

**UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS**

“HERRAMIENTAS DE COMPUTACIÓN GRID”
www.bdigital.ula.ve

AUTOR: JOSÉ IGNACIO PÉREZ CALDERAS

TUTOR: HERBERT HOEGER

Mérida – Venezuela

Diciembre 2005

AGRADECIMIENTOS

- A mi profesor tutor Herbert Hoeger, por ser excelente guía y por su apoyo y colaboración en la realización de este trabajo de investigación.
- A los ingenieros Vanessa Hamar, Víctor Mendoza y Yubiryn Ramírez por su colaboración.
- Al Consejo de Computación Académica de la Universidad de los Andes por permitirme trabajar en sus laboratorios de computación.

www.bdigital.ula.ve

RESUMEN

En este trabajo de investigación se reúne de forma detallada y concisa información referente sobre la computación Grid y sus principales herramientas disponibles para implantarla. La computación Grid es una tecnología reciente propuesta por Ian Foster y Carl Kesselman a mediados de los años noventa. Está enfocada principalmente en el acceso remoto de recursos computacionales compartidos. La idea es permitir a las personas en organizaciones e instituciones compartir el poder cómputo, base de datos, instrumentos y otras herramientas en línea de forma segura sin sacrificar la autonomía local, utilizando la Internet como una plataforma de servicios de computación.

Las herramientas de la computación Grid se dividen en varios tipos y son las encargadas de realizar diferentes trabajos en un Grid. El presente trabajo se enfoca especialmente en la principal herramienta de software para la computación Grid: el Globus Toolkit. Esta herramienta es un software de código abierto usada para construir sistemas y aplicaciones Grid que proporciona muchos de los servicios básicos dentro de una plataforma Grid como: descubrimiento, acceso, manejo y monitoreo de recursos, manejo de datos, comunicación, portabilidad y seguridad.

ÍNDICE DE CONTENIDO

	Página
PORTADA	i
AGRADECIMIENTOS	ii
RESUMEN	iii
LISTAS DE FIGURAS	xi
INTRODUCCIÓN	1
CAPITULO I	
<u>I DEFINICIÓN DEL PROBLEMA Y OBJETIVOS DE PROYECTO</u>	3
I.1 Definición del Problema	3
I.2 Objetivos del Proyecto	3
I.2.1 Objetivo General	3
I.2.2 Objetivos Específicos	3
CAPITULO II	
<u>II MARCO TEÓRICO</u>	4
II.1 Conceptos sobre algunos tipos de computación	4
II.1.1 Supercomputación	4
II.1.2 Computación de Alto Rendimiento	4
II.1.3 Computación Paralela	4
II.1.4 Computación Distribuida	5
II.1.5 Metacomputación	5
II.1.6 Computación Punto a Punto	6

II.2 Computación Grid	6
II.2.1 Definición	6
II.2.2 Historia de la computación Grid	9
II.2.3 Antepasados de la computación Grid	11
II.2.4 Tecnologías claves en el desarrollo de la computación Grid	12
II.2.5 Características de un sistema de computación Grid	13
II.2.6 Propiedades de la computación Grid	14
II.2.7 Elementos necesarios para la computación Grid	15
II.2.8 Analogía de la computación Grid con el Grid de la red eléctrica	15
II.2.9 ¿Por qué la computación Grid?	17
II.2.10 ¿Cómo será la computación Grid?	18
II.2.11 ¿En qué estado se encuentra la computación Grid actualmente?	20
II.2.12 Funcionamiento de la computación Grid	21
II.2.13 Arquitectura de la computación Grid	26
II.2.14 Principales áreas de aplicación de la computación Grid	27
II.2.15 Ventajas y Desventajas de la Computación Grid	28

CAPITULO III

<u>III PRINCIPALES HERRAMIENTAS DE COMPUTACIÓN GRID</u>	30
III.1 Herramientas de computación Grid	30
III.2 Condor	30
III.2.1 Definición de Condor	31

III.2.2 ¿Cómo funciona Condor?	32
III.2.3 Ventajas de Condor	33
III.3 Globus Toolkit	34
III.3.1 Definición de Globus Toolkit	35
III.3.2 ¿Cómo funciona Globus Toolkit?	36
III.4 GridPort	36
III.4.1 Definición de GridPort	37
III.4.2 ¿Cómo funciona Gridport?	38
III.5 Portable Batch System	38
III.5.1 Definición de PBS	39
III.5.2 ¿Cómo funciona PBS?	40
III.5.3 Ventajas de PBS	40
III.6 Sun Grid Engine	41
III.6.1 Definición de Sun Grid Engine	42
III.6.2 ¿Cómo funciona Sun Grid Engine?	42
III.6.3 Ventajas de Sun Grid Engine	43

CAPITULO IV

<u>IV HERRAMIENTA DE SOFTWARE GLOBUS TOOLKIT</u>	44
IV.1 Globus Toolkit 4.0	44
IV.2 Objetivos de Globus Toolkit 4.0	45
IV.3 Servicios de Globus Toolkit 4.0	45
IV.4 Componentes más importantes de Globus Toolkit 4.0	46
IV.5 Arquitectura de Globus Toolkit 4.0	47

IV.6 Componentes de Globus Toolkit 4.0	48
IV.6.1 Seguridad	49
IV.6.1.1 Manejo de Credenciales	49
IV.6.1.2 Autorización y Autenticación de Servicios Pre – Web	50
IV.6.1.3 Autorización y Autenticación de Servicios Web	51
IV.6.1.4 Servicio de Delegación	52
IV.6.1.5 Servicio de Autorización de Comunidad	53
IV.6.2 Manejo de Datos	53
IV.6.2.1 Servicio de Localización de Réplica	53
IV.6.2.2 Protocolo de Transferencia de Archivo Grid	54
IV.6.2.3 Transferencia de Archivo Seguro	55
IV.6.2.4 Arquitectura de Servicios Grid Abiertos - Integración y Acceso a Datos	57
IV.6.2.5 Servicio de Réplica de Datos	57
IV.6.3 Manejo de Ejecución	58
IV.6.3.1 Manejo de Asignación de Recursos Grid de Servicios Pre - Web	58
IV.6.3.2 Manejo de Asignación de Recursos Grid de Servicios Web	59
IV.6.3.3 Manejo de Espacio de Trabajo	59
IV.6.3.4 Protocolo de Control de Teleoperaciones Globus	60
IV.6.3.5 Estructura Planificadora de Comunidad	61
IV.6.4 Servicios de Información	61

IV.6.4.1 Sistema de Descubrimiento y Monitoreo de Servicios	
Pre – Web	61
IV.6.4.2 Servicio Disparador	62
IV.6.4.3 Servicio Índice	62
IV.6.4.4 Sistema de Descubrimiento y Monitoreo de Servicios	63
Web	
IV.6.5 Componentes de Ejecución Común	63
IV.6.5.1 Extensible Entrada y Salida	63
IV.6.5.2 Librerías Comunes de C	64
IV.6.5.3 Centro de Servicios Web Java	64
IV.6.5.4 Centro de Servicios Web C	65
IV.6.5.5 Centro de Servicios Web Phyton	65
IV.7 Ventajas y Desventajas de Globus Toolkit 4.0	66
IV.7.1 Ventajas	66
IV.7.2 Desventajas	67

CAPITULO V

V INSTALACIÓN Y CONFIGURACIÓN BÁSICA DE GLOBUS

<u>TOOLKIT 4.0</u>	68
V.1 Requerimientos para la instalación de Globus Toolkit 4.0	68
V.1.1 Requerimientos de hardware	68
V.1.2 Requerimientos de software	68
V.2 Instalación de Globus Toolkit 4.0	70
V.2.1 Crear la cuenta del usuario normal	70

V.2.2 Crear la cuenta del usuario y grupo globus	71
V.2.3 Crear el directorio donde se instalará Globus Toolkit 4.0	71
V.2.4 Definir variables de entorno	72
V.2.4.1 Definir la variable GLOBUS_LOCATION	72
V.2.4.2 Definir la variable JAVA_HOME	73
V.2.5 Descargar el instalador de Globus Toolkit 4.0	74
V.2.6 Desempaquetar el instalador de Globus Toolkit 4.0	74
V.2.7 Proceso final de la instalación de Globus Toolkit 4.0	74
V.3 Configuración básica de Globus Toolkit 4.0	75
V.3.1 Configuración básica de la seguridad de Globus Toolkit 4.0	76
V.3.1.1 Establecer variables de entorno	76
V.3.1.2 Obtener los certificados	77
V.3.1.3 Configuración GSI	81
V.3.1.4 Solicitud del certificado para la computadora	82
V.3.1.5 Firma del certificado de la computadora	82
V.3.1.6 Solicitud del certificado de usuario	83
V.3.1.7 Firma del certificado de usuario	84
V.3.1.8 Hacer accesible los certificados de la computadora al container	85
V.3.1.9 Autorización del usuario	87
V.3.1.10 Verificando el inicio del Proxy	89
V.3.2 Sobre la configuración básica realizada de Globus Toolkit 4.0	90

CONCLUSIONES Y RECOMENDACIONES	92
REFERENCIAS BIBLIOGRÁFICAS	96
REFERENCIAS RECOMENDADAS	99
GLOSARIO	101

www.bdigital.ula.ve

LISTA DE FIGURAS

	Página
Figura IV.1 Componentes de Globus Toolkit 4.0	48
Figura V.1 Información de la computadora	78
Figura V.2 Dirección de correo electrónico	79
Figura V.3 Tiempo de vida del certificado	80
Figura V.4 Clave o contraseña	80
Figura V.5 Configuración GSI	81
Figura V.6 Archivos con extensión pem	85
Figura V.7 Nuevos archivos con extensión pem	86
Figura V.8 Información sobre la instalación realizada	87
Figura V.9 Crear archivo grid-mapfile	88
Figura V.10 Archivo grid-mapfile	89
Figura V.11 Ejecución del Proxy	90

INTRODUCCIÓN

Hoy en día es prácticamente imposible hacer ciencia sin el uso de las computadoras. Los proyectos científicos que se vienen desarrollando en los últimos años abordan objetivos cada vez más ambiciosos y complicados que requieren la resolución de problemas computacionales complejos, tanto por el volumen de los cálculos a realizar como por el tamaño y complejidad de la información a procesar. Una sola computadora, un clusters de computadoras o una supercomputadora, no es suficiente en algunos casos para los cálculos y procesamientos de datos científicos que se quieren hacer y, a pesar que las computadoras son cada vez más rápidas, no son lo suficientemente rápidas para ir al ritmo de las demandas de los científicos.

Como resultado de esto, los científicos frecuentemente enfrentan situaciones cada vez más difíciles, muy costosas y en algunas ocasiones imposible de resolver con la tecnología de computación actual. La computación Grid surgió como una solución a esto.

Esta tecnología es una infraestructura de hardware y software que pretende utilizar Internet como una plataforma de servicios de computación y no sólo como fuente de información, como es en la actualidad. Proporcionará un gran poder de cómputo y espacio de almacenamiento para guardar de forma segura cualquier tipo de información. Permitirá a las personas en organizaciones e instituciones compartir el poder cómputo, base de datos, instrumentos y otras herramientas en línea de forma segura sin sacrificar la autonomía local.

Este trabajo de investigación pretende familiarizar a las personas con la computación Grid y sus herramientas disponibles para implementarla, realizando un análisis a través de una evaluación y de un estudio sobre estos. Se proporcionará información en español sobre los objetivos, alcances, limitaciones, ventajas y desventajas de la computación Grid y sus herramientas.

Se buscará información a través de Internet y referencias bibliográficas sobre la computación Grid y sus principales herramientas.

Se documentará de forma detallada lo referente a la principal herramienta de computación Grid (Globus Toolkit).

Se instalará, configurará y se probará su principal herramienta (Globus Toolkit).

Se proporcionarán conclusiones y recomendaciones sobre la computación Grid y sus principales herramientas.

Se proporcionará las referencias bibliográficas utilizadas, las direcciones de los sitios Web que se utilizaron y una lista de sitios Web recomendados por ser particularmente interesantes.

Por ultimo, se proporcionará un glosario sobre los términos utilizados en este trabajo de investigación.

CAPÍTULO I

I. DEFINICIÓN DEL PROBLEMA Y OBJETIVOS DEL PROYECTO

I.1 Definición del Problema

Familiarizarse con la computación Grid y las herramientas disponibles para implementarla. Mediante su evaluación y estudio, realizar un análisis de las mismas con el fin de indicar sus objetivos, alcances y limitaciones.

I.2 Objetivos del Proyecto

I.2.1 Objetivo General

- Estudiar la Computación Grid y sus principales Herramientas.

I.2.2 Objetivos Específicos

- Buscar información a través de Internet y referencias bibliográficas referente a la computación Grid y sus principales herramientas.
- Documentar, en español, lo referente a la herramienta Globus Toolkit 4.0.
- Instalar, configurar y probar su principal herramienta (Globus Toolkit 4.0).
- Dar conclusiones y recomendaciones sobre la computación Grid y sus principales herramientas.

CAPÍTULO II

II. MARCO TEÓRICO

II.1 Conceptos sobre algunos tipos de computación

II.1.1 Supercomputación

La supercomputación fue definida en los años ochenta para hacer referencias a la utilización de computadoras de gran tamaño, muy rápidas y extremadamente caras utilizadas para realizar cálculos complejos o sofisticados que no se podían realizar en computadoras comunes. Estas supercomputadoras estaban equipadas con procesadores que eran capaces de ejecutar una cantidad de operaciones por segundo mucho mayor que cualquier otra computadora común [1].

II.1.2 Computación de Alto Rendimiento

Proporciona un mejor y mayor rendimiento de cómputo que el que se puede obtener mediante la utilización de computadoras personales comunes y se obtenía el poder de cómputo a través de supercomputadoras. Hoy en día mucha de la computación de alto rendimiento se realiza en clusters de computadoras [2].

II.1.3 Computación Paralela

La computación paralela es el cómputo en varias unidades de procesamiento, es decir, la utilización de múltiples procesadores para resolver una tarea común. Se divide el problema en trozos más pequeños y se asignan trozos a los procesadores [3].

El atributo principal de la computación paralela es que sus aplicaciones han sido escritas para usar algoritmos que pueden particionar o dividir los trabajos y así procesarlos por separado.

II.1.4 Computación Distribuida

La computación distribuida es una forma de procesamiento de la información en la que el trabajo se realiza en computadoras separadas y conectadas a través de una red de comunicaciones de forma organizada para trabajar en objetivos comunes [1 y 3].

La computación distribuida aprovecha las redes de comunicación para unir múltiples equipos y sistemas para lograr sus objetivos [3]. Esto lo hace de forma transparente al usuario, es decir, el usuario no se da cuenta de que hay múltiples computadoras realizando el trabajo.

II.1.5 Metacomputación

La Metacomputación es un nombre que se da a un tipo particular de computación distribuida. Fue muy popular a comienzos de los años noventa, la cual se encargaba de conectar los centros de supercomputación con lo que era en el momento las redes de alta velocidad [2].

II.1.6 Computación Punto a Punto

Es una rama de la computación actualmente muy popular y es usada para transferir y compartir archivos y datos sin tener que pasar por un servidor central directamente [2].

