

## GRID COMPUTING. PANORAMA Y ACTIVIDADES EN ARGENTINA

**Diego Fernández Slezak<sup>\*</sup>, Pablo Turjanski<sup>\*</sup>, Julio Monetti<sup>†</sup>, Leandro Bertogna<sup>#</sup>,  
Martín Silva<sup>†</sup>, Carlos García Garino<sup>†</sup> y Guillermo Marshall<sup>\*</sup>**

<sup>\*</sup> Laboratorio de Sistemas Complejos, Departamento de Computación,  
Facultad de Ciencias Exactas, Físicas y Naturales, Universidad de Buenos Aires

<sup>†</sup> LAPIC, Instituto Tecnológico Universitario, Universidad Nacional de Cuyo  
Casilla de Correo 947, (5500), Mendoza, Argentina

<sup>#</sup> Area de Sistemas, Facultad de Economía y Administración, Universidad Nacional del Comahue  
Buenos Aires 1400, (8300), Neuquen, Argentina

e-mail: dfslezak, pturjanski,@dc.uba.ar, jmonetti,msilva@uncu.edu.ar,  
mlbertog@uncoma.edu.ar, cgarcia@itu.uncu.edu.ar, marshallg@arnet.com.ar,

**Palabras Clave:** GRID Computing, Sistemas Distribuidos, Computación de Alto Rendimiento

**Resumen.** *La tecnología denominada Grid Computing o GRID es un entorno persistente que permite la integración de recursos informáticos administrados por diversas organizaciones geográficamente alejadas, asumiendo la ausencia de una ubicación central de control. GRID ha recibido un considerable impulso en los países centrales e incluso en algunos países de América Latina como Brasil, pero es relativamente poco conocida en nuestro país. La puesta en marcha de un entorno GRID para permitir que se compartan de manera segura y eficaz los recursos a disposición de la GRID es una tarea de gran envergadura. En este trabajo se presenta la tecnología, se introducen sus antecedentes y se muestran distintas aplicaciones y experiencias realizadas por los grupos de LSC, LAPIC y Universidad del Comahue.*

## 1 INTRODUCCION

La computación científica de alto rendimiento (HPC) ha cambiado dramáticamente en los últimos años permitiendo avances a un ritmo antes inimaginable. HPC reúne la riqueza de los avances tecnológicos no solamente en el área del hardware y de la velocidad de cómputo cruda, sino también en la tecnología de base de datos, en la visualización y su infraestructura, en la algoritmia computacional eficiente, en las redes de alta velocidad y en el acceso remoto a través de portales con nuevas tecnologías de servicio de web. La complejidad de la infraestructura para que estas componentes lógicas funcionen en conjunto de manera eficiente supera la capacidad de un usuario científico y debe ser encarada por especialistas en software. El objetivo último es desarrollar una interfase que simplifique la tarea del científico y que permita el desarrollo de una ciencia más compleja. Aquí surge el Grid Computing<sup>1,2</sup> o GRID o grid middleware que es un software que provee en forma consistente API's (Application Program Interface), para interactuar con diferentes tipos de recursos. El desarrollo del grid middleware es un campo virgen absolutamente prioritario y al cual, los países centrales dedican recursos cuantiosos. Grid Computing es el nuevo paradigma de web del próximo decenio, de allí su relevancia. Para integrar los recursos de cómputos descriptos, se crean los primeros portales GRID (LSC, LAPIC y Laboratorio Área de Sistemas de la U.N. Comahue) del país cuyos objetivos es simplificar el acceso de usuarios a los recursos de cómputo.

Las especificaciones de la tecnología GRID están a cargo de la Globus Alliance<sup>3</sup>, comunidad de organizaciones e individuos que desarrollan tecnologías básicas que definen al GRID y que permiten a un usuario compartir poder de cómputo, bases de datos, instrumentos y otras herramientas 'on-line' a través de fronteras corporativas, institucionales y geográficas, sin sacrificar su autonomía local.

El Open Grid Service Architecture (OGSA) define una arquitectura común, estándar, abierta y basada en servicios para aplicaciones de GRID. El objetivo de OGSA es estandarizar prácticamente todos los servicios que se pueden encontrar en estas aplicaciones (servicios de manejo de recursos, servicios de seguridad, etc.) especificando un conjunto de interfaces para estos servicios. En esta arquitectura un servicio de GRID es simplemente una extensión de un servicio web. El Open Grid Service Infrastructure (OGSI) brinda una especificación formal y técnica de estos servicios.

