

DESCRIPCIÓN DEL DESARROLLO Y ALCANCES DE LA COMPUTACIÓN GRID

PEDRO PABLO PEDRAZA MOGOLLON

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE ELECTRICA Y ELECTRONICA
ESPECIALIZACION EN TELECOMUNICACIONES
BUCARAMANGA

2005

DESCRIPCIÓN DEL DESARROLLO Y ALCANCES DE LA COMPUTACIÓN GRID

PEDRO PABLO PEDRAZA MOGOLLON

TRABAJO DE GRADO

Director

Julio Cesar Baracaldo

Ingeniero de Sistemas

Especialista en Auditoria de Sistemas

Candidato a Magíster en Ciencias de la Información y la Comunicaciones

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE ELECTRICA Y ELECTRONICA
ESPECIALIZACION EN TELECOMUNICACIONES
BUCARAMANGA

2005

A Dios, por permitirme existir junto a una gran familia, por todo lo bueno y malo que he vivido.

A mis Abuelos que aunque no están todos conmigo me han enseñado a valorar lo que tengo.

A mis Padres por darme la confianza, el apoyo, la comprensión en los buenos y malos momentos. Por motivarme y respaldarme en la búsqueda de mis metas y por las buenas enseñanzas que he recibido de ellos a lo largo de toda mi vida.

A mis hermanos y su familia por buscar siempre lo mejor para mí al darme la comprensión y su apoyo durante toda mi vida.

AGRADECIMIENTOS

El autor expresa su agradecimiento a:

LA UNIVERSIDAD INDUSTRIAL DE SANTANDER, SU CUERPO ADMINISTRATIVO Y DOCENTES

MARLY ANGULO, Ingeniero de Sistemas. Por su orientación en la elección del proyecto y su colaboración durante su desarrollo.

JULIO CESAR BARACALDO, Ingeniero de Sistemas. Por su desinteresada y valiosa colaboración durante el desarrollo del proyecto.

Y en general a todas aquellas personas que de una u otra manera hicieron parte de la buena marcha del proyecto al colaborar en su realización.

GLOSARIO

ADI	Abstract Device Interface
Agente	Es cualquier entidad que percibe su entorno a través de sensores y actúa sobre su entorno mediante efectores
API	Application Programming Interface
ASP Provider	Proveedor de Servicio de Almacenamiento - Application Service
C/S	Cliente/Servidor
C2D	Computing-to-Data ratio. Relación entre capacidad de procesamiento necesaria para utilizar un trabajo, y la cantidad de datos o código que se desean distribuir
CGI	Common Gateway Interface
CoGKits	Commodity Grid Kits
CSCW	Computer Supported Cooperative Work
DAIS	Data Access and Integration
FTP	File Transfer Protocol
GASS	Global Access to Secondary Storage
GFLOPS	Unidad de medida para referirse a Miles de Millones de instrucciones en coma flotante
GGF	Global Grid Forum
Globus Empresas Project	Proyecto desarrollado por varias universidades, y administraciones públicas de los Estados Unidos.
GPDK	Grid Portal Development Kit
GPT	Globus Packaging Technology
GRAM	Globus Resource Allocation Manager
GRB	Grid Resource Broker

GSI	Grid Security Infraestructure
GSS	Global Service Security
HBM	Heartbeat Monitor
HPC	High Performance Computing
HTC	High Throughput Computing
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secured
Intranet	Conjunto de computadores que comparten recursos bajo un ambiente web pero a nivel interno de la organización
ISP	Internet Service Provider
JSP	Java Servlet Pages
LDAP	Lightweight Directory Access Protocol -
MPI	Message Passing Interface
MDS	Monitoring and Discovery Service
MPICH	Message Passing Interface
OGSA	Open Grid Service Architecture
OV	Organización Virtual
P2P	Peer to Peer
QoS	Quality Of Service
RSL	Resource Specification Language
SDK	Software Development Kit
SI	Sistema de Información
SMP	Sistemas de Multiprocesamiento
SSH	Secure Shell
SSL	Secure Socket Layer (Protocolo que suministra seguridad a aplicaciones de protocolos (HTTP, FTP, Telnet, NNTP)
SSP	Proveedor de Servicio de Aplicaciones - Storage Service Provider
TI	Tecnología de la Información
UNICORE	UNiform Interface to COmputer REsources
VPN	Red Virtual Privada – Virtual Private Network

TITULO. Descripción del Desarrollo y Alcances de la Computación Grid*

PEDRO PABLO PEDRAZA MOGOLLON**

Global Grid Forum

Open Grid Service Architecture

Open Grid Security Infraestructure

Globus Toolkit

Con el mejoramiento de las características de los computadores se busca satisfacer las necesidades de procesamiento y almacenamiento de los usuarios e investigadores en sus labores cotidianas. Este objetivo no ha sido alcanzado porque a medida que aumenta la capacidad de las maquinas aumentan los requerimientos de los usuarios.

La Computación Grid permite perfeccionar conceptos ya existentes como Computación distribuida, procesamiento paralelo; para maximizar los recursos disponibles. Además, con esta tecnología se busca compartir recursos de forma controlada y reglamentada aprovechando al 100% la capacidad de procesamiento de los computadores y diferentes dispositivos existentes dentro de una organización al dividir las grandes tareas de procesamiento en pequeñas tareas que son enviadas desde un computador central a los diferentes equipos que conforman la Grid, para ser procesadas en el tiempo libre de estos computadores; luego de procesadas se devuelven los datos obtenidos para dar solución al proceso. Para el desarrollo de esta tecnología empresas como IBM, HP, SUN han

* Monografía

* Especialización en Telecomunicaciones - Facultad Eléctrica y Electrónica – Julio Cesar Baracaldo

invertido gran cantidad de recursos de forma organizada. Hay un grupo especial encargado de estandarizarla para poder ser usada masivamente al cual se ha denominado Global Grid Forum. El principal problema al que se enfrenta esta tecnología es el establecimiento de políticas de acceso a los recursos por parte de usuarios quienes no forman parte de la organización.

La computación Grid permite a la gente ahorrar tiempo en las labores de procesamiento, dinero porque ellos no tienen que invertir en equipos de alto procesamiento ni en equipos a los cuales se pueda tener acceso mediante la red sin importar su ubicación, Aunque esta tecnología no se ha logrado poner en funcionamiento esta tecnología ya hay proyectos en el área financiera, medicina, física y otras, que han obtenido buenos resultados y que espera su popularidad como Internet pueda venir en corto tiempo.

TITULO. Descripción del Desarrollo y Alcances de la Computación Grid*

PEDRO PABLO PEDRAZA MOGOLLON**

Global Grid Forum

Open Grid Service Architecture

Open Grid Security Infraestructure

Globus Toolkit

With the improvement of the characteristics of the computers, it looks for to satisfy the necessities of indicting and storage of users and investigators in the quotidian tasks. This objective has not been reached because according that increase the capacity of the machines it increases the investigations of the users.

The Computing Grid permits to perfect concepts already existing as Distribute Computation, similar indicting to maximize the available resources. With this technology it looks for to share resources in a controlled way and regulated taking advantages the hundred per cent the capacity of indicting of the computers and different mechanisms existing inside an organization to divide the big tasks of indicting in small tasks that are sent since a central computer to the different equipments that take part of the Grid, to be processed in the free time of these computers. Afterwards indicted, gotten data a came back to solve the process. To development of this technology companies as IBM, HP, SUN have invert big amount of resources in an organized way. There is a special group who is under take of this technology to be used massively which has named Global Grid Forum. The main problem of this technology is the establishment of access politics to the resource by means of users who do not take part of the organization.

The Calculation Grid lets people to save time in process tasks, money because they do not have to invert equipment of high process, neither equipment which you can access though the web without taking care of its place. Although this technology has not been functioned yet, there are already projects in areas, as financial, medicine, physics and others, which have gotten good results and its popularity can came in a short time as the Internet.

* Monografía

* Especialización en Telecomunicaciones - Facultad Eléctrica y Electrónica – Julio Cesar Baracaldo

LISTA DE FIGURAS

	pág
Figura 1. Evolución Tecnológica	11
Figura 2. Fruto de la Necesidad	16
Figura 3. Origen y Desarrollo del Grid Computing	17
Figura 4. Aplicaciones de la computación Grid	18
Figura 5. Arquitectura de la Grid	24
Figura 6. Paralelo entre Arquitectura Grid y Protocolo TCP	30
Figura 7. Características de un Grid Computacional	31
Figura 8. Arquitectura del Middleware	36
Figura 9. Evolución tecnológica en las herramientas	38
Figura 10. Capas de la Arquitectura Globus GT3	40
Figura 11. Funcionamiento del Globus Toolkit	42
Figura 12. Capas de la Arquitectura Globus GT4	43
Figura 13. Comparativo entre la Arquitectura Actual y la Arquitectura Grid para S.I.	52
Figura 14. O/V participantes en la red IRIS de diseño y desarrollo	59

CONTENIDO

	pág.
INTRODUCCION	3
1. MARCO TEORICO	5
1.1 ANTECEDENTES	5
1.1.1 Computación distribuida	5
1.1.2 Procesamiento simétrico y/o paralelo	7
1.1.3 Clustering	7
1.1.4 Trabajo Colaborativo – CSCW	7
1.2 COMIENZO DEL GRID	8
1.2.1 Problemática inicial para surgimiento de Grid	8
1.2.2 Fundamentos de Grid	9
1.2.3 Fisiología de la Grid	10
1.2.4 Anatomía de la Grid	12
1.2.5 Proyecto SETI@home	13
1.2.6 Proyecto Condor	14
1.3 DEFINICION DE COMPUTACION GRID	14
1.4 DESARROLLO DE LA COMPUTACION GRID	16
1.5 ÁREAS DE APLICACIÓN DE LA COMPUTACION GRID	17
1.5.1 Supercomputación distribuida	18
1.5.2 Sistemas distribuidos en tiempo real	19
1.5.3 Proceso intensivo de datos	21
1.5.4 Entornos virtuales de colaboración	21
1.5.5 Servicios puntuales	21
2. ARQUITECTURA GRID	23
2.1 Propuesta de Ian Foster	23

	pág.
2.2 Especificaciones técnicas	32
3. HERRRAMIENTAS Y ESTRATEGIAS ACTUALES	35
3.1 Plataformas de Grid	35
3.2 TOOLKITS	38
3.2.1 Globus	39
3.2.2 Legion	42
3.2.3 Unicore	43
3.2.4 Gridbus	44
3.3 PLANIFICADORES	45
3.2.1 Condor-G	47
3.2.2 Nimrod-G	47
3.4 ENTORNOS DE DESARROLLO	48
3.5 PORTALES GRID	48
3.6 IBM	50
3.7 ORACLE	51
3.8 Sun Microsystems	51
4. APLICACIONES ACTUALES	56
5. VENTAJAS DE LA COMPUTACION GRID	60
6. DESVENTAJAS DE LA COMPUTACION GRID	65
7. EL FUTURO DE LA COMPUTACION GRID	66
8. IMPACTO DE LA COMPUTACIÓN GRID EN EL SECTOR ACADÉMICO EN COLOMBIA	68
5. CONCLUSIONES	75
BIBLIOGRAFIA	79

INTRODUCCION

Desde la aparición de los equipos de cómputo el hombre ha encontrado en la tecnología informática una pieza clave en el camino para encontrar la solución a muchos de sus problemas. Con la solución han llegado también nuevos problemas pues la mente del hombre nunca esta conforme y busca aprovechar cada vez mas todo lo que se encuentra a su alrededor.

Los requerimientos computacionales del ser humano cada vez son mayores, debido principalmente al nivel de dependencia cada vez mayor en el computador para las tareas cotidianas y todo lo relacionado con el manejo mismo de la información.

Las metas investigativas del hombre cada vez son mayores y ha puesto en evidencia la insuficiencia de equipo tecnológico para poder procesar los grandes requerimientos de potencial computacional y el almacenamiento de la gran cantidad de datos que generan.

Los recursos necesarios para el procesamiento de los datos necesarios para o producto de las investigaciones, no siempre están ubicados en un mismo sitio por razones económicas, de espacio físico, seguridad o infraestructura. Además, en muchas ocasiones los recursos computacionales con que cuenta una organización, suelen estar subutilizados.

Para lograr una mejor utilización de los recursos propios y ajenos en busca de satisfacer las grandes necesidades de procesamiento y almacenamiento de datos de las organizaciones y en especial de los investigadores, se ha venido dando

origen a un nuevo concepto llamado Grid. Bajo este concepto de Grid se busca administrar de forma eficiente y como un único sistema, los recursos propios y ajenos de usuarios ubicados en forma cercana o dispersos en diversos lugares de la geografía mundial, distribuyendo pequeñas tareas dentro de sus computadores para ser realizadas cuando no se encuentren utilizando el 100% de su capacidad de procesamiento o utilizando sus capacidades extras de almacenamiento.

1 MARCO TEORICO

El término Grid nace como analogía con la expresión “Power Grid” usado en las redes de distribución eléctrica domestica, donde un usuario de forma transparente se conecta a la red cuando y donde lo necesita, sin tener en cuenta el origen. La Computación Grid (Malla de Computadores) se fundamenta bajo un paradigma similar y aparece a mediados de los años 90 con los estudios de Ian Foster y Carl Kesselman.

1.1 ANTECEDENTES

Las raíces de esta tecnología datan desde que se crearon las primeras redes de computadores que buscaban aprovechar al máximo los recursos disponibles en la red. En 1995 se estableció una conexión para compartir recursos entre 17 centros de alto rendimiento en los Estados Unidos. Luego se continuó con estos estudios pero se incremento el número de equipos conectados llegando a 30,000 en 1999, este concepto fue llamado Grid Entropía.

1.1.1 Computación Distribuida. El interés por aprovechar el 100% de los recursos con que se cuenta, ha dado origen a este concepto el cual no es nuevo. A principios de los años 70's cuando se inicio el desarrollo de Internet a través de Arpanet, se contaba con programas como el Creeper y Reaper-Ran que buscaban aprovechar los recursos disponibles. En 1973 Xerox creo un programa que se ejecutaba sobre la red Ethernet y asimilando las veces de un gusano recorría las máquinas usando los recursos que no estaban disponibles¹.

¹ http://gsyc.escet.urjc.es/simple_com/phd-thesis-es/node12.html

La computación distribuida es utilizada para resolver grandes problemas y consiste en dividir un proceso en pequeños procesos a los que se llaman tareas, las cuales van a ser distribuidas en todos los computadores dispersos que se encuentran conectados para compartir aplicaciones, datos y otros recursos de red para funcionar como un único sistema. La utilización de este concepto busca resolver problemas existentes en la computación como²:

- *Potencia de cálculo*: La unión de varios computadores no será superada por una supercomputadora, es decir que aunque la supercomputadora sea muy robusta no igualará la unión de los recursos de muchos computadores que al unirse cada vez más incrementaran su potencial y semejaran una supercomputadora.
- *Tiempo*: La computación distribuida permite que investigaciones o procesos que requieren manejo y procesamiento de gran cantidad de información se haga de forma más rápida, lo que permite obtener resultados en menos tiempo.
- *Almacenamiento*: Al reducir tiempo en el procesamiento de información se incrementa la cantidad de datos que se obtienen, estos datos no necesariamente tienen que ser almacenados en una sola maquina, debido a que su tamaño obligaría a tener una gran infraestructura de almacenamiento aumentando los gastos y el tiempo de acceso.
- *Dinero*: Debido a los altos costos, no todas las organizaciones tienen la capacidad de adquirir máquinas robustas con configuraciones de alto desempeño que les permitan asumir proyectos que requieran este tipo de computadoras. Mediante la computación distribuida las empresas tienen acceso a muchos computadores que son subutilizados y que al interconectarse pueden llegar a tener características superiores a las de un supercomputador, ahorrando costos en adquisición de equipos. mantenimiento, infraestructura, administración.

² <http://www.biologia-en-internet.com/default.asp>

➤ **Colaboración:** El intercambio de información entre organizaciones con proyectos afines permiten que sus investigaciones avancen mucho más rápido, al evitar la redundancia de procesos.

1.1.2 Procesamiento simétrico y/o paralelo. Este tipo de procesamiento permite dividir un proceso en diversos procesos, donde cada uno de estos pequeños procesos son asignados a uno de los procesadores instalados en la maquina. El procedimiento anterior permite que estos pequeños procesos se ejecuten simultáneamente en el tiempo.

1.1.3 Clustering. Es un sistema de procesamiento en el cual se conectan varios computadores de forma coordinada. Estos computadores están conectados de forma física y programática, los cuales realizan tareas comunes y son vistos como un solo computador dentro de la red. Una característica importante de este sistema es que en el caso de presentarse una falla en uno de estos computadores se activa automáticamente un proceso llamado *failover* que tiene como función conmutar la carga de trabajo a otro computador de forma transparente al usuario permitiendo que el proceso continúe normalmente³.

1.1.4 Trabajo Colaborativo – CSCW. Este tipo de sistema cooperativo soportado por un computador, busca generar ambiente de trabajo al unir grupos de personas con fines similares, que se unen para compartir información, ayuda en procedimientos, toma de decisiones y otras actividades, sin tener que disponer de servidores, solamente cuentan con un servidor que se encarga de realizar un índice en el cual se registra el usuario, su ubicación y la información que va a compartir.

³ www.microsoft.com/spanish/MSDN/studiantes/computadores/paralelas/clustering.asp

Este servidor se encarga de mostrarle al usuario que usuario(os) tiene la información que busca para que escoja con quien la va a solicitar. Dentro de este tipo de sistemas se pueden destacar P2P (Peer to Peer) muy usado hoy en día para el intercambio de archivos multimedia y música digital a través de Internet. En este sistema cada máquina de usuario final es un servidor para la red.