II.2 Computación Grid

II.2.1 Definición

La computación Grid es una nueva tecnología que incluye software y hardware como paquetes de software diseñados por diferentes compañías e instituciones, entre ellos se encuentra el middleware y los programas planificadores de tareas, y hardware como computadoras, supercomputadoras, servidores, clusters de computadoras, sistemas de almacenamientos de datos y redes de altas velocidades.

Esta tecnología pretende utilizar Internet como una plataforma de servicios de computación y no sólo como fuente de información como es en la actualidad. Incluirá gran poder de cálculo y de cómputo, gran capacidad de almacenamiento de datos y la eliminación del factor distancia de las comunicaciones entre sistemas de computación que se encuentran alejados geográficamente unos de otros, ya que esta tecnología además de proveer redes de alta velocidades también tendrá un gran ancho de banda en sus comunicaciones.

La computación Grid fue propuesta por Ian Foster y Carl Kesselman a mediados de los años noventa. La idea de la computación Grid está enfocada principalmente en el acceso remoto de recursos computacionales compartidos, es decir, recursos

computacionales dispersos geográficamente que son compartidos entre diferentes organizaciones e instituciones alrededor del mundo.

La computación Grid es una infraestructura de hardware y software que provee acceso seguro, consistente, transparente y de bajo costo a las capacidades computacionales de última tecnología [4].

La computación Grid es computación distribuida llevada a un nivel multiorganizacional. Se distingue de la computación distribuida por enfocarse en compartir recursos a gran escala, proporciona aplicaciones innovadoras y de alto rendimiento [3].

La computación Grid es equivalente a un gran computador virtual. Incorpora recursos computacionales, incrementa la cantidad y la capacidad de recursos disponibles. Los integrantes del Grid aportan y comparten recursos. Crea la ilusión de que se está manejando un único computador fácil y transparente de usar [5].

La computación Grid se define simplemente como computación distribuida tomada al próximo nivel evolutivo. La meta es crear la ilusión de una simple, sin embargo grande y poderosa computadora virtual, manejando una gran colección de sistemas heterogéneos conectados que comparten varias combinaciones de recursos [6].

La computación Grid ha sido llamada como la análoga de la Web pero para proveer recursos de computación. Se define como una capa de software y de servicios que se localiza sobre los sistemas operativos y enlaza a diferentes sistemas permitiendo compartir diferentes tipos de recursos computacionales, agregando la capacidad de compartir de manera amplia el poder computacional, las aplicaciones y espacio de almacenamiento a las habilidades actuales de la Web para compartir texto y multimedia, pueden resolverse los problemas que requieren uso intensivo de recursos computacionales, los equipos pueden trabajar más de lo esperado alcanzando sus propios límites y la colaboración puede llegar a ser más intensa [7].

Considerando que la Web es un servicio para compartir la información a través de la Internet, la computación Grid es un servicio para compartir poder de cómputo y capacidad de almacenamiento de datos de la Internet. La computación Grid va más allá de una simple comunicación entre computadoras, y su objetivo finalmente es convertirse en la red global de computadoras dentro de un enorme recurso computacional [2].

La computación Grid puede ser definida como la aplicación simultánea de recursos de muchas computadoras en una red a la solución de un problema, un problema que requiere un gran número de ciclos de procesamiento o acceso a muchas cantidades de datos [8].

La computación Grid consiste en compartir recursos heterogéneos (basados en distintas plataformas, arquitecturas de equipos, programas y lenguajes de programación), situados en diferentes lugares y pertenecientes a diferentes dominios de administración sobre una red que utiliza estándares abiertos [9].

La computación Grid está enfocada en el compartimiento de recursos y en la solución de problemas en organizaciones virtuales dinámicas. La clave es la habilidad para negociar los recursos compartidos entre los proveedores y consumidores y luego usar los recursos reunidos para algún propósito en particular [10].

Esta nueva tecnología llamada computación Grid es un nuevo paradigma de la computación distribuida propuesto en los trabajos publicados por Ian Foster y Carl Kelsseman a mediados de los años noventa. Es llamada por muchos el futuro de la computación y les permitirá a las personas acceder a un enorme recurso computacional sólo con tener un navegador Web con conexión a Internet [11].

II.2.2 Historia de la computación Grid [2]

En la historia de la computación Grid muchas de las ideas básicas han estado de una forma u otra ligada a la historia de la computación, ya que la idea original y principal de la computación Grid es compartir el poder de la computación. Esto se remonta a los años sesenta y setenta cuando la ciencia de la computación estaba en sus comienzos y el compartir el poder computacional era esencial, ya que en

ese tiempo la computación fue dominada por los enormes computadores centrales que abarcaban hasta una habitación por su gran tamaño y hubo la necesidad de compartirlos entre organizaciones e instituciones. Aplicando esta idea original es lo que hoy se conoce como la computación Grid.

También en el año de 1965 los diseñadores del sistema operativo Multics antepasado del sistema operativo Unix, presentaron un enfoque de la computación como una unidad, como se enfoca la computación Grid actualmente.

Esto nos indica que las ideas básicas de la computación Grid siempre han estado presente a lo largo de toda la historia de la computación. Los que reunieron y trabajaron en estas ideas fueron Ian Foster del laboratorio Nacional de Argonne y Carl Kesselman de la Universidad del Sur de California, quienes le dieron un nombre a esta nueva tecnología. Iniciaron sus trabajos en el laboratorio Nacional de Argonne en el año de 1997 con el trabajo de “Construcción de un Grid computacional”, seguido en el año de 1998 con el trabajo de “El Grid: Modelo para una nueva infraestructura de Computación”.

También Ian Foster y Carl Kelsseman en el año de 1997 publicaron un trabajo llamado “Globus: un Toolkit de la infraestructura de la metacomputación” relacionado con el Globus Toolkit que hoy en día es el corazón de la computación Grid y en un futuro seguramente será el estándar que permitirá la comunicación en la computación Grid.

II.2.3 Antepasados de la computación Grid [2]

Se puede decir con certeza que el antepasado de la computación Grid estuvo presente desde el inicio de la ciencia de la computación. Pero el antepasado inmediato de la computación Grid fue la metacomputación usada en proyectos de computación alrededor de los años noventa para interconectar los centros de supercomputación de Estados Unidos y compartir el poder de procesamiento de múltiples supercomputadoras.

Algunos proyectos de metacomputación desarrollados en Estados Unidos como el FAFNER e I-WAY influyeron en el desarrollo de algunos de los proyectos claves de la computación Grid que hoy se encuentran en curso.

El proyecto FAFNER tenía como objetivo descomponer en factores números muy grandes interconectando supercomputadoras usando sólo las redes existentes. Para lograr este objetivo se requirió un uso intensivo de cómputo y se dividió el problema en trozos más pequeños. Muchas de estas técnicas desarrolladas para dividir y distribuir un problema de cómputo grande eran precursoras de la tecnología usada para SETI@home, el cual fue uno de los primeros proyectos de computación en usar características de la computación Grid.

El proyecto I-WAY tuvo como objetivo desarrollar un corredor de recurso computacional conceptualmente muy similar a los corredores de recursos computacionales desarrollados para la computación Grid hoy en día. Además, este proyecto influyó fuertemente en el desarrollo del proyecto Globus que es la

base de la mayoría de actividades de la computación Grid, y actualmente es el proyecto desarrollado más importante de la computación Grid.

De seguro muchos otros proyectos también colaboraron de forma directa o indirectamente en lo que hoy se conoce como la computación Grid, ya que muchas tecnologías han ayudado en el desarrollo y evolución de esta nueva rama de la computación.

II.2.4 Tecnologías claves en el desarrollo de la computación Grid [2]

Hay unas tecnologías que nacieron y evolucionaron durante los años noventa que actualmente hacen posible la computación Grid y su entorno. Algunas de estas tecnologías son:

- La Internet: La cual comenzó a funcionar en 1993 y juega un papel crucial en las comunicaciones de la computación Grid.
- Sistema Operativo Linux: Desarrollado inicialmente por Linus Thorvalds en el año de 1991 y luego desarrollado por miles de programadores de todo el mundo. Actualmente se siguen incorporando cada vez más programadores en el desarrollo del sistema operativo Linux, el cual es de código abierto. Estas son algunas razones por las cuales el sistema operativo Linux es preferido por muchas plataformas de pruebas Grid.

- Clusters de Computadoras: En 1994 se hizo el primer clusters de computadoras de bajo costo en la NASA. Se usaron tarjetas de red Ethernet para que las comunicaciones entre las computadoras fueran de alta velocidad y así emular el poder de cómputo de las supercomputadoras. Estos clusters están disponibles actualmente y la computación Grid confía en ellos para obtener un gran poder de cómputo.
- Lenguaje de Programación Java: Este es un lenguaje de programación que fue desarrollado por los ingenieros de la compañía Sun Microsystems en el año de 1995 y su principal característica es ser independiente de la computadora que se está ejecutando. No es el único lenguaje de programación que usa la computación Grid, pero es muy popular en esta rama de la computación y está muy bien adaptado a la filosofía de la computación Grid.

II.2.5 Características de un sistema de computación Grid [3]

De forma más clara, un sistema es Grid si cumple con los siguientes puntos:

- Donde el control y la coordinación de todos los recursos se hacen de forma descentralizada, es decir, la coordinación de los recursos no está sujeta a un control centralizado.
- Proporciona servicios específicos de calidad en términos de tiempo de respuesta, rendimiento, disponibilidad y seguridad.

- Utiliza protocolos e interfaces estándares y abiertos.

Entonces, la computación Grid es definida de acuerdo a los servicios y aplicaciones que proporcione y que lo haga de una forma descentralizada, no según su arquitectura, y además que utilice estándares abiertos.

II.2.6 Propiedades de la computación Grid [3]

Los servicios que ofrece la computación Grid deben tener ciertas propiedades específicas que cumplan con las características de la computación Grid.

Estas propiedades son las siguientes:

- Extenso: Abarca muchos recursos computacionales, en muchos lugares.
- Constante: Disponible todo el tiempo, las 24 horas del día, los 365 días del año.
- Confiable: Debe utilizar canales seguros.
- Simple: Que sea fácil de usar, no se necesita ser un experto en computación para utilizarla.
- Dinámico: Descubrir, agregar y eliminar recursos.
- Transparente: Que no importe donde se ejecutan los programas ni donde se almacena la información o los datos.
- Económico: Aprovecha y utiliza los recursos actuales de la organización o institución y además que sea muy fácil agregar nuevos recursos.

II.2.7 Elementos necesarios para la computación Grid [12]

Los elementos que se necesitan para el correcto y buen funcionamiento de los servicios que prestará la computación Grid son los siguientes:

- De Internet.
- De la Web.
- Conexión las 24 horas del día, los 365 días del año y de alta velocidad.
- Software especializado.
- Computadoras, supercomputadoras, servidores y clusters de computadoras.
- Sistemas de almacenamiento.
- Seguridad.
- Estándares.

II.2.8 Analogía de la computación Grid con el Grid de la red eléctrica [2]

El termino Grid en inglés es usualmente utilizado para hacer referencia a la red eléctrica. El Grid computacional toma este nombre de una analogía con el Grid de la red eléctrica.

Al igual que la red eléctrica, la computación Grid intenta copiar los servicios en términos de que cualquier persona pueda conectarse a ella desde cualquier lugar y beneficiarse de estos.

Relación de la computación Grid con el Grid de la red eléctrica:

- **Infraestructura:** El Grid de la red eléctrica se encarga de conectar diferentes tipos de centrales eléctricas con un hogar de un usuario determinado a través de estaciones de transmisión, centrales eléctricas, transformadores y líneas de corrientes, para que el usuario obtenga la corriente eléctrica que necesita.

Igual ocurrirá en la computación Grid, ésta se encargará de conectar recursos de computación como computadoras personales, supercomputadoras, servidores, elementos de almacenamientos y además proporcionará los elementos necesarios para accederlos, para que el usuario tenga acceso a dichos recursos computacionales.

- **Transparencia:** En el Grid de la red eléctrica el usuario no necesita preocuparse dónde y cuándo se genera la corriente eléctrica que usa. Igual será en la computación Grid, el usuario no se tendrá que preocupar qué computadora procesará su solicitud ni dónde se encuentran los datos que se necesitan, porque el middleware se encargará de asignar el trabajo al mejor recurso y buscará la manera de localizar los datos que se necesitan.
- **Extenso:** El Grid de la red eléctrica es extenso, ya que la electricidad se encuentra disponible en todas partes y el usuario puede acceder a ella simplemente a través de un enchufe eléctrico.

Igual ocurrirá con la computación Grid, los recursos computacionales serán accedidos por diferentes tipos de plataformas y sistemas simplemente usando un navegador Web con conexión a Internet.

- Utilidad: El Grid de la red eléctrica es una utilidad, ya que el usuario solicita electricidad y la obtiene, y paga por lo que obtiene.

Del mismo modo será la computación Grid, el usuario solicitará poder de cómputo o capacidad de almacenamiento y lo obtendrá, y también pagará por lo que obtendrá.

II.2.9 ¿Por qué la computación Grid? [13]

Hoy en día es prácticamente imposible hacer ciencia sin el uso de las computadoras. Los proyectos científicos que se vienen desarrollando en los últimos años abordan objetivos cada vez más ambiciosos y complicados que requieren la resolución de problemas computacionales complejos, tanto por el volumen de los cálculos a realizar como por el tamaño y complejidad de la información a procesar. Una sola computadora, un clusters de computadoras o una supercomputadora, no es suficiente en algunos casos para los cálculos y procesamientos de datos científicos que se quieren hacer y, a pesar que las computadoras son cada vez más rápidas, no son lo suficientemente rápidas para ir al ritmo de las demandas de los científicos.

Como resultado de esto, los científicos frecuentemente enfrentan situaciones cada vez más difíciles, muy costosas y en algunas ocasiones imposible de resolver,

además de no lograr algunas metas científicas con la tecnología de computación actual.

La computación Grid surgió como una solución a esto. No hay que preocuparse dónde almacenar los datos, ya que proveerá de espacio casi infinito para guardar de forma segura cualquier tipo de información, además de tener el poder de cómputo casi infinito disponible para su organización o institución siempre que se necesite. Además, permite colaborar fácil y eficazmente con los colegas alejados geográficamente compartiendo de forma segura los recursos, datos, procedimientos y resultados, logrando todo esto con los recursos actualmente disponibles y a bajo costo.

II.2.10 ¿Cómo será la computación Grid? [2]

La computación Grid hoy en día está evolucionando cada vez más en su desarrollo. Hay muchas personas interesadas en esta tecnología y mucho dinero invertido en ella, lo cual nos indica que ésta tecnología es muy prometedora en un futuro.

Ahora Imaginemos cómo podrá ser la computación Grid cuando ya se haya desarrollado completamente.

Imaginemos muchas computadoras (varios millones) entre las cuales tenemos computadoras comunes, servidores, supercomputadoras, periféricos y otros dispositivos, situados en todo el mundo, pertenecientes a muchas personas e instituciones diferentes y todos estos sistemas conectados por Internet como lo

están actualmente la mayoría de las computadoras. Hasta ahora todo esto existe y no es nada nuevo. Esta es la base para el desarrollo de la computación Grid.

Entonces imaginemos que existen unas herramientas de software que puedan hacer que todas las computadoras conectadas a Internet se puedan comunicar entre sí y funcionar como una sola y enorme súper poderosa computadora virtual de capacidades y recursos computacionales casi infinitos. Todo esto es lo que puede llegar a ser la computación Grid.

