

IMPLANTACIÓN DE UNA INFRAESTRUCTURA DE COMPUTACIÓN GRID EN
LA UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA, UTILIZANDO EL
MIDDLEWARE GLOBUS TOOLKIT

MAYRA ALEJANDRA GONZÁLEZ RAMÍREZ
VANESA LUCIA HERNÁNDEZ RODRÍGUEZ

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
FACULTAD DE INGENIERÍA DE SISTEMAS
TELECOMUNICACIONES Y TECNOLOGÍAS WEB
BUCARAMANGA

2009

IMPLANTACIÓN DE UNA INFRAESTRUCTURA DE COMPUTACIÓN GRID EN
LA UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA, UTILIZANDO EL
MIDDLEWARE GLOBUS TOOLKIT

MAYRA ALEJANDRA GONZÁLEZ RAMÍREZ
VANESA LUCIA HERNÁNDEZ RODRÍGUEZ

Trabajo de grado presentado para optar el título de:
Ingeniería de Sistemas

DIRECTOR:

Msl. Roberto Carvajal Salamanca

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
FACULTAD DE INGENIERÍA DE SISTEMAS
TELECOMUNICACIONES Y TECNOLOGÍAS WEB
BUCARAMANGA

2009

Nota de aceptación:

Presidente del Jurado

Jurado

Jurado

	pág.
INTRODUCCIÓN	12
1. CLÚSTER	14
1.1 DEFINICIÓN DE CLÚSTER	14
1.2 FUNCIONAMIENTO DE UN CLÚSTER	15
1.3 DISTRIBUCIÓN DE LINUX Y PAQUETES PARA LA CREACIÓN DE CLÚSTER	16
2. GRID	19
2.1 CARACTERÍSTICAS	21
2.2 TOPOLOGÍAS	22
2.2.1 IntraGrid (InnerGrid)	22
2.2.2 ExtraGrid (OuterGrid)	23
2.2.3 InterGrid	24

2.3	PARA QUÉ SIRVE EL GRID DE CÓMPUTO	24
2.3.1	Supercomputación distribuida.	25
2.3.2	Sistemas distribuidos en tiempo real.	26
2.3.3	Proceso intensivo de datos.	26
2.3.4	Servicios puntuales.	26
2.3.5	Entornos virtuales de colaboración.	27
2.4	VENTAJAS Y DESVENTAJAS	27
2.4.1	Ventajas y requisitos.	27
2.4.2	Desventajas.	29
2.5	MIDDLEWARE	30
2.5.1	Globus® Toolkit.	30
2.5.1.1	Herramientas Web de software libre para Globus Toolkit 4.0.	35
3.	PRODUCTO FINAL	38
3.1	CONSTRUCCIÓN DE LAS SALA-CLÚSTER	38
3.2	INTERCONEXIÓN DE DOS SALA-CLÚSTER PARA CREAR UN	

INTRAGRID	39
3.3 CONEXIÓN Y COMUNICACIÓN DE DOS SERVIDORES EN LA RED	39
3.4 CONFIGURACIÓN Y PRUEBA DE COMPONENTES DE GLOBUS	40
4. CONCLUSIONES	42
5. RECOMENDACIONES PARA TRABAJOS FUTUROS	46
BIBLIOGRAFÍA	47

LISTA DE TABLAS

pág.

Tabla 1. Comparación de herramientas Web de Software Libre de Globus

Toolkit.

36

LISTA DE FIGURAS

	pág.
Figura 1. Funcionamiento de un Grid de Cómputo	21
Figura 2. Intragrid, Extragrid e InterGrid	23
Figura 3. Servicios Globus Toolkit versión 4 (GT4)	33
Figura 4. Arquitectura esquemática de GT4 y varios de sus componentes	34
Figura 5. Capacidades del contenedor GT4	35

LISTA DE ANEXOS

	pág.
Anexo A. Marco conceptual sobre middleware Globus Toolkit 4.0 para el montaje de una IntraGrid	51
Anexo B. Marco conceptual sobre la descripción y comparación de herramientas Web de Software libre utilizadas para Globus Toolkit 4.0	87
Anexo C. Documentación de Errores	112

RESUMEN

La Universidad Autónoma de Bucaramanga cuenta con infraestructura tecnológica, esencialmente computadores personales de escritorio que están a disposición de los estudiantes, docentes y directivos. Estos computadores son constantemente utilizados en el día y parte de la noche, pero permanecen ociosos el resto del tiempo. Teniendo en cuenta lo anterior y que además existen proyectos en la universidad que requieren alto poder de cómputo, se propone una solución de bajo costo que aproveche los recursos de la universidad, a través de la utilización de clústeres de computadores y su interconexión a través de redes dedicadas para conformar un Grid de computación. Así mismo, este tipo de solución, relativamente nueva en nuestro medio, abre paso a un nuevo campo de investigación y desarrollo, especialmente en el área del software libre, pilar fundamental para este tipo de soluciones.

Esta solución no solo brinda beneficios a la universidad, si no que a su vez ofrece al sector público y a organizaciones la posibilidad de manejar y administrar la información de procesos financieros, matemáticos y científicos (siendo evidente que este tipo de información tiende a crecer exponencialmente cada día, incrementando así, la necesidad de capacidad, procesamiento y velocidad de transmisión); también ofrece la posibilidad de obtener más poder de cómputo de manera económica, siempre y cuando se utilicen equipos ya existentes en las organizaciones, aprovechando el tiempo ocioso de los mismos para conformar un sistema multiprocesador virtual que permita realizar las tareas fuertes en esos momentos.

Además de traer diversos beneficios sobre distintas áreas de la investigación científica y tecnológica del país, la computación Grid, atraería la atención de otras organizaciones internacionales del mismo tipo posibilitando la integración con otras instituciones regionales, nacionales y Latinoamericanas, gracias al enorme potencial de intercambio y gestión de recursos que ofrece esta tecnología.

Palabras claves: Clúster, Grid, Procesamiento distribuido, Globus Toolkit.

Línea de investigación: Telecomunicaciones y Tecnologías WEB

INTRODUCCIÓN

Actualmente, la Universidad Autónoma de Bucaramanga (UNAB) mantiene una infraestructura de cómputo grande y costosa, destinada solo para atender las labores académicas y administrativas propias de la organización, pero que durante muchas horas del día y especialmente de la noche permanecen ociosas. Así mismo, es evidente también que la UNAB realiza cada vez y con mayor frecuencia proyectos de investigación que requieren gran poder de cómputo y de almacenamiento, los cuales necesitan de poderosas y costosas computadoras que resuelvan sus necesidades.

En cuanto a las necesidades de cómputo, existe una solución de bajo costo que utiliza el poder de cómputo ocioso para crear un sistema multiprocesador virtual que permite cumplir con estas necesidades; por ello se propuso el desarrollo de un proyecto de pregrado que permitiera avanzar en la realización de una implantación de una infraestructura computacional GRID en la Universidad Autónoma de Bucaramanga con el fin de apoyar y servir de herramienta productora de proyectos de investigación, principalmente a aquellos enfocados en el manejo y procesamiento de informes XBRL .

Para implantar una infraestructura como esta, se requiere el uso de equipos de cómputo organizados en forma de Clústeres, los cuales permiten que varios computadores conectados en una red funcionen como una sola máquina agilizando la ejecución de procesos y tareas ya que estas se efectúan de forma paralela, disminuyendo el tiempo de respuesta y logrando obtener resultados en tiempos más razonables. A su vez, para integrar los Clústeres existentes y formar un Grid, se requiere de equipos de cómputo con herramientas software

previamente instaladas y configuradas que integren toda la capacidad de procesamiento distribuida formando una gran malla computacional que permita: división, envío y ejecución de las tareas en cada uno de los nodos disponibles en la Grid y posteriormente el ensamble y almacenamiento de los trabajos realizados, todo esto de forma transparente para los usuarios.

