BioGrid Project (2005)].
- **DAMIEN** Distributed Applications and Middleware for Industrial use of European Networks: Fue un proyecto de la Information Society (IST-2000-25406), en que trabajaron los países de Alemania, Francia y España, duro aproximadamente 3 años, el objetivo principal fue definir e implementar una arquitectura Middleware para supervisión de sistemas genéricos, además ampliar los tradicionales estándares HPC y herramientas a las propiedades de la computación GRID [Damien Project]. Adicionalmente estas herramientas fueron probadas en ambientes industriales usando aplicaciones de producción diaria.
- **Datagrid** -Grid @ CERN : Este proyecto fue financiado por la Unión Europea del 2001 al 2003. El objetivo era construir la nueva generación en infraestructura de computación para proveer computación intensiva y análisis de bases de datos compartidas a gran escala, de centenares de TeraBytes o PentaBytes, a través de la extensa distribución de comunidades científicas [The DataGrid Project].

¿Y LATINOAMÉRICA?

América latina no podía ser la excepción ante este auge de la tecnología GRID. Pero si en los países desarrollados la GRID Computing esta en formación, en nuestros países (Latinoamérica) apenas esta en gestación, por ello se vienen adelantando varios proyectos y alianzas con instituciones que tienen experiencia en la implementación de tecnologías GRID. En la

actualidad se cuenta con **CLARA Cooperación Latino Americana de Redes Avanzadas** que es una organización no gubernamental sin fines de lucro, la cual estimula la cooperación regional en actividades educativas, científicas y culturales, además promueve la integración directa con las comunidades científicas de Europa [CLARA (2006)], esta organización trata de integrar una red regional de telecomunicaciones de la más avanzada tecnología para interconectar a las redes académicas nacionales de la región, hasta el momento tiene los siguientes miembros:

- Argentina - Red Teleinformática Académica-RETINA
- Bolivia - Agencia para el Desarrollo de la Sociedad de la Información en Bolivia-ADSIB
- Brasil - Red Nacional de Enseñanza e Investigación - RNP
- Colombia - Red Nacional Académica de Tecnología Avanzada - RENATA
- Costa Rica - Red Nacional de Investigación - CR2Net
- Cuba - RedUniv
- Chile - Red Universitaria Nacional - REUNA - GREUNA
- Ecuador - Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado - CEDIA
- El Salvador - Red Avanzada de Investigación, Ciencia y Educación Salvadoreña - RAICES
- Guatemala - Red Avanzada Guatemalteca para la Investigación y Educación - RAGIE
- Honduras - Universidad Tecnológica Centroamericana - UNITEC
- México - Corporación Universitaria para el Desarrollo de Internet - CUDI
- Nicaragua - Red Nicaragüense de Internet Avanzada - RENIA
- Panamá - Red Científica y Tecnológica - RedCyT
- Paraguay - Arandu
- Perú - Red Académica Peruana - RAAP
- Uruguay - Red Académica Uruguaya - RAU
- Venezuela - Red Académica de Centros de Investigación y Universidades Nacionales - REACCIUN

Otro proyecto importante para el trabajo en GRID Computing, que se adelanta en Latino América actualmente es **EELA** (E-infraestructure Shared between Europe and Latin America), en el cual participan 10 países con sus redes académicas y diferentes instituciones entre universidades y centros de investigación ver figura 2; el objetivo es llevar las e-Infraestructuras de los países latinoamericanos al nivel de explotación de los países Europeos. La

EELA se beneficia del estado maduro del proyecto ALICE (América Latina Interconectada Con Europa) y de la red CLARA [EELA].

7 CONCLUSIONES

Como hemos visto la tecnología GRID en Latinoamérica esta emergiendo con grandes apoyos en infraestructura y en poco tiempo se han interconectado varios países de la región, pero de que serviría toda una E-Infraestructura, si comunidades de personas provenientes de distintas disciplinas no las utilizaran para correr sus aplicaciones, por ello en los diferentes países se adelantan programas, seminarios, etc, con el fin de presentar proyectos de grid, las infraestructuras de redes que los soportan y ejemplos de cómo estos proyectos e infraestructuras están cambiando la forma de hacer ciencia a nivel mundial [Seminario de Grid Computing (2007)]. Este tipo de eventos sirven para que más entidades (Universidades, Centros de investigación, etc.) se vinculen y fortalezcan día a día la tecnología GRID, alimentándola con proyectos de investigación productivos que utilicen todos sus beneficios.