Si todo esto sobre la computación Grid llegase a desarrollarse completamente en un futuro, cualquier persona conectada al Grid que estuviese interesada en ejecutar una aplicación específica, ya no necesitará instalar en su computadora el programa necesario para ejecutar dicha aplicación, sino colocar la solicitud de ejecutar la aplicación en el Grid el cual se encargaría de buscar una máquina remota que tuviera el programa necesario para ejecutar la aplicación y que estuviera disponible. En el caso de no encontrar esta computadora, buscará una o un conjunto que estén ociosas y le instalará el programa y ejecutará la aplicación sin tener que especificar mucho ya que el Grid se encargará de buscar la mejor opción. Este es un ejemplo de una de las muchas cosas que podrá hacer la computación Grid si se llega a desarrollar completamente. Otro ejemplo sería que alguien que esté interesado en hacer un cierto tipo de trabajo en conjunto con otras personas que se encuentren alejadas geográficamente y estén dispersas alrededor del mundo, el Grid uniría sus computadoras para que todos trabajen

como si en verdad estuvieran en una red local sin tener que preocuparse de muchas claves o contraseñas.

II.2.11 ¿En qué estado se encuentra la Computación Grid actualmente?

Comparando la computación Grid de hoy con la de cómo podría llegar a ser si se llegase a desarrollar completamente, aún están muy limitadas sus capacidades y virtudes, esto se debe a que todavía le falta por desarrollarse y evolucionar más. Aún falta que los recursos computacionales estén disponibles en cualquier lugar, en cualquier momento, en forma confiable y segura, que sea transparente y fácil de usar.

En realidad, todavía no hay un único Grid en el mundo, como ocurre en el caso de la Web o al menos un Grid dominante y todavía no se sabe si lo habrá. Sin embargo, se va en buen camino.

Hay muchos Grids evolucionando y desarrollándose alrededor del mundo, algunos privados, otros públicos, algunos dentro de una región o país, algunos de dimensiones globales, algunos dedicados a un problema científico en particular y algunos de propósito general.

Lo cierto es que la computación Grid cada vez se vuelve más popular y suma más adeptos en diferentes partes de la sociedad. No solo los científicos e ingenieros, sino hombres de negocios y compañías comerciales están interesados en el desarrollo de la computación Grid. Debido a esto, actualmente se está invirtiendo

mucho dinero en el desarrollo y evolución del Grid. Por lo cual, la computación Grid es una de las tecnologías más prometedoras en el campo de la computación, y no solo es una moda, sino va más allá y seguramente en unos años esta tecnología superará las limitaciones actuales [2].

II.2.12 Funcionamiento de la computación Grid

Dentro del funcionamiento de la computación Grid hay unos aspectos muy importantes que son las bases que harán que esta tecnología se desarrolle y evolucione hasta llegar a ser lo que se espera.

Lo más importante para esta ciencia de la computación es compartir recursos en una escala global, ésta es la esencia de la computación Grid. Un aspecto crítico es la seguridad, ya que debe haber un nivel muy alto de seguridad para que así haya confianza entre los proveedores de recursos y los usuarios del Grid, que de seguro muchas veces no se sabrá quienes son los proveedores o los usuarios que usan los recursos.

También aparte de la seguridad, la computación Grid necesita tener estándares de manera que se puedan comunicar y entender todas estas computadoras que se encuentran alrededor del mundo. Además, todas estas computadoras conectadas al Grid deben estar proporcionando recurso eficaz y rápidamente. Entonces la computación Grid comenzará realmente a ser rentable.

Los aspectos más importantes para el funcionamiento de la computación Grid se describen más detalladamente a continuación:

Compartir Recursos: [2]

Esta es la principal idea de la computación Grid. Se accederá al Grid para usar recursos remotos que permitan resolver problemas que no se pueden hacer con una sola computadora o con un conjunto de computadoras. El Grid proporcionará acceso remoto a software, datos, computadoras, poder de cómputo, espacio de almacenamiento y otros tipos de dispositivos que no le pertenezcan al usuario normal que entre al Grid y además, balanceará todos estos recursos con un mecanismo para asignar trabajo eficaz y automáticamente para así satisfacer todas las demandas asignando el trabajo del usuario al recurso más adecuado para resolverlo.

El mayor reto es la implementación de la computación Grid, ya que la mayoría de estos recursos que proveerá el Grid pertenecen a muchas personas e instituciones diferentes, los cuales existen dentro de dominios administrativos diferentes y están sujetos a diferentes políticas de control y seguridad de acceso, por eso se debe usar un estándar para que se entiendan y se comuniquen todos los recursos. Ese estándar es el software llamado middleware, que es Globus Toolkit y se prevee que en un futuro sea el estándar de la computación Grid, tal como hoy en día lo es el protocolo TCP/IP para Internet.

También para compartir recursos de manera óptima en el Grid, deben existir conexiones de alta velocidad y así evitar los cuellos de botella en las comunicaciones. Este es otro factor importante cuando se comparten recursos y debe ser resuelto para así hacer posible el Grid Global.

Además debe existir algún tipo de ventaja o beneficio para los dueños de los recursos interesados en compartirlos en el Grid y deben estar confiados y seguros de cualquier usuario del Grid que use sus recursos. Así mismo el dueño del o los recursos pondrá condiciones sobre el uso de éstos especificando los límites en los cuales los recursos pueden ser usados y cuándo.

Todo el manejo de los recursos computacionales será de una forma descentralizada, no habrá un servidor central dentro del Grid que controlará todo, sino muchas organizaciones virtuales alrededor del mundo encargadas del descubrimiento y manejo de los recursos y de procesar todas las solicitudes de los usuarios del Grid [6].

Seguridad: [2]

Este es el aspecto más crítico dentro de la computación Grid, ya que cuando se comparten recursos y se maneja información remotamente, debe existir un alto nivel de seguridad. Los requerimientos de seguridad son fundamentales en el diseño del Grid.

Debe haber componentes de seguridad básica como la autenticación, autorización, encriptación y política de acceso dentro de Grid.

Sin estos componentes de seguridad el Grid y los que lo usan estarán expuestos a riesgos y peligros.

En el Grid debe existir un mecanismo de autenticación que informe sobre la identidad de un usuario o recurso que esté dentro del Grid y consistirá en el proceso de verificación y confirmación de la identidad de un usuario dado, que es quien dice quién es o de un recurso dado.

La parte de autorización es la que se encarga de decidir quién está autorizado para utilizar Grid y a cuáles recursos puede acceder, mediante un certificado de autoridad especial y único. Estos certificados de autoridad se generan para que el usuario pueda acceder al Grid y pueda trabajar en él, tienen un tiempo determinado de uso y además incluye una clave privada la cual es muy importante porque es la que se le solicita al usuario para el acceso físico y/o remoto al Grid y debe estar bien protegida de los ataques de hackers.

En lo referente a la encriptación, ésta es la base de seguridad cuando se comparten o se transmiten datos entre dos o más computadoras. Debe estar para proporcionar seguridad a los datos, ya que sin la adecuada encriptación es posible que mientras los datos estén siendo transmitidos y/o durante su almacenamiento en otra computadora sobre el Grid, el dueño de esta computadora o algún hacker pueda leerlos.

Y por último se encuentra la política de acceso encargada de proporcionar acceso a los recursos, es decir, cuáles y en que cantidad se comparten. Los proveedores de recursos deben definir transparente y cuidadosamente lo que es compartido, qué se permite compartir y las condiciones bajo las cuales se comparte.

Estándares Abiertos:

Finalmente tener un estándar abierto en la computación Grid es indispensable, ya que sin él los recursos no podrán ser compartidos en una escala global como se quiere hacer con el Grid.

Esto puede parecer iluso después de todo, porque muchas compañías de software hacen sus ganancias precisamente porque ellos no comparten sus estándares con otros. Sin embargo, por la misma naturaleza del Grid, que es el compartir, generalmente se percibe en todos el mismo interés común para usar estándares abiertos [2].

El encargado de desarrollar estándares para la computación Grid es el Foro Grid Global. El estándar actualmente del Grid es el OGSA (Arquitectura de Servicios Grid Abiertos). Este se enfoca en la integración de servicios a través de organizaciones virtuales alrededor del mundo y concuerda en que el estándar para proporcionar servicios en el Grid debe ser Globus Toolkit, una infraestructura de código abierto que proporciona muchos de los servicios básicos necesarios para construir las aplicaciones del Grid como: descubrimiento, acceso, manejo y monitoreo de recursos, manejo de datos, comunicación y seguridad. En un futuro el proyecto de software Globus Toolkit, que actualmente se encuentra en la

versión 4.0, seguramente se convertirá en el estándar de las comunicaciones del Grid y así unificará todos los recursos dentro del Grid [3 y 14].

II.2.13 Arquitectura de la computación Grid [15]

La arquitectura de la computación Grid está formada por varias capas, la capa más alta se enfoca al usuario Grid y la capa más baja se enfoca a las computadoras y redes.

La capa más baja es la capa de infraestructura en donde se encuentran las redes que se encargan de proporcionar la comunicación de todos los recursos del Grid junto con las computadoras, sistema de almacenamiento de datos, periféricos, diferentes tipos de instrumentos y dispositivos, que son los que proveen recursos al Grid.

Después de la capa de infraestructura le sigue la capa de conectividad que está formada por los protocolos estándar de comunicación y seguridad. Esta capa proporciona el intercambio de información entre los recursos y provee mecanismos para identificar un usuario o un recurso específico. Algunos elementos de esta capa son el protocolo TCP/IP que usa Internet y aquí es usado para la comunicación entre los recursos, el protocolo SSL que proporciona seguridad y privacidad en las comunicaciones y los certificados X.509 para la autenticación, que proporciona mecanismos seguros para verificar la identidad de usuarios y recursos para brindar seguridad al Grid.

La siguiente capa es la capa de recurso. Esta capa es la encargada de informar sobre un recurso en particular, su estado y estructura, además controla y maneja el recurso.

Luego le sigue la capa de servicios colectivos que permite gestionar la interacción de un conjunto de recursos como descubrir y ubicar recursos compartidos, despachar trabajos a los recursos, monitorear y diagnosticar cualquier falla que ocurra y manejar replicas de datos.

Por último, se encuentra la capa de aplicación que es la capa que se utilizará para acceder al Grid que incluye todas las aplicaciones del usuario (ciencia, ingeniería, negocios y finanzas), portales y herramientas de desarrollo. Ésta es la capa que los usuarios del Grid verán y se accederá sólo con tener un navegador Web con conexión a Internet.

II.2.14 Principales áreas de aplicación de la Computación Grid [3]

Actualmente la mayoría de las aplicaciones de la computación Grid son científicas y de ingeniería, en futuro seguramente darán un salto a las aplicaciones empresariales y comerciales.

Algunas áreas donde interviene la computación Grid son las siguientes:

- En la investigación científica, médica e ingeniería.
- En la colaboración remota.
- En el control de instrumentos y dispositivos.
- En la gestión y análisis de datos.

II.2.15 Ventajas y Desventajas de la computación Grid

Ventajas de la computación Grid:

- Gran poder de cálculo y de cómputo.
- Gran capacidad de almacenamiento de datos.
- Colaboración remota entre diferentes organizaciones e instituciones.
- La eliminación del factor distancia.
- Incrementa la calidad, cantidad y la capacidad de los recursos que poseen las organizaciones e instituciones.
- Hace posible que recursos heterogéneos se entiendan entre ellos.
- Puede ahorrar tiempo y dinero.

Desventajas de la computación Grid:

- No existe actualmente un estándar para que se puedan comunicar y entender todos los recursos computacionales disponibles.
- La implementación técnica de la computación Grid hoy en día es complicada y requiere de muchos detalles técnicos para que funcione.
- Que mucha gente con objetivos y puntos de vista diferentes se pongan de acuerdo en cómo compartir recursos no es tarea fácil.

www.bdigital.ula.ve

CAPÍTULO III

III. PRINCIPALES HERRAMIENTAS DE COMPUTACIÓN GRID

III.1 Herramientas de Computación Grid

La computación Grid utiliza varias herramientas de software que se encargan de realizar los diferentes trabajos dentro del Grid. Estas herramientas se pueden dividir en: programas planificadores de tareas, portales y el middleware. Este último software es el Globus Toolkit que en un futuro seguramente se convertirá en el estándar de facto para la computación Grid.

A Continuación se explicarán las herramientas de computación Grid actualmente disponibles que se consideran más importantes y significativas:

III.2 Condor [16]

Condor es un software desarrollado por el proyecto de investigación Condor de la Universidad de Wisconsin que comenzó en 1988. Inicialmente se instaló como un sistema de producción de poder de cómputo en el departamento de computación de la Universidad, con el objetivo de reunir todos los recursos computacionales de dicho departamento.

Actualmente el software Condor se sigue utilizando en toda la Universidad de Wisconsin y alrededor del mundo. Muchas organizaciones en la industria, gobierno y académica están usando Condor.

El proyecto Condor tiene como objetivos desarrollar, implementar, y evaluar mecanismos y políticas que apoyen la computación de alto rendimiento en grandes conjuntos de recursos computacionales distribuidos que estén disponibles.

Este proyecto ha desarrollado el software Condor y actualmente sigue desarrollando nuevas versiones del software. Este software permite a científicos e ingenieros incrementar su rendimiento de cómputo.

El software Condor está disponible libremente para su uso. Es técnicamente un software de código abierto y está bajo una licencia BSD [17].

Condor soporta los sistemas operativos Linux, Unix y Windows [17].

Condor puede ser usado para construir ambientes de computación Grid que van más allá de los límites administrativos de una organización o institución.

La tecnología del Condor permite que múltiples instalaciones de cómputo que usan Condor trabajen juntas. Incorpora mucha de las metodologías y protocolos de la computación basada en Grid.

III.2.1 Definición de Condor [16]

Condor es un sistema de software que se encarga del manejo de carga especializado para trabajos de cómputo intensivo. Es definido como un programa planificador de tareas. Proporciona varios mecanismos para despachar un trabajo o tarea en una cola de trabajo, políticas de planificación, esquemas de prioridad,

manejo y monitoreo del recurso. El usuario de Condor despacha sus trabajos seriales o paralelos y Condor se encarga de colocarlos dentro de una cola de trabajo seleccionando cuándo y dónde ejecutar los trabajos en base a una política determinada. Cuidadosamente monitorea el proceso e informa al usuario sobre la culminación del mismo.

Condor también posee un mecanismo que le permite aprovechar eficazmente el CPU de una computadora que no se está usando, es decir, se aprovecha del poder del CPU de las computadoras que se encuentre es un estado ocioso. De esta manera Condor utiliza los recursos de computación que son desperdiciados y les da un buen uso.

Además, Condor extiende los recursos disponibles a todos los usuarios.

III.2.2 ¿Cómo funciona Condor? [16]

El software Condor puede configurarse para que funcione de varias maneras dependiendo de lo que el usuario desea hacer en su organización o institución. También va a depender de los recursos computacionales disponibles que el usuario de Condor posea.

Condor puede ser usado para manejar clusters de computadoras dedicadas para propósitos específicos.