Esta solución ofrece muchos beneficios entre los cuales se destaca el incremento en la velocidad de procesamiento, incremento de la confiabilidad de ejecución de las aplicaciones y el aprovechamiento de los recursos de la organización. Por otro lado, las herramientas software y sistemas operativos necesarios para el proyecto, cuentan con licencia GNU y se encuentran disponibles en la Internet para su libre descarga y utilización.

1. CLÚSTER

En muchas ramas de las ciencias la complejidad de los problemas que se estudian requieren el acceso a una supercomputadora, para el procesamiento de datos, siendo estas máquinas tan poderosas que pueden desarrollar varios miles de millones de operaciones por segundo, algunos fabricantes como IBM se dieron la tarea de crear PCs con capacidades superiores a los convencional ofreciéndolos a precios exagerados, para solucionar el problema de costos y aumentar el poder de cómputo Thomas Sterling, Donald Becker y otros científicos de la NASA iniciaron el proyecto clúster y comenzaron a desarrollar el primer clúster de PCs, este tuvo una eficiencia de 70 megaflops (millones de operaciones de punto flotante por segundo). Ahora la tecnología clúster es cada vez más estudiada y utilizada en el mundo por estudiantes, científicos e investigadores, y se encuentra en constante evolución¹.

1.1 DEFINICIÓN DE CLÚSTER

Un clúster consiste en un grupo de computadoras de relativo bajo costo conectadas entre sí mediante un sistema de red de alta velocidad (Gigabit de fibra óptica por lo general) y un software que realiza la distribución de la carga de trabajo entre los equipos. Por lo general, este tipo de sistemas cuentan con un centro de almacenamiento de datos único. Su ventaja fundamental es la mejor relación coste/rendimiento y sus inconvenientes, la dificultad de programación y de mantenimiento.

¹ LIZARRA, Carlos. Clústeres de Linux [en línea]. Universidad de Sonora (México) [en línea]. <URL: http://clústeres.fisica.uson.mx/clústeres_de_Linux.htm>. [Consulta: Febrero 22 de 2009.]

La base, en términos de ingeniería, detrás del concepto formal de computación clúster, es la conocida Ley de Amdahl, la cual fue publicada en 1967 y se reconoce hoy como documento seminal de la computación paralela².

1.2 FUNCIONAMIENTO DE UN CLÚSTER.

La tecnología clúster está dividida en dos partes fundamentales, la primera es un sistema operativo robusto y estable elaborado para hacer modificaciones al kernel de Linux con compiladores y aplicaciones especiales que permitan que los programas que corran el SO puedan tomar las ventajas que brinda un clúster. Y la segunda es la interconexión de hardware entre los nodos del clúster, se han creado diversas interfaces para la comunicación pero las mas común es conectarse a través de una red Ethernet dedicada de alta velocidad para comunicarse, asignar tareas recursos etc.

La tecnología de Clústeres de Alto Rendimiento para Linux más conocida es el la tecnología Beowulf. Esta tecnología se desarrolló en la NASA y puede proporcionar potencial de cómputo del tipo de una supercomputadora utilizando computadoras personales sencillas.³

² PÉREZ, John. Desarrollo de un clúster computacional para la compilación de algoritmos en paralelo en el Observatorio Astronómico [en línea]. <http://www.usergioarboleda.edu.co/civilizar/revista6/Desarrollo_de_un_clúster_computacional_para_la_compilacion_de_algoritmos_en_paralelo_en_el_Observatorio_Astronomico.doc> [Consulta: Febrero 22 de 2009].

³ UNIVERSIDAD DE SONORA. Clúster de Linux [en línea]. <URL: http://clústeres.fisica.uson.mx/clústeres_de_Linux.htm > [Consulta: Febrero 22 de 2009]

1.3 DISTRIBUCIÓN DE LINUX Y PAQUETES PARA LA CREACIÓN DE CLÚSTER

El problema de interconectar computadoras para ejecutar cálculos en paralelo es antiguo pero recién ahora está tomando mayor importancia a nivel corporativo donde el esquema cliente servidor requiere servidores cada vez más potentes, básicamente en el tema de servicios Web y minería de datos.

A nivel de investigación se requiere gran capacidad de cómputo para aplicaciones que van desde simulaciones del medio ambiente, biomedicina, física de partículas, estadísticas, criptografía, etc. La potencia de las computadoras es cada vez mayor, pero al mismo tiempo siguen apareciendo retos computacionales más grandes.

Beowulf es el nombre que recibe en general las técnicas y procedimientos para usar Linux en modo de clúster computacional.

Hay librerías y API's al respecto, siendo las más conocidas las MPI (Message Passing Interface) y las PVM (Parallel Virtual Machine) para el desarrollo de aplicaciones en clúster. Una variación interesante es el uso de los módulos MOSIX, los mismos actúan básicamente a nivel del sistema operativo que a nivel de la aplicación, es decir, son parte del kernel, finalmente las aplicaciones se montan sobre ésta capa y corren de manera distribuida sobre una gran cantidad de equipos, sin necesidad de hacer cambios en las aplicaciones.

En cambio, en clústeres tipo Beowulf se requiere que cada aplicación o problema sea programado especialmente para ello con ayuda de las librerías MPI y/o PVM. Con MOSIX se obtiene una capa adicional sobre la cual los procesos corren en diversas computadoras, cada proceso es atendido por una computadora, se hace

un balance de carga derivando cada proceso nuevo a la computadora con menos carga. Las técnicas de clústeres Beowulf logran en cambio que un único proceso corra en simultáneo en varios sistemas.

Si se trata de resolver un problema que implique la ejecución de miles de procesos pequeños entonces tal vez sea más sencillo resolverlo con MOSIX, pero si se trata de un solo proceso que necesita una gran cantidad de cómputo entonces sería mejor el uso del esquema Beowulf lo cual implica el desarrollo de la aplicación respectiva y adaptarla a las librerías de clústering, es decir, las MPI o PVM.

Una herramienta bastante popular para el desarrollo de un clúster Beowulf es OSCAR:

- OSCAR es un paquete de RPM's, scripts de perl, librerías, herramientas y todo lo que es necesario para construir y usar un clúster Linux de tamaño modesto. El acrónimo de OSCAR es Open Source Clúster Application Resources.

Primero, OSCAR es un proyecto open source. Cada componente de OSCAR esta disponible bajo una de las bien conocidas licencias de software libre (GPL). El objetivo de este proyecto es hacer clústeres fáciles de construir, fáciles de mantener y fáciles de usar. En otras palabras, este paquete contiene los recursos necesarios para aplicar clústeres a problemas de cómputo de alto desempeño. OSCAR, no es un estándar nuevo.

No es un intento de abarrotar una aproximación particular de clústeres a la comunidad que utiliza el cómputo de alto desempeño. Más aun, OSCAR es una foto instantánea de las actuales y mejor conocidas prácticas de cómputo en clústeres. Los creadores de OSCAR han estudiado que trabajos de cómputo en

clústeres alrededor del mundo y los han recolectado en un simple paquete integrado.

El objetivo es que con OSCAR sea fácil para la gente reproducir satisfactoriamente las técnicas de clústeres en sus propios sitios. Los proyectos en Linux tienen esencialmente implementada la misma idea. El grupo creador de este proyecto le ha dado continuidad a la actualización de paquetes y haciendo seguras sus prácticas actuales para construir y utilizar clústeres. Otro aspecto único de este proyecto son sus orígenes con la colaboración de vendedores de hardware, de software y laboratorios de investigación.

- ROCKS CLÚSTER (originalmente llamado NPACI Rocks) es una distribución de Linux para clústeres de computadores de alto rendimiento. Rocks se basó inicialmente en la distribución Red Hat Linux, sin embargo las versiones más modernas de Rocks están basadas en CentOS, con un instalador anaconda modificado, que simplifica la instalación 'en masa' en muchos computadores. Rocks incluye muchas herramientas (tales como MPI) que no forman parte de CentOS pero son los componentes integrales que hacen un grupo de ordenadores en un clúster.