En este trabajo se ha empleado la distribución Globus Toolkit® 3.2, que implementa la arquitectura y las especificaciones citadas. Actualmente se dispone de la versión 4.0 con la cual se han realizado pruebas preliminares. Un número creciente de proyectos y empresas utilizan el Globus Toolkit® para desarrollar el potencial de esta herramienta.

La implementación de un proyecto GRID debe satisfacer condicionamientos de muy diversa índole tanto de hardware como de software, requiriendo una infraestructura que en general excede o se encuentra en los límites de los presupuestos universitarios. Es por ello que antes de tomar una decisión del esquema definitivo de GRID a utilizar es necesario estudiar distintas opciones desarrollando diseños preliminares que ayuden a la elección óptima.

En el punto 2 de este trabajo se presenta una breve reseña de los antecedentes GRID, se continúa con la descripción y el modelo conceptual de la Tecnología en los puntos 3 y 4

respectivamente. Los estudios y experiencias que se han llevado a cabo se presentan en el punto 5, y finalmente se presentan las conclusiones del trabajo.

## 2 ANTECEDENTES DE GRID COMPUTING

Durante la década de los ochenta se realizó una intensa investigación, desarrollo y despliegue de *hardware*, *software* y aplicaciones de computación paralela. Durante la Conferencia de Supercomputación realizada en San Diego en 1995 por IEEE y ACM, Ian Foster lanzó la iniciativa I-WAY que consistió en una metacomputadora que a través de 11 redes de alta velocidad conectó 17 sitios con recursos de computación, durante la cual se procesaron alrededor de sesenta aplicaciones diferentes en ciencias e ingeniería. Debido al éxito de I-WAY, la agencia DARPA del Ministerio de Defensa de Estados Unidos decidió otorgar fondos para el proyecto *Globus*<sup>3</sup>, que liderado por Ian Foster y Karl Kesselman, diseñó un conjunto de herramientas que han dado lugar a lo que hoy se conoce como “Grid Computing” o Mallas Computacionales<sup>1,2</sup>. La tecnología se ha difundido enormemente en los últimos años, numerosas empresas como SUN, IBM, Intel o Hewlett Packard, etc., se han interesado en la misma y en Europa se la considera una iniciativa estratégica dentro del programa de la Sociedad de la Información<sup>4</sup>.

El desarrollo de GRID ha crecido de manera impactante en los últimos años. Iniciativas tanto académicas: TERAGRID<sup>5</sup> y NORDUGRID<sup>6</sup> como comerciales IBM Toolbox®<sup>7</sup> y Oracle 10g®<sup>8</sup> han demostrado que es posible compartir recursos de manera sencilla con la comunidad. El éxito de estas nuevas tecnologías radica en que han podido resolver las necesidades imperiosas de poder de cómputo en el desarrollo, por ejemplo, de nuevos materiales o en el estudio de la biología molecular, por citar ejemplos paradigmáticos. Algunas de las aplicaciones clave de GRID son la explotación de recursos de cómputos subutilizados, por ejemplo, el uso remoto en horas de la noche de millones de PC's que se encuentran ociosas, o la utilización paralela de la capacidad ociosa de CPU a través de la computación masivamente paralela. Ejemplo en nuestro país de esto último es el proyecto Cluster Educativo<sup>21</sup> implementado por el Laboratorio de Sistemas Complejos (LSC), Departamento de Computación (DC), FCEN, UBA que consiste en la instalación de un Cluster utilizando las máquinas de los laboratorios de alumnos del DC que se encuentran ociosas gran parte del tiempo, en especial durante la noche.

## 3 DESCRIPCION DE GRID COMPUTING

GRID Computing es una tecnología que plantea el acceso a recursos de computación de diferente tipo, empleando recursos geográficamente distantes, sobre la base de un esquema de confianza.

GRID toma su nombre de la red eléctrica (Power GRID) ya que una forma de introducir esta tecnología es plantear una analogía entre la red eléctrica y una red de recursos de computación. Resulta natural encender en el hogar cualquier dispositivo eléctrico sin preocuparnos del origen de la energía que se consume, por otra parte resulta difícil de

concebir las sociedad actual si la energía eléctrica no se dispusiera de una manera abundante y permanente.

La energía eléctrica se puede generar de diversas maneras: centrales nucleares, parques eólicos, plantas hidráulicas, etc. La misma se transporta y se pone a disposición de los usuarios que generalmente desconocen el origen y la localización geográfica de las fuentes de energía.

