

COMPUTACION GRID

**JUAN MANUEL GUERRERO ESTRADA
DAVID EDUARDO JIMENEZ BACCA**

**UNIVERSIDAD TECNOLOGICA DE BOLIVAR
FACULTAD DE INGENIERIA
PROGRAMA DE INGENIERIA ELECTRICA Y ELECTRONICA
CARTAGENA
2011**

COMPUTACION GRID

**JUAN MANUEL GUERRERO ESTRADA
DAVID EDUARDO JIMENEZ BACCA**

**ASESOR
MAGISTER EN CIENCIAS COMPUTACIONALES
EDUARDO GOMEZ VASQUEZ**

**UNIVERSIDAD TECNOLOGICA DE BOLIVAR
FACULTAD DE INGENIERIA
PROGRAMA DE INGENIERIA ELECTRICA Y ELECTRONICA
CARTAGENA
2011**

**A mi familia, por su incondicional apoyo,
A mi abuelita que desde lo más alto me apoyo
Y me acompañó a alcanzar este gran logro en mi vida**

D.E.J.B

**Doy gracias a mi familia por el apoyo incondicional que me brindaron en
el proceso de esta dura carrera,
A mis padres y hermano que siempre sacaron lo mejor de mí para
alcanzar esta meta.
Y a Dios todo poderoso quien me brindo la sabiduría para cumplir con
este gran sueño.**

J.M.G.E

CONTENIDO

	PÁGINA
LISTA DE TABLAS	I
LISTA DE FIGURAS	II
GLOSARIO	III
RESUMEN	IV
INTRODUCCIÓN	1
JUSTIFICACIÓN	2
1. DESCRIPCIÓN DEL PROBLEMA	4
2. OBJETIVOS	6
2.1 GENERALES	
2.2 ESPECIFICOS	
3. PASADO, PRESENTE Y FUTURO DE LA COMPUTACIÓN GRID	7
3.1 INTRODUCCIÓN A LA COMPUTACIÓN GRID	8
3.2 LA EVOLUCIÓN DE LA COMPUTACION GRID	11
3.2.1 Comienzos de la Red	12
3.2.2 Un modelo de comunidad Grid	15
3.2.3 Grid Service	19
3.3 LOS CONCEPTOS Y COMPONENTES DE LA GRID	21
3.3.1 Los Tipos de Recursos	21
3.3.2 Computación	21
3.3.3 Almacenamiento	22
3.3.4 Las Comunicaciones	24
3.3.5 El Software y las Licencias	26
3.3.6 El Equipo Especial, Capacidades, Arquitecturas, y Políticas	26
4. COMPUTACIÓN GRID Y OTRAS TECNOLOGÍAS	28
5. APLICACIONES DE LA COMPUTACIÓN GRID	36
5.1 ACTUALIDAD EN EL AMBITO INTERNACIONAL	36
5.2 A NIVEL DE LATINOAMERICA	51

5.3 A NIVEL NACIONAL	52
6. CONCLUSIONES	59
BIBLIOGRAFIA	60

LISTA DE TABLAS

	PÁGINA
Tabla 1. Comparación entre tecnologías GRID y P2P.	30
Tabla 2. Comparación entre tecnologías GRID, Clúster y P2P (Supercomputador).	32
Tabla 3. Proyectos utilizando la Computación Grid en Estados Unidos y Europa.	39

LISTA DE FIGURAS

	PÁGINA
Figura 1. Grid resources linked together for neuroscientist Mark Ellisman's.	8
Figura 2. Infraestructura Grid que involucra Hardware y Software.	10
Figura 3. Entorno computacional para el análisis de datos en tiempo real tomadas en la fuente de fotones avanzada.	12
Figura 4. Capas de la arquitectura del Modelo de Comunidad Grid.	16
Figura 5. Arquitectura Grid.	17
Figura 6: Data Striping.	24
Figura 7. Diagrama P2P, Topología Estrella.	29
Figura 8. Concepto de Computación GRID básica.	29
Figura 9. Esquema de la Tecnología Redes de sensores sin cable (WSN).	33
Figura 10. Aplicación del GRID en la bioinformática.	40
Figura 11. Proyecto SETI@home.	42
Figura 12. PPDG	45
Figura 13. Prospecto de GriPhyN	46
Figura 14. Estructura del EU DataGRid en sus bloques de componentes de trabajo	49
Figura 15. Arquitectura de Cross Grid	50