Rocks es una de las distribuciones más empleadas en el ámbito de clústeres, por su facilidad de instalación e incorporación de nuevos nodos. Otra de sus grandes facilidades es que incorpora gran cantidad de software para el mantenimiento y monitorización del clúster, lo que a su vez podría suponer en algunos casos una limitación.

En MOSIX se puede visualizar una gran cantidad de nodos como si fueran uno solo, administrados por una consola maestra. Existen varias distribuciones Linux que ya vienen con MOSIX integrado al Kernel y permiten tener en poco tiempo un

clúster. Mosix es un conjunto de herramientas administrativas que permite que un clúster o un Grid se vean como una única computadora con múltiples procesadores, al igual que en un sistema SMP.

Mosix es implementado como una capa de virtualización del sistema operativo que provee a los usuarios y sus aplicaciones un SSI con un entorno de tiempo de ejecución de Linux, permitiendo ejecutar aplicaciones sin la necesidad de modificarlas o enlazarlas con alguna librería especial. De esta forma, en un sistema Mosix los usuarios pueden crear sus aplicaciones basadas en la creación de múltiples procesos, los cuales a través de la búsqueda de recursos, en forma transparente, por parte de Mosix pueden automáticamente migrar, a los distintos nodos para mejorar la performance global sin cambiar el ambiente de tiempo de ejecución de los procesos migrados.

2. GRID

El concepto de Grid fue introducido oficialmente por Ian Foster y Carl Kesselman en 1998, aunque no era la primera vez que se hablaba de acceso por demanda de datos y servicios computacionales. El enfoque inicial de computación Grid, definía que los usuarios (clientes y aplicaciones) ganan acceso a recursos de cómputo, como procesadores, almacenamiento, información y aplicaciones, a medida que lo necesitan sin importancia o conocimiento de la localización de estos recursos o las tecnologías, hardware, sistemas operativos, detrás de éstos.

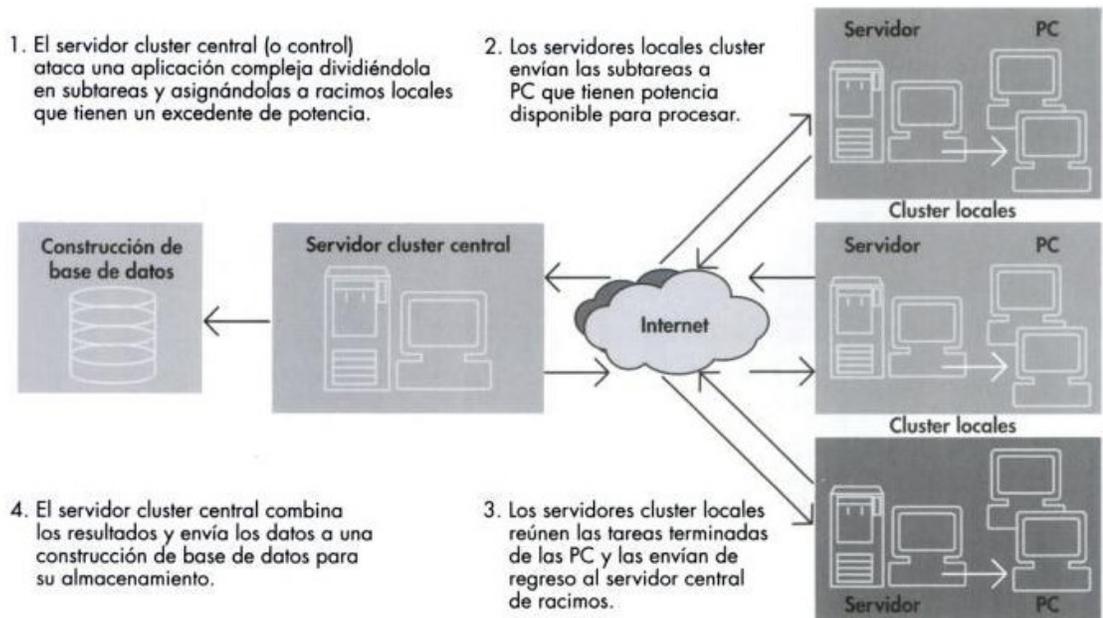
Desde el momento en el que los primeros computadores comenzaron a conectarse a Internet, surgió la idea de unir la potencia inutilizada de cada uno para abordar problemas a los que sólo podían enfrentarse las supercomputadoras pertenecientes a organizaciones gubernamentales, universidades o grandes multinacionales.

El término Grid se refiere a una infraestructura que permite la integración y el uso colectivo de ordenadores de alto rendimiento, redes y bases de datos que son propiedad y están administrados por diferentes instituciones. Puesto que la colaboración entre instituciones envuelve un intercambio de datos, o de tiempo de computación, el propósito del Grid es facilitar la integración de recursos computacionales. Universidades, laboratorios de investigación o empresas se asocian para formar Grid para lo cual utilizan algún tipo de software que implemente este concepto.

Llamamos Grid al sistema de computación distribuido que permite compartir recursos no centrados geográficamente para resolver problemas de gran escala. Los recursos compartidos pueden ser ordenadores (PC, estaciones de trabajo, supercomputadoras, PDA, portátiles, móviles, entre otros dispositivos), software,

datos e información, instrumentos especiales (radio, telescopios, entre otros) o personas/colaboradores.

Figura 1. Funcionamiento de un Grid de Cómputo



Fuente: El futuro de los Negocios. Lawrence J.

2.1 CARACTERÍSTICAS

- Alta disponibilidad. con la nueva funcionalidad, si un servidor falla, se reasignan los servicios en los servidores restantes;
- Reducción de costes: con esta arquitectura los servicios son gestionados por "granjas de recursos". Ya no es necesario disponer de "grandes servidores" y podremos hacer uso de componentes de bajo coste. Cada sistema puede ser configurado siguiendo el mismo patrón.

Se relaciona el concepto de Grid con la nueva generación de Internet. El nuevo protocolo de Internet IPv6 permitirá trabajar con una Internet más rápida y accesible. Una de las ideas clave en la superación de las limitaciones actuales de Internet Ipv4 es la aparición de nuevos niveles de servicio que harán uso de la nueva capacidad de la red para intercomunicar los ordenadores.

Este avance en la comunicación permitirá el avance de las ideas de Grid de Cómputo al utilizar como soporte la altísima conectividad de Internet.

2.2 TOPOLOGÍAS

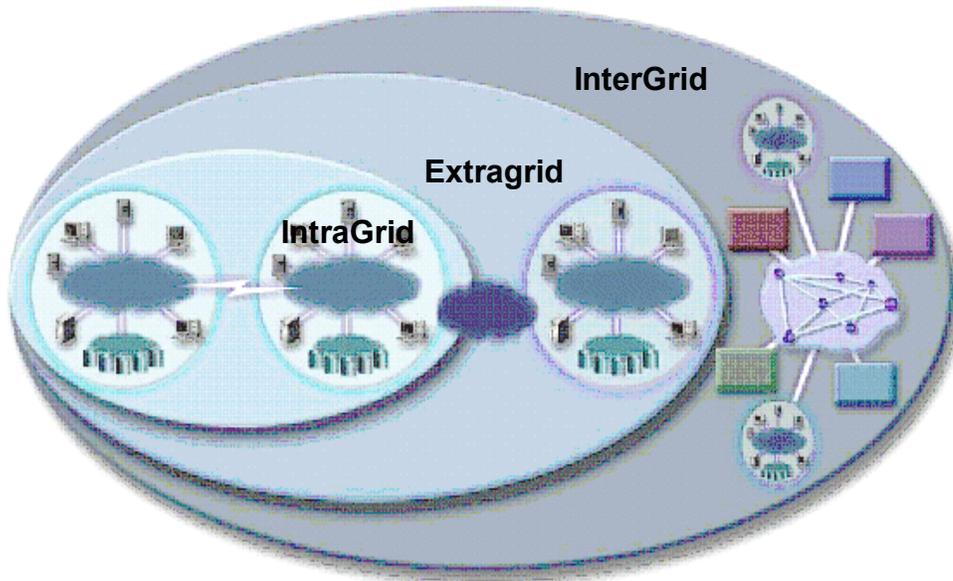
Las topologías Grid hacen referencia a los distintos tipos de Grid que podemos encontrar de acuerdo a los servicios que éste ofrezca y su arquitectura física.