Sin embargo para que se pueda disponer de energía de manera eficiente y económica es necesario realizar diferentes transacciones entre los generadores y los distribuidores, controlar y facturar las mismas, medir la energía que consumen los usuarios, etc., y, además, controlar y medir la red propiamente dicha, las redes de transmisión, transformadores, etc.

Los pasos anteriores permiten que muchos usuarios anónimos, en el última etapa del sistema conecten sus equipos a la red, simplemente enchufando a la misma el dispositivo en cuestión. Un usuario de Internet que enciende el equipo en su hogar u oficina realiza una tarea análoga. Sin embargo dicha etapa, relativamente pasiva, necesita, en términos computacionales, disponer de los mecanismos de *generación y distribución* de recursos.

Grid Computing, a diferencia de la conocida Internet, ofrece recursos de cómputo, dispositivos de almacenamiento, acceso a información, etc., de una manera segura y en ausencia de un control centralizado.

Una definición bastante aceptada de Grid es la propuesta por Globus<sup>3</sup>: *Mediante Grid Computing se permite que organizaciones virtuales compartan recursos geográficamente distribuidos a la vez que se proponen objetivos comunes, suponiendo la ausencia de un sitio de control central localizado, a partir de la base de una relación de confianza existente.*

Es necesario entonces disponer de un catálogo de recursos, así como coordinar de manera eficiente y segura de dichos recursos y disponer de mecanismos eficientes de uso de los mismos.

#### 4 MODELO CONCEPTUAL DE GRID

El conjunto de usuarios y desarrolladores de la tecnología ha propuesto un modelo por capas que define la *Anatomía de la Grid*, y define la funcionalidad de cada una de las capas<sup>9</sup>, el cual se muestra en la figura 1.

El trabajo mencionado define las funciones y características de cada una de las capas. Un resumen de las mismas, basado en un trabajo previo de los autores<sup>10</sup> se lista a continuación:

La Capa inferior del modelo conceptual de Grid denominada *Fabric* provee la descripción de recursos a ser compartidos a lo largo de la malla (procesadores, discos, sensores, etc). En cada caso un recurso puede implementar protocolos internamente, pero esta propiedad no es observada por la arquitectura GRID, como así tampoco las capacidades de administración y políticas de uso que recaigan sobre el mismo en forma local. Existe una fuerte interdependencia entre las funciones implementadas en el nivel de *Fabric* y las capacidades del recurso a ser compartido.

La capa *Connectivity* define las comunicaciones a nivel de núcleo de la aplicación y también especifica los protocolos de autenticación requeridos para realizar transacciones dentro de la malla. Los protocolos de comunicación permiten el intercambio de datos entre

los distintos recursos, y los protocolos de autenticación proveen los mecanismos de criptografía seguros para el reconocimiento de usuarios y recursos.

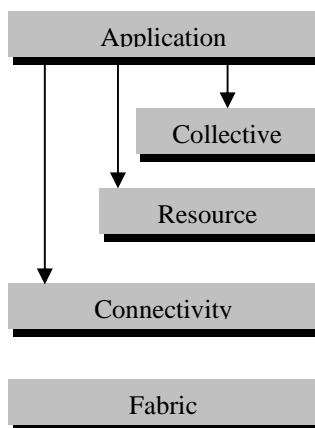


Figura 1: Modelo de Capas de una Arquitectura GRID<sup>9</sup>

La capa *Resource* se basa en los protocolos de comunicación encontrados en la capa de conectividad para realizar tareas de negociación segura, iniciación, monitoreo y control de uso de los distintos recursos compartidos. Los protocolos ubicados en esta capa se relacionan con recursos individuales, ignorando temas de estado global. Se pueden distinguir dos clases primarias de protocolos en esta capa: a) protocolos de información y b) protocolos de administración.

Mientras la capa anterior se enfoca en interacciones con recursos individuales la capa denominada *Collective* no está asociada con ningún recurso en particular, sino que especifica protocolos y servicios de naturaleza global, capturando interacciones establecidas dentro de un grupo o colección de recursos. Los componentes de esta capa podrán implementar una amplia variedad de comportamientos sobre los recursos al compartirlos, sin introducir nuevos requerimientos, ya que esta capa reutilizará la funcionalidad propuesta por las capas de recursos y conectividad.

Finalmente la capa *Application* comprende las aplicaciones que utiliza el usuario final que opera en una organización virtual. Las aplicaciones se construyen en función de los diferentes servicios disponibles en las capas inferiores, mediante llamadas a las mismas. De esta manera se puede acceder a servicios que permiten la administración de recursos, acceso a datos, descubrir servicios, etc.