También puede ser configurado para usar sólo computadoras que se encuentren en un estado ocioso, es decir, que no se esté usando el teclado ni el ratón de la

computadora. Condor detecta cuando una computadora ya no está disponible, es decir, detectando cuando una tecla se oprime en el teclado o cuando se mueve el ratón. Si Condor no ha terminado un trabajo en una computadora cuando ésta deja de estar disponible, Condor se encarga de migrar su trabajo de una forma transparente a una computadora que se encuentre en un estado ocioso y continuar con el trabajo que estaba realizando.

Condor cuenta con un mecanismo flexible para los trabajos y recursos, en el cual los trabajos pueden indicar requisitos para su ejecución y también los recursos pueden especificar requisitos sobre los trabajos que van a ejecutar.

III.2.3 Ventajas de Condor

- Software de código abierto y disponible libremente para su uso.
- Permite incrementar el poder de cómputo disponible.
- Funciona para manejar recursos dedicados (manejo de clusters de computadoras dedicadas) y también recursos no dedicados (se aprovecha de las computadoras ociosas).
- Soporta gran variedad de sistemas operativos.
- Amplia documentación para su uso.

III.3 Globus Toolkit [18]

El Globus Toolkit es una herramienta de software diseñado especialmente para construir sistemas y aplicaciones Grid. Actualmente es desarrollado por la Alianza Globus. Esta alianza está formada por una comunidad de organizaciones e individuos que desarrollan tecnologías fundamentales para la computación Grid, permitiendo a las personas compartir el poder cómputo, base de datos, instrumentos y otras herramientas en línea de forma segura, a través de organizaciones e instituciones sin sacrificar la autonomía local.

La primera versión de Globus Toolkit fue lanzada en 1998 y actualmente se encuentra disponible la versión 4.0 basada en un nuevo estándar abierto de servicios Grid.

Actualmente esta herramienta de software es muy usada en proyectos de computación Grid y cada vez hay más personas en organizaciones e instituciones que están empezando a utilizarla para su propia causa.

Muchas compañías e individuos alrededor del mundo están trabajando en el desarrollo de Globus Toolkit y además se ha invertido, y se sigue invirtiendo mucho dinero por lo que se está convirtiendo en la herramienta de software estándar de facto de la computación Grid.

El Globus Toolkit se ha desarrollado y ha evolucionado rápidamente gracias a una estrategia de código abierto muy parecida a la del sistema operativo Linux. Esto se debe a que la comunidad de código abierto aporta a Globus Toolkit grandes y rápidas mejoras en su desarrollo y evolución.

III.3.1 Definición de Globus Toolkit

El Globus Toolkit es una herramienta de software de código abierto usada para construir sistemas y aplicaciones Grid [18].

El Globus Toolkit también se conoce como el software middleware de la computación Grid. Se encuentra en la capa intermedia del Grid entre los sistemas operativos y las aplicaciones Grid finales. Es el encargado de prestar diferentes servicios dentro de una plataforma Grid [15].

Esta herramienta incluye software de servicios como descubrimiento, acceso, manejo y monitoreo de recursos, manejo de datos, detección de fallas, comunicación, portabilidad y seguridad. El Globus Toolkit está formado por un conjunto de componentes que pueden ser usados independientemente o en conjunto para el desarrollo de aplicaciones Grid [18].

III.3.2 ¿Cómo funciona Globus Toolkit? [18]

El Globus Toolkit está diseñado especialmente para ser usado por arquitecturas Grid, integración de sistemas y diseños de aplicaciones Grid, para reducir la cantidad de trabajo requerida para lograr los objetivos y asegurar un nivel básico de interoperabilidad entre sistemas y aplicaciones Grid.

Además, el Globus Toolkit también fue creado para que funcione como medio de comunicación entre diferentes recursos que se encuentren alejados geográficamente, eliminando los obstáculos que impiden la colaboración conjunta entre organizaciones e instituciones, prestando servicios básicos, interfaces y protocolos que permitan a usuarios acceder a recursos remotos de una forma transparente y segura.

www.bdigital.ula.ve

III.4 GridPort [19]

El GridPort es una herramienta de software que se viene desarrollando desde 1999 y ha facilitado el desarrollo de portales de computación Grid que utilizan tecnologías estándares Web. Actualmente se encuentra disponible la versión 3.0 que proporciona un conjunto de herramientas y servicios Grid que pueden ser accedidos a través de un navegador Web estándar.

El principal objetivo de GridPort es el desarrollo y uso de servicios Web facilitando la operación dinámica de portales y aplicaciones.

La herramienta de software GridPort se encuentra disponible libremente para su uso y está bajo una licencia BSD [20].

La herramienta de software GridPort soporta fácilmente la integración de servicios externos tales como Globus Toolkit y programas planificadores de tareas.

III.4.1 Definición de GridPort

Es una herramienta de computación Grid definida como un portal Grid. Está formada por una colección de tecnologías diseñadas para ayudar en el desarrollo de portales de ciencia y de aplicaciones Grid como portales de usuarios, interfaces de aplicaciones Grid y portales de educación [20].

Esta herramienta de software también se puede definir como un middleware de alto nivel que se encuentra entre los servicios Grid de bajo nivel como Globus Toolkit y la capa de interfaz de aplicación como los portales Web, sobre la infraestructura de computación Grid subyacente [21].

El GridPort está formado por una arquitectura abierta que es diseñada para ser flexible y capaz de soportar nuevos servicios y tecnologías Grid que puedan emerger en un futuro [20].

El GridPort proporciona servicios interactivos y seguridad a los usuarios. Integra los siguientes componentes: cuentas de usuarios, mecanismos para la autenticación, manejo de archivos, despacho de trabajo, secuencia de trabajo y descubrimiento de servicios Grid y Web [19].

III.4.2 ¿Cómo funciona GridPort? [19]

El GridPort integra un API que es accedido a través de un navegador Web estándar para proporcionar un conjunto de herramientas y servicios a los usuarios, permitiendo múltiples portales para el uso de estas herramientas y servicios dentro de un solo ambiente.

El GridPort usa esta tecnología para permitir a los portales presentar información estática y dinámica sobre todos los recursos en el Grid. Maneja archivos y datos, ejecuta procesos y comandos, despacha, monitorea y anula trabajos.

También el GridPort implementa el uso de un modelo de servicios Grid que proporciona un servidor de aplicaciones Grid, permitiendo acceso remoto de clientes para el uso de los recursos Grid disponibles remotamente.

III.5 Portable Batch System (PBS) [22]

Es una herramienta de software que se creó con el objetivo principal de solucionar el manejo de carga de trabajo para sistemas de computación de alto rendimiento y para manejar clusters de computadoras Linux y Unix.

PBS fue diseñado originalmente por la NASA, a mediados de los noventa, porque requería un software para el manejo de recursos. Los sistemas de manejo de recursos existentes eran inadecuados para las grandes y modernas computadoras y clusters que poseía la NASA. Gracias a esto se desarrolló la herramienta de software PBS.

La herramienta de software PBS es líder en la solución de sistemas de computación de alto rendimiento. También es el software estándar para manejo de clusters de computadoras Linux.

Actualmente los desarrolladores de PBS ofrecen dos versiones del software:

- OpenPBS: versión original sin soporte técnico y disponible libremente para su uso.
- PBS Pro: versión profesional de calidad empresarial que está disponible por un cierto precio y con soporte técnico [23].

III.5.1 Definición de PBS [22]

PBS es una herramienta de software encargada de manejar la carga de trabajos en sistemas computacionales de alto rendimiento.

Además, esta herramienta de software permite el manejo de recursos y planificar la ejecución de trabajos, proporcionando una política planificadora totalmente configurable. PBS es definido como un programa planificador de tareas.

La herramienta de software PBS proporciona las siguientes características: sistema planificador de trabajos, información de trabajos, nivelación de carga de trabajos, aprovechamiento de los ciclos de procesamiento de computadoras ociosas, control de acceso y seguridad.

III.5.2 ¿Cómo funciona PBS? [22]

El objetivo de la herramienta de software PBS es proporcionar controles adicionales sobre inicialización y planificación de la ejecución de trabajos, permitiendo que los trabajos se ejecuten entre varias computadoras conectadas a una red.

PBS permite definir e implementar una política para la ejecución de los trabajos, como definir qué tipos de recursos y cuánto de cada recurso puede ser usado por los trabajos.

Además, proporciona un mecanismo que el usuario puede usar para asegurar que un trabajo tenga acceso a los recursos requeridos para ejecutarse.

PBS está formado por varios componentes como el servidor y el cliente, que funcionan estableciendo una interacción entre ellos. El cliente realiza una solicitud a un servidor y el servidor se encarga de realizar el trabajo del cliente.

El servidor de PBS proporciona servicios de crear, ejecutar, modificar o eliminar trabajos del cliente, dependiendo de la solicitud del cliente.

III.5.3 Ventajas de PBS

- Software disponible en dos versiones, una versión libre y otra comercial.
- Manejo de carga de trabajo.

- Manejo de recursos.
- Incrementa el poder de cómputo disponible.
- Proporciona servidores y clientes.

III.6 Sun Grid Engine (SGE) [24]

Es un proyecto desarrollado por la comunidad de código abierto. El principal objetivo de este proyecto es desarrollar y proporcionar una herramienta de software de código abierto, que facilite la adopción de soluciones de computación distribuida y promueva el desarrollo de estándares abiertos para el manejo de recursos distribuidos. Es patrocinado y apoyado por la compañía Sun Microsystems.

El proyecto Sun Grid Engine proporciona una herramienta de software para facilitar el manejo de recursos distribuidos que necesitan los grandes requerimientos de la computación Grid.

Actualmente el proyecto Sun Grid Engine ha desarrollado la versión 6.0 que se encuentra disponible libremente para su uso.

La herramienta de software Sun Grid Engine ha sido usada y se continúa usando por muchas organizaciones e instituciones en una gran variedad de actividades académicas y de investigación.

Esta herramienta de software puede ser integrada con el Globus Toolkit para que Globus Toolkit pueda despachar trabajos para ser ejecutados por el software Sun Grid Engine.

III.6.1 Definición de Sun Grid Engine

Sun Grid Engine es una herramienta de software para el manejo de recursos distribuidos en ambientes de red heterogéneos que se encarga de agregar poder de cómputo y lo entrega como un servicio de red [24]. Sun Grid Engine es definido como un programa planificador de tareas, desarrollado y distribuido con una licencia de código abierto [25].

Sun Grid Engine provee las funciones de un manejador de recursos distribuidos como mecanismos para despachar trabajos, balanceo de cargas, estadísticas de trabajos despachados, manejo dinámico de los recursos, protocolos de seguridad, administración de los recursos y monitoreo de los procesos [25].

III.6.2 ¿Como funciona Sun Grid Engine? [24]

La herramienta de software Sun Grid Engine fue diseñada para prestar un servicio de manejo de recursos computacionales en varios entornos de computación que posean recursos limitados y necesiten gran poder de cómputo, como en organizaciones e instituciones donde un gran número de usuarios estén realizando trabajos que necesiten de un gran número de recursos y de poder de cómputo para procesarlos.

Para lograr esto, Sun Grid Engine coloca los trabajos en un sitio óptimo y balancea la carga de un conjunto de computadoras conectadas en red, permite a los usuarios generar y colocar trabajos en cola de los que puedan ser ejecutados en el momento, se asegura que los trabajos son ejecutados con la prioridad establecida y proporciona a todos los usuarios una parte justa del acceso a los recursos en un cierto plazo de tiempo.

III.6.3 Ventajas de Sun Grid Engine

- Software de código abierto y disponible libremente para su uso.
- Código fuente disponible para modificarlo.
- Facilita el manejo de recursos distribuidos.
- Aumenta el poder de cómputo disponible.
- Permite fácil integración con Globus Toolkit.

CAPÍTULO IV

IV. HERRAMIENTA DE SOFTWARE GLOBUS TOOLKIT

IV.1 Globus Toolkit 4.0

Herramienta de software de arquitectura abierta y código abierto usada para construir sistemas y aplicaciones Grid.

Desarrollada actualmente por la alianza Globus y muchas otras alianzas y organizaciones alrededor del mundo [18].

Globus Toolkit es el middleware de la computación Grid que se encuentra en la capa intermedia del Grid, entre los sistemas operativos y las aplicaciones Grid finales. Es el encargado de prestar diferentes servicios dentro de una plataforma Grid [15].

La herramienta de software Globus Toolkit está basada en tecnologías estándar, posee un alto nivel de seguridad en todos los servicios que presta y está implementada íntegramente en el lenguaje de programación Java [26].

Globus Toolkit se ha convertido en la herramienta de software más popular y utilizada por sistemas de computación Grid. Actualmente esta herramienta es utilizada por la mayoría y los más importantes proyectos Grid que están en desarrollo alrededor del mundo. Por ejemplo, es utilizada por TeraGrid, que es el proyecto de computación Grid más importante dentro de los Estados Unidos, así

como por el Grid Europeo. Actualmente estos proyectos de computación Grid están utilizando la versión 2 de Globus Toolkit, esto se debe a que las nuevas versiones que han salido presentan cambios muy drásticos en toda su estructura, con respecto a las versiones anteriores, y esto implicaría realizar muchos cambios en las plataformas Grid. Se está esperando que la herramienta Globus Toolkit se estabilice y que el cambio de una versión a otra no sea muy drástico.

IV.2 Objetivos de Globus Toolkit 4.0 [18]

La herramienta de software Globus Toolkit es diseñada y desarrollada para ser usada en las siguientes áreas:

- Manejo de arquitecturas Grid.
- Integración de sistemas.
- Diseño de aplicaciones Grid.

IV.3 Servicios de Globus Toolkit 4.0 [3]

La herramienta de software Globus Toolkit incluye los siguientes servicios:

- Descubrimiento de recursos.
- Acceso a recursos.
- Manejo de recursos.
- Monitoreo de recursos.

- Manejo de datos.
- Detección de fallas.
- Comunicación.
- Portabilidad.
- Seguridad.

IV.4 Componentes más importantes de Globus Toolkit 4.0 [26]

Los principales componentes de Globus Toolkit desde sus primeras versiones y hasta la última versión disponible que es la 4.0, han sido los siguientes:

- GRAM (Grid Resource Allocation Manager): Manejador de Asignación de Recursos Grid.
- MDS (Monitoring and Discovery System): Sistema de Descubrimiento y Monitoreo.
- GridFTP (Grid File Transfer Protocol): Protocolo de Transferencia de Archivo Grid.
- GSI (Grid Security Infrastructure): Infraestructura de Seguridad Grid.

Un nuevo componente de Globus Toolkit 4.0 es el Web Services (Servicios Web) que viene incluido en la última versión de Globus Toolkit ya que está orientada a prestar servicios Web (TallerGrid/GT4).

IV.5 Arquitectura de Globus Toolkit 4.0 [26]

La arquitectura del Globus Toolkit 4.0 se encuentra estructurada en varios niveles o capas que a continuación se describen:

- Infraestructura:
 - Recursos: elementos de cómputo, almacenamiento e instrumentación.
- Seguridad:
 - Servicios de seguridad: autenticación, autorización, delegación y encriptación.
- Servicios de bajo nivel:
 - Descubrimiento, acceso, manejo y monitoreo de los recursos.
 - Manejo de datos, manejo de réplicas, transferencias de datos y acceso a datos.
- Servicios de alto nivel:
 - Servicios Web.