Existen tres tipos de Grid, los cuales se explican a continuación. (Ver Figura 3)

2.2.1 Intragrid (InnerGrid). Esta arquitectura es la más sencilla de las tres topologías: existe en una sola organización (red privada) y posee los servicios básicos de un Grid. La organización puede tener un número de computadores que compartan un dominio de seguridad común. Las principales características de un IntraGrid son: un solo proveedor de seguridad, un ancho de banda alto en la red privada (LAN) y siempre disponible. Un IntraGrid provee una relativa fijación estática de los recursos computacionales y la habilidad de compartir fácilmente datos entre sistemas Grid⁴.

⁴ IBM Redbooks, Introduction to Grid Computing with Globus [en línea] <URL: <http://www.redbooks.ibm.com/abstracts/sg246895.html?Open>> [Consulta: Febrero 22 de 2009]

Figura 2. Intragrid, Extragrid e InterGrid.



Fuente: Introduction to Grid de Cómputo with Globus, IBM.

2.2.2 ExtraGrid (OuterGrid). Estos tipos de Grid también se conocen como “partner Grids”, porque se extienden fuera del dominio de administración de la empresa, conectándose a dos o más ExtraGrid. Se ocupan de compartir recursos, incluyendo a aquellos socios u organizaciones externas con las cuales ya se ha establecido una relación de negocios. Involucra generalmente más de un proveedor de seguridad, un nivel de complejidad de manejo alto y cuenta con conectividad remota (WAN).

2.2.3 InterGrid. La interGrid permite compartir recursos computacionales y datos de almacenamiento a través de una red pública, en colaboración con otras empresas. La información o datos de la InterGrid son en principio global y públicos. Así mismo las aplicaciones, deben ser modificadas por una audiencia global y deben cumplir con unos requisitos o características principales que las clasifiquen como tales. Estas características incluyen seguridad dispersa, múltiples organizaciones dentro de ella y conectividad remota presente (WAN). Una InterGrid necesita integración dinámica de aplicaciones, recursos y servicios⁵.

2.3 PARA QUÉ SIRVE EL GRID DE CÓMPUTO

El concepto de Grid de Cómputo da idea de una gran potencia de cálculo y almacenamiento, y parece un gran avance en las ciencias de la computación.

Existen además una multitud de aplicaciones reales que hacen uso de mini Grids, casi todas ellas están centradas en el campo de la investigación en el terreno de las ciencias físicas, médicas y del tratamiento de la información.

Se ha hablado mucho de lo que ofrece y de las posibilidades del Grid de Cómputo pero todavía no se ha definido de manera concreta el término.

Un Grid de cómputo es una infraestructura hardware y software que suministra al que la utiliza acceso seguro, consistente, penetrante y barato, a unas elevadas capacidades computacionales.

⁵ Joshy Joseph, Mark Ernest y Craig Fellenstein. Evolution of grid computing architecture and grid adoption models [en línea]. <URL: <http://www.research.ibm.com/journal/sj/434/joseph.html>> [Consulta: Febrero 22 de 2009].

El concepto infraestructura se utiliza porque un Grid es un conjunto de recursos (ciclos de CPU, datos, sensores, etc.), y todos esos recursos necesitan una interconexión hardware y un control software para que estén ensambladas en un Grid.

Esta infraestructura debe proporcionar a los usuarios un servicio seguro a todos los niveles: capacidad de cómputo, de integridad de datos, de seguridad de acceso, etc. El servicio debe ser consistente, basado en estándares, y de esta manera el acceso y las operaciones sobre el Grid estarán definidas por dichos estándares evitando la heterogeneidad. La idea de penetración no es tanto la posibilidad de acceder a cualquier recurso del Grid como que el Grid llega a cualquier sitio, de esta manera se asegura que una vez conectado desde cualquier punto puede extraer del Grid toda la potencia que requiera. Por último el acceso y uso del Grid debe tener un coste económico que le haga atractivo para que su utilización se universalice.

Para terminar de definir el concepto se desgrana cada uno de los posibles campos de aplicación de la tecnología Grid. El análisis lleva a definir cinco grandes áreas de trabajo determinadas por las necesidades de cálculo, espacio para el almacenamiento de los datos y tiempo de respuesta. Las áreas son:

2.3.1 Supercomputación distribuida. Dentro de esta área se encuentran aquellas aplicaciones cuyas necesidades es imposible satisfacer en un único nodo. Estas necesidades se producen en instantes de tiempo determinados y consumen muchos recursos, por lo que se dice que son puntuales e intensivas. Ejemplo de este tipo de aplicaciones son las simulaciones, las herramientas de cálculo numérico, los procesos de análisis de datos, la extracción de conocimiento de almacenes de datos, etc.

2.3.2 Sistemas distribuidos en tiempo real. En este tipo de aplicaciones se consideran aquellas que generan un flujo de datos a alta velocidad que debe ser analizado y procesado en tiempo real.

Ejemplo de este tipo de aplicaciones son los experimentos de física de alta energía, control remoto de equipos médicos de alta precisión y precio, todos los procesos de la denominada e-Medicine, el tratamiento de imágenes para la visión artificial, etc.

2.3.3 Proceso intensivo de datos. Esta área se centra en aquellas aplicaciones que hacen un uso intensivo del espacio de almacenamiento. Las necesidades de almacenamiento de este tipo de aplicaciones desbordan la capacidad de almacenamiento de un único nodo y los datos son distribuidos por todo el Grid. Además de los beneficios por el incremento de espacio, la distribución de los datos a lo largo del Grid permite el acceso a los mismos de forma distribuida.

Ejemplos de este tipo de aplicaciones son todos los sistemas gestores de bases de datos distribuidas.

2.3.4 Servicios puntuales. En esta área, se olvida el concepto de potencia de cálculo y capacidad de almacenamiento, para centrarse en recursos que una organización puede considerar como no necesarios. De esta manera el Grid ofrece a la organización esos recursos sin que la organización deba desarrollarlos por sí misma.

Ejemplos de este tipo de aplicaciones son aquellas que permiten acceder a hardware muy específico (equipos costosos de medida o de análisis de muestras) para la realización de labores a distancia.

2.3.5 Entornos virtuales de colaboración. Esta área está relacionada directamente con el concepto de Teleinmersión, de manera que se utilizan los enormes recursos computacionales del Grid y su naturaleza distribuida para generar entornos virtuales 3D distribuidos.

2.4 VENTAJAS Y DESVENTAJAS

2.4.1 Ventajas y requisitos. Los requisitos que debe cumplir cualquier Grid son:

- Los datos deben compartirse entre miles de usuarios con intereses distintos.
- Se deben enlazar los centros principales de supercomputación, no sólo los PC.
- Se debe asegurar que los datos sean accesibles en cualquier lugar y en cualquier momento.
- Debe armonizar las distintas políticas de gestión de muchos centros diferentes.
- Debe proporcionar seguridad.

Y los beneficios que se obtienen:

- Facilitar el pronto retorno de las inversiones.
- No necesitar de toda una nueva infraestructura para que funcione.
- Facilitar poder de computación / precio muy barato.
- Brindar el poder de un supercomputador.
- Utilizar software gratuito y usar código fuente abierto.
- Brindar transparencia para el usuario que participa en el Grid.
- Proporciona un mecanismo de colaboración transparente entre grupos dispersos, tanto científicos como comerciales.
- Posibilita el funcionamiento de aplicaciones a gran escala.
- Facilita el acceso a recursos distribuidos desde nuestros PC.

Estos beneficios tendrán repercusión en muchos campos:

- Medicina (imágenes, diagnosis y tratamiento).
- Bioinformática (estudios en genómica y proteómica).

- Ingeniería (diseño, simulación, análisis de fallos y acceso remoto a instrumentos de control).