Se suele confundir CSCW con Groupware, pero “La forma más sencilla para responder es decir que Groupware es la tecnología y CSCW es la actividad y los métodos y prácticas”⁴

1.2 COMIENZO DEL GRID

1.2.1 Problemática inicial para el surgimiento de Grid. Con el desarrollo de la computación en red, se abren nuevas oportunidades para la realización de nuevos proyectos de investigación y desarrollo en sectores como el industrial, financiero, científico y de otras ramas a nivel personal como el intercambio de archivos, imágenes, música entre otras. La utilización de recursos en estas áreas da origen a grandes volúmenes de datos a ser procesados y almacenados. La capacidad de los equipos de cómputo no era lo suficientemente capaz para realizar estas tareas, se tenía que recurrir a máquinas de configuración robusta trayendo como consecuencia la demora en las investigaciones pues el acceso a estos dispositivos no era fácil debido a problemas como tamaño y costo.

Junto a la necesidad de obtener capacidad y velocidad de procesamiento nace otro problema que es donde almacenar la cantidad de datos que se generan cada día. Por otra parte los PC's utilizados en el hogar y en las organizaciones no son

⁴ <http://www.aet.es/faqs/faq17.htm>

explotados al 100% de su capacidad, demostrando que se necesita de maquinas mas robustas pero a la vez, se sub-utilizan recursos de maquinas de escritorio.

Al establecerse las redes de cómputo dentro de las organizaciones como un medio que permite compartir recursos (Impresora, escáner), intercambiar archivos, se da origen a la posibilidad de explotar los recursos existentes en la organización. Luego con la aparición de Internet se extiende la posibilidad de intercambiar archivos, compartir recursos entre equipos personales, redes. Sin embargo, la compartición de recursos debe estar controlada y reglamentada, en el sentido de determinar quien esta compartiendo qué, a quien se lo esta compartiendo y bajo que condiciones. Todos aquellos entes que se ven cobijados bajo estas reglas dan origen a lo que se conoce como **Organización Virtual** (OV), que es la encargada de generar la productividad.

Las organizaciones virtuales requieren recursos variados, tener relaciones de compartición de recursos y servicios muy flexibles como se realiza actualmente en los sistemas Cliente/Servidor o las redes P2P y niveles de control entre los que se pueden mencionar el control de acceso, delegación y aplicación de políticas.

A pesar de la interconexión vía Internet, se continúa con la raíz del problema que es poder compartir recursos en forma coordinada y dinámica.

1.2.2 Fundamentos de Grid. Los fundamentos de la Grid se pueden resumir en tres aspectos.

1- Compartición de recursos a gran escala. Poder compartir una serie de recursos de forma coordinada entre los usuarios, sin tener en cuenta si la infraestructura para realizar este procedimiento es local o esta ubicada en otra parte del mundo. Entre estos recursos podemos nombrar:

- Computadores que se encargan de proporcionar potencia de cálculo y almacenamiento.
 - Redes de comunicaciones que permiten acceder a nuevos recursos interconectando diversos sistemas.
 - Instrumentos generalmente de carácter científico necesarios en algunas aplicaciones.
 - Datos que se necesario compartir por una comunidad para lograr sus objetivos
 - Instrumentos de medida.
- 2- Organización de recursos distribuidos de varias organizaciones. Los recursos pueden pertenecer a una o varias organizaciones ubicadas en diferentes sitios que tengan sus propios administradores donde se determina que compartir, a quien y las condiciones bajo las cuales se realiza este procedimiento.
- 3- Recurso heterogéneo. Cuando diversas organizaciones pueden disponer de la multitud de elementos para ser compartidos con otros departamentos

1.2.3 Fisiología de la Grid. En el pasado la computación se realizaba bajo un ambiente centralizado principalmente con una estructura cliente/servidor, con la aparición de nuevas tendencias, recursos e infraestructura se dio origen a nuevas herramientas como el Internet que permitía a los usuarios interconectarse desde cualquier lugar del planeta (ver figura 1). La aparición de nuevos desarrollos de software unido a las capacidades de Internet dio cabida a la necesidad de utilizar todos los recursos distribuidos en la red, no solo la red Internet sino los recursos distribuidos en la red interna de la organización⁵.

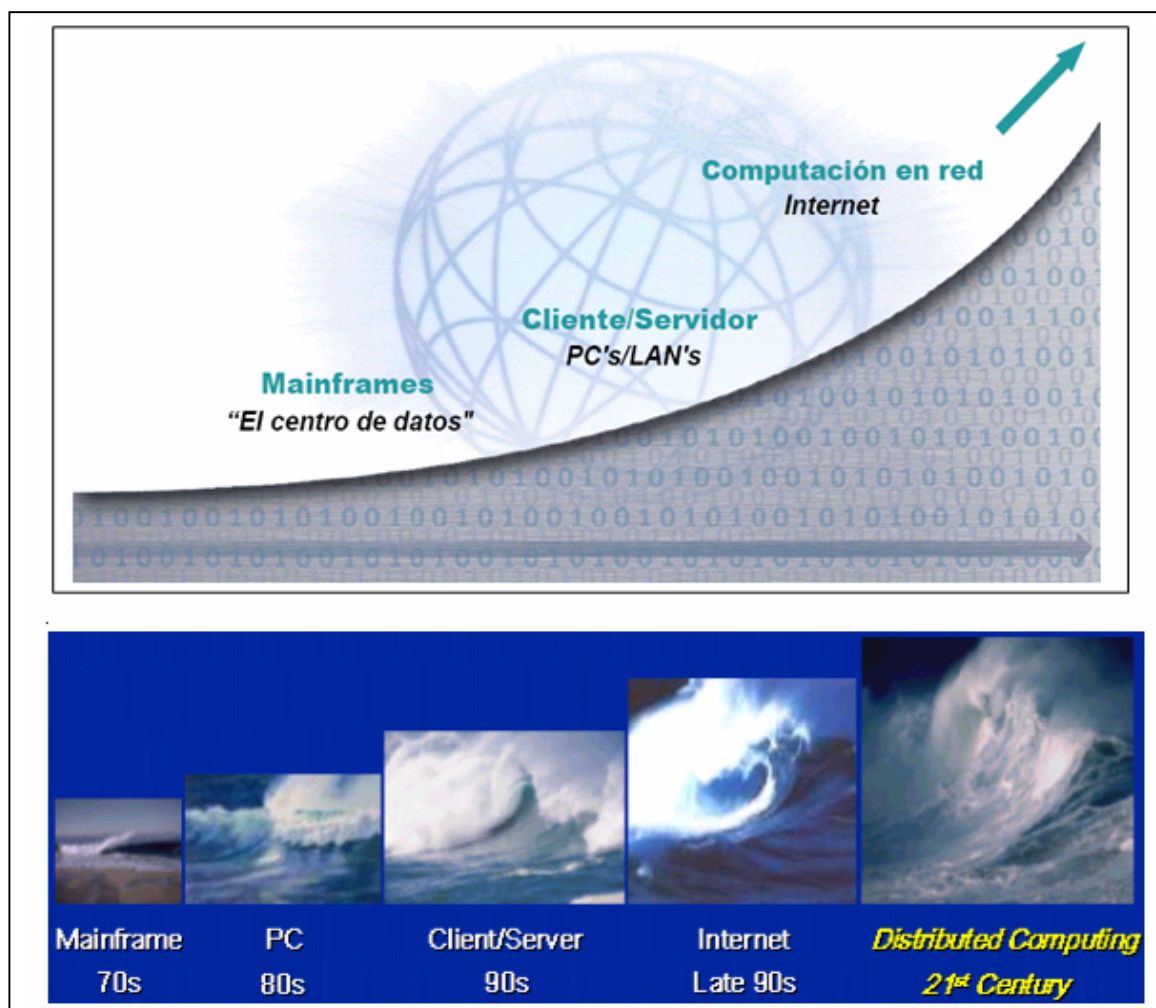
Para poder realizar esta función se dio origen al Globus Toolkit, que es una herramienta que pone en servicio una serie de protocolos, servicios y bibliotecas del software que apoyan la Grid. El toolkit se encamina a solucionar los

⁵ <http://www.globus.org/research/papers/physiology>.

problemas existentes en la Grid como seguridad, descubrimiento de información, ubicación de los recursos, dirección de los datos, comunicación.

La Grid ha tomado gran fuerza en el ambiente de las redes distribuidas como lo es Internet, prestando cada vez más atención a la seguridad de la Grid. Para fortalecer este aspecto se desarrollaron protocolos como el GSI que usa el certificado X.509, generando protección a la autenticación de los usuarios y a la comunicación.

Figura 1. Evolución Tecnológica



Fuente. www.Gridsystems.com

1.2.4 Anatomía de la Grid. La Grid se ha destacado como una tecnología orientada a compartir los recursos disponibles de forma coordinada, dinámica flexible y segura entre individuos, instituciones⁶ y multi-institucionales.

Las organizaciones virtuales permiten a sus miembros compartir recursos de forma que puedan alcanzar una meta compartida. Estas organizaciones difieren entre si dependiendo de el número y tipo de participantes, el tipo de actividad, los recursos que son compartidos.

Para poder cumplir con el objetivo de la Grid como lo es compartir recursos se debe tener en cuenta la existencia de estándares que permitan la extensibilidad, interoperabilidad, portabilidad y compartición del código. Todo esto se logra teniendo una arquitectura de protocolos, esto trae como beneficio la generación de un estandarización de los servicios y a la postre el fortalecimiento de la Grid.

La interoperabilidad entre los participantes de la Grid, permite que los mecanismos usados en el proceso sean los mismos para todos sin tener en cuenta la plataforma ni los ambientes de programación. Se debe tener en cuenta que los protocolos son los encargados de establecer la interacción entre los componentes de la Grid más no los encargados de poner en funcionamiento estos componentes.

Es importante destacar algunos aspectos claves que se obtienen dentro de la Grid como son:

- Calidad del servicio
- Flexibilidad de los costos variables
- Menos complejidad
- Mayor productividad

⁶ <http://www.globus.org/research/papers/anatomy.pdf>

- Disminución de tiempo
- Hacer vs. Comprar

1.2.5 Proyecto SETI@home. Nace en la Universidad de Berkeley en 1999 con el objetivo de buscar vida extraterrestre con base en las grabaciones obtenidas en el telescopio de Arecibo en Puerto Rico⁷. En el servidor los datos recibidos se dividen en segmentos a los que se denominarán unidades de trabajo. Para lograr esta meta este proyecto buscaba aprovechar toda la capacidad inoficiosa de los computadores que se conectan a Internet.

Inicialmente se debe conseguir la autorización del dueño del computador para instalar un software que sirve como protector de pantalla. Este protector se activa cuando el usuario no está usando la máquina para en ese momento empezar a procesar los datos que le son enviados. Luego de procesar los datos se envían los resultados al servidor para poder recibir un nuevo paquete. En caso que el usuario necesite usar su equipo, basta con empezar a trabajar y el proceso entra en receso.

El éxito de este proyecto ha estimulado la creación de otros proyectos @home (en casa), como:

Genome@home	Descifrar el Genoma Humano y su aplicación
Folding@home	Estudio del plegamiento de las proteínas
FightAIDS@home	Estudio para descubrir la cura del SIDA
LHC@home	Estudio para construir el acelerador de partículas
Predictor@home	Estudio de enfermedades relacionadas con proteínas

⁷ www.seti@home.com

1.2.6 Proyecto Condor. Iniciado en la universidad de Wisconsin-Madison. Este proyecto hace hincapié en mostrar que la cantidad de tiempo ocioso en que permanecen los equipos de cómputo personales es alto.

Con el fin de utilizar este periodo de tiempo y alcanzar un rendimiento de procesamiento alto, “Condor proporciona dos funciones importantes. Primero, hace recursos disponibles más eficientes poniendo las máquinas ociosas para trabajar. En segundo lugar, amplía los recursos disponibles para los usuarios, funcionando bien en un ambiente de la propiedad distribuida.”⁸

Condor ofrece tres ventajas muy importantes, la primera los trabajos pueden producir puntos de comprobación lo que permite reiniciar la ejecución a partir de un punto específico, la segunda pueden realizar llamadas del sistema alejado y la tercera puede realizar tareas de modo serial o paralelo. Condor coloca las tareas en una cola y con base en una política determina cuando y donde se realizara el proceso.

Este proyecto permite a una organización pequeña o grande programar el sistema de manera que se inicie la ejecución en las horas de la noche cuando la mayoría de los funcionarios no se encuentra y se detenga en la mañana antes de iniciar labores.

1.3 DEFINICION DE COMPUTACION GRID

La Computación Grid es la agregación de recursos computacionales distribuidos geográficamente como máquinas, datos, software, junto con redes de alta velocidad y cualquier tipo de recurso necesario para realizar un proceso bajo el concepto de computación distribuida para ser utilizados como si fuera un supercomputador.

⁸ <http://www.cs.wisc.edu/condor/>

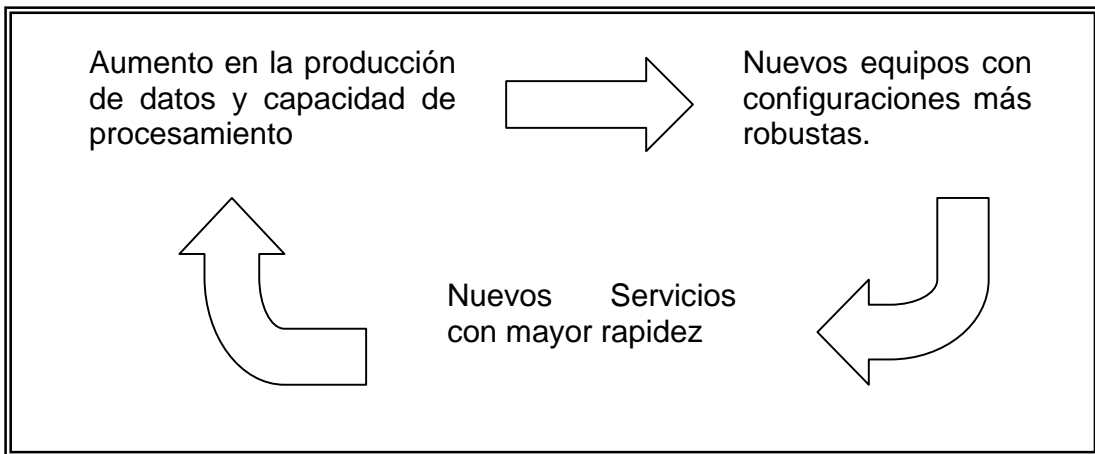
Se necesita una interconexión de hardware junto a un control de software para que un dispositivo forme parte de una Grid con servicios heterogéneos distribuidos entre distintas organizaciones que dan origen a un metacomputador⁹. Además ofrecen aplicaciones que permitan el acceso eficaz a los recursos computacionales, basados en recursos individuales que son compartidos.

Grid nace como fruto de una necesidad (ver Figura 2), para designar una infraestructura distribuida dedicada a la ejecución de aplicaciones científicas y de ingeniería. Tiene la idea clave de trabajo en equipo pero de forma independiente, como un escenario informático ideal en el que los ciclos de baja actividad CPU y espacio de almacenamiento de millones de sistemas informáticos distribuidos a lo largo y ancho de la red mundial, funcionen como una reserva. Permitiendo que estos recursos sean aprovechados por quien lo necesite en un algún momento. Grid se basa en el paradigma de la red eléctrica como un servicio al que se puede acceder de forma transparente.

En el año 2000 la empresa Boeing empieza a utilizar los PC's de su Intranet corporativa como una Grid para realizar los cálculos necesarios en el diseño de los motores de sus aviones, al igual que lo hace Intel para diseñar sus microchips.

⁹ Metacomputador hace referencia a la unión de muchos computadores como si fuera uno ofreciendo gran capacidad de cálculo y múltiples servicios

Figura 2. Fruto de la Necesidad



Fuente. El Autor

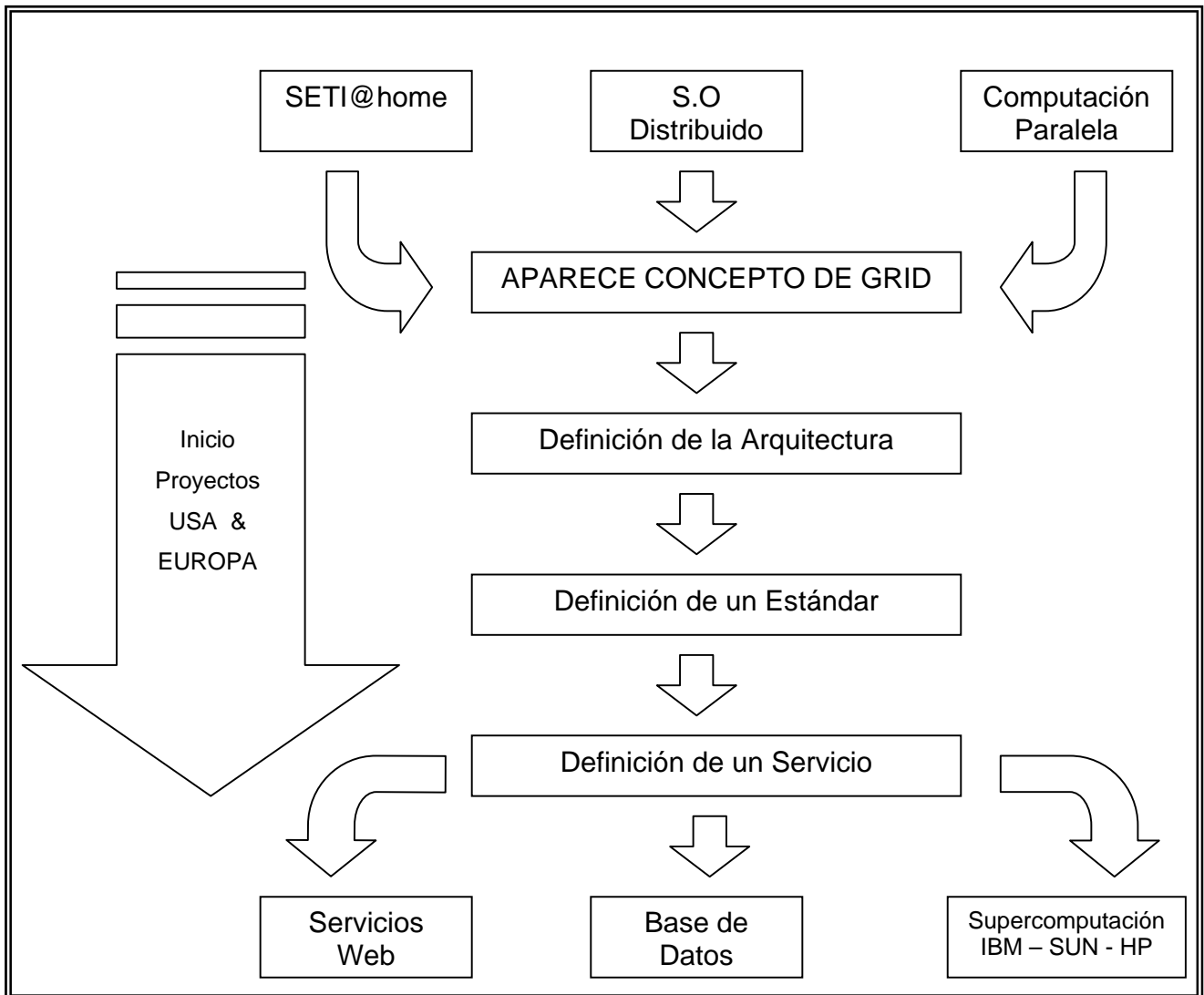
1.4 DESARROLLO DE LA COMPUTACION GRID

Esta tecnología tiene muchas posibilidades de aplicación como el almacenamiento, ayuda a sistemas operativos, aplicaciones, procesos, dispositivos de entrada y salida, manejo de datos. Pero lo más importante es que responde a un problema como es “la compartición coordinada y dinámica de recursos y la resolución de problemas en las organizaciones virtuales multinstitucionales dinámicas”¹⁰