## 5 APLICACIONES

Como se ha señalado en este trabajo en nuestro país hay muy pocos centros en los cuales se desarrolle el estudio y la investigación de GRID Computing. En este apartado se presentan las experiencias que se han llevado a cabo en LAPIC, ITU, UNCUyo, LSC, DC, FCEN, UBA y en el Área de Sistemas de la Facultad de Administración de la Universidad del Comahue. Los laboratorios y prototipos que se describen, si bien son experiencias autónomas de cada centro, se han desarrollado de manera coordinada en el contexto de los proyectos PICTR

184<sup>11</sup> y Sub4/PAV 127<sup>12</sup> del programa FONCyT de la ANPCyT.

Una dificultad que surge en nuestro país respecto de GRID es la falta de redes de alta velocidad que conecten universidades y centros de investigación. Al presente se dispone de un túnel IP entre el LSC y el LAPIC y a la brevedad se espera contar con conexiones dedicadas entre ambos centros mediante la Red Académica Avanzada de RETINA. El grupo de la Universidad Nacional del Comahue en la provincia del Neuquén se conecta al CRIBABB (Centro de Investigaciones Básicas y Aplicadas de Bahía Blanca) dependiente del CONICET a través de red pública y desde allí accede la Red Académica Avanzada de RETINA.

### 5.1 Prototipo de GRID en entorno LAN

Una posibilidad de interés que permite GRID es el empleo de las computadoras de una organización de manera mancomunada para disponer de potencia de cálculo a gran escala.

Para investigar esta posibilidad en LAPIC se instaló un prototipo en el Laboratorio de Redes de la Carrera Redes y Telecomunicaciones del ITU, en donde se cuenta con 9 Microcomputadoras tipo PC equipadas con un microprocesador Duron a 900Mhz, 256 MB de memoria RAM, Disco Rígido de 20 GB, placa de red FastEthernet y los puertos de E/S típicos de estos equipos. Los equipos funcionan con el sistema operativo Linux Red Hat 9.0.

Del conjunto de estaciones disponibles se aislaron dos, como se muestra en la figura 2, que se dispusieron en una red privada, y a las cuales se les instaló el Globus Toolkit 3.2, con el fin de realizar diferentes experiencias con la herramienta que se estudia. En una de las máquinas se habilitó la capacidad de generar certificados digitales para garantizar la seguridad del acceso. Una tercera estación se habilitó con el objeto de verificar parámetros de compilación e instalación del Toolkit.

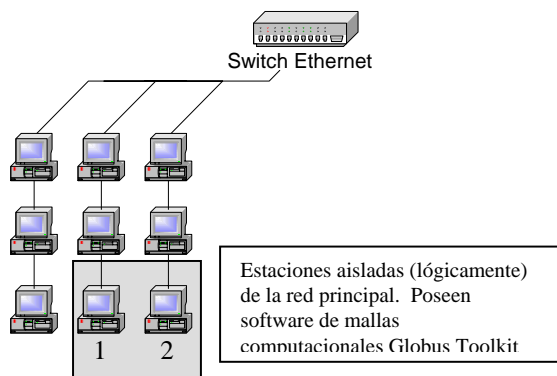


Figura 2: Una malla computacional simple con dos nodos. Nodo 1: servidor Grid; Nodo 2; cliente Grid.

La disposición de los equipos en forma local ha permitido depurar diferentes problemas de instalación y puesta en marcha. También se realizaron diferentes pruebas y aplicaciones sencillas, entre ellas cabe citar la remisión de trabajos entre los distintos *hosts* participantes en el experimento o bien el empleo de aplicaciones incluidas en el Toolkit, por ejemplo *GridFTP*.

Las pruebas se realizaron con una configuración muy sencilla, empleando dos de las estaciones, en las cuales se dispuso un servidor y un cliente de Globus respectivamente. Se estudió con detalle la seguridad del Grid, con el objeto de conocer las facilidades disponibles en la distribución, como resulta la creación de certificados digitales de usuario y host.

En general el ensayo puede considerarse exitoso, ya que si bien el prototipo utilizado funcionó en un entorno de red local, se pudo realizar la instalación y puesta en marcha, emplear recursos remotos y de esta manera acceder a las posibilidades de la tecnología analizada. Se han podido detectar dificultades en la instalación del producto, en la compilación de aplicaciones, recursos de infraestructura necesarios, etc.