Recursos naturales y medio ambiente (previsión meteorológica, observación del planeta, modelos y predicción de sistemas complejos).

2.4.2 Desventajas. No obstante, la computación Grid presenta algunos inconvenientes que deben solucionarse.

- Recursos heterogéneos: la computación Grid debe ser capaz de poder manejar cualquier tipo de recurso que maneje el sistema, sino resultará totalmente inútil.
- Descubrimiento, selección, reserva, asignación, gestión y monitorización de recursos son procesos que deben controlarse externamente y que influyen en el funcionamiento del Grid.
- Necesidad de desarrollo de aplicaciones para manejar el Grid, así como desarrollo de modelos eficientes de uso.
- Organizativos: dominios de administración, modelo de explotación y costes, política de seguridad.

La tecnología derivada del Grid abre un enorme abanico de posibilidades para el desarrollo de aplicaciones en muchos sectores. Por ejemplo: desarrollo científico y tecnológico, educación, sanidad, y administración pública.

2.5 MIDDLEWARE

El Middleware es un software puente que ofrece un conjunto de servicios que hacen posible el funcionamiento de aplicaciones distribuidas sobre plataformas heterogéneas y dispersos geográficamente; trabaja en las capas de aplicación y las capas inferiores (sistema operativo y red). Los middleware para Grid más utilizados son: Globus Toolkit, gLite y otros como Grid, OGSA-DAI, Gedeon, rocks.

2.5.1 Globus® Toolkit. Es un software de código abierto fundamental para la construcción de Grids que permite a los usuarios de una forma segura y homogénea compartir poder de cómputo, datos y otras herramientas sin importar la ubicación geográfica. Globus es uno de los middlewares más utilizados para la tecnología Grid y gracias a éste se han creado y dirigido diversos e importantes proyectos en el mundo.

Globus Toolkit ha crecido con la iniciativa Open Source, similar al proyecto GNU/Linux, lo que ha permitido que los usuarios lo utilicen con más entusiasmo y le hagan mejoras, para hacer a Globus cada vez más atractivo. Hasta el momento hay 3 versiones de Globus Toolkit: la 1.0 de 1998, la 2.0 de 2002 y la 4.0 con nuevos y mejores paquetes que las anteriores. Basta con estas 3 versiones para decir que Globus Toolkit es el estándar por defecto para la construcción de Grid, aunque no el único.

La Figura 3 muestra de manera general todos los servicios disponibles en Globus Toolkit versión 4, que incluye servicios de seguridad, administración de datos, ejecución de administración, servicios de información y tiempos de ejecución comunes.

Las herramientas que se han generado con Globus Toolkit y sus versiones, se han utilizado alternadamente para desarrollar una amplia gama de infraestructuras “Grid” y aplicaciones distribuidas. Se resumirán a continuación las características principales del último lanzamiento, la Web de servicios basada en GT4, que proporciona mejoras significativas sobre los lanzamientos anteriores en términos de robustez, funcionamiento, utilidad, documentación, rendimiento de los estándares, y funcionalidad.

El software por sí mismo, tiene un conjunto de librerías y programas direccionado a problemas comunes cuando se construyen sistemas de servicios y aplicaciones distribuidas. Provee una variedad de componentes y de capacidades, incluyendo las siguientes:

- Un conjunto de implementación de servicios enfocado a la administración de su infraestructura.
- Herramientas para la construcción de nuevos servicios Web en Java, C y Python.
- Una infraestructura de seguridad basada en poderosos estándares.
- APIs de clientes (en diferentes lenguajes) y programas de líneas de comandos para tener acceso a estos servicios y capacidades.
- Documentación detallada de sus variados componentes, interfaces, y cómo pueden ser usados para construir aplicaciones.

Estos componentes permiten un ecosistema rico de elementos y de herramientas de construcción o interoperación con los componentes GT y una amplia variedad de usos dentro muchos dominios.

GT4 hace un extenso uso de mecanismos de servicios Web para definir sus interfaces y la estructura de sus componentes. Los servicios Web proveen flexibilidad, extensibilidad, y mecanismos basados en XML, para descubrir e invocar servicios de red; adicionalmente su documentación orientada de protocolos está bien adaptada a un conjunto de interacciones que muchos afirman son preferibles para sistemas distribuidos robustos.

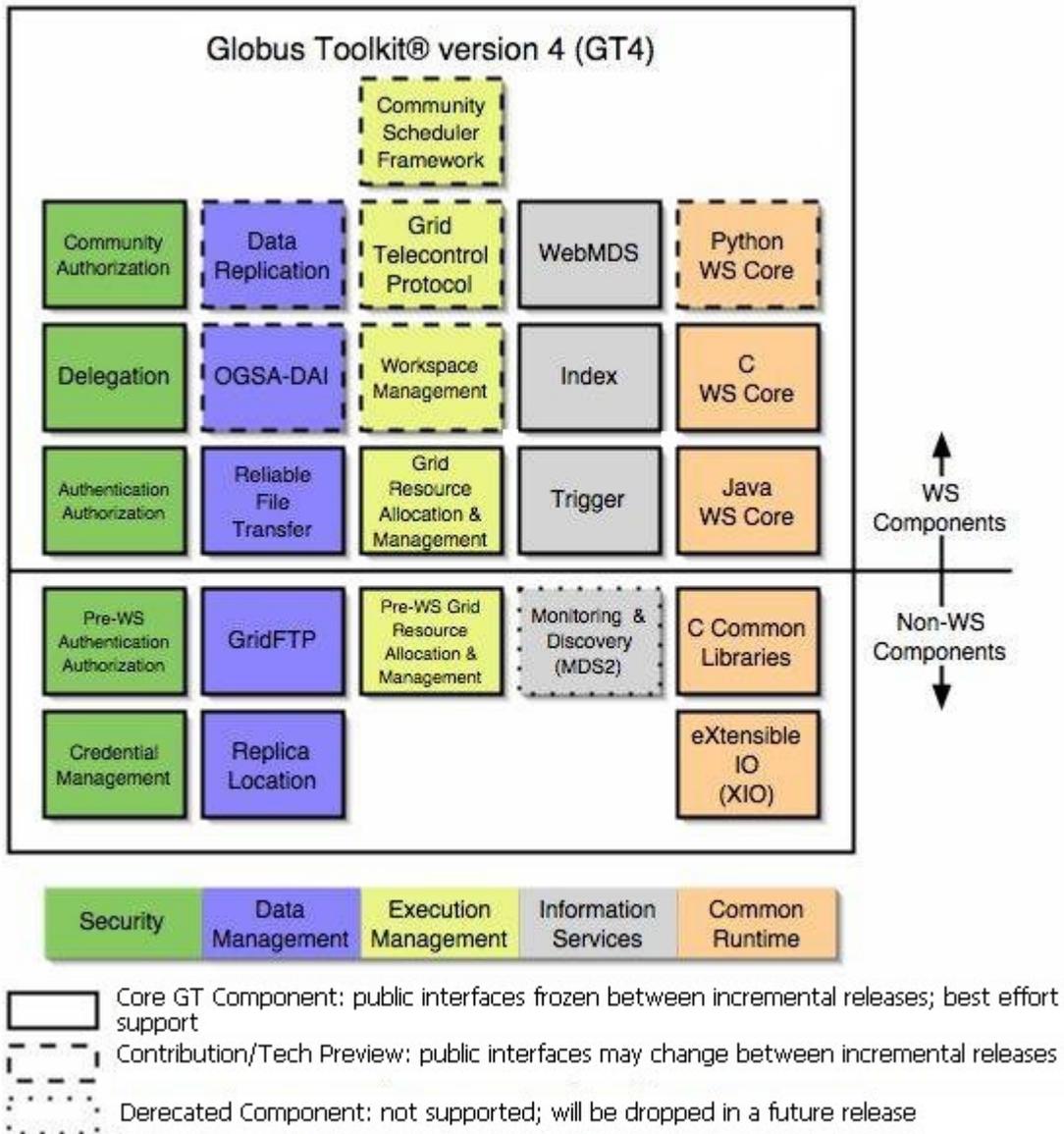
Estos mecanismos facilitan el desarrollo de arquitecturas orientadas a servicios, sistemas y aplicaciones estructuradas como servicios de comunicación, en donde se describen las interfaces de servicios, las operaciones son invocadas, el acceso asegurado, entre otros.