En el proceso de desarrollo de esta tecnología (ver Figura 3), se ha creado el Globus Project, que es un proyecto que involucra un grupo de investigadores con la tarea de desarrollar los conceptos de Computación Grid. Esta tarea de investigación no solo involucra personal académico sino que en la parte comercial a recibido gran apoyo, empresas como IBM, ORACLE, HP, SUN

¹⁰ <http://www.globus.org/research/papers/anatomy.pdf>

Figura 3. Origen y Desarrollo de la Computación Grid



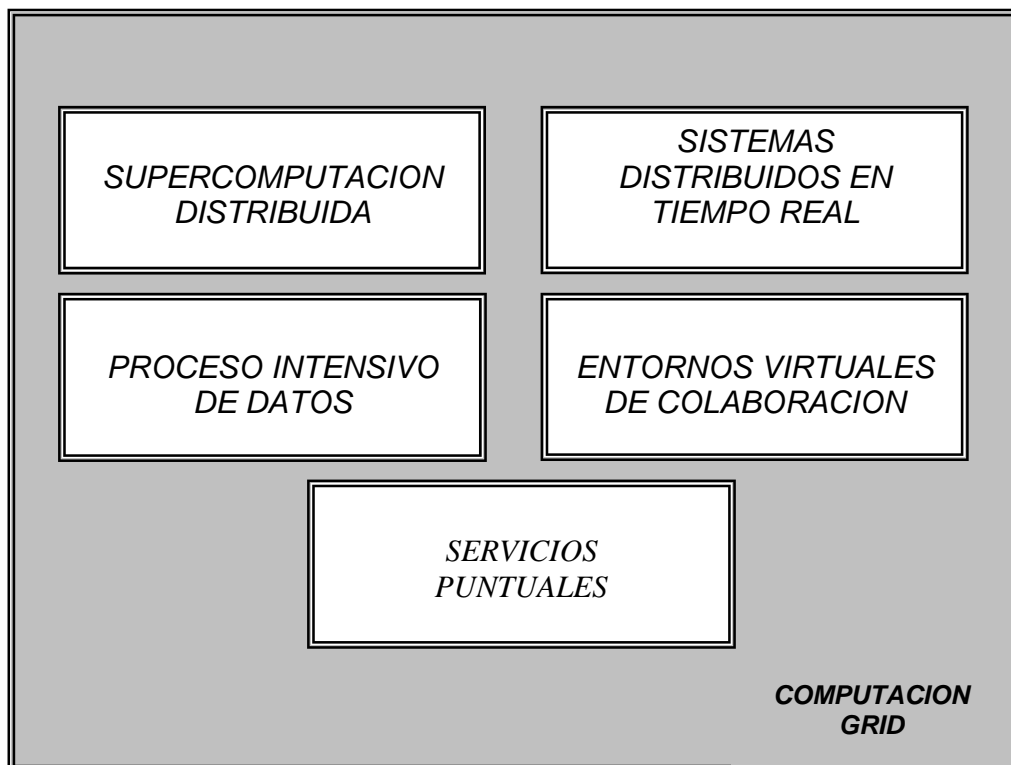
Fuente. El Autor

1.5 ÁREAS DE APLICACIÓN DE LA COMPUTACION GRID

Para solucionar las necesidades de cálculo, espacio de almacenamiento y tiempo de respuesta que presentan los sistemas de procesamiento, se han definido cinco

grandes áreas de trabajo en la Computación Grid¹¹ (ver figura 4)

Figura 4. Aplicaciones de la computación Grid



Fuente. El Autor

1.5.1 Supercomputación distribuida. Dentro de esta área se incluyen todas aquellas aplicaciones cuyas necesidades son imposibles de satisfacer en un único nodo. Estas necesidades requieren el uso de grandes recursos que no están disponibles localmente. Un sistema distribuido debe permitir añadir nuevos servicios como sean necesarios para gestionar todos los recursos.

Es importante considerar que un sistema distribuido consta de dos partes:

¹¹ http://www.ceditec.etsit.upm.es/grid_computing.php

- Nivel físico. Compuesto por equipos de computo distribuidos geográficamente pero que esta interconectados mediante un a red de comunicaciones de cualquier tipo.
- Nivel Lógico. Compuesto por un conjunto finito de procesos y canales que lo comunican.

Esta clase de modelos deben poseer un sistema síncrono para poder lograr un buen funcionamiento y así resolver varios problemas al tiempo con las siguientes características:

- Potencia de cálculo: Ninguna computadora es capaz de procesar la cantidad de datos que se pueden procesar en miles de equipos.
- Tiempo. El tiempo se reduce directamente proporcional a la cantidad de equipos conectados
- Almacenamiento: No existe dentro de la organización una máquina que almacene el volumen de datos que son procesados
- Dinero. No todas las organizaciones están en capacidad de adquirir equipos de alto desempeño
- Colaboración. Poder compartir las investigaciones y avances entre maquinas ubicados a lo largo del planeta.

1.5.2 Sistemas distribuidos en tiempo real. Este tipo de aplicaciones son aquellas que dan origen a un alto flujo de datos a alta velocidad y que se deben analizar inmediatamente para continuar con el proceso. Dentro de estas

aplicaciones podemos citar algunos ejemplos donde se aplica como: tratamiento de imágenes, control remoto de equipos médicos de alta precisión, lectura de instrumentos de medida experimentos de alta energía, sistemas de reservas de líneas aéreas, redes bancarias, experimentos de física de alta energía, control remoto de un recurso no trivial (microscopio, equipo médico).

Para lograr una ejecución exitosa en un sistema distribuido se requieren una serie de servicios que provean la funcionalidad que se requiere.

- **Autenticación**, con este servicio el usuario puede certificar realmente quien es, para poder brindarle los recursos que puede y quiere utilizar. Este servicio es necesario también para la utilización de un recurso y así poder llevar un registro de la hora de petición o ejecución, la clase de servicio entre otras.
- **Scheduling**, servicio de planificación de los recursos, para que sea eficiente la utilización de los recursos en este tipo de entornos distribuidos y sea equitativo entre todos los usuarios. En este servicio se pueden aplicar diversas políticas de administración dependiendo de aspectos como la necesidad de los usuarios, las aplicaciones, el estado dinámico de los recursos.
- **Descubrir Recursos**, este recurso es complementario con el scheduling y es de gran importancia puesto que en un sistema distribuido se pueden agregar y quitar recursos en cualquier momento lo que conlleva a que se debe realizar esta tarea de forma dinámica
- **Caracterización de Recursos**, brinda la información puntual de cada uno de los recursos que están a disposición dentro de la Grid.
- **Reserva Anticipada**, este servicio tiene como objetivo poder ejecutar un grupo de recursos en los que normalmente no es posible hacerlo, para tener éxito debe trabajar junto con el servicio de planificación

- **Acceso a Datos Remotos**, necesario para poder obtener los datos que se necesitan en el momento en que se esta realizando el proceso, se debe tener en cuenta que en algunas ocasiones se necesitara la repetición de algunos datos, razón por la cual aparece el servicio de réplica. Este servicio permite además realizar copias de datos que es muy costoso su acceso o difícil su transporte y que sea conveniente tener en una localización más próxima.

- **Transferencias Rápidas** es de gran importancia y de mucha planificación. Se debe contar con recursos que permitan el acceso de forma más rápida de lo acostumbrado a los datos que son requeridos en procesos en tiempo real

- **Monitorización** es clave para poder controlar la correcta ejecución de todos los trabajos, al igual que controlar la correcta ejecución de los servicios que se encuentren disponibles y corriendo correctamente para su utilización.

1.5.3 Proceso intensivo de datos. Son aquellas aplicaciones que hacen un uso intensivo de los medios físicos de almacenamiento. Las necesidades de almacenamiento son tan grandes que sobrepasan la capacidad en un solo punto, por tanto se debe distribuir esta necesidad a través de la red. Esta distribución trae como beneficio un acceso mucho más rápido a esto datos. En este tipo de aplicaciones los encontramos plasmados en los gestores de Bases de Datos distribuidas.

1.5.4 Entornos virtuales de colaboración. En este sector se encuentra todo lo relacionado con el concepto de Tele inmersión. Este concepto genera entornos virtuales distribuidos en 3D.

1.5.5 Servicios puntuales. Este aspecto es muy diferente a lo que tiene que ver con la velocidad, capacidad de procesamiento o el almacenamiento, se hace

referencia al hecho de conseguir recursos que la organización no quiere comprar o desarrollar porque no los va explotar o son muy costosos. Dentro de este tipo de aplicaciones podemos encontrar todas aquellas que permitan a la organización acceder a recursos de hardware específicos (estos recursos por lo general son equipos muy costosos dedicados a la toma y muestra de datos al igual que equipos de medida muy desarrollados y que ameritan grandes distancias). Se debe tener en cuenta que en este punto no se hace referencia a potencia computacional y no se necesita procesamiento en tiempo real

2. ARQUITECTURA GRID

Con el objetivo de lograr una completa interoperabilidad entre los diferentes fabricantes de dispositivos y tecnologías de red se han generado un conjunto de entidades, normas y protocolos que permitan lograr una estandarización de esta nueva tecnología.

2.1 PROPUESTA DE IAN FOSTER.

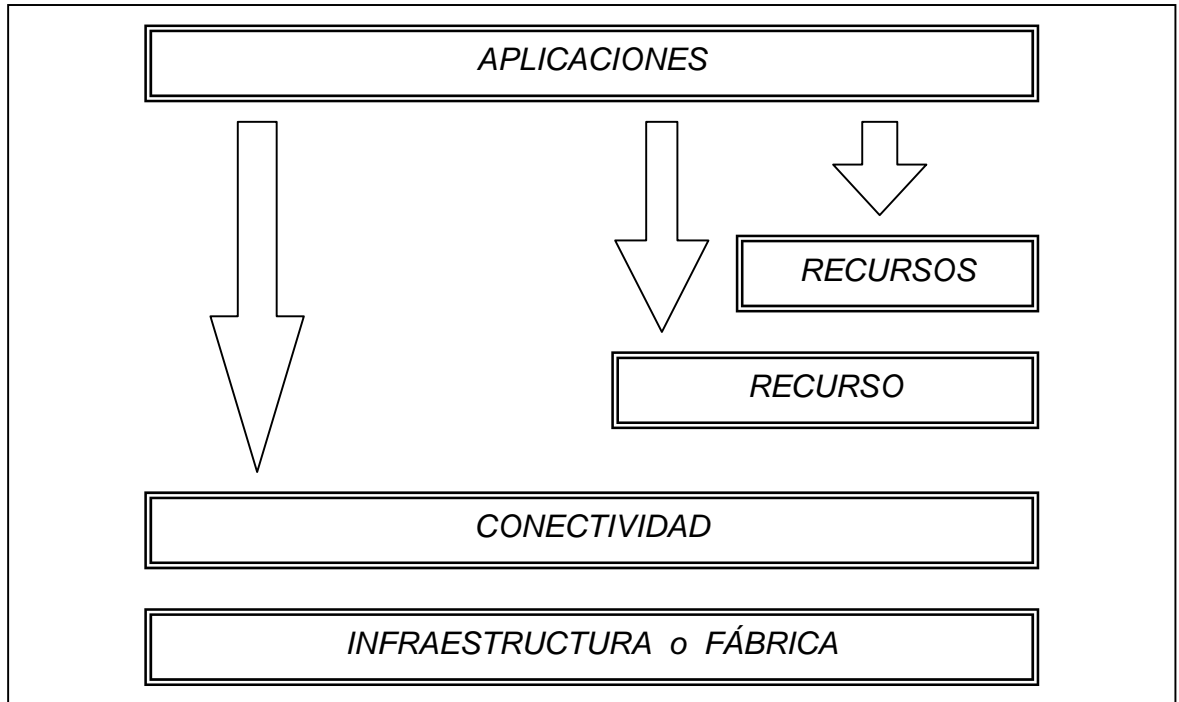
Foster es considerado como líder en el estudio de la computación distribuida y paralela, que a mediados de los años 90's propone un nuevo paradigma en la computación moderna al cual definió como Computación Grid. Definió esta tecnología como "Un Grid computacional es una infraestructura hardware y software que suministra al que la utiliza acceso seguro, consistente, penetrante y barato, a una elevadas capacidades computacionales como se muestra en la figura 7"¹². En la figura 5 se puede apreciar la arquitectura que propone Ian como guía para esta tecnologías y la cual se ha adaptado como el modelo a seguir por la comunidad científica que investiga y desarrolla esta tecnología, en la figura 7 se puede observar las características que según Ian debe tener esta tecnología.

Ian Foster, Carl Kesselman y Steve Tuecke¹³, plasman el concepto del Grid a través del desarrollo de cinco niveles (ver Figura 5) que funcionan como un modelo de arquitectura en capas.

¹² http://www.ceditec.etsit.upm.es/grid_computing.php

¹³ <http://www.globus.org/research/papers/anatomy.pdf> Ian Foster, Carl Kesslman, Steven Tueeke, Anatomy of the Grid:Enabling Scalable Organizations

Figura 5. Arquitectura de la Grid



Fuente. The Anatomy of the Grid – Enabling Scalable Virtual Organizations.

Esta arquitectura esta basada en el uso de protocolos, que son los encargados de establecer las reglas básicas que permitirán a los usuarios establecer la comunicación entre ellos y los recursos.

Esta arquitectura define cinco niveles a saber:

1- Nivel infraestructura o Fábrica. Suministra los componentes que serán compartidos, en este nivel se encuentran todos los recursos computacionales, como los equipos de cómputo, Clusters, supercomputadores, sistemas de almacenamiento, dispositivos de medición, bases de datos y todo lo que tenga que ver con la parte Hardware. Se debe incluir en este nivel toda su infraestructura

física y todos sus mecanismos de control y gestión. En esta capa se modelan los recursos accesibles como:

- Recursos computacionales (Computadores personales)
- Cluster
- Sistemas de Almacenamiento de datos
- Sistemas distribuidos de archivos
- Instrumentos de laboratorio
- Instrumentos de medida y calibración

Esta capa debe brindar la certeza de saber con que recursos se cuenta para su utilización. En el caso de los recursos computacionales es muy importante tener en cuenta aspectos como:

- Proveer la información sobre el Hardware y Software disponible, el estado presente y pasado en lo que se refiere a carga, utilización, disponibilidad, frecuencias de acceso.
- Posibilidad de monitorear los procesos que se están ejecutando al igual que el estado físico de los recursos asociados al proceso
- Poder hacer una reserva de un recurso
- Estar en capacidad de informar la disponibilidad de los recursos
- Informar el espacio disponible de almacenamiento, el porcentaje de espacio libre para almacenamiento, carga de procesamiento del dispositivo etc.

En estos recursos se debe tener la posibilidad de almacenar no solo ficheros simples sino también gran cantidad de datos. Lo que involucra tener una interfaz ágil de servicios de transferencia de archivos, que permita en los casos en que no se logre completar el proceso no solo el restablecimiento de la conexión sino también la retransmisión desde el último paquete correcto recibido.

En cuanto a los recursos de red se puede proveer información no solo sobre la carga actual del dispositivo sino también sobre predicciones futuras, en este aspecto cabe destacar que se deben crear políticas de priorización para un mejor aprovechamiento de los recursos y mejorar el tiempo de algunos procesos.

Es muy importante contar con un servicio que permita la administración del espacio de almacenamiento para poder así restringir este servicio a un usuario o grupo de usuarios teniendo en cuenta políticas previamente establecidas.

2- Nivel de conectividad. Permite establecer la comunicación entre los recursos computacionales al contener en este nivel los protocolos de comunicación y seguridad necesarios para esta comunicación. Dentro de estos protocolos podemos encontrar TCP/IP (Usado para las comunicaciones vía Internet), Protocolo SSL, Certificados X.509, Los nuevos protocolos que serán usados en redes de alta velocidad que se encuentran en la etapa de prueba.

En esta capa vale la pena resaltar la analogía que existe con la capa de transporte en la pila de protocolos TCP/IP. Donde encontramos protocolos estándar de la comunicación como IP en la capa de Internet, ICMP y TCP y UDP de la capa de transporte.

Se espera implementar en la arquitectura Grid servicios que provean la posibilidad de establecer la comunicación a cualquier middleware o aplicación que este por arriba de su nivel.

En este nivel es importante resaltar servicios de verificación de identidad de recursos y usuarios para poder prestar seguridad y establecer las comunicaciones previa autenticación mutua entre ellos. Para fortalecer estos servicios de seguridad es necesario el inicio de sesión único.