A manera de resumen cabe señalar que actualmente se dispone de la versión 4.0 de Globus, que facilita muchísimo la instalación, prácticamente no presenta dificultades y el tiempo para instalar el paquete es relativamente bajo, dado que puede obtenerse precompilado. Por otro lado esta versión obliga a rescribir y rediseñar todas las aplicaciones en el contexto de los denominados *Web Services*. Si bien ya se dispone de hardware más potente en LAPIC, para desarrollar pruebas con GRID es suficiente con un PC del tipo de los reseñados en este apartado.

## 5.2 Laboratorio Virtual para GRID

En el contexto de la colaboración que se desarrolla entre LSC y LAPIC, en Buenos Aires se llevó a cabo una experiencia con el objeto de estudiar la tecnología GRID y poner en marcha un laboratorio virtual<sup>13</sup> como diseño preliminar de un laboratorio de dichas características.

Con el propósito de analizar el middleware, solucionar problemas puntuales de instalación y configuración, y de esta forma definir una instalación tipo, se utilizó una máquina de escritorio bajo plataforma Windows XP® en la cual se creó una máquina virtual utilizando el programa Microsoft VirtualPC®. En dicha máquina se instaló Linux Debian (versión estable) y el Globus Toolkit® 3.2. A partir de las experiencias obtenidas se implementó el laboratorio virtual.

Durante la primera etapa del trabajo, “experiencias con el nodo inicial”, se presentaron algunas dificultades similares a las encontradas en LAPIC, ya que la correcta instalación de Globus Toolkit® requiere tener preinstalados una gran variedad de paquetes del sistema operativo Linux. Si bien Linux Debian permite una enorme flexibilidad en la elección de herramientas o paquetes adicionales a instalar en el sistema, la documentación de la versión instalada del Globus Toolkit® no provee una lista detallada de los paquetes necesarios, lo que obliga al instalador a seleccionar a priori los paquetes a instalar, a realizar instalaciones de prueba y naturalmente a detectar a través de sucesivos ensayos y errores los paquetes faltantes.

Fueron necesarios, en promedio, más de una docena de ensayos de instalaciones fallidas, cada uno de los cuales requiere más de 4 horas de compilación, a lo que se debe agregar el tiempo de detección de errores y búsqueda e instalación de paquetes faltantes, para obtener una idea de la magnitud de la tarea. Afortunadamente, dicha tarea fue facilitada a través de la creación de sucesivas copias de seguridad de la máquina virtual, que permitieron volver a

instancias anteriores y así superar con mayor rapidez los errores no esperados.

Como resultado de las sucesivas iteraciones, se cuenta con un manual de instalación, de Globus Toolkit® 3.2 bajo el sistema operativo Linux Debian, que documenta los pasos a seguir y los paquetes a instalar e indicando los errores más frecuentes de instalación.

Como segunda etapa del diseño del Laboratorio de GRID Virtual, se utilizaron 3 máquinas de escritorio, nuevamente bajo plataforma Windows XP®. En cada máquina real se crearon 3 máquinas virtuales. Cada máquina virtual fue instalada siguiendo la instalación tipo de la etapa inicial. El esquema se presenta en la Figura. 3.

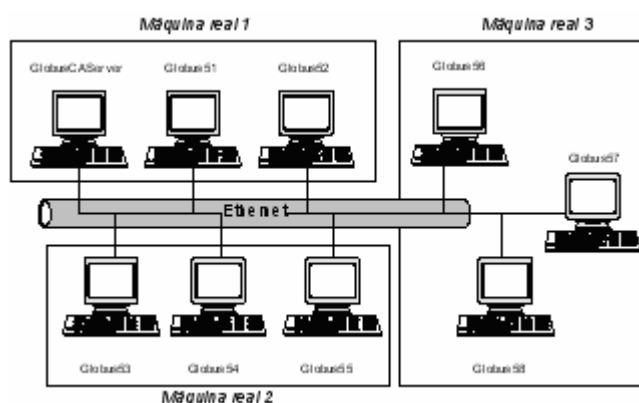


Figura 3: Esquema de laboratorio de GRID Virtual – GRID Virtual UBA

En cada una de las máquinas reales se realizó una instalación modelo de máquina virtual, la cual fue clonada para cada uno de los nodos adicionales. En cada máquina clonada fue necesario realizar pequeños cambios de configuración que identifican cada máquina unívocamente, entre los cuales podemos destacar, el cambio de identificador virtual de placa de red, nombre de equipo de red y servicios activos que definen la funcionalidad de cada nodo (por ejemplo, GlobusCAServer como servidor de archivos, compartiendo una carpeta con los paquetes necesarios de instalación).