IV.6 Componentes de Globus Toolkit 4.0 [18]

En la Figura IV.1 se observan todos los componentes que forman parte de Globus Toolkit 4.0:

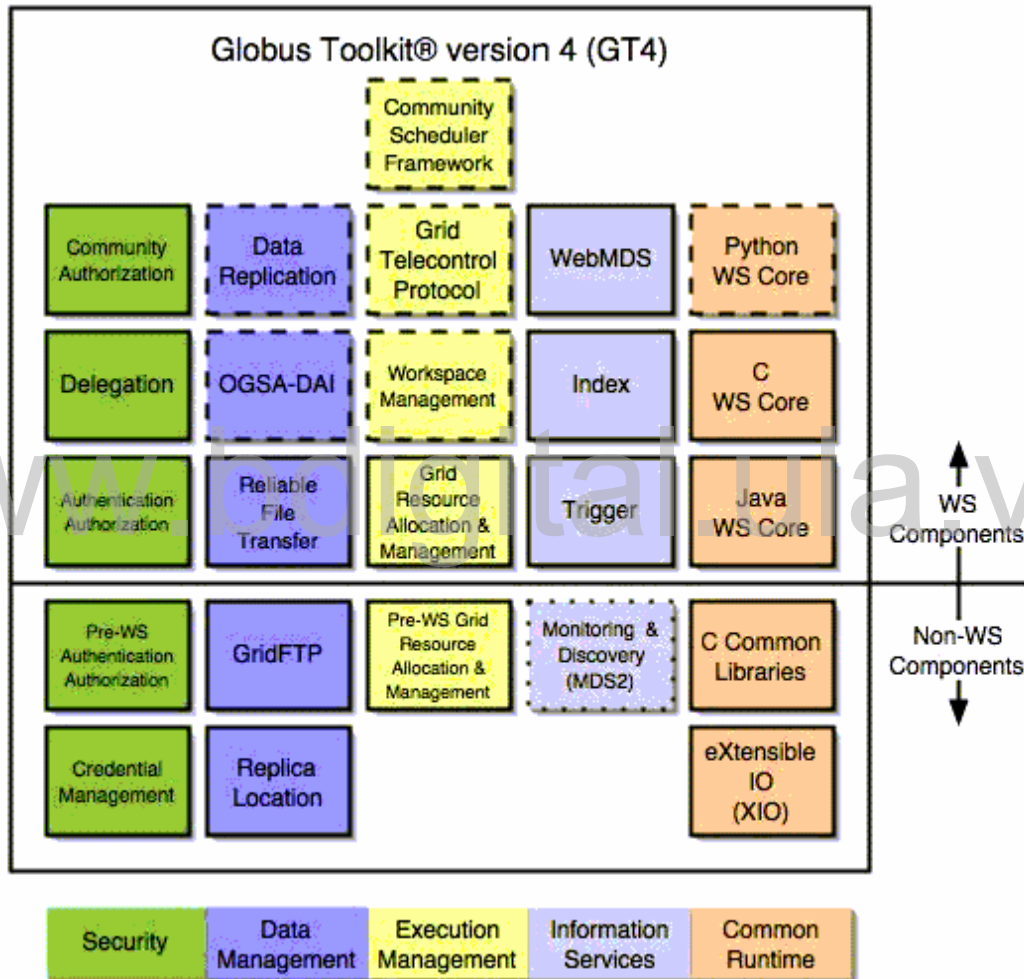


Figura IV.1 Componentes de Globus Toolkit 4.0

Los componentes de Globus Toolkit 4.0 se dividen en los siguientes niveles:

- Seguridad (Security).
- Manejo de Dato (Data Management).
- Manejo de Ejecución (Execution Management).
- Servicios de Información (Information Services).
- Componentes de Ejecución Común (Common Runtime).

IV.6.1 Security (Seguridad)

IV.6.1.1 Credential Management (CM): Manejo de Credencial

El componente CM está formado por dos subcomponentes: SimpleCA y MyProxy.

SimpleCA:

Es un paquete de software que proporciona una simple y sencilla forma de certificación de autoridad, con el propósito de emitir credenciales para los usuarios y servicios de Globus Toolkit 4.0.

Cada usuario y servicio en el Grid es identificado a través de un certificado de autoridad, que contiene información importante para identificar y autenticar a usuarios o servicios.

MyProxy:

Es un depósito de credenciales en línea. Almacena y recupera credenciales Proxy X.509, que se encarga de proteger a las entidades de una red al tiempo que proporciona acceso a un determinado recurso, protegidas por una clave o contraseña disponibles a través de una red.

El depósito de credenciales MyProxy elimina la necesidad de copiar manualmente los archivos de credenciales y las claves privadas entre las computadoras.

MyProxy también proporciona una capacidad de delegación a través del Proxy, la cual permite al usuario reducir el número de veces que debe proporcionar su clave o contraseña al utilizar distintos recursos del Grid. Esto se logra al crear un servicio Proxy.

www.bdigital.ula.ve

IV.6.1.2 Pre - Web Services Authentication and Authorization (Pre-WSAA):

Autorización y Autenticación de Servicios Pre- Web

El componente Pre-WSAA proporciona APIs (Interfaz de Programas de Aplicación) y herramientas de software para la autenticación, autorización y manejo de certificados.

El API de autenticación está formado por tecnologías PKI (Infraestructura de Clave Publica) que es una combinación de elementos de hardware y software, políticas y procedimientos, que permiten asegurar la identidad de los usuarios en un intercambio de datos, además la autenticación usa un mecanismo de delegación basado en certificados Proxy X.509.

La autorización está formada por dos APIs. El primer API proporciona una autorización genérica de control de acceso basado en credenciales del cliente (credenciales Proxy X.509). El segundo API proporciona una lista control de acceso de nombres de usuarios locales.

Además de la autenticación y autorización, el componente Pre-WSAA proporciona APIs de un nivel más bajo y herramientas para manejar, descubrir e indagar certificados.

IV.6.1.3 Web Services Authentication and Authorization (WSAA):

Autorización y Autenticación de Servicios Web

El componente WSAA está formado por dos subcomponentes: Message and Transport Level Security (Seguridad a Nivel de Transporte y Mensaje) y Authorization Framework (Estructura de Autorización).

Message and Transport Level Security (MTLS):

El componente MTLS implementa seguridad estándar de servicios Web y la especificación de conversación segura de servicios Web, para proporcionar protección para los mensajes SOAP que permiten que dos objetos en diferentes procesos puedan comunicarse por medio de datos XML. Las características incluyen autenticación del remitente, encriptación del mensaje, protección de integridad del mensaje, réplica de protección y prevención de ataques de hackers.

La autorización y autenticación de servicios Web de seguridad a nivel de transporte proporciona un canal seguro para el transporte de mensajes, es decir, soporta HTTPS, que es la versión segura de HTTP (Protocolo de Transferencia de Hipertexto) para el transporte seguro de mensajes.

Authorization Framework (AF):

El componente AF proporciona una estructura para la autorización de nivel de container, que es donde se localizan los servicios Grid. Permite módulos de autorización con interfaces bien definidas para ser asociada con varias entidades, por ejemplo, servicios en el container.

También, este componente proporciona autorización a través de listas de control de acceso, como el archivo grid-mapfile, el cual proporciona los datos de los usuarios que son autorizados para el uso de los servicios Grid.

IV.6.1.4 Delegation Service (DS): Servicio de Delegación

El componente DS proporciona una interfaz para la delegación y renovación de credenciales para un entorno múltiple. Esto permite que una sola credencial delegada pueda ser compartida a través de invocaciones múltiples de servicios en entornos múltiples, es decir, una sola credencial de delegación puede ser usada y reusada a través de múltiples solicitudes de servicios.

IV.6.1.5 Community Authorization Services (CAS): Servicios de Autorización de Comunidad

El Componente CAS permite que una organización virtual exprese una política con respecto a los recursos distribuidos a través de otros sitios.

Un servidor de servicios de autorización de comunidad emite aserciones a los usuarios de organizaciones virtuales, accediéndole los derechos de los recursos de forma controlada. Los servidores reconocen y hacen cumplir las aserciones.

Este componente es diseñado para ser extensible para múltiples servicios.

IV.6.2 Data Management (Manejo de Datos)

IV.6.2.1 Replica Location Service (RLS): Servicio de Localización de Réplica

El componente RLS es uno de los servicios que se encarga del manejo de datos en ambientes Grid, permitiendo el registro y descubrimiento de réplicas de archivos de datos.

El componente RLS es una herramienta que proporciona la capacidad de rastrear una o más réplicas de archivos de datos en ambientes de computación Grid. Los usuarios o servicios registran archivos en el servicio de localización de réplica, para más tarde solicitar la búsqueda de replicas a los servidores del servicio de localización de réplica.

Un servicio de localización de réplica mantiene y proporciona acceso al camino de información sobre la réplica como: ubicación lógica y física, nombre lógico y físico.

IV.6.2.2 GridFTP: Protocolo de Transferencia de Archivo Grid

El componente GridFTP es un protocolo de transferencia de datos confiable, seguro y de alto rendimiento, desarrollado para ser utilizado en ambientes de computación con un gran ancho de banda en redes de área amplia.

El protocolo GridFTP está basado en el protocolo estándar de Internet para la transferencia de datos: FTP (Protocolo de Transferencia de Archivo).

Características del protocolo de transferencia GridFTP:

- Seguridad GSI: mecanismos de seguridad estándar en ambientes Grid, basado en tecnologías PKI (Infraestructura de Clave Pública).
- Transferencia de terceros: donde un usuario establece una transferencia con dos servidores.
- Acceso parcial a archivos.
- Notificación del estado actual de la transferencia: especifica si la transferencia se ha realizado con éxito o si ha fallado.
- Soporte para transferencias de archivos grandes.

- Reutilización del canal de dato: el canal de dato puede permanecer abierto y ser reutilizado si la siguiente transferencia tiene la misma fuente, el mismo destino y las mismas credenciales.
- Instrumentación integrada: indicadores de funcionamiento.
- Transferencias paralelas: flujos paralelos de datos.
- Control del tamaño del Buffer TCP (Protocolo de Control de Transmisiones).
- Cómputo en el servidor.
- Basado en estándar.
- Además, proporciona librerías de desarrollo y APIs para los que deseen agregar funcionalidad GridFTP a sus aplicaciones.

IV.6.2.3 Reliable File Transfer (RFT): Transferencia de Archivo Seguro

El componente RFT es un servicio de transferencia seguro que se encarga de la administración y control de las transferencias de archivos. Usa mensajes SOAP estándar sobre HTTP para enviar y manejar transferencias de terceros.

Este componente de servicio de transferencia de archivo seguro también proporciona una interfaz para controlar varias características de transferencias sobre el canal de control de GridFTP como: el tamaño del buffer TCP y los flujos paralelos de datos.

El recurso de transferencia segura es creado por un usuario enviando una solicitud de transferencia al servicio RFT. El recurso es creado después que el usuario es correctamente autorizado y autenticado.

La implementación del servicio RFT proporciona operación para el control y manejo de las transferencias.

Características del Servicio RFT:

- Eliminación de archivos: elimina un conjunto de archivos y directorios.
- Respaldo de transferencia: cuando una transferencia falla es recuperada.
- Transferencia Toda o Nada.
- Permiso de transferencia: los permisos del archivo se restaurarán en el destino una vez que el archivo es transferido con éxito.
- Transferencias concurrentes.
- Transferencias recursivas.
- Flujos paralelos de datos.
- Tamaño del buffer TCP.
- Autenticación.
- Autorización.
- Soporte para transferencia de archivos de tipo binario y ASCII.
- Reporte de errores y fallas.

IV.6.2.4 Open Grid Services Architecture - Data Access and Integration (OGSA-DAI): Arquitectura de Servicios Grid Abiertos - Integración y Acceso a Datos

El componente OGSA-DAI proporciona una estructura de servicio de datos Java para acceder e integrar recursos de datos como: archivos, bases de datos XML y bases de datos relacionales en ambientes de computación Grid.

Este componente puede cambiar en futuras versiones de Globus Toolkit.

Características del componente OGSA-DAI:

- Acceso a datos.
- Múltiples recursos de datos, accedido por un solo servicio.
- Permite la transformación de datos.
- Facilita la integración de datos de varias fuentes.
- Permite agregar o extender la funcionalidad dentro del OGSA-DAI.

IV.6.2.5 Data Replication Service (DRS): Servicio de Réplica de Datos

El componente DRS permite a los usuarios del Grid identificar una colección de archivos de datos deseados existente en su entorno de computación Grid para hacer réplicas locales de esos archivos de datos y transferirlos a una o más fuentes locales y registrar las nuevas replicas en el servicio de localización de réplica.

El servicio de réplica de datos también proporciona un recurso llamado replicador (Replicator Resource) que se encarga de informar sobre el estado de la transacción de la actividad de réplica solicitada y permite a los usuarios monitorear el estado del recurso. También, depende de una interfaz a bases de datos estándar JDBC (Conectividad de Bases de Datos Java).

Además, este servicio de réplica de datos proporciona APIs para crear e iniciar operaciones de réplica, soporta transporte seguro, conversación segura y comunicación de mensajes seguros.

Este componente puede cambiar en futuras versiones de Globus Toolkit.

IV.6.3 Execution Management (Manejo de Ejecución)

IV.6.3.1 Pre - Web Service Grid Resource Allocation Management (Pre-WS GRAM): Manejo de Asignación de Recursos Grid de Servicios Pre - Web

El componente Pre-WS GRAM proporciona una interfaz para solicitar y usar recursos de sistemas remotos para la ejecución y control de tareas remotas. El uso más común de este componente es el control y ejecución de tareas y aplicaciones remotas.

Se diseñó principalmente para proporcionar una interfaz uniforme y flexible para sistemas planificadores de tareas, como: Condor, Portable Batch System, Sun Grid Engine y Load Share Facility.

Además, selecciona bajo qué cuenta la tarea remota se ejecutará. También, proporciona flujos de datos de salida y de errores durante la ejecución de las tareas.

IV.6.3.2 Web Service Grid Resource Allocation Management (WS GRAM):

Manejo de Asignación de Recursos Grid de Servicios Web

El componente WS GRAM proporciona un conjunto de servicios Web que permiten localizar, despachar, monitorear y cancelar tareas en recursos computacionales Grid disponibles.

El WS GRAM no es un planificador de tareas, sino un conjunto de servicios y clientes que se comunican con diversos planificadores de tareas usando un protocolo común.

Además, el componente WS GRAM realiza operaciones confiables, maneja credenciales, selecciona la cuenta bajo la cual la tarea remota se ejecutará, publicación del código de salida de la tarea y limpieza de archivos y directorios después de la ejecución de la tarea.

IV.6.3.3 WorkSpace Management (WM): Manejo de Espacio de Trabajo

El componente WM permite a un cliente Grid crear y manejar dinámicamente un espacio de trabajo. Está implementado actualmente como una cuenta Unix en un sitio remoto.

Además, permite que un cliente Grid tenga autorización para crear cuentas individuales o grupos de cuentas. Proporciona un servicio de cuenta que permite que el cliente Grid tenga autorización para manejar propiedades de las cuentas individuales, tal como política de acceso a las cuentas o tiempo de vida de las cuentas.

El servicio de espacio de trabajo consiste en protocolos implementados para la creación y manejo de espacios de trabajos. La creación y manejo de espacio de trabajos puede ser implementada de diferentes formas, según políticas y preferencias del sitio, a través de un comando Unix o de un archivo gridmap, donde la asignación de cuentas es autorizada basada en una lista de control de acceso.