GLOSARIO

- **Middleware:** Es un software de conectividad que ofrece un conjunto de servicios que hacen posible el funcionamiento de aplicaciones distribuidas sobre plataformas heterogéneas.
- **Software:** Se refiere al equipamiento lógico o soporte lógico de una computadora digital, y comprende el conjunto de los componentes lógicos necesarios para hacer posible la realización de tareas específicas.
- **Hardware:** Corresponde a todas las partes físicas y tangibles de una computadora, sus componentes eléctricos, electrónicos, electromecánicos y mecánicos; sus cables, gabinetes o cajas, periféricos de todo tipo y cualquier otro elemento físico involucrado.
- **Tecnología:** se utiliza para definir a los conocimientos que permiten fabricar objetos y modificar el medio ambiente, con el objetivo de satisfacer las necesidades humanas. De acuerdo a la Real Academia Española, la tecnología es el conjunto de teorías y técnicas que permiten el aprovechamiento práctico del conocimiento científico. Cabe destacar que, en forma errónea, se utiliza la palabra tecnología como sinónimo de tecnología informática, que es aquella que permite el procesamiento de información por medios artificiales y que incluye todo lo relacionado con las computadoras.
- **Máquina:** Aparato o conjunto de aparatos capaces de efectuar un trabajo o de llevar a cabo una función, ya sea dirigida por un operador, ya sea de forma autónoma; cuyo funcionamiento posibilita aprovechar,

dirigir, regular o transformar energía o realizar un trabajo con un fin determinado.

- **Programa:** El término programa de acuerdo a distintos diccionarios se ha definido de diversas formas como: que programa se define como un plan y orden de actuación, organización del trabajo dentro de un plan general de producción y en unos plazos determinados, o como la secuencia precisa de instrucciones codificadas en un ordenador para resolver un problema así como la declaración previa de lo que se piensa hacer en alguna materia u ocasión, o el anuncio o exposición de las partes de que se han de componer ciertas cosas o de las condiciones a que ha de sujetarse. Todas estas definiciones coinciden en que el programa se realiza previo a la acción.

- **Homologar:** Equiparar, poner en relación de igualdad o semejanza dos cosas; Contrastar una autoridad oficial el cumplimiento de determinadas especificaciones o características de un objeto o de una acción.

- **Acceso remoto:** En informática, se considera Administración Remota a la funcionalidad de algunos programas que permiten realizar ciertos tipos de acciones desde un equipo local y que las mismas se ejecuten en otro equipo remoto.

- **Entorno Computacional:** Es un sistema complejo que puede llegar a estar constituido por millones de componentes electrónicos elementales.

- **Estado Del Arte:** Dícese del trabajo de algo que se está mostrando, que es lo que se ha hecho en dicho proceso y resalta lo último que se ha alcanzado.