La clonación de máquinas virtuales permitió rápidamente contar con el laboratorio virtual con un bajo costo en tiempo y esfuerzo.

Una vez instalado el laboratorio virtual se configuraron los nodos para que funcionen en su conjunto como un sistema GRID Virtual. Para ello, se seleccionó una máquina particular (GlobusCAServer) como servidor de credenciales para los permisos de usuarios y equipos. Posteriormente se emitieron los certificados correspondientes.

### 5.3 Interconexión de los equipos de LSC y LAPIC

Los prototipos desarrollados en los puntos 5.1 y 5.2 se interconectaron mediante un Túnel IP-IP<sup>14</sup> entre ambos centros.

Un túnel es un canal de comunicación creado en una red de computadoras que encapsula un paquete de datos de un protocolo de comunicación dentro de un segundo protocolo. En el



caso del protocolo IP<sup>15,16</sup> se encapsulan direcciones privadas dentro de direcciones IP públicas y de esta forma se accede en forma transparente a redes privadas remotas.

En este caso, como se indica más abajo, mediante esta facilidad se conectaron los equipos Chizito en UBA, y el master del prototipo del LSC en Mendoza.

Con el objetivo de mantener la conexión a través de un Túnel IP-IP con el cluster Beowulf de la Universidad de Cuyo Mendoza, en el laboratorio GRID Virtual de la UBA se agregó un nuevo nodo en una máquina física, nuevamente con un sistema operativo Linux Debian versión *stable* y Globus Toolkit® 3.2. El esquema de laboratorio virtual puede observarse en la Figura. 5.

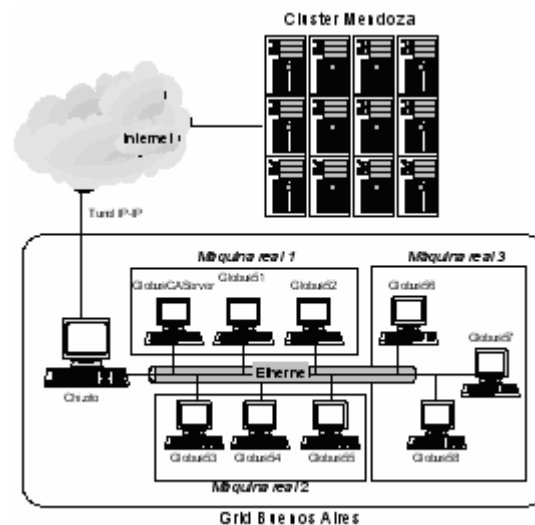


Figura 4: Esquema del GRID Buenos Aires con Túnel al prototipo del LSC en Mendoza

## 5.4 Laboratorios Remotos

Esta aplicación tiene como objetivo la interconexión de equipamiento y acceso remoto, como destinatarios tiene CRIBABB con sus institutos asociados, Planta Piloto de Química (PLAPIQUI), Instituto Argentino de Oceanografía (IADO), Investigaciones Bioquímicas de Bahía Blanca (INIBIBB) y Centro de Recursos Renovables de la Zona Semiárida (CERZOS), además del departamento de Ciencias de la Computación de la Universidad Nacional del Comahue.

El proyecto surgió ante la necesidad de agrupar recursos tecnológicos heterogéneos dentro y fuera de las organizaciones involucradas, y mediante una computadora virtual, lograr un poder de cálculo mayor que el obtenido individualmente. Este proyecto también persigue el objetivo de establecer relaciones con otros emprendimientos internacionales, que tiene como fin lograr una cooperación en el ámbito científico sobre la tecnología Grid entre Europa y Latinoamérica, y el proyecto Red CLARA, que es una iniciativa para las redes académicas

avanzadas en Latinoamérica.

Al considerar sobre qué tipo de aplicaciones era preferible desarrollar nuestras experiencias, pensamos en problemas adecuados para procesamiento paralelo. La búsqueda de aplicaciones para prueba se orientó entonces a programas de generación de imágenes sintéticas o rendering, con el asesoramiento del grupo local de investigación en gráficas del área de Ciencias de la Computación de la Universidad del Comahue. Se utilizó la versión 0.97 de Tachyon<sup>18</sup>, la cual ofrece una gran variedad de opciones de compilación para diferentes plataformas. En nuestra plataforma Linux, Tachyon fue compilado con soporte para librerías de pasaje de mensajes MPICH<sup>19</sup>.