No se requiere que el usuario ingrese su identificación cada vez que va a acceder a un recurso, pero si es necesario que sus privilegios viajen con la solicitud del recurso. Las características anteriores se hacen necesarias debido a que Grid es un sistema distribuido y un mismo trabajo puede viajar por multitud de elementos hasta alcanzar aquel en que desea ejecutar

Hay que tener en cuenta que al tener que viajar un trabajo para ser procesado, el tiempo de respuesta va a incrementar por lo que se establece un tiempo de vida para su ejecución dentro del sistema. Por todo esto el sistema de renovación de credenciales se encontraría en esta capa.

Esta capa debe desarrollar los servicios de integración con servicios de seguridad locales. Las relaciones de confiabilidad permiten autorizar una serie de usuarios con la autorización de uno de ellos que hace las veces de punto de contacto. La característica anterior es muy importante en este tipo de sistemas, y es requerida para poder compartir recursos entre usuarios de organizaciones virtuales.

3- Nivel de Recurso. Tiene como función gestionar un único recurso permitiendo tener el control y la información sobre éste. En este punto se tienen una serie de protocolos y herramientas software que permiten obtener información puntual de este recurso o de la aplicación sobre la que este trabajando. Entre las funciones a las que se puede tener acceso están: arranque del proceso, parada y fin del proceso, monitorización del recurso, administración del recurso desde la contabilización de su uso y la auditoria de su función.

En esta capa se lleva a cabo la iniciación de las transacciones que sean necesarias para la realización del trabajo como son la localización y recuperación de los datos necesarios para la ejecución de la tarea.

En esta capa también encontramos los protocolos y servicios de monitorización de servicios para determinar que están disponibles o aplicaciones que ya se han enviado y se requiere conocer su estado. Junto con el servicio anterior debe realizarse el control de los trabajos de manera que pueda chequearse su estado y en determinado caso establecer métodos y políticas cuando deba reiniciarse, mover su localización o cancelarlo.

Establecer los servicios de accounting, a través de los cuales podemos tener un registro estadístico de las aplicaciones que se han ejecutado en un recurso específico, que usuarios accedieron a un recurso y durante cuanto tiempo. Todo esto permite establecer que servicios son los más requeridos, costos, calidad del servicio, y otras características que serán fundamentales para la creación de políticas que mejorar el acceso y operacionalidad de los recursos.

Los protocolos de transmisión de datos, proveen acceso a los datos y comunicación de alto rendimiento. Buscando transmitir información de forma confiable como lo realiza el protocolo FTP.

4- Nivel de Recursos. Involucra una mayor cantidad de recursos a los que se llamará *conjunto de recursos*, donde se pueden encontrar directorios que dan la localización de un recurso específico.

En este nivel se sitúan servicios como los de Co-Allocation, Brokering y Scheduling, los cuales proveen los métodos para localizar los recursos que se adaptan a las necesidades del usuario conociendo algunos datos sobre ellos y además intentado optimizar la asignación global mediante métodos de búsqueda o brokering. Con los servicios de co-allocation se pueden hacer reservas simultáneas de los recursos necesarios para la realización de las diferentes tareas.

Para la consecución correcta del trabajo también es importante hacer una planificación del uso de los recursos mediante la aplicación de scheduler distribuidos en sus tres tipos como son:

- Planificador de trabajo. Se encarga de maximizar la cantidad de trabajo realizado en una unidad de tiempo (job scheduler)
- Planificador del recurso. Busca maximizar el uso del recurso
- Planificador de la aplicación. Divide la aplicación en tareas que serán asignadas a los recursos para su ejecución y vigilancia del desarrollo de su actividad.

El planificador de trabajo y el de recursos, se encaminan a la eficiencia del sistema mientras el planificador de aplicación se encamina por la eficiencia de la aplicación.

La monitorización y el diagnóstico antes y después de la elaboración de distintas tareas que han sido distribuidas, establecer el cálculo de los costos de utilización de varios recursos, el nivel de uso de los recursos y sobre todo el acceso a datos distribuidos que gestionan la replicación de los datos.

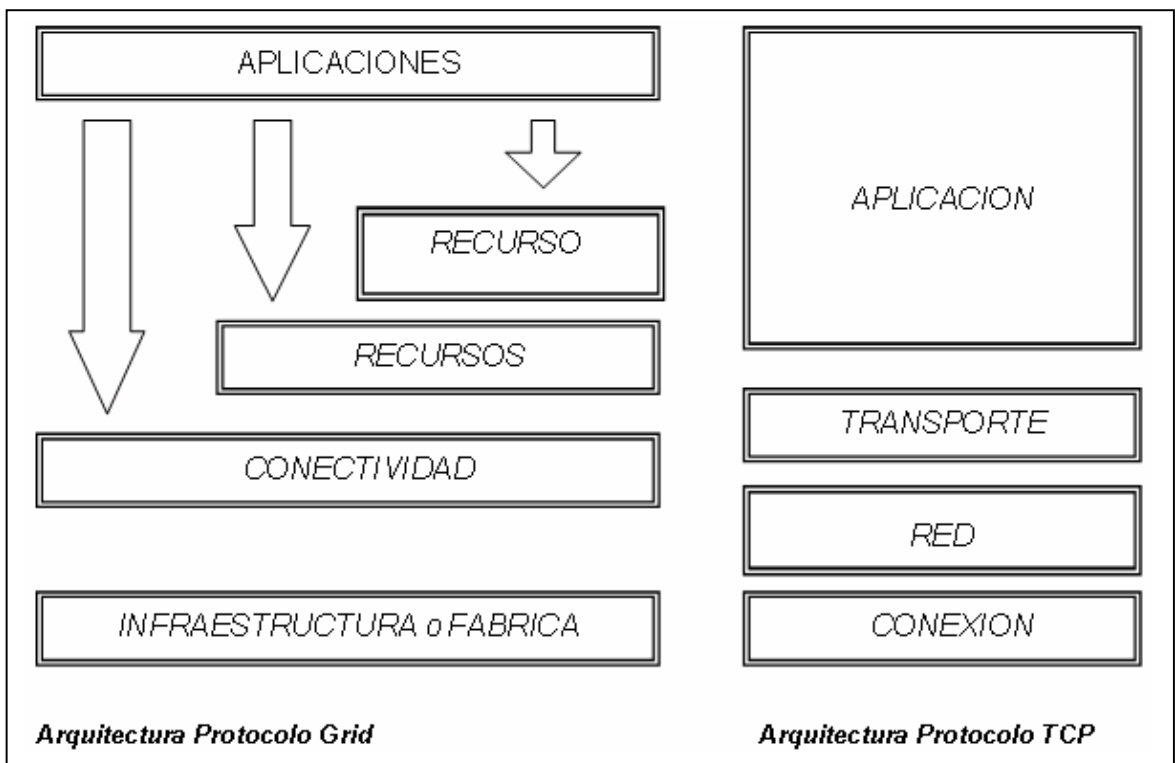
Dependiendo de la organización, los elementos de los sistemas de información se puede obtener información fiable y fiel a la realidad. También pueden existir servicios de replicación de datos, cuya función es la de realizar una copia en el sistema que solicita el trabajo luego de traer aquellos datos necesarios para poder realizar la ejecución de un programa y así optimizar el desarrollo de esta tarea.

5- Nivel de Aplicación. Se centra en la definición de protocolos que se encargan de permitir a las aplicaciones acceder a la infraestructura de la Grid mediante las distintas capas. No necesariamente se debe acceder a todas las capas para hacer

uso de alguna de ellas, en algunas ocasiones se puede hacer acceso directamente a la infraestructura de la Grid.

Como se puede observar existe una gran relación entre la arquitectura de la Grid con la arquitectura empleada en el protocolo usado en Internet como lo muestra la figura 6. Estas dos arquitecturas aunque no son iguales tienen una estructura similar y por que no tenerla si Grid esta diseñada para trabajar sobre la plataforma de Internet

Figura 6. Paralelo entre Arquitectura Grid y Protocolo TCP

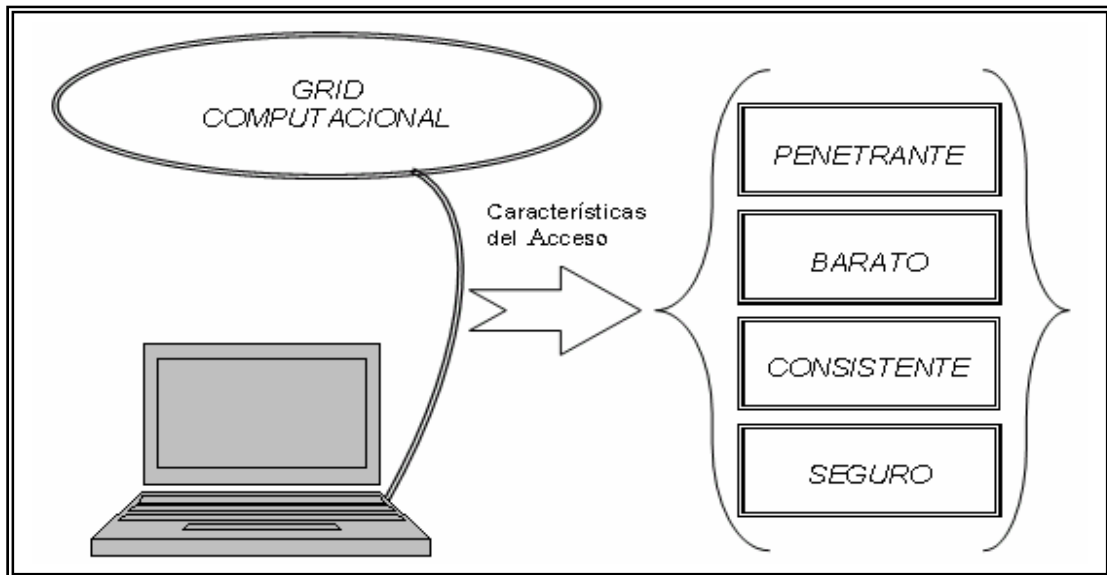


Fuente: The Anatomy of the Grid – Enabling Scalable Virtual Organizations.

Ian Considera que la Computación Grid va a lograr su desarrollo mundial al aprovechar la infraestructura ya instalada de Internet, sobre la cual se realizaran algunas modificaciones para explotar todos estos beneficios al fusionar las

herramientas existentes en la Computación Grid con junto a la evolución de Internet la cual denominan Internet2

Figura 7. Características de un Grid Computacional



Fuente. El Autor

Esta infraestructura debe proporcionar a los usuarios un servicio:

- Consistente. Que este disponible para realizar de forma efectiva las tareas que se requieran.
- Seguro. Que garantice que todos aquellos procesos de cálculo y demás que son realizados en la Grid cuenten con la certeza de su proceso.
- Penetrante. Que permita llegar a múltiples sectores de la vida del hombre para así maximizar su desarrollo.
- Barato. Que sea accesible a toda la comunidad, debido a que el costo del servicio es vital en la masificación de su uso

La Computación Grid proporciona una forma transparente de ejecutar el trabajo que se desea, inicialmente se debe probar la autenticidad del usuario. La realización de la tarea se ejecuta mediante el uso de agentes que se encargan de

encontrar los recursos disponibles, asegurar un acceso confiable a los datos, monitorear las tareas antes de la ejecución, durante y después de los procesos, y en la medida de las condiciones estar en la capacidad de paralelizar el trabajo encomendado.

2.2 ESPECIFICACIONES TÉCNICAS

➤ Open Grid Services Architecture - OGSA. Conjunto de servicios creados para lograr estandarizar todas las aplicaciones dentro de la Grid. El OGSA son creados por el GGF (Global Grid Forum), es una organización fundada con el fin de crear un estándar para las aplicaciones Grid. OGSA que es una estructura de alto nivel con ocho servicios específicos que permite compartir recursos entre Organizaciones Virtuales¹⁴.

- 1- Servicios de Infraestructura. Permiten establecer la comunicación sin barreras entre los recursos que se están compartiendo.
- 2- Servicios de Gestión de Recursos. Permiten monitorear, reservar, desplegar y configurar recursos dentro de la Grid.
- 3- Servicios de Datos. Permiten el movimiento de datos
- 4- Servicios de Contexto. Buscan optimizar los recursos de la Grid al describir los requerimientos y uso de políticas por cada cliente.
- 5- Servicios de Información. Permiten conocer información sobre el estado del Grid y cada uno de sus recursos.
- 6- Servicios de Gestión Propia. Permiten mantener con mayor automaticidad la conexión con los niveles de estado de servicio con el fin de reducir los costos y complejidad de la gestión del sistema.

¹⁴ <http://www.globus.org/research/papers/ogsa.pdf>

- 7- Servicios de Seguridad Refuerzan los servicios de seguridad existentes en las Organizaciones Virtuales, permite la autenticación y autorización de usuarios para poder realizar la compartición de recursos en forma segura.
- 8- Servicios de Gestión de Ejecución. Permiten el flujo de acciones de trabajo simples y complejas como colocación, aprovisionamiento y gestión del ciclo de vida de las tareas.

El GGF ha creado un grupo llamado DATA que se encarga de estudiar y estandarizar todo lo relacionado con temas como Replicación, Metadata, Movimiento de datos. De igual manera, se ha creado un grupo específico para todo lo relacionado con Bases de Datos en entornos Grid al que se denomina Working Group DAIS que tiene como función principal identificar soluciones efectivas que permitan administrar de forma autónoma Bases de Datos existentes en la Grid

“Los servicios de OGSA suministran una capa de virtualización encima del recurso de computo para poder crear la administración dinámica de las organizaciones virtuales”¹⁵

➤ Open Grid Services Infraestructure - OGSi. Estos servicios están encaminados a la parte de seguridad y al igual que el OGSA se crearon por el GGF. Dentro de las principales funciones del OGSi se pueden enunciar¹⁶.

- 1- Crea, nombra y administra el tiempo de vida de las instancias de servicios
- 2- Ofrece el servicio de declarar e inspeccionar el estado de los datos
- 3- Notifica de forma asíncrona el cambio en el estado de los datos

¹⁵ Manual de Implementación Globos Toolkit 3.2.1 Grid Experimental, Julio C Baracaldo, Oscar Fajardo, Gastón Mejía. Julio 2004

¹⁶ HERNANDEZ BRAVO, Ángel. Programa de Doctorado: Sistemas Distribuidos. Proyecto Globus, Enero 2004

- 4- Representación y administración de colecciones de instancias de servicios.
- 5- Servicio sobre invocación de fallos

El OGSÍ define como construir un servicio Grid especificando las pautas para crear, manejar e intercambiar la información entre los Servicios Grid. También se encarga administrar el acceso de los usuarios a los servicios mediante los servicios de autenticación, enviar los datos de forma encriptada gracias a códigos de llave pública que hacen difícil el acceso a la información sin autorización. Un servicio Grid es un servicio Web que define como un cliente interactúa con un servicio Grid.

3. HERRRAMIENTAS Y ESTRATEGIAS ACTUALES

Para poder construir y utilizar un Grid es necesario conocer y utilizar una serie de herramientas representadas en bibliotecas de funciones, aplicaciones, protocolos, organizaciones, fabricantes, etc.

3.1 PLATAFORMAS DE GRID

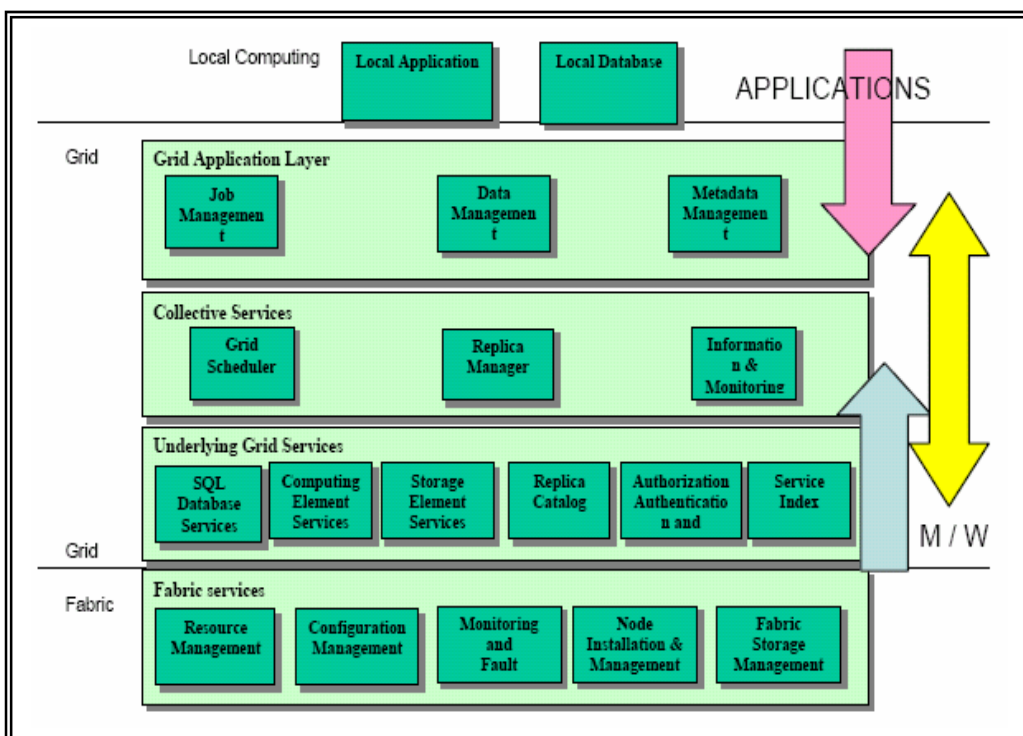
Para el funcionamiento del sistema basado en Computación Grid se debe disponer de una serie de programas instalados sobre un determinado número de maquinas denominadas sistemas cliente, al igual que sobre uno o mas servidores dedicados al control. También se puede disponer de clientes dotados de programas que permitan enviar tareas a listas con los recursos requeridos cuando usuarios se conecten. De tal manera que al momento de conectarse, su equipo (cliente) recibe tareas que se completaran a medida que se vaya disponiendo la maquina de recursos. El programa ejecutándose sobre el cliente detecta el momento en que el sistema particular entra en desuso, notifica al servidor que el sistema está disponible para procesamiento solicitando la tarea. El cliente recibe a continuación la tarea del servidor y ejecuta el software cuando dispone de los ciclos CPU, enviando al servidor una vez finalizada la realización de la tarea. Esta interconexión se hace por medio de un conjunto de protocolos e interfaces abiertos que permitan gestión de recursos, gestión remota de procesos, librerías de comunicación, seguridad y soporte a monitorización.