- **Redes Globales:** Son aquellas que se extienden sobrepasando las fronteras de las ciudades, pueblos o naciones. Los enlaces se realizan con instalaciones de telecomunicaciones públicas y privadas, además por microondas y satélites.
- **Algoritmo:** Es un conjunto pre-escrito de instrucciones o reglas bien definidas, ordenadas y finitas que permite realizar una actividad mediante pasos sucesivos que no generen dudas a quien deba realizar dicha actividad.
- **Usuario:** Es la persona que utiliza o trabaja con algún objeto o que es destinataria de algún servicio público, privado, empresarial o profesional.
- **Datos:** es una representación simbólica (numérica, alfabética, algorítmica etc.), un atributo o una característica de una entidad. Los datos son hechos que describen sucesos y entidades. No tienen ninguna información. Puede significar un número, una letra, o cualquier símbolo que representa una palabra, una cantidad, una medida o una descripción. El dato no tiene valor semántico (sentido) en sí mismo, pero si recibe un tratamiento (procesamiento) apropiado, se puede utilizar en la realización de cálculos o toma de decisiones.
- **Comunicaciones:** es el proceso mediante el cual se puede transmitir información de una entidad a otra. Los procesos de comunicación son interacciones mediadas por signos entre al menos dos agentes que comparten un mismo repertorio de signos y tienen unas reglas semióticas comunes.

RESUMEN

La monografía "COMPUTACIÓN GRID", responde a la necesidad de investigación, divulgación y formación en esta tecnología basada en el aprovechamiento de recursos subutilizados por gran cantidad de equipos conectados a la Red, permitiendo abordar y solucionar tareas que son demasiado exigentes para una sola máquina y además, el trabajo cooperativo a nivel global.

Fue elaborado entre los meses de noviembre de 2010 y mayo del año 2011 por estudiantes del Programa de Ingeniería Electrónica en La Universidad Tecnológica de Bolívar, que cursan El Minor en Telecomunicaciones requisito para optar al título.

El trabajo integra la consulta, recopilación, comparación y concreción de datos sobre el tema con la descripción final o tratado en el cual se muestra las bases teóricas y evolución histórica de la computación GRID con el fin de informar sobre los avances, ventajas y posibilidades que provee para que sean conocidos y evaluados. Paralelamente, la monografía busca que dicha tecnología sea implementada en nuevas aplicaciones que repotencialicen su funcionalidad.

La metodología de trabajo fue la pertinente con la investigación bibliográfica y los objetivos responden a la necesidad de tener una fuente de consulta sobre "Computación GRID" de fácil comprensión y libre acceso. Esta necesidad fue arrojada por la selección temática realizada entre las opciones propuestas por el director del programa.

Palabras Claves:

- Tecnología
- Computación Grid
- Conexión

INTRODUCCION

La presente monografía se enmarca como trabajo de grado para la obtención del título de Ingeniero Electrónico por parte de los autores. Es requisito de la universidad la realización de un trabajo de investigación de este tipo para la aprobación del Minor en Telecomunicaciones.

El trabajo consta de la descripción sistemática de la Computación Grid a través de la historia con la presentación de comparaciones y ejemplos de aplicaciones y funcionamiento. Esto implica que su finalidad sea divulgarla para que las comunidades científicas y comerciales a nivel global hagan uso de ella.

En primera instancia la monografía parte puntualizando los conceptos básicos y evolución de la *computación GRID*, seguido de una comparación con otras tecnologías. Posteriormente se hace énfasis en las aplicaciones de la tecnología GRID en el ámbito global, latinoamericano y nacional. Finalmente se describen las conclusiones.

Se presentan también tablas, gráficos y análisis que dinamizan la comprensión sobre el tema.

JUSTIFICACION

Los acercamientos computacionales para resolver problemas han probado su valor en casi cada campo del esfuerzo humano. Los computadores son utilizados para modelar y simular complejos problemas de la ciencia y la ingeniería, diagnosticar condiciones médicas, controlar equipo industrial, pronosticar el tiempo y muchos otros propósitos. Aunque todavía algunos problemas retadores pueden exceder nuestra habilidad para resolverlos, estamos usando los computadores mucho menos de lo que realmente debería ser.

Un claro ejemplo de ello, son las investigaciones que se realizan en las universidades, en las cuales se hace un extenso uso de los computadores para estudiar unos aspectos y otros no, generando desequilibrio, de hecho es posible que los usen para estudiar el impacto de los cambios del suelo en la biodiversidad, pero no para realizar las planificaciones de la ciudad como seleccionar las rutas para las nuevas carreteras o proyectar las nuevas zonas de ordenamiento.