Las aplicaciones para usuarios finales se refieren únicamente a dominios específicos con operaciones tales como tasación de listas o análisis de una secuencia de genes: Este tipo de computación últimamente requiere la manipulación y administración de infraestructura de dispositivos físicos como: computadores, sistemas de almacenamiento e instrumentación.

GT4 proporciona un conjunto de servicios de Infraestructura Grid que implementa interfaces de administración computacional, almacenamiento, y otros recursos. En muchos despliegues o versiones de Globus (ej. TeraGrid, Open Science Grid, LHC Computing Grid, China Grid, APgrid), estos servicios son desarrollados para soportar un rango de aplicaciones, cada una de las cuales ejecuta su propio código de aplicaciones específicas que se encargan de realizar estos servicios.⁶

⁶ Foster Ian, Globus Toolkit Version 4: Software for Service-Oriented Systems [en línea]. <URL:<http://www.globus.org/alliance/events/sc05/GT4.pdf>> [Consulta: Febrero 22 de 2009].

Figura 3. Servicios Globus Toolkit versión 4 (GT4)



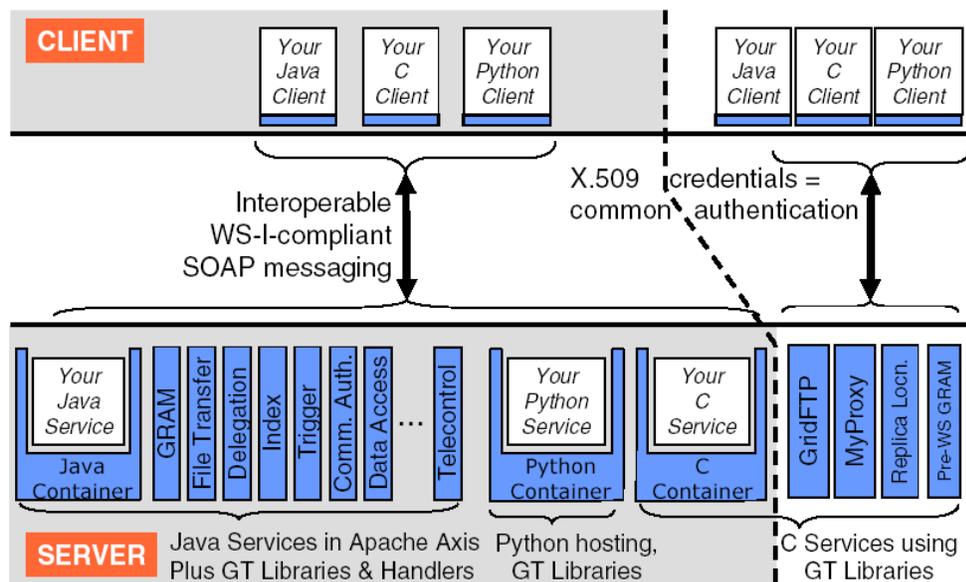
Fuente: Globus Toolkit Version 4: Software for Service-Oriented Systems.

En la Figura 4, se muestran varios aspectos de la Arquitectura esquemática de GT4 y varios de sus componentes. Destacando lo siguiente:

- Un conjunto de implementaciones de servicios (el fondo de la mitad de la Figura).
- Tres contenedores que se pueden utilizar para recibir desarrollos de usuarios escritos en Java, Python y C.
- Un sistema de librerías que permite que los programas en JAVA, Python y C invoquen operaciones tanto de GT4, como de servicios de desarrollo para usuarios.

Es importante observar que con GT4 los clientes pueden interactuar con diferentes servicios de la misma forma, lo cual facilita la construcción de sistemas complejos, facilita su interoperabilidad y anima la reutilización del código.

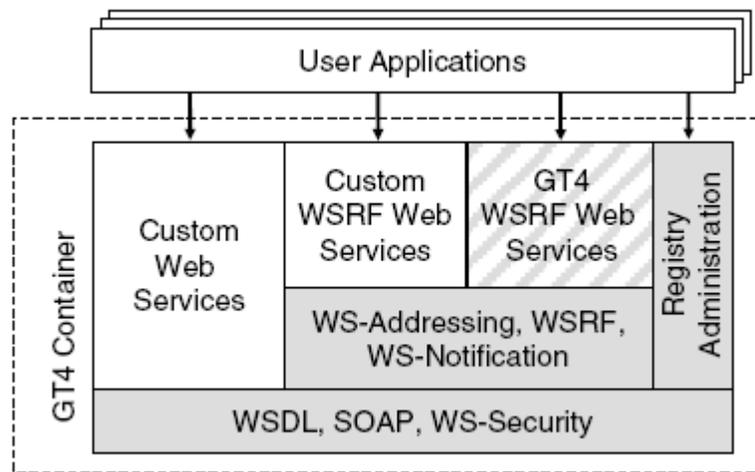
Figura 4. Arquitectura esquemática de GT4 y varios de sus componentes.



Fuente: Globus Toolkit Version 4: Software for Service-Oriented Systems.

Para finalizar, en la Figura 5, se muestran las capacidades o habilidades con que cuenta la nueva versión de Globus Toolkit (GT4) en cuanto a aplicaciones de usuario.

Figura 5. Capacidades del contenedor GT4



Fuente: Globus Toolkit Version 4: Software for Service-Oriented Systems.

2.5.1.1 Herramientas Web de Software libre para Globus Toolkit 4.0. Globus Toolkit cuenta con pocas herramientas Web de software libre con las que se facilita la implementación de nuevos modelos de desarrollo de aplicaciones en Grid ofreciendo una interfaz de programación estándar (Ver Anexo B). A continuación se muestra una tabla de comparación entre alguna de las herramientas más utilizadas para Globus TK.

Tabla 1. Comparación de herramientas Web de Software Libre de Globus Toolkit.

	GRAM	SAGA	DRMAA	MAGO
Servicios Web	Utiliza el modelo WRSF.	Utiliza el modelo OGSA.	No utiliza un entorno particular.	Utiliza el modelo OGSA.
Estructura	Componentes.	Módulos.	Cadenas simples de caracteres.	Multicapas.
Topología	Estilo concurrente.	Estilo rejillas		Topología malla, distribuida y paralela
Características	Capacidad de monitorear los trabajos. Cuenta con todos los API del gestor de trabajo.	Capacidad de gestionar los trabajos. Tiene gestor de trabajos pero no cuenta con todos los API.	Tiene una sesión dedicada a los trabajos (monitorear, gestionar, entre otros). Cuenta con todos los API del gestor de trabajo.	Invocación dinámica
Lenguaje de programación	Sola API para Java y C.	API para diferentes lenguajes de programación.	API para diferentes lenguajes de programación.	Solo cuenta con un lenguaje de programación.

Tabla 1. (Continuación).

	GRAM	SAGA	DRMAA	MAGO
--	-------------	-------------	--------------	-------------

Administración	Flexible.	Simple.	Oculto el gestor de recursos.	Estándar.
Infraestructura	Cuenta con autenticación u autorización simple.	Oculto detalles de infraestructura.		Simplifica detalles de infraestructura tecnológica.
Ficheros	Cuenta con transferencia de ficheros.	Fichero de sentencia específicas.	Maneja una sesión que permite independencia.	Se basa en requerimientos no funcionales.
Tiempo de ejecución	Bajo.	Promedio.	Alto.	Alto pero con facilidad de ejecución.
Base de datos	Guarda proceso de ejecución en una estructura lógica.	Base de datos para guardar proceso de ejecución.	Los procesos de ejecución se guardan en ficheros.	Utiliza dataGrid.
Interfaz	Interfaz de usuario con servicios Web. Interfaz en entorno distribuido.	Interfaz estándar.	Interfaz de aplicaciones directas. Interfaz homogénea y portable.	Interfaz heterogénea y autónoma
Seguridad	Usa clave única.	Maneja servicio de gestión de seguridad.		Garantiza seguridad y confiabilidad.