La Computación Grid requiere el uso de software que se encargue de dividir y separar en miles de pequeñas partes un programa para poder enviar estos datos como tareas a los equipos que forman parte de ese Grid. Además se requiere de una serie de servicios relacionados con aplicaciones, proveedores de recursos y

usuario para gestionar aspectos como seguridad, acceso a los recursos y a datos, instrumentación, contabilidad entre otros

Una de las características de las tecnologías Grid es que, en la relación de pareja cliente-servidor interviene un tercer actor, el **middleware**. Este es uno de los vértices del triángulo virtuoso que posibilita un nuevo tipo de interacción entre las máquinas. El middleware es la infraestructura virtual como un sistema operativo que media entre el cliente, dispositivos y los múltiples servidores, dividiendo el trabajo para que las tareas de unos y de otros se realicen (ver figura 8).

Figura 8. Arquitectura del Middleware



Fuente. www.globus.org

El middleware es el encargado de “ver” cuáles son los computadores y recursos disponibles para realizar una tarea y con qué facilidades cuentan para llevar a cabo el servicio requerido. Una vez que verifica el estado de cada integrante de la

red (cuánta capacidad “ociosa” tienen para participar en la nueva tarea), se ocupa de asignarles diferentes porciones de trabajo. El middleware “arma la agenda”. El resultado de este tipo de organización es una arquitectura dinámica, que permite adecuar la configuración de la red a las necesidades y posibilidades de cada momento, para ganar velocidad de procesamiento y eficiencia.

Todas estas tareas que realiza el middleware se hacen gracias a una herramienta que aunque no es la solución definitiva han demostrado gran capacidad y rendimiento, esta herramienta son los **agentes inteligentes** que pueden definirse como:

- “programas que se encargan de realizar tareas específicas y poseen un grado de inteligencia suficiente para ejecutar parte de la misión encomendada de forma autónoma y para interactuar con su entorno útil”¹⁷.
- “un agente es cualquier entidad que percibe su entorno a través de sensores y actúa sobre ese entorno mediante efectores. Un agente es racional cuando realiza la mejor acción posible a partir de los datos percibidos”¹⁸
- “una agente inteligente es un sistema (hardware o software) situado en un determinado entorno, capaz de actuar de forma autónoma y razonada en dicho entorno para conseguir unos objetivos”¹⁹

Como se puede observar, los agentes son una pieza fundamental dentro del middleware, no solo por las funciones que tienen sino también porque estos entes deben tener la capacidad de **aprender** con la inteligencia a encontrar fuentes de información en los entornos abiertos (medida como grado de sobrecarga informativa), **decidir** que fuentes han sido creadas modificadas o no existen (una medida de la complejidad) y **adaptarse** a los cambios que pueden sufrir los

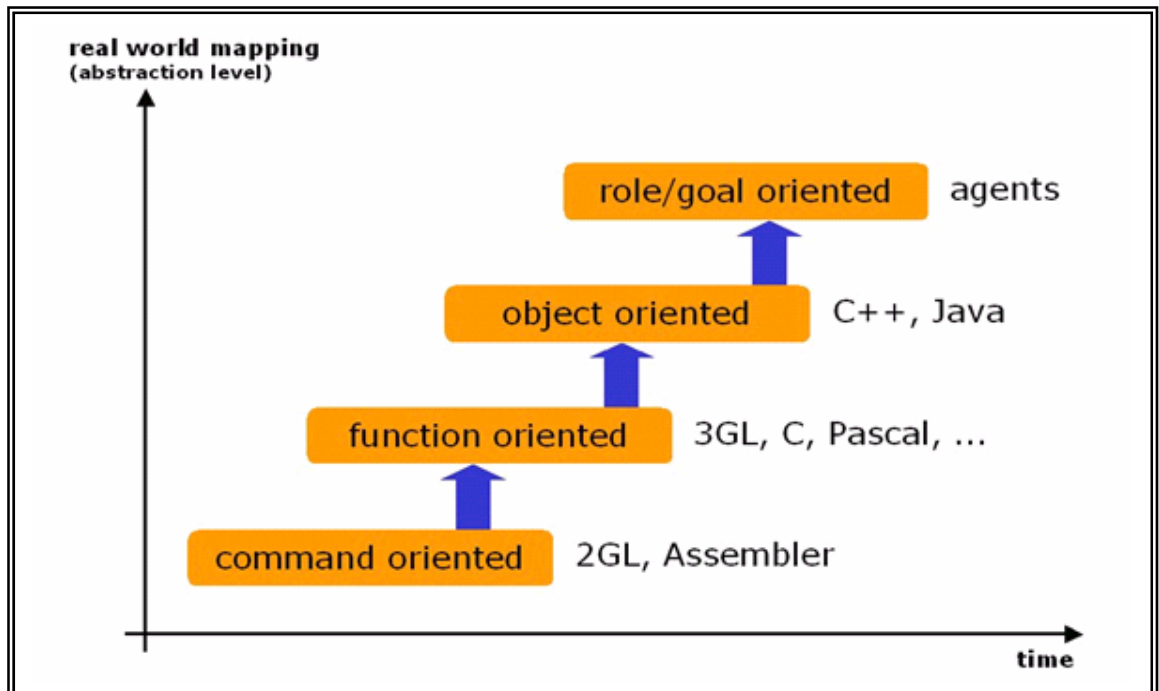
¹⁷ www.sia.eui.upm.es/docent/ai.htm - Definición de agente de Brenner

¹⁸ www.sia.eui.upm.es/docent/ai.htm - Definición de Agente de Russell&Norving

¹⁹ www.sia.eui.upm.es/docent/ai.htm - Definición de Agente de Wooldridge

entornos o servicios a los que acceda. En la figura 9 se observa como los agentes son en el momento el último escalón en lo que tiene que ver con herramientas de diseño y desarrollo.

Figura 9. Evolución tecnológica en las herramientas de diseño y desarrollo



Fuente. www.Gridsystems.com

3.2 TOOLKITS

Es un conjunto de aplicaciones y bibliotecas de funciones que forman la estructura básica de un Grid. Esta herramienta proporciona múltiples servicios como los de seguridad, autenticación, información, comunicación entre recursos, ejecución de aplicaciones, transferencia de datos a los usuarios y programadores.

Existen varias varios Toolkits pero los que más se destacan son²⁰: Legion, Unicote, Globus, Gridbus

3.2.1 Globus Se originó en los laboratorios de IBM en 1996 y es el estándar para la construcción de Grid definido por la implementación del Globus Project. Se concibe como un conjunto de servicios constituidos por un conjunto modular de protocolos y librerías que se pueden ser usados de forma independiente como juntos para desarrollar aplicaciones Grid y herramientas de programación²¹. Sirve como capa de middleware para ocultar la semántica de todo el proceso de visualización en que se apoya la tecnología Grid. Se ha definido a Globus Toolkit versión 3 (GT3) como el estándar. Aunque la versión Beta de Globus 4 (GT4) fue liberada en Febrero 25 de 2005 y se espera la liberación de la versión definitiva para Abril 28 de 2005, en este documento se hará referencia a Globus Toolkit Versión 3.2.1. Esta versión (GT3) al igual que sus sucesoras (GT4) está implementada íntegramente en Java²².

El concepto de Globus para la versión GT3 se ha estructurado en varias capas (Ver figura 10)

- Primera Capa o Capa Inferior, es el núcleo de esta herramienta en ella se encuentras las Fabricas de recursos, los servicios de notificaciones, persistencia y el ciclo de vida.
- Segunda Capa concentra todos los servicios de seguridad (GSI), utilizando una infraestructura de seguridad denominada GSI la cual está basada en el uso de claves públicas encriptadas, certificados X.509, SSL

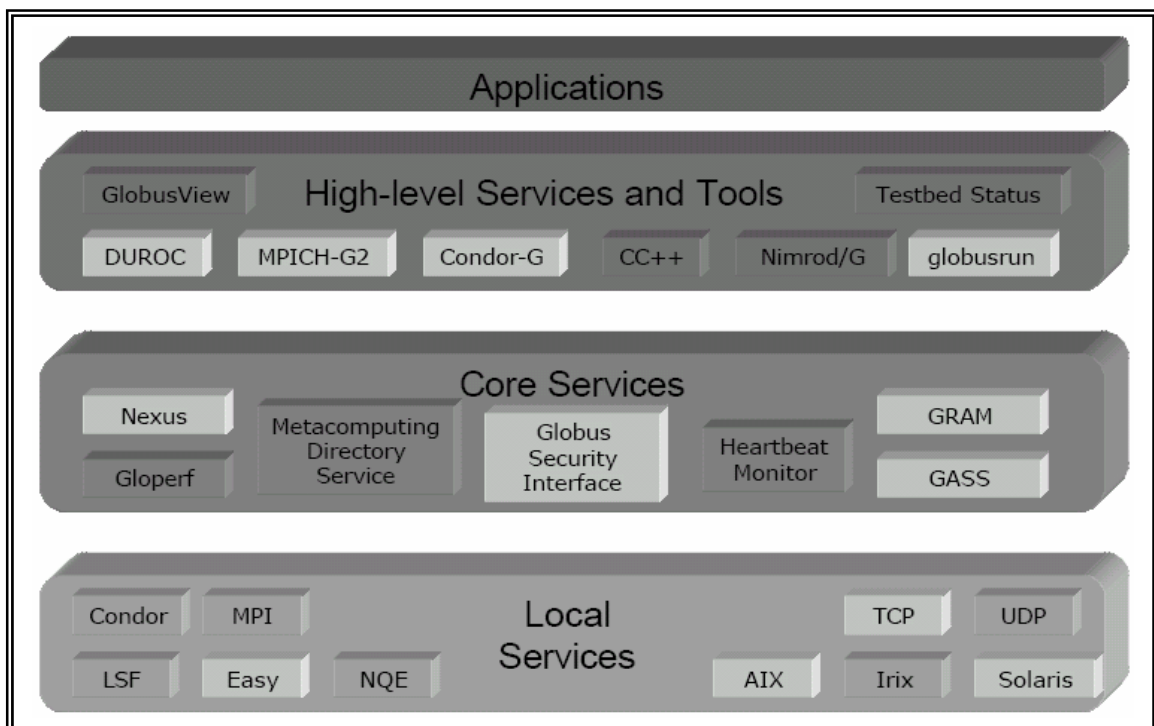
²⁰ TARRIO BARREIRO, Jacobo. Diseño e implementación de un portal basado en tecnologías Grid para el acceso a recursos de súper computación

²¹ <http://arcos.inf.uc3m.es/~fgarcia/sd>

²² www.globus.org//research/papers/anatomy.pdf

- Tercera Capa encontramos servicios como la gestión de trabajos, servicios de directorio y monitorización de los de transferencia de ficheros.
- Cuarta Capa contiene los servicios de gestión de grandes cantidades de datos y los servicios que no son GT3 pero que están basados en esta arquitectura.

Figura 10. Capas de la Arquitectura Globus GT3



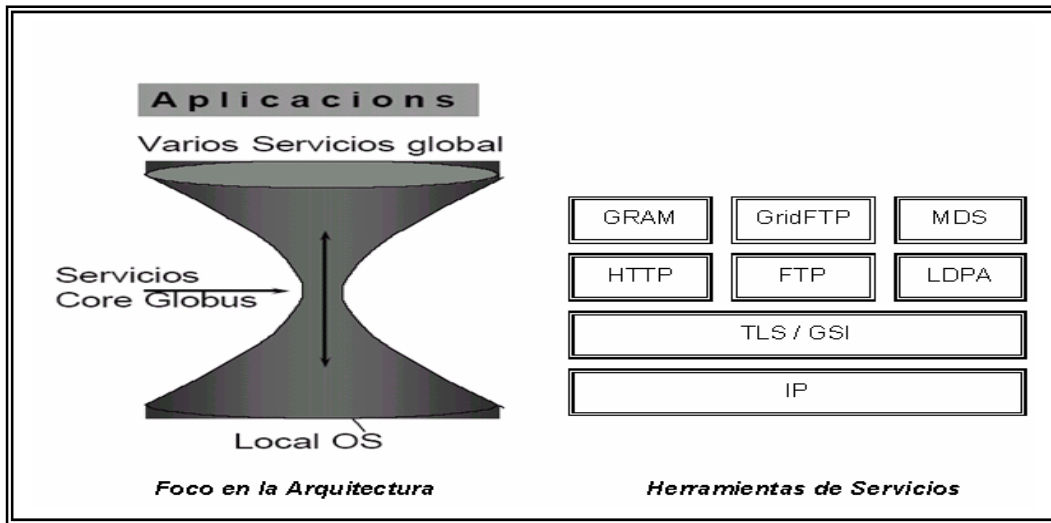
Fuente. www.globus.org

- Globus Resource Allocation Manager (GRAM). Este componente proporciona servicios de gestión de recursos, de creación y procesos, monitorización. Se encarga de Convertir las peticiones expresadas en RSL, en órdenes para los computadores locales.

- Grid Security Infraestructura (GSI). Servicio de autenticación y cifrado, permite la gestión local de los permisos de acceso de los usuarios a todos los recursos del Grid con una única autenticación.
- Monitoring and Discovery Service (MDS). Es un servicio de información del Grid que proporciona un mecanismo uniforme para acceder a la información sobre la configuración de los servidores, la red, los recursos disponibles y ocupados entre otros. Además suministra los datos utilizando un directorio LDAP.
- Global Access to Secondary Storage (GASS). Permite a los programas que se ejecuten en lugares remotos acceder a la información local implementando un conjunto de estrategias de movimiento de datos.
- Nexos y Globus-IO. Proporciona servicios de comunicación para entornos heterogéneos.
- Heartbeat Monitor (HBM). Permite que sean detectadas fallas en los componentes del sistema o de procesos en ejecución por parte de los administradores del sistema y/o los usuarios.

Funcionamiento del Globus Toolkit. Utiliza una serie de herramientas que se aplican en aspectos como la seguridad (GSI, PKI,SSL v3) y servicios (GRAM, MDS, GFTP). Además propone es sistema de servicios del Core como infraestructura básica permitiendo la construcción de soluciones de alto nivel, dominio-específicas (Ver Figura 11)

Figura 11. Funcionamiento del Globus Toolkit



Fuente. www.globus.org/toolkit

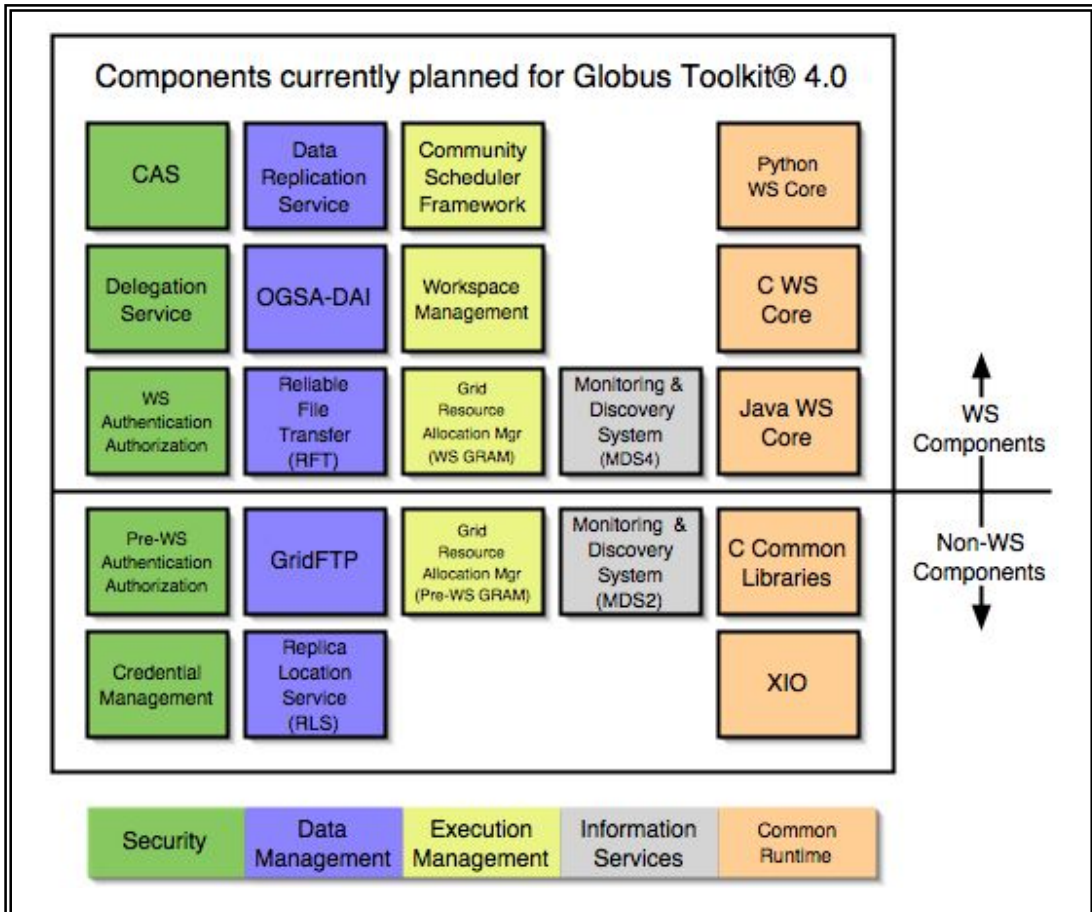
En el momento se encuentra liberada la versión Beta 4 del Globus Toolkit, el cual está enfocado a ser usado mediante los servicios Web para poder explotar la infraestructura ya existente de Internet. Al igual que su predecesor, su estructura es por capas como se aprecia en el Figura 12²³.