Hay varias razones probables para la omisión de los métodos computacionales en la resolución de problemas, entre ellos tenemos la escasez de educación en este campo y la falta de herramientas apropiadas, convirtiendo en un factor importante para el fracaso de algunos proyectos la falta de posibilidades para intercambiar, rastrear y divulgar información y además el bajo promedio de ambientes computacionales para propósitos de cómputo sofisticados [13].

Por lo anterior se hace necesaria la implementación de un sistema computacional donde la mayoría de instituciones tengan acceso y en el cual se pueda transmitir y compartir conocimiento de forma fácil y segura. En estas condiciones, la única solución viable consiste en compartir recursos entre

grupos de forma solidaria hasta reunir recursos suficientes para abordar los problemas experimentales, en otras palabras utilizar la COMPUTACION GRID para la implementación de redes globales para intercambio de información.

1. DESCRIPCION DEL PROBLEMA

La Grid Computing es una tecnología innovadora que permite utilizar de forma coordinada todo tipo de recursos (entre ellos cómputo, almacenamiento y aplicaciones específicas) que no están sujetos a un control centralizado. En este sentido es una nueva forma de computación distribuida, en la cual los recursos pueden ser heterogéneos (diferentes arquitecturas, supercomputadores, clusters) y se encuentran conectados mediante redes de área extensa (por ejemplo Internet).

El termino Grid tiende de mediados de los años 90's, refiriéndose a una propuesta de infraestructura de computación distribuida, para trabajos avanzados de ciencia e ingeniería. Estas se plantearon como un nuevo modelo para resolver problemas de computación masiva, utilizando varias computadoras organizadas en una infraestructura de telecomunicaciones distribuida, cuya finalidad es compartir recursos heterogéneos situados en distintos lugares y pertenecientes a diferentes dominios de administración, sobre una red que utiliza estándares abiertos.

En la actualidad, existen muchos proyectos que han sido desarrollados en esta línea tecnológica, entre ellos Edonkey, Emule o Limewire. Se trata de programas para compartir datos a nivel mundial entre diferentes máquinas. Un punto común entre Grid y las Peer-to-Peer, especialmente en la idea de compartir recursos. Entre las diferencias existentes, se pueden tomar la P2P como más anónimas y generalizadas a ordenadores de usuarios de Internet, mientras que las Grid nacen de una estructura de nodos más controlada y jerarquizada en centros científicos.

Las empresas que han participado en el desarrollo de esta tecnología, están buscando cuanto antes la manera entrar en una etapa de explotación comercial, es por eso que conocer el pasado, presente y las tendencia a las

que se orienta el desarrollo de la computación Grid, puede abrir un abanico de oportunidades para participar en esa explotación comercial.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Presentar un estudio sobre el estado del arte, las ventajas y el uso de la Computación GRID.

2.2. OBJETIVOS ESPECIFICOS

- Analizar las diversas teorías que hacen parte del pasado y presente de la Computación GRID.
- Comparar la Tecnología GRID con otras homologables que permitan dar criterios para su utilización.
- Mostrar las aplicaciones más relevantes de la Computación GRID en el ámbito nacional e internacional.