Nuestra intención fue paralelizar la generación de frames o cuadros de movimientos de una animación sobre los clusters; por ejemplo, si deseamos generar 128 frames de animación de una escena, el proceso de ejecución es generar el archivo XML con los trabajos para cada cluster y enviar a ejecución a un cluster los primeros 64 frames y al mismo tiempo enviar al segundo cluster los próximos 64 frames. Una vez que ambos finalizan su ejecución, se combinan los frames para formar archivos en formato de video MPEG-1, y éstos se recuperan desde cada uno de los clusters combinándolos en uno solo para luego publicarlo en Internet.

Esta aplicación se implementó sobre tres clusters, dos de los cuales se encuentran instalados físicamente en el departamento de Ciencias de la Computación de la Universidad Nacional del Comahue y el tercero en el CRIBABB de Bahía Blanca. Cada cluster tiene aproximadamente cinco computadoras la mayoría de estas Pentium III con 64Mb de RAM, los clusters están instalados con arquitectura Beowulf con la ayuda del software Oscar 4.0<sup>20</sup>. También se contó con máquinas Pentium IV con 1GB de RAM como nodo coordinador de los clusters, en total se utilizaron 14 máquinas reales. Las tres máquinas conectadas a Internet formaban una VPN (red privada virtual) para cubrir aspectos de seguridad y interconexión.

En el aspecto de monitoreo del sistema se vinculó el servicio MDS, monitoreo a nivel de GRID, con datos de la aplicación Ganglia<sup>21</sup> monitoreo a nivel de cluster, con este fin se utilizó la propiedad GLUE provista por el Globus Toolkit 4.0. Esta obtiene información de distintas fuentes; como el planificador, o el sistema de información propio de cluster. De esta manera bajo una misma interfaz obtenemos la información del sistema Grid y los clusters en forma particular desde una interfaz centralizada.

Los resultados obtenidos en las ejecuciones pueden verse en la Figura 5. Las tres primeras columnas muestran los mejores tiempos de los clusters sin la intervención del software de Grid; la dos columnas siguientes, muestran los tiempos de ejecuciones paralelas entre dos y tres clusters respectivamente.

En las ejecuciones de los clusters en conjunto los tiempos disminuyen. En la ejecución de los clusters CRIBABB-G1, los dos más veloces, el tiempo resultado es casi un promedio de los dos mejores tiempos. En cambio si agregamos el tercer cluster y ejecutamos en forma paralela, si bien la diferencia de tiempo decrece sensiblemente es menor que si realizáramos el promedio de los tres.

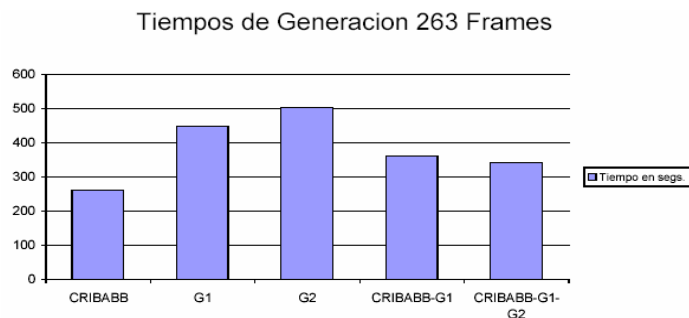


Figura 5: Tiempos de generación de las imágenes

Es necesario remarcar que con esta experiencia no se busca establecer análisis de performance entre las distintas modalidades de ejecución, sino que el objetivo es establecer la factibilidad del uso de la tecnología Grid, tomando en cuenta la sobrecarga que añade este software intermedio sobre entornos heterogéneos dispersos geográficamente.

## 6 CONCLUSIONES

En el trabajo se han descrito diferentes experiencias realizadas con Grid Computing. Las mismas muestran la factibilidad de trabajar con esta tecnología, empleando diferentes recursos de hardware: máquinas virtuales, laboratorios parcialmente disponibles en una organización real y finalmente en un entorno WAN.

Desde el punto de vista de las distribuciones, las experiencias presentadas se han realizado con el Toolkit 3.2 de Globus que resulta relativamente complejo de instalar. Actualmente en los tres laboratorios se ha migrado a Globus 4.0, versión mucho más simple de instalar.

La posibilidad de instalar Grid sobre máquinas virtuales permite avanzar en la investigación de la tecnología, aún cuando no se dispongan de grandes recursos de hardware, y de esa forma se puede avanzar en el diseño de aplicaciones, pruebas de certificados, depuración de la herramienta, etc.