Este componente puede cambiar en futuras versiones de Globus Toolkit.

IV.6.3.4 Globus Teleoperations Control Protocol (GTCP): Protocolo de Control de Teleoperaciones Globus

El componente GTCP es una interfaz de servicio básico para telecontrol, es usado para controlar simulaciones computacionales heterogéneas y físicas en experimentos de ingeniería de terremotos y movimientos telúricos.

También, es usado para controlar los sistemas de adquisición de datos y cámaras de alta resolución durante los experimentos de ingeniería de terremotos. Además, en menor grado se ha usado para controlar la posición y la longitud focal del microscopio de electrón en una aplicación de neurología.

El componente GTCP proporciona dos interfaces de servicio, la primera usada por clientes para controlar instrumentos remotos y simulaciones distantes y la segunda es una interfaz plugin para facilitar la integración de nuevas plataformas de simulación computacional o físicas para el servidor GTCP.

Este componente puede cambiar en futuras versiones de Globus Toolkit.

IV.6.3.5 Community Scheduler Framework (CSF): Estructura Planificadora de Comunidad

El componente CSF es una estructura meta-planificador construida sobre Globus Toolkit 4.0 que proporciona interfaces y herramientas para los usuarios Grid para despachar, controlar y monitorear tareas, crear reservaciones anticipadas y definir políticas planificadoras en el nivel Grid.

Al usar CSF los usuarios Grid pueden tener acceso a diferentes planificadores de tareas, tales como: Condor, PBS, SGE y LSF a través de una sola interfaz.

Este componente puede cambiar en futuras versiones de Globus Toolkit.

IV.6.4 Information Services: Servicios de Información

IV.6.4.1 Pre-Web Service Monitoring and Discovery System (Pre-WS MDS): Sistema de Descubrimiento y Monitoreo de Servicios Pre - Web

El componente Pre-WS MDS proporciona un mecanismo estándar para publicar y descubrir el estado de un recurso e información sobre su configuración. También, proporciona una interfaz flexible y uniforme para la colección de datos por los proveedores de información de bajo nivel.

Además, el sistema de descubrimiento y monitoreo de servicios Pre - Web posee una estructura descentralizada que le permite escalar.

Este componente va a ser eliminado en futuras versiones de Globus Toolkit.

IV.6.4.2 Trigger Service (TS): Servicio Disparador

El servicio disparador recibe pedidos de cualquier recurso que esté en el Grid y si el pedido coincide con las reglas definidas por el administrador, entonces ejecuta ciertas acciones. Un ejemplo del uso del servicio disparador es el envío de un correo electrónico cuando una cola de un recurso computacional se hace más larga del tamaño definido. Este servicio se dispara activando un correo electrónico que es enviado a un administrador para que tome alguna acción para aliviar la distribución de tareas del recurso.

Además, este componente monitorea las propiedades de los recursos en el Grid a través de APIs.

IV.6.4.3 Index Service (IS): Servicio Índice

El componente IS colecciona, monitorea y descubre información de recursos Grid y los publica en un único sitio. Generalmente, se espera que una organización virtual utilice uno o más servicios índices que coleccionan datos sobre todos los recursos Grid disponibles dentro de la organización virtual.

El servicio índice emerge como un grupo de servicios donde las listas registran los recursos junto con información dinámica coleccionada de estos recursos.

IV.6.4.4 Web Service Monitoring and Discovery System (WS MDS): Sistema de Descubrimiento y Monitoreo de Servicios Web

El componente WS MDS permite a los usuarios finales monitorear información a través de una interfaz de un navegador Web estándar sin tener que instalar ningún software adicional en su computadora.

Este componente está implementado como un servlet, que usa una interfaz de plugin para reunir la información de monitoreo y transformarla a XSLT que permite especificar hojas de estilos que aplican un formato de datos complejos XML para la representación en HTML y presentar los datos al usuario en una forma legible.

Los administradores de sitios Web pueden personalizar sus propios despliegues WS MDS usando opciones de formato HTML, configurando diferentes plugin para la colección de datos y transformarlo a XSLT, creando sus propios plugin y transformaciones XSLT.

IV.6.5 Common Runtime: Componentes de Ejecución Común

IV.6.5.1 Extensible IO (XIO): Extensible Entrada y Salida

El componente XIO es una librería de entrada y salida extensible de C desarrollada para Globus Toolkit 4.0.

Este componente proporciona un API que soporta múltiples protocolos como: TCP, UDP, file, HTTP, GSI, TELNET y colas de trabajo.

También XIO proporciona una estructura para el manejo de errores, la entrega asincrónica de mensajes y tiempo de salidas.

IV.6.5.2 C Common Libraries (CCL): Librerías Comunes de C

El componente CCL proporciona una capa de abstracción para tipos de datos utilizados frecuentemente, llamadas al sistema libc que es el conjunto de todas las funciones de librería que están especificadas en los estándares de ANSI C y estructuras de datos usadas en Globus Toolkit 4.0 útiles para las aplicaciones que usa esta herramienta de software.

Es altamente recomendada como librería básica en cualquier aplicación que utilice las herramientas Globus Toolkit.

IV.6.5.3 Java Web Services Core (Java WS Core): Centro de Servicios Web Java

El componente Java WS Core es una implementación de la estructura de recursos de servicios Web (Web Services Resource Framework), que incluye el tiempo de vida de los recursos de servicios Web y las propiedades de los recursos de servicios Web. Además, este componente incluye la notificación de servicios Web, que es una colección de especificaciones que permiten configurar un servicio Web como un elaborador de notificaciones, permitiendo notificar si ocurre algún cambio en el servicio Web.

El Java WS Core proporciona APIs y herramientas para construir servicios Web en Globus Toolkit.

También, este componente tiene soporte básico del servidor y cliente HTTP/1.1 y Tomcat 4.1 y 5.0.

IV.6.5.4 C Web Services Core (C WS Core): Centro de Servicios Web C

El componente C WS Core proporciona un conjunto de herramientas básicas en C para crear servicios Web permitiendo la estructura de recursos de servicios Web y clientes que conforman las especificaciones de los recursos de servicios Web y las notificaciones de servicios Web.

Este componente proporciona algunas características como servicio API y soporte HTTP/1.1.

IV.6.5.5 Python Web Services Core (Python WS Core): Centro de Servicios Web

Python

El componente Python WS Core proporciona una herramienta básica python para crear servicios Web permitiendo la estructura de recursos de servicios Web, probado para interoperar con la estructura de recursos de servicios Web Java.

Este componente puede cambiar en futuras versiones de Globus Toolkit.

IV.7 Ventajas y Desventajas de Globus Toolkit 4.0

IV.7.1 Ventajas

- Herramienta de software para la creación y desarrollo de ambientes de computación Grid.
- Herramienta de software de amplios servicios y funcionalidad para ambientes de computación Grid.
- Herramienta de software de arquitectura abierta.
- Herramienta de software de código abierto.
- Soporte para un gran número de tecnologías estándar.
- Proporciona un alto nivel de seguridad en todos sus servicios.
- Soporte para la integración con otras herramientas de software de computación Grid y para ampliar su desarrollo.
- Herramienta de software ampliamente utilizada en la mayoría y en los más importantes Grid actualmente disponibles alrededor del mundo.
- No se necesita una supercomputadora para que se pueda ejecutar Globus Toolkit 4.0, sino una computadora con características básicas.

IV.7.2 Desventajas

- Requiere de gran variedad de software para conseguir instalar Globus Toolkit 4.0
- Dificil instalación de Globus Toolkit 4.0.
- No posee interfaces gráficas para realizar su instalación, sino a través de líneas de comandos.
- Requiere realizar muchas configuraciones de sus componentes.
- No posee interfaces gráficas para realizar las configuraciones de sus componentes, sino a través de líneas de comandos.
- Algunas configuraciones de sus componentes son difíciles de realizar.
- Manejo técnico de Globus Toolkit 4.0 complejo.
- Se necesita un cierto grado de conocimientos en administración, instalación y configuración de sistemas de computación para poder manejar Globus Toolkit 4.0.
- Se necesita tener conocimientos en sistemas operativos Linux o Unix.
- Una persona sin conocimientos o con conocimientos básicos de computación, le será muy difícil manejar y trabajar con Globus Toolkit 4.0.
- La documentación de Globus Toolkit 4.0 no está muy bien detallada, deja muchas partes al conocimiento previo del usuario.
- No posee soporte técnico.

CAPÍTULO V

V. INSTALACIÓN Y CONFIGURACIÓN BÁSICA DE GLOBUS TOOLKIT 4.0 [18]

V.1 Requerimientos para la instalación de Globus Toolkit 4.0

Para la instalación de la herramienta de software Globus Toolkit 4.0 hay ciertas especificaciones de software y componentes de hardware.

V.1.1 Requerimientos de hardware

La herramienta de software de Globus Toolkit 4.0 está creada sobre los lenguajes de programación Java y C para sistemas operativos Linux y Unix, por lo tanto se requerirá de una computadora capaz de soportar y ejecutar estas plataformas.

Además, la computadora debe tener las siguientes características en sus componentes: velocidad del procesador de 1Ghz o mayor, memoria RAM 256 Megabytes mínima y espacio en el disco duro disponible de 500 Megabytes mínimo.

V.1.2 Requerimientos de software

Se debe tener el siguiente software en la computadora donde se va a instalar Globus Toolkit 4.0:

- El programa instalador de Globus Toolkit 4.0 descargado del sitio Web del proyecto Globus.

- J2SE 1.4.2 SDK o una versión mayor: Edición estándar de la plataforma Java 2 que proporciona un completo ambiente para el desarrollo de aplicaciones Java.
- Ant 1.5.1 o una versión mayor: Herramienta diseñada en Java, usada normalmente en programación durante la fase de compilación. Tiene la ventaja de no depender del sistema operativo en que se está ejecutando.
- Compilador C: Programa que traduce el código fuente de un programa escrito en el lenguaje de programación C y C++ a código ejecutable.
- GNU Tar: Programa que es utilizado para crear o extraer archivos tar.
- GNU sed: Herramienta para el manejo de archivos de texto, usada para filtrar texto de un archivo.
- zlib 1.1.4: Librerías de compresión y descompresión de archivos.
- GNU Make: Herramienta que controla la generación de archivos ejecutables y otros archivos no fuentes de un programa.
- Sudo: Comando que permite a un usuario ejecutar comandos con permisos del administrador del sistema (root).
- Base de datos que reconozca JDBC (Conectividad de Bases de Datos Java), por ejemplo PostgreSQL 7.1 o una versión mayor: Sistema de base de datos relacional.

V.2 Instalación de Globus Toolkit 4.0

Para la instalación de Globus Toolkit 4.0 se necesitará tener tres cuentas de usuarios disponibles en el sistema donde va a ser instalado, además se debe tener un directorio disponible donde se va instalar todo el software de Globus Toolkit 4.0.

Las cuentas de usuarios que se necesitan son: una cuenta del administrador del sistema (root), una cuenta de un usuario normal y una cuenta globus la cual debe pertenecer al grupo globus.

La cuenta del administrador del sistema (root) se necesitará para crear las cuentas del usuario normal, la del usuario y grupo globus, además de crear el directorio donde va a ser instalado Globus Toolkit 4.0 y realizar otros trabajos en el proceso de instalación y configuración de Globus Toolkit 4.0.

La cuenta del usuario normal se usará para ejecutar los programas clientes de Globus Toolkit 4.0.

La cuenta globus se usará para instalar, configurar y realizar diferentes tareas de administración de Globus Toolkit 4.0.

V.2.1 Crear la cuenta del usuario normal

Como root, ejecutar la siguiente línea de comando para crear la cuenta del usuario normal con el nombre de ignacio:

```
root# adduser -m ignacio
```

Se creará una cuenta de usuario con el nombre ignacio, este usuario tendrá su directorio principal en */home/ignacio/*.

V2.2 Crear la cuenta del usuario y grupo globus

Como root, ejecutar la siguiente línea de comando para crear la cuenta del usuario globus:

```
root# adduser -m globus
```

Se creará una cuenta de usuario con el nombre globus, este usuario tendrá su directorio principal en */home/globus/*.

Como root, ejecutar la siguiente línea de comando para crear el grupo globus:

```
root# groupadd globus
```

Se creará en el sistema un grupo llamado globus el cual va pertenecer el usuario globus.

Para cambiar al usuario globus al grupo globus se debe ejecutar la siguiente línea de comando:

```
root# usermod -g globus globus
```

V.2.3 Crear el directorio donde se instalará Globus Toolkit 4.0

Como root, crear un directorio donde se va a instalar Globus Toolkit 4.0 con la siguiente línea de comando:

```
root# mkdir /usr/local/globus-4.0.0
```

Se ha escogido el directorio ***/usr/local*** para crear el directorio donde se instalará Globus Toolkit 4.0, pero podría ser cualquier otro sitio donde el usuario globus sea propietario y tenga permiso de lectura y escritura sobre el directorio de instalación.

Como al crear el directorio ***/usr/local/globus-4.0.0*** se hizo como el usuario root, el usuario globus no es propietario y no tiene permiso de lectura y escritura en el directorio. Como root, cambiar de propietario y de grupo el nuevo directorio creado al usuario globus y al grupo globus respectivamente:

```
root# chown globus /usr/local/globus-4.0.0/
```

Cambia de propietario el nuevo directorio creado.

```
root# chgrp globus /usr/local/globus-4.0.0/
```

Cambia de grupo el nuevo directorio creado.

V2.4 Definir variables de entorno

Para la instalación de Globus Toolkit 4.0 se deben definir las variables GLOBUS_LOCATION y JAVA_HOME en el sistema, que son las variables que contiene el camino donde va a estar la instalación de Globus Toolkit 4.0 y donde está instalado el software de Java respectivamente.

V.2.4.1 Definir la variable GLOBUS_LOCATION

Como usuario root, globus y normal (usuario ignacio), ejecutar la siguiente línea de comando para definir la variable GLOBUS_LOCATION, la cual va a contener el camino completo donde se va a instalar Globus Toolkit 4.0.

Además, la variable `GLOBUS_LOCATION` necesita estar ya definida en las tres cuentas para cuando se vaya a realizar la configuración básica de Globus Toolkit 4.0:

```
export GLOBUS_LOCATION=/usr/local/globus-4.0.0/
```

Para verificar si se definió correctamente, ejecutar la siguiente línea de comando:

```
echo $GLOBUS_LOCATION
```

Como resultado, debe aparecer el camino completo del directorio donde va a ser instalado Globus Toolkit 4.0. En nuestro caso será el siguiente camino:

```
/usr/local/globus-4.0.0/
```

V.2.4.2 Definir la variable `JAVA_HOME`

También, se debe asegurar que la variable `JAVA_HOME` esté definida en el sistema donde se va a ser la instalación de Globus Toolkit 4.0, la cual debe contener el camino completo del directorio donde está instalado el software de Java.

Para comprobar si la variable `JAVA_HOME` está definida, ejecute la siguiente línea de comando como root:

```
root# echo $ JAVA_HOME
```

Como resultado debe aparecer el camino completo del directorio donde está instalado el software de Java.