3. PASADO Y PRESENTE DE LA COMPUTACION GRID

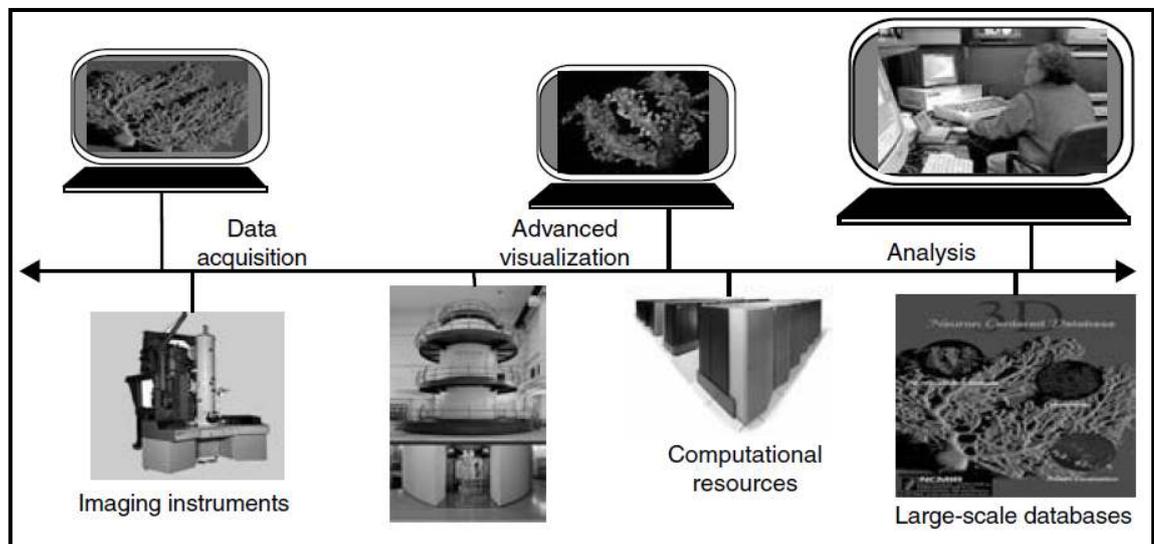
En el transcurrir de la última década, se nota como se ha desarrollado un cambio en la forma de manejar y hacer uso de los recursos informáticos. Este cambio se ha dado gracias a la asimilación de materias primas y componentes de computadoras en red, además del desarrollo de hardware y software mucho más sofisticados y didácticos.

Como es normal, todo cambio trae consigo una serie de consecuencias, en este caso muy favorables dentro de las cuales se destaca la capacidad de utilizar eficaz y eficientemente los recursos distribuidos para dar cumplimiento a las necesidades y aplicaciones.

Hablar de sistemas distribuidos, es básicamente referirse a una serie de computadoras conectadas y comunicadas entre sí, que tengan la posibilidad de compartir tanto información como recursos entre a cada una de ellas. El diseño y la búsqueda a la solución de los problemas en este tipo de sistemas ha sido tema de investigación por mucho tiempo; como resultado de estas investigaciones se han implementado las “middleware”; herramientas de uso cooperativo que permiten compartir recursos distribuidos geográficamente, y pueden actuar como una sola plataforma de gran escala para la ejecución de aplicaciones paralelas y distribuidas.

A partir de que este enfoque de la computación nació como tema de investigación atravesando por su desarrollo y aplicaciones, ha recibido muchos nombres pero en la década de los 90's se acuñó el término de *Computación Grid*, con el cual se conoce actualmente.

Figura 1. Grid resources linked together for neuroscientist Mark Ellisman's



Fuente: <http://www.npaci.edu/Alpha/telescience.html>

3.1 INTRODUCCION A COMPUTACION GRID

Luego del desarrollo de mejores tecnologías en los primeros computadores y la posibilidad de conectarse a Internet, nace la idea de aprovechar la potencia inutilizada de cada uno de estos equipos, con la finalidad de abordar problemas que exigían una mayor capacidad y uso de más recursos, los cuales; únicamente eran brindados por supercomputadoras de uso gubernamental, académicos y pocas veces comerciales.

Esta necesidad, dio origen a la tecnología llamada *Computación Grid*, la cual está basada en el aprovechamiento de recursos subutilizados por gran cantidad de equipos conectados a la Red, permitiendo abordar y solucionar tareas que son demasiado exigentes para una sola máquina.