Fuente: Autores del proyecto.

3. PRODUCTO FINAL

3.1 CONSTRUCCIÓN DE LAS SALA-CLÚSTER

Los equipos de las sala-clúster implementadas cuentan con un solo sistemas operativos instalados en cada uno de ellos: el sistema operativo del clúster (Rocks).

Una sala-clúster típica cuenta con un servidor Frontend y dos o más nodos. Esta sala-clúster se utilizará para los fines básicos de una sala de informática (Windows y otras aplicaciones) y además se encontrará a disposición de proyectos que requieran alto poder de cómputo.

Para llevarse a cabo el proceso de construcción de las dos sala-clúster es de gran importancia tener en cuenta ciertos factores como:

- El tiempo que se requiere para realizar el montaje.
- La disponibilidad y tráfico de la red local (LAN) con la que se va a trabajar.
- La capacidad en hardware de los equipos, sobre todo del equipo designado como servidor Frontend (SFE).

3.2 INTERCONEXIÓN DE DOS SALA-CLÚSTER PARA CREAR UN INTRAGRID

El procedimiento realizado para el proceso de interconexión de dos sala-clúster internas, tiene como base la documentación encontrada en la Web acerca de otros proyectos en los cuales se han realizado montajes de Grid, pero con sistemas operativos diferentes a ROCKS, algunos de ellos como UBUNTU y FEDORA.

También existe documentación acerca del manejo, instalación y configuración de Globus Toolkit 4, disponible en el sitio Web de Globus: <http://www.globus.org/toolkit>.

Durante la configuración de Globus se deben llevar a cabo dos grandes pasos para el montaje y funcionamiento de la Intragrid.

- Conexión y comunicación de dos servidores en la red.
- Configuración y prueba de componentes de Globus.

3.3 CONEXIÓN Y COMUNICACIÓN DE DOS SERVIDORES EN LA RED

Como se mencionó anteriormente en el anteproyecto existen dos roles de usuario dentro de un entorno Grid: el Rol de Administrador y el Rol de Donante. Teniendo en cuenta esto, se muestran a continuación tres tipos de usuario que se clasifican dentro de estos dos Roles y que participan en el proceso de instalación y configuración de la Intragrid (Ver Anexo A). Estos son:

- Autoridad Certificante

- Usuario
- Servidor

Antes de empezar a construir el entorno Grid lo primero que debe hacerse es sincronizar el tiempo del sistema de todas las máquinas de la LAN con un servidor de tiempo NTP y verificar que los puertos 8080, 8443 y 2811 estén abiertos.

El siguiente paso es hacer la solicitud de un paquete de certificados desde el equipo designado como CA. Este paquete será usado en los nodos Grid.

Luego, desde el nodo Grid se copia la clave pública de la Autoridad Certificante a ese equipo y se instala el paquete CA.

Hecho esto, se procede a realizar la solicitud de un Certificado de Usuario el cual será aquel que tendrá acceso a la IntraGrid.

Finalmente se debe crear un mapa Grid entre un usuario de Grid autenticado y un usuario local del equipo CA.

Con este paso se culmina el proceso de conexión y comunicación de los dos servidores con lo que se tiene ya un usuario válido que puede acceder desde un nodo Grid al equipo CA a través de la LAN y por medio de Globus.

3.4 CONFIGURACIÓN Y PRUEBA DE COMPONENTES DE GLOBUS

Globus Toolkit es una colección de componentes de código abierto y a partir de la versión 4 este middleware brinda un soporte de servicios Web basado en implementaciones de muchos de sus componentes los cuales se clasifican dentro de cinco categorías:

- Componentes comunes en Tiempo de Ejecución
- Seguridad
- Administración de Datos
- Servicios de Información
- Gestión de Ejecución.

Dentro de estas categorías de servicios, encontramos 5 componentes importantes para el desarrollo y montaje de la IntraGrid que permiten probar y usar el entorno creado, administrando y permitiendo el paso de datos dentro de la Grid (Ver Anexo A).

- Monitoreo y descubrimiento de servicios

MDS4: los servicios MDS se encargan de la recolección, distribución, indexación, almacenamiento y procesamiento de alguna otra información sobre el estado de diversos recursos, servicios y configuraciones del sistema. La información recopilada se usa para descubrir nuevos recursos o servicios o para monitorear el estado del sistema. ⁷

Finalmente, es importante tener en cuenta que estos 5 componentes quedaron instalados en el clúster cuando se llevó a cabo la instalación de Java y el middleware Globus 4.0.4 y es necesario realizar la prueba de funcionamiento y configuración de cada uno de ellos

⁷ Jacob, Bart; Brown, Michael [et al]. Introduction to Grid Computing [Introducción a la computación Grid] [en línea]. Versión para IBM Redbooks. IBM, 27 diciembre 2005 [citado en 1 de mayo de 2009]. Un documento pdf: 3.3 MB. Texto en Ingles. Disponible en Internet: <http://www.redbooks.ibm.com/redbooks/SG246778/wwhelp/wwhimpl/js/html/wwhelp.htm>

4. CONCLUSIONES

Aunque las tecnologías Grid han madurado considerablemente durante los últimos años, las aplicaciones que pueden utilizar con efectividad de estas tecnologías están lejos de todas partes. Los avances en aplicaciones Grid simplemente no han seguido el ritmo de los avances en otros aspectos de la ciber-infraestructura distribuida, tales como los middleware Grid - tanto si se mide por el número de aplicaciones que puede utilizar fácilmente las muchas características avanzadas ofrecidas por la infraestructura distribuida o medido por el número de nuevas aplicaciones capaces de utilizar la infraestructura.

Un obstáculo clave para el desarrollo acelerado y despliegue de aplicaciones Grid es la escasez de abstracciones de alto nivel de programación de aplicación que salva la brecha entre las necesidades de las aplicaciones Grid y las capacidades ofrecidas por el middleware.

Otra forma en que el desarrollo de aplicaciones Grid se está retrasado por la falta de abstracciones adecuadas en el desarrollo de los primeros principios aplicaciones Grid - aplicaciones que pueden aprovechar la heterogeneidad de rendimiento y la respuesta dinámica inherente en la tabla a su favor - que han demostrado ser extremadamente difícil de aplicar desarrollo de la red se ha concentrado en el apoyo a la paralela legado y los códigos de las aplicaciones de racimo como una forma de garantizar la pertinencia científica.

El beneficio del paradigma de la cuadrícula, sin embargo, vendrá de desarrollo de nuevas aplicaciones que se no dependen del modelo homogéneo y relativamente

estática de rendimiento de los recursos heredados de paralelo o legados de racimo.

El procesamiento paralelo sobre redes de computadores es una excelente alternativa para aprovechar los recursos de máquinas y redes de datos para aumentar la capacidad de cómputo, con el propósito de resolver problemas complejos que demandan este potencial.

Como resultado del proyecto realizado se generaron las bases y recomendaciones para la implementación, configuración y administración de Intragrids, esenciales en el proceso de instalación de ExtraGrids o InterGrids y para el proyecto aprobado en el marco de la convocatoria interna a investigadores y docentes 2009 – 2010, Implantación de una Infraestructura computacional GRID en la Universidad Autónoma de Bucaramanga, para el procesamiento de informes XBRL.

En un ente universitario, la IntraGrid brinda además de los beneficios de poder de cómputo, beneficios económicos ya que los costos de implementación son relativamente bajos al utilizar máquinas que ya tiene la universidad.

Durante el proceso de instalación y configuración del Grid se identificaron algunos factores críticos que deben tenerse en cuenta y sobre los cuales se concluye lo siguiente:

- El tiempo de montaje del clúster es variable y depende de la carga que presente en determinados momentos la red a la que pertenecen los equipos

que van a conformar el clúster, lo cual indica que debe disponerse de tiempo suficiente para realizar los montajes, ya que aunque el proceso no es complicado, si conlleva un tiempo considerable.