3.2.2 Legion es un Toolkit basado en objetos, desarrollado en la Universidad de Virginia y muestra al usuario una sola máquina virtual, gracias a la infraestructura que permite que un conjunto heterogéneo de equipos que se encuentran distribuidos geográficamente funcionen de forma coordinada. Este enfoque es orientado a objetos y dificulta la interacción con algunas aplicaciones. Principalmente cuando se promueve la creación de aplicaciones con lenguajes nuevos o adaptados²⁴.

²³ arcos.inf.uc3m.es/~fgarcia/sd/presentaciones/Arquitectura/Seguridad,.ppt

²⁴ <http://www.cs.virginia.edu/~humphrey/papers/PhilosophicalTechnicalComparisonOfLegionAndGlobus.pdf>

Figura 12. Capas de la Arquitectura Globus GT4



Fuente. www.globus.org

3.2.3 Unicore. Esta diseñado para trabajo por lotes, consiste en un conjunto de tareas, con dependencias que indican relaciones temporales o transferencias de datos. La petición de la ejecución de un trabajo indica dónde se deben ejecutar las tareas que la forman y que recursos requieren. Presenta un nivel jerárquico al permitir que una tarea sea subdividida y ejecutada en diferentes sistemas²⁵.

²⁵ <http://www.unicore.org/documents/UNICOREPlus-Final-Report.pdf>

Los objetivos de diseño de esta herramienta incluyen una interfaz de usuario uniforme y fácil de usar, una arquitectura abierta basada en el concepto de tareas abstractas, una arquitectura de seguridad consistente, y mínima interferencia con los procedimientos administrativos locales.

Es utilizada en los centros de supercomputación en Alemania, proporcionando una interfaz gráfica para el acceso seguro e intuitivo a los recursos computacionales, pero con la facultad de ocultar las particularidades de cada sistema con el objetivo de facilitar el desarrollo de aplicaciones distribuidas.

3.2.4 Gridbus. Este proyecto tiene su origen en la Universidad de Melbourne, con el patrocinio de empresas como Australian Research Council, Storage Technology Corporation, sol Microsystems, VPAC, IBM, y sistemas informáticos de Singapur. Gridbus está desarrollando tecnologías de la Grid con herramientas como el (middleware, herramientas, y usos) que entregan una alta calidad de servicios. Incluyen el planificador económico de la Grid, el planificador del racimo (libra), la herramienta que modela la simulación (GridSim), el corredor de la Grid de los datos, GridBank, y las herramientas del GUI para la gerencia del workflow y composición de usos distribuidos.

Gridbus cuenta con el servicio GDM que sirve como registro de todos los servicios con que cuenta la Grid. Con este servicio se puede conocer los servicios que se encuentran activos y el costo que conlleva su utilización. El conocer esta información agiliza la tarea del usuario y de paso da origen a tres aspectos muy importantes dentro del objetivo del Gridbus de alcanzar la mas alta QoS (Calidad de Servicios),

- Disminución de Tiempo
 - Disminución de Costo
 - Optimización del Tiempo-Costo
-

Gridbus al igual que Globus cuenta con una caja de herramientas denominada GridSim la cual proporciona las condiciones necesarias para modelar y simular los recursos al igual que la conectividad de la red, configuraciones y dominios, apoya los servicios para la búsqueda de los recursos y las interfaces que permitan asignar tareas a los recursos²⁶

3.3 PLANIFICADORES

Los planificadores son sistemas con la capacidad de asignar recursos para que los usuarios puedan ejecutar sus tareas, dentro del proceso que sigue un planificador Grid se pueden enunciar once pasos distribuidos en tres fases²⁷

Fase 1: Descubrimiento de Recursos. Se determina que recursos se encuentran disponibles a un determinado trabajo de un usuario. Esta fase se puede dividir en tres pasos

- Paso 1: Filtrado de autorizaciones. Se determina el grupo de recursos en los cuales el usuario está autorizado, al igual que los recursos en los que no está autorizado
- Paso 2: Definición de los requerimientos de las aplicaciones. El usuario debe especificar un grupo mínimo de requerimientos para filtrar. Cuanto más detalles sean incluidos más eficiente será la planificación y el matching de los recursos
- Paso 3: Filtrado de requerimientos mínimos. Una vez ya se dispone de un conjunto de recursos disponibles y de los requerimientos de los trabajos, se hace un filtrado de los requerimientos que no satisfacen las aplicaciones.

²⁶ <http://www.gridbus.org>

²⁷ FERNANDEZ CASINI, Álvaro. Arquitecturas Grid orientadas a la gestión de recursos. Trabajo Investigación Instituto de Física Corpuscular. Diciembre 2004.

Fase 2: Selección del sistema. Después de obtener el grupo de recursos, se decidirá donde se planificará el trabajo. Para esta tarea se necesita información más detallada de los recursos y el estado del sistema

- Paso 4: Recolección de información dinámica. Se debe estar recolectando información de diversos recursos de acuerdo a las necesidades que se vayan presentando.
- Paso 5: Selección y Planificación. Luego de haberse obtenida la información se procede a decidir en qué recursos o grupos de recursos van a ser seleccionados para la ejecución del trabajo.

Fase 3. Ejecución del trabajo. Esta última fase tiene que ver con todos los pasos que hacen posible la ejecución del trabajo gracias a los recursos e información ya obtenidos.

- Paso 6: Reserva (opcional). Para hacer el mejor uso de los recursos en el momento en que se necesitan, es conveniente hacer reserva de ellos, principalmente de los más difíciles de obtener o que tienen mayor demanda.
- Paso 7: Envío del trabajo. Este paso consiste en enviar el trabajo a los recursos para que sean procesados mediante la gestión de protocolos
- Paso 8: Tareas preparatorias. Este paso está muy ligado al anterior, ya que se puede requerir de otros pasos para que se realice correctamente el envío del trabajo
- Paso 9: Ejecución y monitorización. Culminados los pasos anteriores se procede a ejecutar el proceso. La ejecución y monitorización están fuera del control de los planificadores de alto nivel y más relacionada con los planificadores locales. La parte de monitorización puede ser ejecutada por el usuario. El cual puede cambiar su decisión sobre dónde o cómo se está ejecutando.

- Paso 10: Finalización del trabajo. Ya finalizado el trabajo el usuario puede obtener información sobre la ejecución y resultados del proceso.
- Paso 11: Tareas de Limpieza. Se asocia al paso anterior, la misión es obtener los ficheros útiles y limpiarlos restos de la ejecución de la tarea.

Los planificadores normalmente funcionan en cluster de Computadores o en una red local; Existen diversas versiones de planificadores para Grid, capaces de planificar recursos ubicados en áreas geográficas muy lejanas. Dentro de estos planificadores tenemos.

3.3.1 Condor-G. Diseñado inicialmente para aprovechar ciclos de CPU en estaciones de trabajo ociosas, pero se le ha dado una nueva visión como lo es la planificación de tareas por lotes bajo el esquema de colas de trabajo, políticas de planificación, monitorización y gestión de recursos²⁸.

Esta versión de Condor que usa Globus agrega funcionalidades a la herramienta GRAM de Globus. Permite conocer y utilizar recursos distantes gracias a la gestión de recursos, envío de trabajos y monitorización.

Este proyecto fue iniciado en la universidad de Wisconsin–Madison. Hace hincapié en el procesamiento paralelo estableciendo la diferencia entre HPC (High Performance Computing), donde siempre se ha hablado de MPLOPS para medir el rendimiento y HTC (HIGH Throughput Computing) que se ocupa de sistemas capaces de proporcionar gran capacidad sostenida de procesamiento.

3.3.2 Nimrod-G. Herramienta que facilita la realización de experimentos basados en la ejecución del mismo proceso con distintos parámetros de entrada. Usado

²⁸ <http://www.cs.wisc.edu/condor/>

para probar modelos meteorológicos. Globus utiliza la versión Grid de Nimrod para iniciar trabajos en máquinas distantes.

3.4 ENTORNOS DE DESARROLLO

Son conjuntos de bibliotecas de funciones y aplicaciones destinadas a desarrollar aplicaciones para el Grid. Cada Toolkit esta asociada a un entorno de desarrollo. Dentro de este grupo de herramientas están incluidas bibliotecas de programación paralela como MPICH-G2

MPI es al implementación estándar pero se ha desarrollado siendo MPICH su última modificación. La principal característica de MPICH es su portabilidad que se fundamenta en dos capas:

- Capa inferior conocida como ADI (Abstract Device Interface), se encarga de ocultar la mayoría de los detalles dependientes del hardware.
- Capa Superior se encarga de la semántica y la sintaxis de la mayor parte de la biblioteca MPI

Cada implementación distinta de la capa ADI se denomina “dispositivo” y cada dispositivo define una implementación diferente. MPICH-G2 es una extensión de MPICH para poder ser utilizada por Servicios Globus, permitiendo acoplar múltiples máquinas de diferentes arquitecturas para ejecutar aplicaciones MPI

3.5 PORTALES GRID.

Es necesario presentar el Grid de una forma visual agradable y fácil de utilizar por los diferentes usuarios. Una forma fácil de lograr este objetivo es la utilización de

portales Web, a los cuales se puede acceder desde cualquier equipo en cualquier parte mediante un Browser.

Los portales Grid se dividen generalmente en dos:

- Orientados al usuario. Son genéricos y se diseñan para que puedan utilizar los servicios Grid.
- Orientados a la aplicación. Diseñados para que los utilicen una determinada aplicación.

La elaboración de estos portales se hace con base en lenguajes de programación estándar (Commodity). Que pueden acceder a los servicios del Grid, utilizando un conjunto de componentes llamados CoGKits (Commodity Grid Kits).

1- GridPort. Fue creado para el portal NPACI hotpage²⁹. Es un Kit de desarrollo de portales Grid de usuario y de aplicación. Se implementa mediante scripts escritos en lenguaje Perl y son invocados mediante la interfaz estándar CGI.

2- GPDK. Es la combinación de un entorno de desarrollo de nuevos portales al igual que una colección de componentes Java. Construidos con tecnologías JSP y Java Beans para realizar operaciones sobre el Grid como envío de tareas, ficheros, obtención de información y otras.

Los portales GPDK proponen una arquitectura de tres capas: servidor Web, Globus y Grid y se ejecutan sobre el servidor de aplicaciones Tomcat que se pueden configurar tras cualquier servidor Web.

²⁹ <https://hotpage.npaci.edu>

3- Grid Resource Broker Es un portal Web que permite manejar Grids de una manera fácil, segura e independiente de su localización. No requiere la modificación de los programas ya existentes como tampoco la utilización de comandos Globus.

Al igual que GPKD tiene una arquitectura de tres capas:

- En la primera capa reside un servidor al cual se conecta un navegador de forma segura.
- En la segunda capa se encuentra Globus. Que sirve como puente para que el servidor Web acceda a la tercera capa
- En la tercera capa se encuentran los recursos computacionales Grid.

4- GridSphere. “Es un proyecto de creación de un entorno de desarrollo de portales basado en portlets que es un componente que se puede integrar en un portal Web; funcionan como pequeñas aplicaciones Web gestionadas desde un programa contenedor de portlets”³⁰.

Actualmente Java es la tecnología más utilizada para la elaboración de aplicaciones Web, por lo que existe mucha documentación, bibliotecas de funciones, aplicaciones

3.6 IBM

Ha sido una de las empresas que ha visto el potencial de la Computación Grid, por esto ha brindado gran ayuda en el desarrollo de esta tecnología

³⁰ <http://www.gridsphere.org>

desde hace mucho tiempo. Dentro de su sistema operativo AIX incorpora una caja de herramientas de protocolos abiertos. Dentro de los avances que ha conseguido esta el aumento en la volúmenes de transferencia

IBM en 1986 ha sido la primera en crear sus laboratorios herramientas como el Globus que se ha convertido en un estándar para el desarrollo de aplicaciones en la Computación Grid. Todo este apoyo que brinda IBM al desarrollo de la Computación Grid es gracias a las nuevas perspectivas de negocio de la compañía que ha visto como el negocio de los computadores de escritorio esta muy competido con mayor competencia y menos ganancia, pero por otro lado desarrollar una herramienta que dentro de la organización permita usar todos estos computadores de escritorio es muy atractiva para cualquier organización³¹.

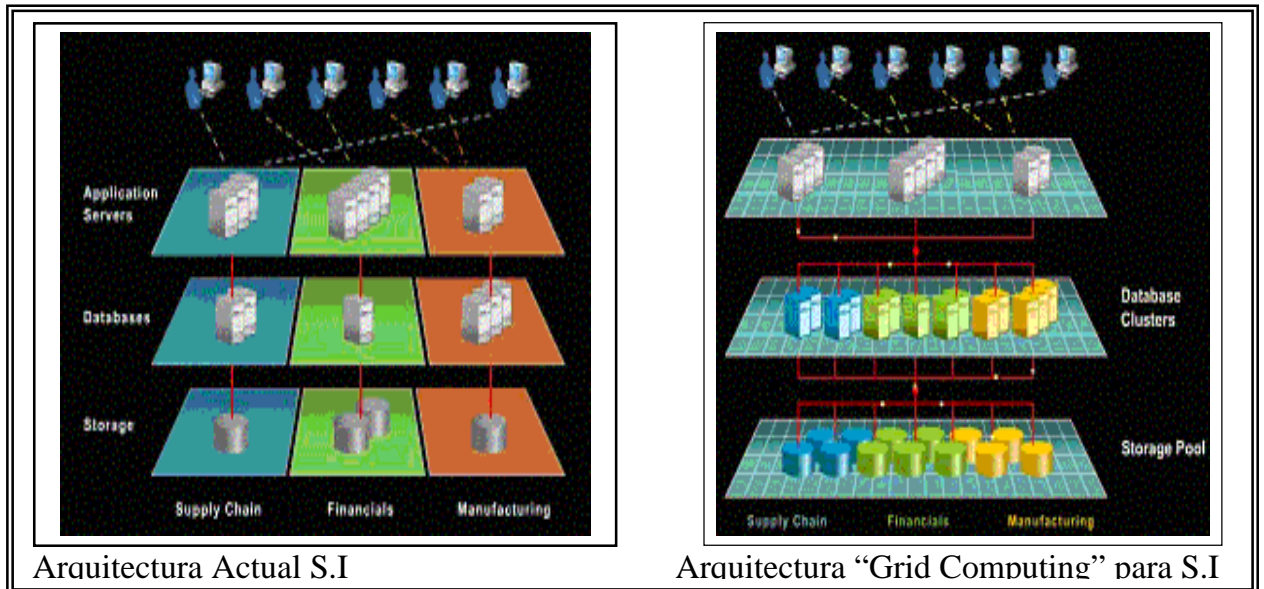
3.7 ORACLE

Ha desarrollado una nueva versión llamada 10g donde pretende crear “granjas/factorías de almacenamiento”, con base en servidores estándares y modulares. A estas granjas se pueden agregar servidores cuando se requiera, esta arquitectura puede ver de manera mas clara mediante la figura 13³².

³¹ www.ibm.com

³² www.oracle.com

Figura 13. Comparativo entre la Arquitectura Actual y la Arquitectura Grid para S.I.



Fuente: www.oracle.com

Entre las características de esta plataforma se puede destacar las siguientes:

- La capacidad de trabajo se podrá reasignar desde la granja de recursos a donde se requiera sin tener que pensar en la capacidad de los sistemas en función de los ciclos de trabajo, logrando así un balanceo dentro del sistema
- Oracle 10g ha mejorado su arquitectura RAC con dos nuevas funciones. La primera de ellas ASP que permite la asignación y reasignación de los servidores a las cargas de trabajo. Facilitando la labor de los clientes al encaminarlos automáticamente a los servicios con solo hacer login, esta característica permite que si un servidor falla Oracle se encarga de

redistribuir los servicios en los servidores restante. La segunda función es IC que permite la configuración, gestión y todas las funciones de gestión de clusters.

- Reducción de costos al no necesitar de grandes máquinas sino del uso de los equipos con los que se cuenta para distribuir las tareas.
- Facilidad de administración al reducir el número de parámetros para la configuración.
- Se ha incorporado un nuevo motor de diagnóstico llamado Automatic Database Diagnostics Motor
- Se ha creado una nueva herramienta que permite almacenar automáticamente información sobre diagnósticos de rendimiento y recomendaciones de tuning a la que han denominado Automatic Workload Repository.
- Esta nueva versión continúa ofreciendo un servicio robusto en entornos con un único servidor como se ha hecho tradicionalmente.
- Se ha aumentado el rendimiento en cerca de un 28% en el gestor de BD en comparación con la versión anterior 9i.

3.8 Sun Microsystems.

Es una de las primeras organizaciones que ha ofrecido colocar a disposición de las organizaciones poder de cómputo y almacenamiento bajo el sistema de

suscripción para poder acceder a su red diseñada para funcionar como un Grid. El nuevo Sun Grid apoyará la nueva generación de las IT, esta herramienta tiene la capacidad de operar en sistemas de alto rendimiento y sistemas operativos como Solaris 10 y N1 Grid Engine.³³

Sun Microsystems a interpretado como una oportunidad, implantar su Sun Grid a diferentes niveles de la industria al punto que ha iniciado el establecimiento de Centros Sun Grid en Estados Unidos (Virginia, Texas y New Jersey), Canadá y Europa (Escocia).

Al igual que IBM, Sun incluye dentro de su sistema operativo SUSE Linux una herramienta diseñada exclusivamente para la aplicación en ambientes Grid.

El apoyo que han brindado las grandes compañías de la computación como HP, Sun Microsystems, IBM, mediante la creación del GGF, han colocado a la cabeza de los desarrollos la herramienta Globus como el estándar para la aplicación de esta tecnología. Se debe tener en cuenta que al tanto Globus como Gridbus poseen herramientas similares para la prestación de los servicios sobre la Grid.