En el caso de no estar definida la variable `JAVA_HOME`, como root, ejecutar la siguiente línea de comando para definirla:

```
root# export JAVA_HOME=/camino_del_directorio_de_java
```

V.2.5 Descargar el instalador de Globus Toolkit 4.0

Para instalar el Globus Toolkit 4.0 se debe descargar primero el instalador del sitio Web de Globus Toolkit disponible en <http://www.globus.org/toolkit> a la cuenta de usuario globus que es de donde se va a instalar el software Globus Toolkit 4.0, ya que es la opción recomendada.

Es nuestro caso se instalará y se configurará la opción de código fuente de Globus Toolkit 4.0.

www.bdigital.ula.ve

V.2.6 Desempaquetar el instalador de Globus Toolkit 4.0

Como globus, ejecutar la siguiente línea de comando para desempaquetar el instalador de Globus Toolkit 4.0:

```
globus$ tar xvfz gt4.0.0-all-source-installer.tar.gz
```

A continuación se descomprimirá y desempaquetará todo el software de instalación de Globus Toolkit 4.0.

V.2.7 Proceso final de instalación de Globus Toolkit 4.0

Como globus, después de desempaquetar el instalador se debe cambiar al directorio **gt4.0.0-all-source-installer** con la siguiente línea de comando, que es el directorio donde están los instaladores de Globus Toolkit 4.0:

globus\$ cd gt4.0.0-all-source-installer

Después, ejecutar las siguientes líneas de comando para terminar el proceso de instalación básica y recomendada:

globus\$./configure --prefix=\$GLOBUS_LOCATION

Este comando ejecuta la configuración por defecto de Globus Toolkit 4.0 a ser instalado.

globus\$ make

Después de ejecutar el comando ***make***, se empezará a instalar todo el software por defecto de Globus Toolkit 4.0, esta instalación puede tomar entre 2 y 3 horas en terminar.

Si la ejecución del comando ***make*** finalizó con éxito, ejecute la siguiente línea de comando para terminar completamente la instalación:

globus\$ make install

Después de haber terminado de ejecutarse el comando ***make install***, la instalación de Globus Toolkit 4.0 ya ha sido completada totalmente.

V.3 Configuración básica de Globus Toolkit 4.0

A continuación se van a realizar varios pasos de configuración que son necesarios para el correcto funcionamiento de Globus Toolkit 4.0.

Esta configuración comprende la configuración básica de seguridad que obtiene los certificados de la computadora y del usuario que va a trabajar con Globus

Toolkit 4.0. Autoriza al usuario para que tenga acceso a Globus Toolkit 4.0 e inicia el Proxy necesario para que el usuario pueda ejecutar trabajos si dispone de recursos computacionales necesarios para realizarlos.

V.3.1 Configuración básica de la seguridad de Globus Toolkit 4.0

Esta configuración es primordial para el correcto funcionamiento de Globus Toolkit 4.0 y su seguridad, ya que si no se realiza, o si se hace de forma incorrecta, no se podrá lograr que Globus Toolkit 4.0 funcione correctamente. Esto se debe a que la configuración de seguridad es el proceso de autorización de la computadora y del usuario que van a utilizar Globus Toolkit 4.0, sin esto el usuario no podrá iniciar el Proxy que es el que permitirá utilizar los recursos computacionales que dispone.

Además, esta configuración es la que proporciona seguridad para el usuario y sus recursos, ya que es la que identifica quién está autorizado para trabajar en Grid y a cuáles recursos puede acceder.

A continuación se realizará el proceso de configuración básica de seguridad de Globus Toolkit 4.0 en el orden correcto.

V.3.1.1 Establecer variables de entorno

Se debe establecer las variables de entorno en el sistema, para que el sistema conozca dónde se encuentran los comandos que se instalaron.

Además, se debe ejecutar el código del script: globus-user-env.sh o .csh.

Como globus, ejecutar la siguiente línea de comando para establecer la variable GLOBUS_LOCATION donde se instaló Globus Toolkit:

Para shell bourne:

```
globus$ export GLOBUS_LOCATION=/usr/local/globus-4.0.0/
```

Para shell csh:

```
globus$ setenv GLOBUS_LOCATION/usr/local/globus-4.0.0/
```

Como globus, ejecutar el siguiente script: \$GLOBUS_LOCATION/etc/globus-user-env.sh o \$GLOBUS_LOCATION/etc/globus-user-env.csh, dependiendo de shell que se esté usando.

Para shell bourne se usa el siguiente script .sh:

```
globus$ .$GLOBUS_LOCATION/etc/globus-user-env.sh
```

Para shell csh se usa el siguiente script .chs:

```
globus$ source $GLOBUS_LOCATION/etc/globus-user-env.csh
```

V.3.1.2 Obtener los certificados

Se debe tener certificados X.509 para utilizar el software de Globus Toolkit 4.0 de manera segura.

Los certificados consisten en:

- Dos archivos: hostcert.pem y hostkey.pem.
- Un directorio adecuado para los servicios seguridad: /etc/grid-security/.
- La computadora debe tener un nombre estático en DNS, no se debe utilizar una computadora que utilice DHCP, es decir, el nombre de la computadora no debe ser dinámico.

Para obtener los certificados, en nuestro caso se va a utilizar el software SimpleCA, que suministra certificados de usuarios y para la computadora donde se instaló Globus Toolkit 4.0, esto es suficiente para los servicios Grid básicos.

Como globus, ejecutar la siguiente línea de comando para ejecutar el script que instalará un nuevo SimpleCA, sólo se necesita ejecutar una sola vez el script por Grid:

`globus$ $GLOBUS_LOCATION/setup/globus/setup-simple-ca`

Al ejecutar esta línea de comando se mostrará información sobre la computadora que se está utilizando, como se muestra en la Figura V.1, la cual debe responder si está o no de acuerdo.

```
The unique subject name for this CA is:
cn=Globus Simple CA, ou=simpleCA-canagua, ou=GlobusTest, o=Grid
Do you want to keep this as the CA subject (y/n) [y]:
```

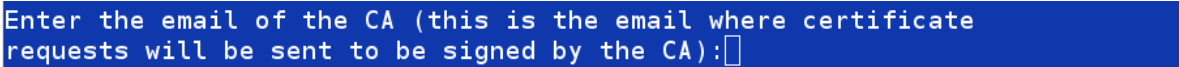
Figura V.1 Información de la computadora

Escriba “y” para dejar la configuración por defecto y además, es la recomendada.

Descripción de los nombres de los componentes:

- cn: representa el nombre común (common name), que identifica este certificado en particular que es el Globus Simple CA.
- ou: representa la unidad organizacional (organizational unit), que identifica este CA de otros CAs creados por SimpleCA por otras personas, el segundo ou especifica el tipo que es de prueba de Globus (GlobusTest).
- o: representa la organización identificada con el Grid.

Luego se mostrará información solicitando el correo electrónico, como se muestra en la Figura V.2.



```
Enter the email of the CA (this is the email where certificate
requests will be sent to be signed by the CA):
```

Figura V.2 Dirección de correo electrónico

Coloque una dirección de correo electrónico real para recibir la solicitud del certificado.

Luego se mostrará información sobre el tiempo de vida del certificado, como se muestra en la Figura V.3.

```
The CA certificate has an expiration date. Keep in mind that
once the CA certificate has expired, all the certificates
signed by that CA become invalid. A CA should regenerate
the CA certificate and start re-issuing ca-setup packages
before the actual CA certificate expires. This can be done
by re-running this setup script. Enter the number of DAYS
the CA certificate should last before it expires.
[default: 5 years (1825 days)]:
```

Figura V.3 Tiempo de vida del certificado

Este es el número de días que el certificado CA es válido. Acepte la opción por defecto, la cual es la recomendada y le dará un tiempo de vida de 5 años del certificado.

Luego se mostrará información solicitando una contraseña, como se muestra en la Figura V.4.

```
Enter PEM pass phrase:
Verifying - Enter PEM pass phrase:
```

Figura V.4 Clave o Contraseña

Escriba una contraseña y luego vuelva a verificar la contraseña, la contraseña no debe contener ningún espacio.

Después de escribir y verificar la contraseña, el proceso de obtención de certificado a través de SimpleCA ha terminado.

V.3.1.3 Configuración Grid Security Infrastructure (GSI)

Para la configuración de GSI se debe ejecutar el siguiente script:
`$GLOBUS_LOCATION/setup/globus_simple_ca_CA_Hash_setup/setup-gsi – default.`

Donde **`CA_Hash`** es una cadena de caracteres única que se genera para la configuración de seguridad de Globus Toolkit 4.0.

Como root, ejecutar la siguiente línea de comando:

```
root# $GLOBUS_LOCATION/setup/globus_simple_ca_CA_Hash_setup/setup-gsi –default
```

La salida de la línea de comando anterior debe ser la misma que se observa en la

Figura V.5.



```
setup-gsi: Configuring GSI security
Making /etc/grid-security...
mkdir /etc/grid-security
Making trusted certs directory: /etc/grid-security/certificates/
mkdir /etc/grid-security/certificates/
Installing /etc/grid-security/certificates//grid-security.conf.fe39cb2b...
Running grid-security-config...
Installing Globus CA certificate into trusted CA certificate directory...
Installing Globus CA signing policy into trusted CA certificate directory..
.
setup-gsi: Complete
```

Figura V.5 Configuración GSI

Al obtener esta salida en la pantalla, se ha completado correctamente la configuración de GSI de Globus Toolkit 4.0.

V.3.1.4 Solicitud del certificado para la computadora

Se debe solicitar certificados para las computadoras y copiar los certificados dentro del directorio adecuado de los servicios de seguridad.

Los certificados deben ser solicitados sólo para computadoras que posean un nombre estático no dinámico y que ya esté instalado Globus Toolkit 4.0.

Como root, ejecute la siguiente línea de comando para solicitar el certificado para la computadora que está utilizando:

```
root# $GLOBUS_LOCATION/bin/grid-cert-request – host “el nombre de la computadora”
```

Después de ejecutar la línea de comando anterior, se crearán los siguientes archivos: hostkey.pem, hostcert_request.pem y el archivo vacío hostcert.pem.

Todos estos archivos se localizarán en el directorio **/etc/grid-security/**

V.3.1.5 Firma del certificado de la computadora

Como globus, ejecutar la siguiente línea de comando para generar el archivo hostsigned.pem:

```
globus$ $GLOBUS_LOCATION/bin/grid-ca-sign -in /etc/grid-security/  
hostcert_request.pem –out hostsigned.pem
```

Después de ejecutar la línea de comando anterior se le solicitará la contraseña que usted escribió en la configuración del SimpleCA.

Luego se creará un archivo llamado `hostsigned.pem` en el directorio actual, que debe ser movido al directorio `/etc/grid-security/` con el nombre de `hostcert.pem` que reemplazará al archivo vacío que va a ser el certificado de la computadora.

Como root, ejecute la siguiente línea de comando:

```
root# mv /home/globus/hostsigned.pem /etc/grid-security/hostcert.pem
```

Asumimos que el archivo `hostsigned.pem` está en el directorio casa del usuario `globus`, si no se encuentra ahí, se debe colocar el camino exacto donde se encuentre.

V.3.1.6 Solicitud del certificado de usuario

Se debe solicitar un certificado de usuario que es el que va a permitir utilizar los servicios clientes de Globus Toolkit 4.0.

Como el usuario normal, en nuestro caso `ignacio`, ejecute la siguiente línea de comando:

```
ignacio$ $GLOBUS_LOCATION/bin/grid-cert-request
```

Después de ejecutar la línea de comando anterior se le solicitará el nombre del usuario, una contraseña y verificar la contraseña.

Luego se crearán los siguientes archivos: `userkey.pem`, `usercontent_request.pem` y el archivo vacío `usercontent.pem`. Estos archivos se localizarán en el directorio oculto `.globus` dentro del directorio casa del usuario normal (usuario `ignacio`), en nuestro caso: `/home/ignacio/.globus`

V.3.1.7 Firma del certificado de usuario

Como globus, ejecutar la siguiente línea de comando para generar el archivo signed.pem:

```
ignacio$ $GLOBUS_LOCATION/bin/grid-ca-sign -in /home/ignacio/.globus/  
usercert_request -out signed.pem
```

Después de ejecutar la línea de comando anterior, se le solicitará la contraseña que usted escribió en la configuración del SimpleCA.

Luego se creará un archivo llamado signed.pem en el directorio actual que debe ser movido al directorio **/home/globus/.globus/** con el nombre de usercert.pem, que reemplazará al archivo vacío el cual va a ser el certificado del usuario.

Como globus, ejecutar la siguiente línea de comando para copiar el archivo al directorio temporal de la computadora y después al directorio oculto **.globus** dentro del directorio casa del usuario ignacio.

Este procedimiento es necesario, ya que el usuario globus no tiene permiso de escritura en el directorio casa del usuario ignacio.

```
globus$ cp signed.pem /tmp/
```

Asumimos que el archivo signed.pem está en el directorio casa del usuario globus, si no se encuentra en el directorio casa del usuario, se debe colocar el camino exacto donde se encuentra.

Como ignacio, ejecutar la siguiente línea de comando:

```
ignacio $ cp /tmp/signed.pem /home/ignacio/.globus/usercert.pem
```

V.3.1.8 Hacer accesible los certificados de la computadora al container

El container debe ser ejecutado como el usuario globus, para esto se deben crear dos archivos que pertenezcan a dicho usuario.

Como root, ubicarse en el directorio */etc/grid-security/* ejecutando la siguiente línea de comando:

```
root# cd /etc/grid-security/
```

Luego ejecutar la siguiente línea de comando para obtener los archivos que terminen en .pem:

```
root# ls -l *.pem
```

Después de ejecutar la línea de comando anterior se le mostrará la siguiente información en la pantalla como se muestra en la Figura V.6.

```
-rw-r--r-- 1 root root 2660 Jul 6 15:05 hostcert.pem  
-rw-r--r-- 1 root root 1384 Jul 6 14:58 hostcert_request.pem  
-r----- 1 root root 887 Jul 6 14:58 hostkey.pem
```

Figura V.6 Archivos con extensión pem

Además de estos archivos .pem, se deben crear dos nuevos archivos: el containerkey.pem y el containercert.pem, copiándolos de los archivos hostkey.pem y hostcert.pem respectivamente, ejecutando las siguientes líneas de comando:

```
root# cp hostkey.pem containerkey.pem
```

```
root# cp hostcert.pem containercert.pem
```

Los nuevos archivos deben ser cambiados de usuario y de grupo a globus, ejecutando las siguientes líneas de comando:

Cambia al usuario globus:

```
root# chown containerkey.pem containercert.pem
```

Cambia al grupo globus:

```
root# chgrp containerkey.pem containercert.pem
```

Para comprobar que el proceso se realizó correctamente, ejecute nuevamente la siguiente línea de comando:

```
root# ls -l *.pem
```

Ahora se le mostrará adicional los dos archivos nuevos que se crearon pertenecientes al usuario y al grupo globus, como se muestra en la Figura V.7.

```
-rw-r--r-- 1 globus globus 2660 Jul 6 15:50 containercert.pem
-r----- 1 globus globus  887 Jul 6 15:50 containerkey.pem
-rw-r--r-- 1 root  root  2660 Jul 6 15:05 hostcert.pem
-rw-r--r-- 1 root  root  1384 Jul 6 14:58 hostcert_request.pem
-r----- 1 root  root   887 Jul 6 14:58 hostkey.pem
```

Figura V.7 Nuevos archivos con extensión pem

Lo cual indica, que el proceso de hacer accesible los certificados de la computadora al container se ha realizado correctamente y ya se puede ejecutar el container de Globus Toolkit 4.0 a través del usuario globus.