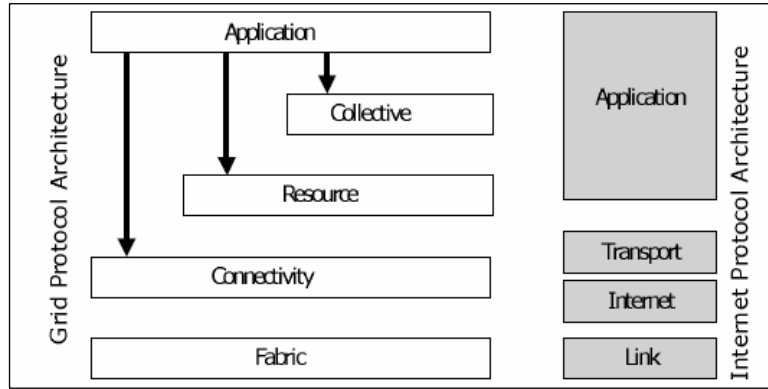


Figura1.

Arquitectura Grid por capas y su relación con la arquitectura de protocolos de Internet



Figura

E-Infrastructure del EELA

2. La

BIBLIOGRAFÍA

Gordon E. Moore.(1965), Cramming more components onto integrated circuits. En: Electronics, Vol 38, Num 8.

Foster, I. and Kesselman, C. (eds.)(1999). The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann.

Foster, I., C. Kesselmann, S. Tuecke (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. En: International Journal of SuperComputer Applications, vol. 15, nº 3.

Foster, I. (2002). What is the Grid? A Three Point Checklist. En: GRIDToday

Fuentes, A., Vázquez, J. L., Huedo, E., Montero, R. S., Llorente I. M..(2005), Benefits Achieved in Bioinformatics by Using Grid Computing Technology within IRISGrid Infrastructure . En: Boletín de RedIRIS, Num 72.

Realizing the Information Future: The Internet and Beyond. National Academy Press, 1994. <http://www.nap.edu/readingroom/books/rtif/>.

Area tematica de bioinformática (2003). Biogrid.En: irisgrid

Fundación Española para la Ciencia y la Tecnología (2004), Libro blanco e-CIENCIA en España. FECYT (Fundación Española para la Ciencia y la Tecnología)©

Doallo, R., Touriño, Heymann, J., E. (2006), Tecnologías Gris. En: CLEI 2006

Condor Project, Condor [High Throughput Computing \(HTC\)](#), [En línea], [con acceso el 20 de febrero de 2007]. Disponible en:
URL:<http://www.cs.wisc.edu/condor/>

CrossGrid Exploitation, Crossgrid, [En línea], 2004, [con acceso el 20 de febrero de 2007]. Disponible en: URL:<http://www.crossgrid.org/main.html>

BioGrid Project, biogrid construction of a Supercomputer Network, [En línea], 2005, [con acceso el 22 de febrero de 2007]. Disponible en:
URL:<http://www.biogrid.jp/>

Damien Project, Damien - Distributed Applications and Middleware for Industrial Use of European Networks, [En línea], [con acceso el 24 de febrero de 2007]. Disponible en:
URL:<http://www.hlrs.de/organization/pds/projects/damien/>

The DataGrid Project, DataGrid, [En línea], [con acceso el 25 de febrero de 2007] Disponible en: URL:<http://eu-datagrid.web.cern.ch/eu-datagrid/>

CLARA, [En línea], 2006, [con acceso el 01 de Marzo de 2007], Disponible en: URL:<http://www.redclara.net/01/01.htm>

EELA E-Infraestructure shared between Europe and Latin America, About EELA, [En línea], [con acceso el 01 de Marzo de 2007]. Disponible en:
URL:http://www.eu-eela.org/eela_about.php

Maria Teresa Dova (2006), El proyecto EELA Status en UNLP. En: 8th EELA Tutorial

Seminario de Grid Computing, Seminario Internacional de Grid Computing en Bogotá-Colombia, [En línea], 2007, [con acceso el 07 de Marzo de 2007]. Disponible en: URL:<http://agamenon.uniandes.edu.co/~esored/egrid/>

SOCIEDAD DE LA INFORMACION

www.sociedadelainformacion.com

Edita:



Director: José Ángel Ruiz Felipe
Jefe de publicaciones: Antero Soria
Luján

D.L.: AB 293-2001
ISSN: 1578-326x