- Los requisitos de hardware son indispensables para el montaje del clúster. Esto permite concluir que no todas las máquinas que disponga la institución pueden ser utilizadas para conformar un clúster de procesamiento. Existen requerimientos mínimos en cuanto a memoria principal de los nodos, espacio en disco duro, interfaces de red y procesador, entre otros, que deben ser atendidos antes de comenzar una instalación.
- La instalación de Globus Toolkit presento una serie de errores que se fueron solucionando y recopilando durante todo el proceso (Ver Anexo C).
- La sincronización del tiempo en el sistema es uno de los factores más importantes para la conformación de un Grid de cómputo toda vez que la generación y envío de certificados a los usuarios del Grid requieren de esta característica. Antes de empezar a construir el entorno Grid lo primero que debe hacerse es sincronizar el tiempo del sistema de todas las máquinas de la LAN con un servidor de tiempo NTP (Network Time Protocol) y verificar que los puertos 8080, 8443 y 2811 estén abiertos, ya que serán asignados a componentes de Globus. Para este proyecto, hubo necesidad de instalar un servidor NTP local.

El middleware (Globus T.K.) utilizado para la conformación del Grid está referenciado de manera escasa en Internet y muy generalizada, lo cual implicó adaptar su funcionalidad al sistema operativo Rocks, constituyéndose esto en uno de principales retos para el desarrollo del proyecto.

Existen herramientas que facilitan la ejecución y el desarrollo de aplicaciones Grid bajo determinadas condiciones, lo deseable sería disponer de un modelo general para el desarrollo de dichas aplicaciones.

5. RECOMENDACIONES PARA TRABAJOS FUTUROS

Para trabajos futuros, apoyándose en los resultados de la implementación de una infraestructura de computación Grid interna (IntraGrid), en la Universidad Autónoma de Bucaramanga, puede llevarse a cabo la realización del proyecto aprobado en el marco de la convocatoria interna a investigadores y docente 2009 – 2010, “Implantación de una Infraestructura computacional Grid en la Universidad Autónoma de Bucaramanga, para el procesamiento de informes XBRL”; a su vez, puede pensarse en la implementación de una infraestructura de computación Grid Externa (ExtraGrid), a través del uso de la RED NACIONAL DE TECNOLOGÍA AVANZADA (RENATA), o de la Internet (InterGrid), entre universidades e instituciones que necesiten computación de alto rendimiento para proyectos de ciencias e ingeniería.

Se deja entonces abierta la puerta y el punto de partida, de lo que podría constituirse en una iniciativa de Grid Nacional de Colombia, sin dejar de lado que la importancia radica no en la infraestructura o plataforma del Grid, sino en las aplicaciones y problemas que se pueden resolver con él.

Uno de los retos es poder establecer políticas, compromisos de uso y tarificación en ExtraGrid en donde varias instituciones colocan sus recursos de cómputo y redes para resolver problemas, que incluso pueden ser no comunes.

BIBLIOGRAFÍA

CONTRERAS, Gerardo. Clúster de alta disponibilidad [en línea].

<URL:http://www.puntogmx.net/wordpress/wpcontent/uploads/2007/04/clúster_ha.pdf> [Consulta: Febrero 22 de 2009]

FOSTER Ian. Globus Toolkit Versión 4: Software for Service-Oriented Systems [en línea]. <URL:<http://www.globus.org/alliance/events/sc05/GT4.pdf>> [Consulta: Febrero 22 de 2009]

Grid Café. Grids internacionales. [En línea] <URL:http://www.gridcafe.org/es-globus-toolkit_ES.html> [Consulta: 01/09/2009]

Grid Computing. Planet.News. [En línea]

<URL:<http://www.gridcomputingplanet.com/news/>> [Consulta: 02/09/2009]

Interfaz Cloud para Servicios de Gestión de Recursos. [En línea]

<URL:<http://pisis.unalmed.edu.co/3CCC/pdf/86.pdf>> [Consulta: Septiembre 2 de 2009]

LIZARRA, Carlos. Clústeres de Linux [en línea]. Universidad de Sonora (México) [en línea]. <URL: http://clústeres.fisica.uson.mx/clústeres_de_Linux.htm> [Consulta: Febrero 22 de 2009.]

Magos-Workflow: Cooperación entre Aplicaciones Grid, Autónomas y Heterogéneas. [En línea] <URL:<http://maomorales.com/src/Cloud.pdf>> [Consulta: Agosto 20 de 2009]

MERKEY, Phil Beowulf History [en línea].
<URL:<http://www.beowulf.org/overview/history.html>> [Consulta: Febrero 22 de 2009]

MARTÍNEZ, Víctor. Tecnología Grid para la Detección de Cáncer de Mama y Cuello Uterino por Medio de Procesamiento de Imágenes En Clarc Conferencia latinoamericana de computación de alto rendimiento (2007: Santa Marta). Bucaramanga. UIS, Colombia. 7.

MORENO HERNÁNDEZ, Diana. Y VECINO VECINO, Lina. Configuración e interconexión de dos salas clúster tipo beowulf en la universidad autónoma de Bucaramanga - UNAB utilizando la arquitectura Rocks. Bucaramanga, 2007. 59 p. Trabajo de grado (Ingeniero de Sistemas). Universidad Autónoma de Bucaramanga. Facultad de ingeniería de sistemas.

OSPINA SANJUÁN, Álvaro y CARVAJAL SALAMANCA, Roberto. Pruebas de conectividad y procesamiento distribuido con clústeres beowulf universitarios usando la red nacional de tecnología avanzada (RENATA). Bucaramanga, 2008. 53 p. Trabajo de grado (Maestría en software libre). Universidad Autónoma de Bucaramanga. Área de software libre.

PÉREZ, John. Desarrollo de un clúster computacional para la compilación de algoritmos en paralelo en el Observatorio Astronómico [en línea].

<URL:http://www.usergioarboleda.edu.co/civilizar/revista6/Desarrollo_de_un_clúster_computacional_para_la_compilacion_de_algoritmos_en_paralelo_en_el_Observatorio_Astronomico.doc > [Consulta: Febrero 22 de 2009].

Programación de aplicaciones paralelas con MPI (Message Passing Interface).
José Miguel Alonso, Facultad de Informática UPV/EHU [en línea].

<URL:<http://www.sc.ehu.es/acwmialj/edumat/mpl.pdf> > [Consulta: Febrero 22 de 2009.]

PVM, Departamento de Matemática Aplicada y Computación [en línea].

<URL:<http://www.mac.cie.uva.es/pvm.html>.> [Consulta: Febrero 22 de 2009.]

SETI@Home. About SETI@Home. [En línea].

<URL:http://setiathome.berkeley.edu/sah_about.php> [Consulta: Febrero 22 de 2009]

UNIVERSIDAD DE ANTIOQUIA. Primer clúster de computación en la Facultad de Ingeniería [en línea].

<URL:<http://ingenieria.udea.edu.co/portal/ingeniemos/versionimpresa/200706/pag6.pdf>> [Consulta: Febrero 22 de 2009.]

UNIVERSIDAD DE MÁLAGA. Arquitectura Clúster. Clústeres. Arquitecturas Distribuidas [en línea].

<URL:<http://www.ac.uma.es/educacion/cursos/informatica/ArqDist/pdfs/04-Clústeres.pdf>> [Consulta: Febrero 22 de 2009]

UNIVERSIDAD DE SONORA. Clúster de Linux [en línea].

<URL:http://clústeres.fisica.uson.mx/clústeres_de_Linux.htm > [Consulta: Febrero 22 de 2009.]

UNIVERSITAT OBERTA DE CATALUNYA. Administración avanzada de GNU/Linux, Joseph Jorba Esteve y Remo Suppi Boldrito [en línea].

<URL:http://www.uoc.edu/posgrado/matricula_abierta/img/871.pdf> [Consulta: Febrero 22 de 2009.]