V.3.1.9 Autorización del usuario

Para autorizar al usuario normal (ignacio) para ejecutar los programas clientes de Globus Toolkit 4.0, se debe crear el archivo `grid-mapfile` en el directorio `/etc/grid-security/` como el usuario `root`, el cual diferencia las cuentas de los usuarios que están autorizados y los que no lo están, para utilizar los servicios de Globus Toolkit 4.0.

Se necesitan los datos del usuario que va a utilizar Globus Toolkit 4.0 y el nombre de la cuenta del usuario para crear el archivo.

Como el usuario normal, ejecutar las siguientes líneas de comando para obtener la información que se necesita:

```
ignacio$ $GLOBUS_LOCATION/bin/grid-cert-info -subject
```

Se debe obtener una salida que suministre información de los datos donde se instaló Globus Toolkit 4.0, el tipo de instalación y el usuario que va utilizar los programas clientes de Globus Toolkit 4.0, como se muestra en la Figura V.8.

```
/O=Grid/OU=GlobusTest/OU=simpleCA-canagua/CN=ignacio
```

Figura V.8 Información sobre la instalación realizada

Para obtener el nombre de la cuenta del usuario, ejecute la siguiente línea de comando:

```
ignacio$ whoami
```

El comando anterior debe retornar el nombre de la cuenta del usuario que está usando para utilizar los programas clientes de Globus Toolkit 4.0, en nuestro caso retorna **ignacio**.

Ahora, se debe crear el archivo grid-mapfile con la información que se obtuvo.

Como root, ejecutar la siguiente línea de comando para agregar la autorización creando el archivo **/etc/grid-security/grid-mapfile**:

```
root# $GLOBUS_LOCATION/sbin/grid-mapfile-add-entry -dn "/O=Grid/
```

```
OU=GlobusTest/OU=SimpleCA-canagua/CN=ignacio" -In ignacio
```

Después de ejecutar la línea de comando anterior se debe obtener una salida parecida a la que se muestra en la Figura V.9.

```
Modifying /etc/grid-security/grid-mapfile ...  
/etc/grid-security/grid-mapfile does not exist... Attempting to create /etc/  
/grid-security/grid-mapfile  
New entry:  
"/O=Grid/OU=GlobusTest/OU=simpleCA-canagua/CN=ignacio" ignacio  
(1) entry added
```

Figura V.9 Crear archivo grid-mapfile

Para comprobar si se agregó la autorización correctamente, ejecute la siguiente línea de comando como root para saber si creó el archivo grid-mapfile:

```
root# cat /etc/grid-security/grid-mapfile
```

Se debe obtener el archivo ya creado y la información que se agregó, una salida parecida a la mostrada en la Figura V.10.

```
"/O=Grid/OU=GlobusTest/OU=simpleCA-canagua/CN=ignacio" ignacio
```

Figura V.10 Archivo grid-mapfile

Al llegar a este punto, la configuración de agregar autorización para el usuario se ha realizado completa y correctamente.

Se puede agregar a cualquier otro usuario del sistema siempre y cuando posea certificados de autorización para el uso de Globus Toolkit 4.0

V.3.1.10 Verificando el inicio del Proxy

Para terminar la configuración básica de Globus Toolkit 4.0, se ejecuta la siguiente línea de comando para iniciar el Proxy para comprobar si funciona correctamente y si la configuración de seguridad se realizó correctamente.

Como el usuario normal (ignacio), ejecute la siguiente línea de comando:

```
ignacio$ $GLOBUS_LOCATION/bin/grid-proxy-init -debug -verify
```

Después de ejecutar la línea de comando se le solicitará la contraseña que escribió en la solicitud de certificado de usuario.

Después de escribir la contraseña correcta se le mostrará la información como se muestra en la Figura V.11.

```
ignacio@canagua ignacio $ $GLOBUS_LOCATION/bin/grid-proxy-init -debug -verify
User Cert File: /home/ignacio/.globus/usercert.pem
User Key File: /home/ignacio/.globus/userkey.pem

Trusted CA Cert Dir: /etc/grid-security/certificates

Output File: /tmp/x509up_u1004
Your identity: /O=Grid/OU=GlobusTest/OU=simpleCA-canagua/CN=ignacio
Enter GRID pass phrase for this identity:
Creating proxy ...+++++++
.....+++++++
Done
Proxy Verify OK
Your proxy is valid until: Thu Jul 7 03:38:33 2005
```

Figura V.11 Ejecución del Proxy

Que informa que el Proxy se creó correctamente en la computadora y además el tiempo hasta cuándo es válido el Proxy. El tiempo de vida del Proxy es de 12 horas por defecto.

3.2 Sobre la configuración básica de Globus Toolkit 4.0 realizada

Al realizar los pasos de configuración básica de Globus Toolkit 4.0 anteriores, sólo se logra el funcionamiento básico de Globus Toolkit 4.0.

Para el completo funcionamiento de un Grid, no sólo basta instalar el software de Globus Toolkit, sino otros software como los programas planificadores de tareas, que se encargan de ejecutar las tareas o trabajos que despachen los usuarios del

Grid. Además, se debe disponer de recursos computacionales necesarios que se encarguen de procesar y realizar las tareas o trabajos de los usuarios.

La herramienta de software Globus Toolkit solo es una parte de muchas en el funcionamiento de un Grid y si no se posee todas las partes necesarias, no se podrá crear una plataforma Grid.

www.bdigital.ula.ve

CONCLUSIONES Y RECOMENDACIONES

En este trabajo de investigación realizado sobre la computación Grid y sus herramientas se puede concluir lo siguiente:

La computación Grid es una tecnología que surgió hace poco tiempo. Sin embargo, se ha desarrollado y evolucionado rápidamente convirtiéndose en una rama de la computación muy prometedora para un futuro. Muchas personas confían que esta nueva tecnología va más allá de una moda y seguramente será la próxima generación de la computación.

La computación Grid surgió como una solución a los problemas que requieren uso intensivo de recursos computacionales, ya que esta tecnología proporcionará un gran espacio de almacenamiento para guardar de forma segura cualquier tipo de información, además de poseer todo el tiempo un gran poder de cómputo disponible para la organización o institución, logrando todo esto con los recursos computacionales actualmente disponibles y a bajo costo.

Actualmente la computación Grid tiene muy limitadas sus capacidades, esto se debe a que todavía le falta por desarrollarse y evolucionar. Aún falta que los recursos computacionales estén disponibles en cualquier lugar, en cualquier momento, en forma confiable y segura, que sea transparente y fácil de usar. Estas son metas que aun se encuentran distantes, pero se va en buen camino. Un ejemplo de esto, es que hoy en día todavía no existe un único Grid en el

mundo, como ocurre en el caso de la Web, ni tampoco un Grid dominante. Sin embargo, cada vez hay más Grids alrededor del mundo que están desarrollándose de una forma muy rápida, lo cual indica que se va camino al Grid global.

Hoy en día la implementación de la computación Grid es complicada y se requiere de muchos detalles técnicos para que funcione. Se necesita de personas capacitadas en administración y configuración de sistemas (Linux y/o Unix) y manejo de recursos computacionales distribuidos (clusters de computadoras), además la documentación que se encuentra disponible actualmente para su implementación no es la óptima y muy poca se encuentra en español. Todo esto es por los momentos, ya que la computación Grid se encuentra en una fase de desarrollo y cada vez se avanza más rápido en su evolución.

Muchas personas están interesadas en el desarrollo y evolución de la computación Grid, se ha invertido y se siguen invirtiendo muchos recursos en ella y seguramente esta tecnología en unos años superará las limitaciones actuales.

Las herramientas de computación Grid realizan diferentes trabajos dentro de una plataforma Grid y se dividen en varios tipos. Cada tipo realiza un trabajo en particular y es una parte primordial dentro del Grid, deben trabajar en conjunto para lograr el correcto funcionamiento del Grid.

Estas herramientas necesitan de recursos computacionales para poder llevar a cabo su implementación, ya que fueron creadas para manejar un gran número de computadoras distribuidas en organizaciones e instituciones. Sin estos recursos las herramientas de computación Grid no lograrán cumplir su tarea para la cual fueron creadas.

Hoy en día, en la computación Grid se viene desarrollando la herramienta de software Globus Toolkit. Esta herramienta se está convirtiendo en el estándar de facto de la computación Grid. El encargado de desarrollar estándares para la computación Grid es el Foro Grid Global. Actualmente el estándar del Grid es el OGSA (Arquitectura de Servicios Grid Abiertos) y concuerda que la herramienta Globus Toolkit sea el estándar para la computación Grid. Esta herramienta ofrece muchos de los servicios básicos dentro de una plataforma Grid como: descubrimiento, acceso, manejo y monitoreo de recursos, manejo de datos, comunicación, portabilidad y seguridad. Las plataformas Grid que no utilicen esta herramienta seguramente se quedarán aisladas al no ofrecer interoperabilidad con otros Grid y por lo tanto serán Grids confinados y limitados.

En cuanto a las recomendaciones sobre la computación Grid y sus herramientas, lo más importante es que cada organización o institución que desee poner en marcha una plataforma Grid, debe realizar un estudio objetivo sobre si en verdad la necesita y si posee los recursos computacionales para implementarla, ya que esta tecnología aprovecha y utiliza los recursos actuales

que poseen las organizaciones e instituciones y de esta forma reduce los gastos de adquisición de nuevos recursos.

Además, se debe saber si habrá algún beneficio en compartir los recursos computacionales de la organización o institución con otras organizaciones e instituciones.

Por ultimo, la principal recomendación es que se debe tener muy en cuenta esta nueva tecnología y sus herramientas, ya que en un futuro seguramente se utilizará en la mayoría de las áreas de investigación, ingeniería, empresarial y comercial, y el Grid será tan notorio como hoy día lo es la Web.

www.bdigital.ula.ve

REFERENCIAS BIBLIOGRÁFICAS

[1] Joanne Woodcock y otros, 2001. Diccionario de Informática e Internet de Microsoft. España. McGraw-Hill.

[2] GridCafe. Disponible en: <http://gridcafe.web.cern.ch/gridcafe/>

[3] Gilberto Díaz, Primer Taller de Adiestramiento de Grid en Venezuela, Introducción a la Computación Grid. Disponible en:
http://www.cecalc.ula.ve/tallergrid/conceptos_grid.pdf

[4] Ian Foster y Carl Kesselman, 1998. The Grid: Blueprint for a New Computing Infrastructure. USA. Morgan Kaufmann Publishers.

[5] Vanessa Hamar, Herbert Hoeger, Víctor Mendoza y Yubirin Ramírez. Grids en la ULA. Disponible en: <http://www.ing.ula.ve/~hhoeger/Grid-ULA-Vzla.pdf>

[6] Luís Ferreira, Viktors Bertstis y otros, 2002. Redbook: Introduction to Grid Computing with Globus. Disponible en:
<http://www.redbooks.ibm.com/redbooks/pdfs/sg246895.pdf>

[7] Bienvenidos a "The Grid". Disponible en:
http://www.usando.info/main_file.php/us_busca/8413/

[8] The Grid. Disponible en:
<http://www-306.ibm.com/e-business/la/evolving/ondemand/technology/grid.shtml>

[9] Computación distribuida. Disponible en:
http://100cia.com/enciclopedia/Computaci%F3n_distribuida

[10] Ian Foster, Carl Kesselman y Steven Tuecke. The Anatomy of the Grid, Enabling Scalable Virtual Organizations. Disponible en:

<http://www.globus.org/alliance/publications/papers/anatomy.pdf>

[11] Borja Sotomayor, 2004. Introducción a la computación Grid. Disponible en:

<http://www.e-ghost.deusto.es/docs/IntroduccionGrid.pdf>

[12] Algunos conceptos de la computación Grid. Disponible en:

<http://www.justiniano.com/noticias/magazine/MAGAZINE94.html>

[13] La computación en Grid se abre camino en el IFIC. Disponible en:

http://ific.uv.es/rei/Paseo/REI_Grid_3.pdf

[14] Foro Grid Global. Disponible en: <http://www.ggf.org>

[15] Santiago Banhero, Mariano Felice, Jorge Gonzales y Pablo Lavallén.

¿Qué es la computación Grid? Disponible en:

http://www.unlu.edu.ar/~tyr/tyr/TYR-trab/2004/computacion_grid-banhero-otros.pdf

[16] Proyecto Condor. Disponible en: <http://www.cs.wisc.edu/condor/>

[17] Wikipedia enciclopedia libre, Condor. Disponible en:

http://en.wikipedia.org/wiki/Condor_cycle_savenger

[18] Proyecto Globus. Disponible en: <http://www.globus.org>

[19] Proyecto GridPort. Disponible en: <https://gridport.net>

[20] Una extensión del proyecto GridPort. Disponible en:

<https://gridport.npaci.edu/>

[21] Freddy Rojas, Primer Taller de Adiestramiento de Grid en Venezuela,

Portales Grid: GridPort 3. Disponible en:

<http://www.cecalc.ula.ve/tallergrid/gridport.pdf>

[22] PBS Professional. Disponible en:

<http://www.altair.com/software/pbspro.htm>

[23] Open PBS. Disponible en: <http://www.openpbs.org>

[24] Proyecto Sun Grid Engine. Disponible en: <http://gridengine.sunsource.net>

[25] Vanessa Hamar, Primer Taller de Adiestramiento de Grid en Venezuela, Grid Engine. Disponible en:

http://www.cecalc.ula.ve/tallergrid/sun_grid_engine.pdf

[26] José Arturo García Monroy, Globus Toolkit. Disponible en:

<http://internetng.dit.upm.es/joe/Art/Globus.pdf>

www.bdigital.ula.ve

REFERENCIAS RECOMENDADAS

[1] GridCafe. Disponible en: <http://gridcafe.web.cern.ch/gridcafe/>. Sitio Web que proporciona gran variedad de información sobre la computación Grid, como conceptos básicos, historia y proyectos Grid explicada de una manera sencilla.

[2] Proyecto Globus. Disponible en: <http://www.globus.org>. Sitio Web oficial de la herramienta de software Globus Toolkit que proporciona información completa sobre toda la herramienta, como conceptos, componentes y manuales para utilizar.

[3] Foro Grid Global. Disponible en: <http://www.ggf.org>. Sitio Web oficial del Foro Grid Global, proporciona toda la información sobre la estandarización de la computación Grid.

[4] Centro de Información de Computación Grid. Disponible en: <http://www.gridcomputing.com>. Sitio Web que proporciona una gran variedad de enlaces a sitios Web relacionados con la computación Grid.

[5] GridStart. Disponible en: <http://www.gridstart.org>. Sitio Web que proporciona gran variedad de información sobre la computación Grid, como noticias, eventos y proyectos Grid.

[6] RedBooks. Disponible en:

<http://www.redbooks.ibm.com/redbooks.nsf/redbooks/>. Sitio Web que proporciona gran cantidad de trabajos sobre la computación Grid y tecnologías informáticas, disponibles en formato Adobe PDF para verlos en línea y/o descargarlos.

www.bdigital.ula.ve