Muchas veces existen recursos ociosos que pueden, en el entorno de una organización, real, dedicarse parcialmente a Grid, como se mostró con la experiencia realizada en entorno Lan. Todas las experiencias desarrolladas en un laboratorio virtual se pueden transferir directamente a esta escala de equipamiento.

Grid Computing combinado en ciertos casos con arquitecturas paralelas, es una tecnología valiosa para el acceso a recursos remotos y para la Computación de Alto Rendimiento, como se muestra en la experiencia del rendering de imágenes, que se desarrolla en el contexto de una red WAN.

En resumen el presente trabajo muestra el estado del arte de Grid Computing en Argentina, mediante una descripción de la tecnología, una presentación de los grupos activos en el tema y, muy especialmente, permite presentar un esquema de colaboración del tipo *e-science* entre los grupos de Buenos Aires, Mendoza y Neuquen, que surge como fruto de una cooperación planteada a partir del interés común por Grid.

Permanecen algunos interrogantes acerca de Grid como resultan la difusión de la misma y su empleo masivo, o bien qué posibilidades de uso prácticas se puede concebir con la actual

infraestructura de comunicaciones en nuestro país.

## 7 AGRADECIMIENTOS

Los autores agradecen a la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), la ayuda económica recibida a través de los proyectos PICTR 184 y Subproyecto 4 del PAV 127, del FONCyT.

## 8 REFERENCIAS

- [1] F. Berman, G. Fox, A. Hey, Grid Computing. Making the global infrastructure a Reality, Wiley, 2003
- [2] Ahmar Abbas, Grid Computing: A practical Guide to Technology and Applications, Charles River Media, 2003.
- [3] Proyecto GLOBUS: [www.globus.org](http://www.globus.org).
- [4] Programa Europeo para la Sociedad de la Información. Comunidad Europea. Disponible en: <http://www.cordis.lu/ist/grids>
- [5] Proyecto TERAGRID: [www.teragrid.org](http://www.teragrid.org).
- [6] Proyecto NORDUGRID: [www.nordugrid.org](http://www.nordugrid.org).
- [7] [http://www-1.ibm.com/grid/solutions/grid\\_toolbox.shtml](http://www-1.ibm.com/grid/solutions/grid_toolbox.shtml).
- [8] <http://www.oracle.com>.
- [9] Ian Foster, Carl Kesselman y Steven Tuecke: The Physiology of the Grid: Enabling Scalable Virtual Organizations. Intl Journal of Supercomputer Applications, 2001. Disponible en [www.globus.org/research/papers/anatomy.pdf](http://www.globus.org/research/papers/anatomy.pdf).
- [10] Julio C. Monetti y Carlos García Garino: Mallas Computacionales. Conceptos y Experiencias, Segundo Congreso Internacional de Informática Cidi – Cuyo 05 , Edición en CD, CIDI-Cuyo, Mendoza, 13 al 15 de Abril de 2005.
- [11] Proyecto PICTR 184, Centro de Computación de Alto Rendimiento. 2004-07 (en curso). Directores: Guillermo Marshall y Carlos García Garino. FONCyT, Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT).
- [12] Proyecto PAV 127, Subproyecto 4, Centro Virtual de Computación de Alto Rendimiento. 2005-06 (en curso). Director: Guillermo Marshall . FONCyT, Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT).
- [13] D. Fernández Slezak et al., Construcción de un laboratorio de GRID Virtual para HPC., presentado durante el Sexto Simposio Argentino de Tecnología, AST, JAIIO 2005, Rosario 1 y 2 de Septiembre. Publicado en CD, SADIO, 2005.
- [14] W. Simpson, *RFC 1853 "IP in IP tunneling"*, Network Working Group, 1995.
- [15] J. Reynolds, J. Postel, *RFC 1700 "Assigned Numbers"*, Network Working Group, 1994.
- [16] J. B. Postel, *RFC 791 "Internet Protocol (IP)"*, DARPA Internet Program, 1981.
- [17] Y. Rekhter et al, *RFC 1597 "Address Allocation for Private Internets"*, Network Working Group, 1994.

- [18] MPICH. <http://www-unix.mcs.anl.gov/mpi/mpich/>
- [19] John Stone, "*An Efficient Library for Parallel Ray Tracing and Animation*". Intel Supercomputer Users Group Conference. 1995
- [20] Oscar 4.0. <http://oscar.openclustergroup.org/>
- [21] Massie, M., B. Chun, and D. Culler. The Ganglia Distributed Monitoring: Design, Implementation, and Experience. in *ParallelComputing*. May 2004