El Grid Forum organismo encargado de estandarizar los servicios y protocolos que son usados para poder llevar a la practica los conceptos de Computación Grid junto con el apoyo que han brindado las grandes compañías del sector de la computación como HP, Sun Microsystems, IBM, SGI, Microsoft, Fujitsu, Hitachi y NEC han seleccionado la herramienta Globus Toolkit como es estándar a seguir cuando se trata de crear proyectos grid actuales, y esto no quiere decir que las otras herramientas no permitan la creación de estos proyectos, pero la gran mayoría de Toolkits existentes ha sido desarrollada en el seno de instituciones académicas.

³³ APC-news.com

La arquitectura sobre la que se basa Globus (GT3) y las nuevas características (GT4) permite a los programadores orientar sus herramientas hacia una implementación más transparente con los servicios Web que conocemos actualmente en Internet. Esta nueva característica es una de las principales fortalezas de la herramienta que la han convertido en el estándar.

Globus al igual que las demás herramientas enfoca un nivel de su estructura a la parte de la seguridad pero en Globus se puede obtener mayor seguridad, gracias al desarrollo de protocolos que brindan a los usuarios la tranquilidad saber que la información que viaja por la red se encuentra encriptada con un alto nivel de seguridad gracias al manejo de claves públicas y certificados.

Existen herramientas con características más desarrolladas que Globus (GT3) en algunos aspectos como el caso de Unicore que está diseñado para el trabajo por lotes, pero al igual que el mercado de cualquier herramienta, la acogida por parte de la comunidad científica y de negocios han permitido recopilar las debilidades de Globus en su versión 3 para lograr mejorarlas en su versión 4 (GT4) y ponerla a disposición de entidades académicas y empresariales.

Globus es en el momento la herramienta Toolkit con la mejor arquitectura en el área herramientas de soporte y seguridad para la creación de nuevos proyectos basados en la Computación Grid. Gracias a herramientas como OGSA Y OGSF que se encargan de generar y soportar la gran mayoría de recursos necesarios por los usuarios que desean implementar esta nueva herramienta. Es importante destacar que Globus a diferencia de otros Toolkits permite la escalabilidad y aceptación de las diversas plataformas existentes. El hecho de que las casas productoras de software y hardware desarrollen sus nuevos productos orientados a soportar las nuevas herramientas incorporadas en GT4 muestra la fortaleza que tiene esta herramienta sobre las otras de su rama.

4. APLICACIONES ACTUALES

Empresas, organizaciones, universidades muy reconocidas en el campo de la computación y la investigación han iniciado el desarrollo de aplicaciones Grid entre ellas podemos destacar.

➤ La organización Europea para la Investigación Nuclear CERN en Ginebra Suiza, ha planteado como solución a sus problemas de almacenamiento y procesamiento de sus datos la utilización de las tecnologías Grid; de forma que pueda aprovechar los recursos informáticos distribuidos geográficamente por todo el mundo. No solo en el área nuclear sino también en otras como Altas Energías, Área de la Salud, Bioinformática, Bioingeniería, Astrofísica, Automoción y otras están planteando la necesidad de utilizar tecnologías Grid como solución a sus problemas.

➤ InnerGrid, es el software desarrollado por GridSystems para aplicar a entornos académicos, empresariales e investigativos. Divide en pequeñas unidades los datos y cálculos ajustándose dinámicamente a los agentes instalados en los equipos que constituyen la red. En el caso en que una máquina deje de estar disponible, sus procesos son reasignados automáticamente a otro equipo. Finalizadas todas las tareas el equipo central recopila los datos de todas las máquinas y genera un resultado final.

- Conector Excel, es una extensión de InnerGrid y se encarga de distribuir el cálculo de las celdas entre los equipos disponibles en la Intranet, sin modificar la forma de trabajo del usuario final y así poder resolver el problema en minutos. Muestra los resultados en la misma hoja de cálculo. Este sistema ha sido puesto en funcionamiento en el departamento de modelado de riesgos de mercado del Banco Santander Central Hispano, reduciendo procesos de 1 hora a tan solo 1 minuto manteniendo intactos la programación de las hojas e interfaz del usuario.
- La universidad Politécnica de Valencia (UPV) cuenta con una red Grid interdepartamental con la compañía española GridSystems.

Esta red integra el servidor de la serie ALTIX 3000 de Silicon Graphics, con 48 procesadores Itanium 2, el cluster de IBM con 64 nodos, biprocesador Xeon y con aproximadamente 3000 ordenadores y estaciones de trabajo disponibles en las aulas de informática y diversos departamentos de la UPV bajo una capacidad de multiplataforma (Unix, Linux, y Windows)³⁴.

Este logro permite a usuarios e investigadores avanzar en sus procesos investigativos y múltiples beneficios como tolerancia a fallos, facilidad de uso para el usuario final, así como tener desarrollados conectores específicos para diferentes aplicaciones dentro y fuera de la universidad.

- El Instituto de Física Corpuscular IFIC cuenta con un grupo de ordenadores para el Grid (GOG) y se compone de 192 PC's Athlon (134 procesadores en el IFIC y 58 en el instituto de Ciencia Molecular con

³⁴ www.upv.edu.es

formato 2U para rack comunicados con FastEthernet mediante equipos de comunicación de Gigabit ethernet CISCO. Cada PC posee un procesador AMD Athlon de 1.2 Ghz y de 1.4 Ghz con 1 Gbyte de memoria SDRAM y un disco duro UDMA100de 40 GBytes³⁵

➤ World Community Grid es una organización que encaminara todo el potencial de Grid para analizar los códigos genéticos que subyacen en enfermedades como el SIDA, Alzheimer, Cáncer, mejorar el estudio de pronósticos de desastres naturales y respaldar estudios que sirvan para preservar el abastecimiento de alimentos y agua potable para el mundo. Organizaciones científicas, educativas y filantrópicas como Institute of Health, la Nacional Science Fundación, las fundaciones Duke, Markle, MacArthur, el Mayo Clinic Collage of Medicine, La universidad de Oxford, La organización Mundial de la Salud, el Programa de Desarrollo de Naciones Unidas, United Davices líder en soluciones Grid e IBM son los encargados de apoyar este proyecto. World Community Grid esta alimentado por tecnología IBM, que incluye sistemas IMB eServer p630, x345 y Shark Enterprise Storage Server de IBM con software de base de datos DB2 y los sistemas operativos AIX y Linux

➤ Sun Microsystems posee en las instalaciones de Foro Motor's Empine han Transmisión Groups una Grid que cuenta con 500 workstation Sun Blade con procesador dual o SAAB Automovile Ab que cuenta con 100 workstation Sun. En este sitio se dedican a realizar las pruebas de aerodinámica de los nuevos modelos. La herramienta utilizada para esta actividad es Sun ONE Grid Engine (Capaz de ejecutar 7000 Grids con unas

³⁵ www.ific.uv.es

47 CPU's por Grid)³⁶ Este aplicativo puede encontrarse en la pagina de Sun (www.sun.com) de distribución gratuita y corre bajo Linux o Sun OS, si se desea tener mayores servicios se puede adquirir la versión comercial.

➤ IrisGrid. Se ha constituido en España , tiene como fundamento aportar los estándares necesarios para construir un Grid de investigación dentro de España, coordinando con la diferentes O/V interesadas (ver Figura 14)

Figura 14. O/V participantes en la red IRIS



Fuente. www.rediris.es/irisGrid

³⁶ www.sun.com

5. VENTAJAS DE LA COMPUTACION GRID

- Tolerante a Fallos.
- Permite ahorrar costos.
- Fácil de usar y administrar.
- Flexibilidad en la potencia de uso.
- Tiene el poder de un supercomputador.
- Se puede iniciar con inversiones mínimas.
- No hay gastos de soporte y mantenimiento.
- Flexible para el desarrollador y la institución.
- Se trata de una solución altamente escalable.
- No se necesita la compra de un supercomputador.
- Escalable de acuerdo a las necesidades del usuario.
- No hay pérdida de la infraestructura por obsolescencia.

- La integración de sistemas y dispositivos heterogéneos.
- Mejorar la colaboración entre las organizaciones diversificadas.
- Elimina los cuellos de botella de algunos procesos de computación.
- Funcionará con todas las arquitecturas, Windows, Linux, Unix, Aix y otras.
- Disminuye los costos informáticos gracias a una mejor utilización de los recursos.
- Permite afrontar la solución a nuevos procesos que no se podían realizar por su complejidad.
- Ahorro en inversiones de infraestructura para los nuevos equipos al igual que en equipos de refrigeración.
- No hay necesidad de comprar hardware adicional, para posibilitar el incremento de la potencia de cómputo.
- Resuelve muchos problemas actuales que requieren gran cantidad de cálculos sin necesidad de modificar las aplicaciones.
- La capacidad de los Computadores sube entre 50 y 80%. Hoy, en muchos casos sólo son 20% porque cada Computador nada más acepta una tarea.

- Nunca queda obsoleta, como ocurre con los grandes equipos debido a su capacidad dinámica de modificar el número y características de sus componentes

- Incrementa la productividad, proporcionando a usuarios finales un acceso sin restricción a los recursos informáticos de datos y de almacenamiento que necesitan.

- Se pueden enviar tareas para que sean ejecutadas en aplicaciones que están disponibles sólo en algunos equipos, esta clasificación conlleva ahorro en los costos de licencias.

- Debido al concepto de Computación Grid, de estándares abiertos permite crear una infraestructura única y unificada, liberando a las organizaciones de tecnologías de la información del peso de administrar sistemas no integrados, reduciendo de esta forma la supervisión.

- Dentro de Grid los usuarios comparten los mismos fines y necesidades permitiendo la agrupación de ellos mediante la formación de organizaciones virtuales O/V. Definiendo una serie de privilegios de uso y de acceso a los recursos Grid, manteniendo sus propias políticas de seguridad y gestión de recursos no solo dentro de su propia organización sino para la interacción con otras.

- Facilita la colaboración y promueve la flexibilidad de las operaciones, las ofertas de Grid pueden agrupar no solo recursos tecnológicos distintos, sino también a la gente. Facilitando que el personal pueda compartir, acceder y gestionar información, la tecnología Grid puede hacer que las

organizaciones sean capaces de mejorar la colaboración en todas las unidades del negocio y geográficas para dar soporte a estrategias de globalización.

➤ Efectivamente escalable para favorecer distintos tipos de demandas con Grid los negocios pueden crear infraestructuras flexibles y elásticas que gestionen rápidamente las fluctuaciones y demandas de los clientes permitiendo el acceso instantáneo a los recursos de computación y datos que respondan a las necesidades del negocio. La capacidad de resolver problemas del negocio complejos más rápido significa que las organizaciones pueden moverse más rápidamente y ganar ventajas competitivas en el mercado

➤ Mejora el tiempo de los resultados para nuevos productos y servicios aumentando la productividad y la colaboración, las organizaciones mejoran el tiempo en la obtención de resultados. Tanto si estos resultados incluyen llevar un nuevo producto al mercado más rápidamente, resolver un complejo problema de negocio más pronto, realizar un análisis de datos en profundidad para lanzar un nuevo servicio, la Grid ofrece a las compañías acelerar el tiempo de procesamiento y mejorar su posicionamiento en el mercado.

➤ Mantener las inversiones de capital maximizando la productividad, la utilización eficaz de los recursos existentes es una de las claves para minimizar los costos. Ayudando a asegurarse la utilización óptima de las capacidades informáticas, la tecnología puede ayudar a las empresas a exceso de costos para infraestructura.

➤ Según Meter Zangl, subdirector de la Dirección General de Sociedad de la Información de la Comisión Europea, “En las próximas décadas no solamente serán los investigadores quienes usen esas redes: la tecnología de redes puede revolucionar la forma en que se desarrollan las actividades empresariales presentando cinco ventajas claras:

- 1- Mejor Calidad de los productos y servicios
- 2- Reducir el costo total de propiedad
- 3- Permitir la creación y la prestación de servicios en cualquier momento y lugar
- 4- Contribuir a la próxima generación de Internet
- 5- Proporcionar una infraestructura verbal básica”

6. DESVENTAJAS DE LA COMPUTACION GRID

- La ausencia de un modelo unificado de almacenamiento de datos es un problema potencial.
- Se requiere la interacción de personas y organizaciones, lo que puede generar dificultad para su desarrollo.
- La interoperabilidad del software, ya que se requiere de la creación de normas que permitan estandarizar el software de administración y demás aplicativos necesarios para el funcionamiento de la red. Hay que tener en cuenta que no solo va a existir una única red sino muchas redes interconectadas.

7. EL FUTURO DE LA COMPUTACION GRID

La Computación Grid ha despertado en todo el mundo un alto interés al punto que la Comisión Europea ha decidido invertir 52 millones de euros en su primera fase de una inversión total de 120 millones de euros en proyectos Grid hasta el 2006. Iniciando con el desarrollo de doce (12) proyectos que trabajan en áreas distintas con recursos informáticos en red. Buscando que Europa sea quien lidere el campo de estas tecnologías Grid. “La visión de la Unión Europea incluye una dimensión social mayor, que por ejemplo Estados Unidos, donde se presta más atención a la capacidad de las supercomputadoras. Europa ha de centrarse en ofrecer a toda la comunidad toda la gama completa de dispositivos técnicos. Tenemos una excelente capacidad de investigación en Europa, pero muy poca actividad en términos de comercialización. Las redes no son simplemente otra tecnología, van a cambiar la forma de vivir y de trabajar”³⁷

La cantidad de investigaciones que cada día se van profundizando, el mundo de los negocios que cada vez más tiene como aspecto fundamental el conocimiento de la información para tener los mejores beneficios, provocan que se reciban con beneplácito nuevas herramientas que mejoren la calidad de la información y la rapidez de su acceso. Hay que tener en cuenta que las organizaciones sociales, políticas y todas en general han convergido a un desarrollo global lo que hace que no existan barreras de raza, cultura.

A largo tiempo el desarrollo de esta tecnología traerá grandes beneficios a la comunidad que se reflejarán en mejor calidad de vida al poder descubrir cura a enfermedades, ahorro de tiempo y dinero, se llegará a un punto realmente cercano a lo que se busca con el proceso de globalización, sin barreras políticas. En el

³⁷ Wolfgang Boch Responsable de las Tecnologías Grid en la Comisión Europea

corto tiempo permitirá a las organizaciones ahorrar dinero al aprovechar los recursos existentes y de paso conseguir otros recursos fuera de la organización de una manera más económica y rápida.

La gente común y corriente también va a recibir beneficios, al reducirse los costos de investigaciones se reduce en forma directa el costo de los productos permitiendo que más cantidad de personas accedan a ellos y mejoren su calidad de vida.

Los beneficios de esta tecnología son muchos pero con ellos llegan también nuevas amenazas, las organizaciones van a poder realizar mejores productos en menos tiempo, la información va a estar más expuesta por esto las medidas de seguridad deben desarrollarse mucho más, se podrán realizar investigaciones en áreas que se considera delicadas como el estudio del ADN, este punto tiene varias implicaciones, se podrá conocer como estamos compuestos pero también se puede querer jugar a ser DIOS.

Para conseguir explotar todos los beneficios que aporta esta tecnología se debe cambiar la mentalidad de las organizaciones, las cuales aún tienen miedo de compartir sus recursos.

8. IMPACTO DE LA COMPUTACIÓN GRID EN EL SECTOR ACADÉMICO EN COLOMBIA.

La computación Grid ha tenido su origen principalmente en los campos Científico y académico. Han sido las universidades en compañía o con el apoyo de grandes empresas, las que han venido logrando desarrollos importantes en este campo.

En nuestro país son pocas las personas y entidades que han acogido este concepto para su estudio y desarrollo, aunque ya lleva varios años de estudio en el extranjero. Es importante tener en cuenta que la tecnología llega siempre retrasada.

En el país como se había mencionado, son pocas las instituciones que tienen grupos de trabajo dedicados al estudio de este campo, dentro de ellas se destaca en el campo académico la Universidad Nacional en Bogotá que cuenta con un grupo académico que ha estudiado este tema y que además ha realizado el montaje de una Grid experimental dentro de su institución, denominada UNGRID.

Siendo las universidades focos de estudio e investigación, pero que en el país carecen de apoyo económico para emprender sus proyectos de investigación, la aplicación de computación Grid aporta muchos beneficios en aspectos como son: el costo de inversión en equipos de procesamiento y almacenamiento, tiempo en el desarrollo de los proyectos. Para lograr este objetivo es necesario divulgar y profundizar esta tecnología de manera que pueda ser aplicada.

En la ciudad de Bucaramanga, y mas específicamente en la Universidad Industrial de Santander (UIS), esta tecnología ha sido muy poco divulgada, son muy pocas las personas que la conocen. La acogida de esta tecnología seria muy provechosa, debido a que la universidad cuenta con una red interna de alta

velocidad distribuida en subredes, un alto número de equipos de cómputo personales distribuidos en todos los departamentos con una configuración aceptable y en muchos sitios muy actualizada. Al combinar estos aspectos con la calidad de la educación que se encuentra en el claustro universitario y el apoyo que se recibe para la investigación se puede llegar a obtener buenos resultados.

Se debe destacar que la aplicación de esta tecnología no va encaminada solamente al área de sistemas, sino a muchas áreas de la universidad tanto en lo académico como en lo administrativo y en especial en el campo investigativo.

En la parte académica puede ayudar a que se aumenten las investigaciones en áreas como la química, física, petróleos, diseño de piezas industriales y otros campos al permitir realizar procesos de cálculo y simulación de forma rápida. En la parte académica permiten realizar diversas tareas como el manejo de información de alumnos, notas, generación de informes, almacenamiento de información.

Para implementar esta tecnología en la UIS y lograr resolver un problema de almacenamiento o procesamiento en el área administrativa por ejemplo, sería necesario seguir los siguientes pasos.

- Montar una red Lan que permita interconectar los equipos que van a formar parte de la Grid.
- Las características básicas del equipo que será configurado como servidor debe ser Pentium III 550 Mhz, 64 y 1 GB de espacio en disco
- Descargar una herramienta Toolkit que permita la instalación de la plataforma Grid, en este caso se podría utilizar Globus por ser la herramienta estándar para aplicaciones Grid. Esta herramienta se puede instalar sobre Linux o sobre otro sistema operativo para lo cual se debe descargar los complementos del sitio Web de Globus.

- Instalar y compilar GPT, herramienta que permite instalar y desinstalar los diferentes componentes de Globus
- Instalar Globus, preferiblemente como administrador del sistema “root”
- Configuración de Globus en aspectos como usuarios y host para establecer los aspectos de seguridad y acceso.
- Instalar la herramienta
- Realizar el desarrollo del portal con las herramientas que provee el Toolkit.
- Configurar las diferentes tareas para las cuales se desea utilizar la Grid en el equipo servidor y en el de los usuarios.

Como se puede observar los costos necesarios para el montaje de una plataforma Grid son muy bajos a comparación de los beneficios. De la misma manera el tiempo empleado para la instalación y configuración no es muy alto y algo muy importante es que las herramientas necesarias para poder realizar la instalación se pueden obtener de forma gratuita de la Web.

Esta tecnología puede ser usada en otras facultades en las cuales se requiere realizar simulaciones cálculos que involucran gran capacidad de procesamiento,

La configuración de la Grid puede involucrar no solo un edificio o escuela sino que puede contener mayor cantidad de equipos. De igual manera como se puede instalar este servicio para

Otra aplicación en la parte académica a nivel universitario es aprovechar el convenio existente entre las universidades de la ciudad “UNIRED” y mediante la pagina Web aprovechar los diversos equipos con que cuenta cada una de las universidades para proyectos de investigación, compartir información de bases de datos de proyectos, material bibliográfico de forma más segura.

Un aspecto muy importante es dar a conocer esta tecnología y profundizar en su conocimiento, de forma que se pueda en un futuro próximo realizar montajes de pequeñas Grid experimentales y constatar su potencial.

Con el conocimiento y uso de este tipo de tecnología, los centros universitarios dispondrán de una herramienta que le permitirá alcanzar no solo beneficios a nivel académico sino que además con el desarrollo de investigaciones se pueden obtener beneficios de tipo económico gracias a los logros alcanzados en las investigaciones.

No hay que olvidar que esta tecnología es el siguiente paso en el desarrollo de los alcances de Internet en su misión por globalizar la información y los diferentes dispositivos dispersos por todo el mundo y que pueden llegar a ser utilizados desde cualquier parte.

En la parte comercial e industrial, la Computación Grid aún no ha tenido la connotación suficiente debido a varios factores como:

- Es una tecnología que en el momento se encuentra en desarrollo por parte del sector académico principalmente.
- Las herramientas actuales como el Globus, Manejadores de Bases de Datos como Oracle utilizadas para poder realizar el montaje de esta tecnología, aún no permiten una comunicación satisfactoria con todas las herramientas necesarias para el buen funcionamiento de una Grid.
- Dentro de los conceptos fundamentales de esta tecnología esta la necesidad de compartir recursos físicos e información entre las diferentes organizaciones virtuales, el compartir estos recursos involucra el llegar a un acuerdo entre estas organizaciones para saber que se va a compartir y en que condiciones para poder

establecer una serie de políticas. El tener que establecer que llegar a estos acuerdos es bastante difícil debido a que las organizaciones son temerosas de compartir.

➤ La mentalidad de la gran mayoría de empresarios colombianos, no se arriesgan a invertir en tecnologías que se encuentran en proceso de desarrollo, debido a que tienen una concepción consumista que los obliga a adquirir productos que ya han sido probados con éxito en otros países y los libera de correr riesgos con esta inversión.

➤ La falta de personal con conocimiento en el área que permita orientar este tipo de proyectos involucra inversión no solo en equipos sino en capacitación de profesionales para que realice la tarea de desarrollo y de personal para que conozcan lo que se está realizando y así evitar su rechazo a la nuevo sistema.

No obstante las limitantes que presenta la utilización de la computación Grid en Colombia, existen entidades las cuales poseen no solo la necesidad de utilizar esta clase de herramientas sino también los recursos e infraestructura para realizarlos

En sector financiero el manejo de grandes volúmenes de información para ser procesada al igual que para almacenar, Las entidades bancarias son las que más invierten dinero en la parte tecnológica y por ende un buen candidato a utilizar esta tecnología.

Es importante destacar que en el área financiera ya se cuenta con aplicaciones que han tenido éxito como Conector Excel empleada en Banco Santander Central Hispano.

En Colombia la capacidad de procesamiento que aporta el uso de la Computación Grid podría brindar a las entidades bancarias ahorro en tiempo en la elaboración de procesos de generación de estados financieros cuando se presentan cierres, en procesos en los cuales se debe realizar reliquidaciones o ajustes a sus ahorradores y deudores. En la realización de cálculos estadísticos basados en posibles movimientos de las tasas de interés del mercado entre otras aplicaciones.

En la parte estatal entidades como la DIAN podrían utilizar esta tecnología para poder lograr el objetivo que se han trazado hace varios años como es lograr realizar un procedimiento que permita cruzar la información que reportan las personas en la declaración de renta con la información que reportan las entidades comerciales y demás empresas. Esta tarea involucra una gran capacidad de procesamiento y traería grandes beneficios en la parte económica al permitir conocer a la entidad el estado de cada uno de las personas y empresas registradas en sus bases de datos.

Uno de los principales obstáculos para que no se aplique esta tecnología en este caso es que desde hace varios años la entidad viene desarrollando sin éxito un software propietario que cumpla esta tarea.

El ISS es otra entidad estatal que tiene la necesidad debido a que el volumen de información que tiene que procesar y sobre todo almacenar es alto por lo cual la utilización de esta clase de tecnología le permitiría distribuir de forma segura la carga de datos almacenados dentro de su red y realizar sus procesamientos en menos tiempo. En este caso es bastante difícil que se aplique en el momento debido a los costos de inversión en capacitación y que por ser una entidad estatal el trámite burocrático para su aprobación puede ser demorado.

Otra área en la que tiene cabida el uso de la Computación Grid es el estudio del clima y la simulación de posibles comportamientos de corrientes de aire. En este caso la necesidad radica en el nivel de procesamiento que involucra realizar una simulación al tener que tomar gran cantidad de posibles factores en diferentes posiciones de la geografía. Como beneficio de aplicar la Computación Grid a este campo esta la disminución en el tiempo de procesamiento de las simulaciones y por ende el poder prevenir algunos desastres; lo que se vería reflejado en salvar vidas y en algunos casos también ahorro de dinero al poder tomar algunas medidas preventivas para salvar bienes y cosechas.

En el sector privado esta tecnología se puede aplicar en el diseño industrial en áreas que involucran el diseño de piezas mecánicas permitiendo realizar simulaciones en corto tiempo y de esta manera determinar el comportamiento que tendría bajo diferentes condiciones si se fabrica.

Las empresas prestadoras de servicios podrían utilizar esta tecnología trayendo como beneficio mayor rapidez en los procesos de facturación y de esta forma brindar al usuario información referente a su pago en menos tiempo. Lo anterior brinda a la organización el beneficio de recibir en menos tiempo el pago por los servicios que presta logrando un mejor rendimiento en su flujo de caja.

CONCLUSIONES

Las necesidades computacionales actuales han dado origen a uno de los retos más importantes en el área de la computación y busca dotar a la comunidad de una segunda revolución informática con efectos superiores a los ya alcanzados por Internet y la Web (World Wide Web) al dotarla de múltiples servicios sobre una red extensa y rápida como lo es Internet.

La tendencia a futuro será la agrupación de recursos informáticos en diferentes centros distribuidos geográficamente optimizando los recursos disponibles. Estos recursos se encargarán de guardar datos y pondrán a disposición del usuario potencia de procesamiento y acceso a una red para poder abordar diferentes temas científicos (ámbito multidisciplinario) de primera línea,

El concepto de Computación Grid se ha desarrollado inicialmente en el área académica con una perspectiva netamente científica y técnica. Organizaciones como Sun Microsystems han empezado a ofrecer al público algunos servicios bajo el concepto de Computación Grid de infraestructura para procesamiento datos y zonas de almacenamiento, estos servicios a través de Internet a un costo de US\$ 1 por CPU por hora, mientras que en la suscripción de capacidad de almacenamiento la oferta es de US\$ 1 por GB por mes

Esta tecnología ha sido recibida con entusiasmo por la comunidad científica al punto que se ha dado origen a una organización llamada Globos Alliance (www.globus.org), dentro de la cual se ha creado una serie de agrupaciones entre las que se pueden citar

- Global Grid Forum
- EGrid European Grid Forum
- La comunidad Grid Asia -Pacífico

La computación Grid ha surgido como una herramienta nueva, robusta y fácil de administrar, a un costo menor que las tradicionales y permite principalmente satisfacer las necesidades de procesamiento y almacenamiento que los clientes requieren en sus aplicaciones más complejas.

La Computación Grid Permite compartir y optimizar los recursos computacionales disponibles de una organización, consiguiendo la reducción de costos y mayores beneficios en áreas como la seguridad, rapidez y elasticidad de su estructura informática. Su rápida aceptación por parte de las organizaciones hace pensar que serán una característica típica del futuro tecnológico.

Organizaciones dedicadas a la generación de productos especializados en el manejo de grandes volúmenes de datos y procesamiento; ya están adoptando en sus productos características que permitan aprovechar los beneficios de la Grid debido a que en el momento no se puede acceder fácilmente a ellos usando esta nueva tecnología.

La Computación Grid será virtualmente dominante, al estar permanentemente disponible; permitiendo que sus servicios se adapten a las necesidades de los usuarios bajo los principios de sencillez, transparencia, economía y fácil administración.

La estrecha relación entre el desarrollo de la tecnología de hardware y software, el aumento en la capacidad de almacenamiento, velocidad de transmisión y la disminución en los costos de producción, han permitido y seguirán haciendo que el flujo de datos entre equipos de cómputo aumente logrando un mayor avance en

los desarrollos Grid. Gracias a estos avances se ha podido determinar que es más importante tener estándares abiertos que código abierto

Las últimas versiones de herramientas que aplican los conceptos de Computación Grid se enfocan en utilizar sus servicios de la misma forma que Internet lo hacen los Web Service, que son aplicaciones que utilizan estándares XML, UDDI y SOAP. Tienen la desventaja en Internet de no poder crear varias instancias de un mismo servicio y luego cuando ya no son requeridas eliminarlas, tampoco han sido incluidas en su desarrollo servicios de apoyo como son las notificaciones, servicios de persistencia, gestión del ciclo de vida entre otros. Todas estas falencias del Web Service se espera sean resueltas con la versión 4.0 de Globus, donde se crean instancias individuales para un servicio y no una única instancia para múltiples usuarios

Es necesario que se diseñen aplicaciones Web sencillas, transparentes a los usuarios para que puedan acceder los distintos servicios Grid disponibles. Se ha considerado que el umbral apropiado para poder plantearse la solución a un problema mediante un servicio Grid es de $c2d > 1.5GFH/GB$. Habitualmente, la carga computacional suele estimarse por GFLOPS, que en una hora serán 3.6 billones de instrucciones (1 GFH). La cantidad de datos (o código) que es necesario mover se suele medir en GigaBytes (GB).

No todas las aplicaciones comerciales se pueden implementar a través de Grid. Además, el licenciamiento actual de software no contempla la ejecución distribuida como algo habitual aunque las nuevas versiones del software están permitiendo esta característica dentro de su infraestructura.

La aceptación de Grid en el área de los negocios ha sido bien recibida, pero todavía no se concibe bajo su principio fundamental que es la globalización de los

recursos. En estas primeras etapas de desarrollo de aplicaciones se ha visto como las organizaciones han preferido instalar IntraGrids, que tiene el mismo principio de la Computación Grid pero solo utiliza los recursos propios de la organización.

BIBLIOGRAFIA

LIBROS

Ian Foster and Carl Kesselman. Globus: A metacomputing infrastructure toolkit. International Journal of Supercomputer Applications, 11(2):115-128 1997

TESIS

FERNANDEZ CASINI, Alvaro. Arquitecturas Grid orientadas a la gestión de recursos. Trabajo Investigación Instituto de Física Corpuscular. Diciembre 2004.

HERNANDEZ BRAVO, Angel. Programa de Doctorado: Sistemas Distribuidos. Proyecto Globus, Enero 2004

Manual de Implementación Globus Toolkit 3.2.1 Grid Experimental , Julio C Baracaldo, Oscar Fajardo, Gastón Mejía. Julio 2004

TARRIO BARREIRO, Jacobo. Diseño e implementación de un portal basado en tecnologías Grid para el acceso a recursos de súper computación. Proyecto de Grado Facultad de Informática. Universidad de la Coruña. 2003

REVISTAS

IEEE. Grid Services for Distributed System Integration. Vol 35 No. 6 Junio 2002. p 37-46

INTERNET

[www. ific.uv.es](http://www.ific.uv.es) (Diciembre 15 de 2004)

www.sun.com (Diciembre 15 de 2004)

www.ibm.com (Diciembre 15 de 2004)

www.oracle.com (Diciembre 15 de 2004)

www.uptv.edu.es (Diciembre 15 de 2004)

<http://www.ggf.org> (Diciembre 15 de 2004)

<http://gridshere.org> (Diciembre 18 de 2004)

<http://sara.unile.it/grb> (Diciembre 19 de 2004)

www.globusworld.org (Diciembre 18 de 2004)

www.seti@home.com (Diciembre 15 de 2004)

<http://www.gridbus.org> (Diciembre 15 de 2004)

www.mcs.anl.gov/~foster (Diciembre 1 de 2004)

<http://terra.com/noticias> (Diciembre 5 de 2004)

<https://hotpage.npaci.edu> (Diciembre 19 de 2004)

<http://www.gridsphere.org> (Enero 5 de 2005)

<http://www.cs.wisc.edu/condor/> (Enero 10 de 2005)

<http://arcos.inf.uc3m.es/~fgarcia/sd> (Enero 15 de 2005)

www.sia.eui.upm.es/docent/ai.htm (Enero 25 de 2005)

<http://www.ietf.org/rfc/rfc/rfc3280.txt> (Enero 20 de 2005)

<http://www.mcs.anl.gov/~jms/ggf-sched> (Enero 15 de 2005)

http://ific.uv.es/rei/Paseo/REI_Grid_3.pdf (Noviembre 18 de 2004)

www.globus.org/research/papers/anatomy.pdf (Febrero 1 de 2005)

<http://www.biologia-en-internet.com/default.asp> (Febrero 7 de 2005)

<http://www.globus.org/research/papers/ogsa.pdf> (Diciembre 28 de 2004)

<http://arcos.inf.uc3m.es/~fgarcia/sd/presentaciones/> (Enero 15 de 2005)

http://www.ceditec.etsit.upm.es/grid_computing.php (Enero 11 de 2005)

[posiciona.webcindario.com/Clustering/**clustering**.html](http://posiciona.webcindario.com/Clustering/clustering.html) (Enero 24 de 2005)

<http://irisgrid.rediris.es/coord/fig2/gt2003-irisgrid-intro.pdf> (Diciembre 3 de 2004)

[grid.ifca.unican.es/**cursos**/presentaciones/DB_GRID.ppt](http://grid.ifca.unican.es/cursos/presentaciones/DB_GRID.ppt) (Enero 9 de 2005)

[www.pcm.gob.pe/portal_ongei/publica/**proyectos**/4819.pdf](http://www.pcm.gob.pe/portal_ongei/publica/proyectos/4819.pdf) (Diciembre 2 de 2004)

<http://www.risklab-madrid.uam.es/seminarios/seminarios.html> (Noviembre 15 de 2004)

http://gsyc.escet.urjc.es/simple_com/phd-thesis-es/node12.html (Febrero 14 de 2005)

[www.grid.ifca.unican.es/English/Presentations/uc/**Crossgrid**.pp](http://www.grid.ifca.unican.es/English/Presentations/uc/Crossgrid.pp)
(Diciembre 5 de 2004)

<http://www.unicore.org/documents/UNICOREPlus-Final-Report.pdf>
(Abril 1 de 2005)

http://www.jatonribes.com/seti/about_seti/about_seti_at_home_1.html

(Abril 1 de 2005)

http://www.consumer.es/web/es/especiales/2004/10/11/110178_3.php

(Diciembre 8 de 2004)

<http://www.ejournal.unam.mx/compuysistemas/vol03-03/CYS03302.pdf>

(Diciembre 19 de 2004)

[arcos.inf.uc3m.es/~fgarcia/sd/presentaciones/ Arquitectura/Seguridad,.ppt](http://arcos.inf.uc3m.es/~fgarcia/sd/presentaciones/Arquitectura/Seguridad,.ppt)

(Diciembre 29 de 2004)

<http://www.monografias.com/trabajos14/aplicacion-distrib/aplicacion-istrib.shtml>

(Enero 17 de 2005)

[www.microsoft.com/spanish/MSDN/estudiantes/computadores/paralelas/**clustering**.asp](http://www.microsoft.com/spanish/MSDN/estudiantes/computadores/paralelas/clustering.asp) (Febrero 25 de 2005)

<http://www.lawebdelprogramador.com/cursos/enlace.php?idp=2660&id=48&texto=Linux> (Diciembre 15 de 2004)

<http://www.microsoft.com/spanish/MSDN/estudiantes/computadores/paralelas/clustering.asp> (Diciembre 15 de 2004)

<http://www.cs.virginia.edu/~humphrey/papers/PhilosophicalTechnicalComparisonOfLegionAndGlobus.pdf> (Abril 15 de 2005)

www.rediris.es/jt/jt2003/archivo-jt/SALAA/05112003/SesionI/PresJT2003-Jacob-GRID-5Nov-SalaB-18h00.ppt (Diciembre 27 de 2004)

<http://www.globus.org/research/papers/anatomy.pdf> Ian Foster, Carl Kesselman, Steven Tuecke, Anatomy of the Grid:Enabling Scalable Organizations (Noviembre 15 de 2004)

<http://www.globus.org/research/papers/physiology.pdf> Ian Foster, Carl Kesselman, Steven Tuecke, The Physiology of the Grid an open Grid Services Architecture for Distributed Systems Integration (Noviembre 15 de 2004)