La *Computación Grid*, como todo avance tecnológico presenta varios beneficios, entre ellos: [1]:

- Flexibilidad para llenar las necesidades cambiantes

- Alta calidad a bajo costo
- Facilita el pronto retorno de inversiones
- No necesita de una nueva infraestructura para implementar su uso
- Facilita el poder de computación
- Poder de supercomputación
- Utilizar software gratuito y código de fuente abierto
- No precisar hardware adicional, para posibilitar el incremento de la potencia del computo
- Transparencia para el usuario que participa en la Grid

Básicamente, el enfoque tomado por la *Computación Grid* es el acceso remoto a recursos computacionales y pretende ser un paradigma de desarrollo no centrado en una tecnología concreta [2].

Esto permite que la tecnología Grid esté incluida dentro del marco del concepto de computación distribuida, donde igualmente se aplican y utilizan conceptos como sistemas operativos distribuidos, programación multiprocesador, redes de computadoras, programación paralela, etc.

La evolución de esta tecnología se ve afianzada a través de la estandarización, bajo la cual se encuentra definida la arquitectura, servicios y requisitos, entre otros recursos fundamentales para su aplicación.

Por estas características, la Computación Grid, se considera como una gran solución a problemas de distribución de información y escasez de recursos ya que brinda la posibilidad de resolver tareas haciendo uso de información y recursos que se encuentran en ubicaciones geográficas lejanas con respecto al lugar donde se trabaja.

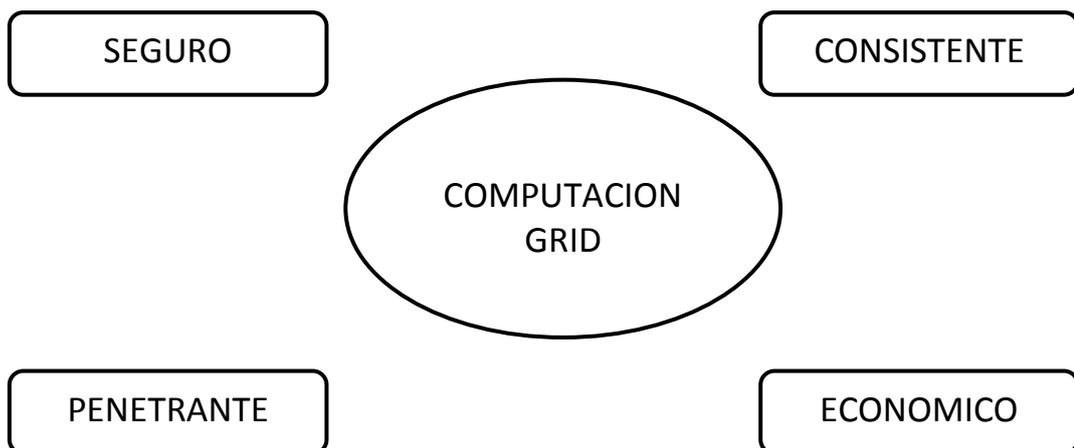
Si bien anteriormente se ha hablado de la Grid como una tecnología, vale aclarar que, se define como una infraestructura, ya que es un conjunto de

recursos, tanto físicos como lógicos, que necesitan de interconexión de hardware y control de software. Es importante resaltar que esta infraestructura debe brindar a quien la utilice altos niveles de seguridad en calidad de cómputo, seguridad de acceso, integridad de datos, etc.

Se considera entonces, que la Computación Grid es una infraestructura que involucra hardware y software capaz de brindar al usuario un acceso seguro, consistente, penetrante, a bajo costo, y con grandes capacidades computacionales [3].

Se asume también, que la Computación Grid debe ofrecer consistencia al usuario, lo cual implica que debe estar basada en estándares que definan el acceso y operaciones lo que permite evitar la heterogeneidad, garantizando la facilidad de conexión y acceso a cualquier punto. De manera continua, esta infraestructura debe facilitar que luego de acceder se pueda obtener toda la potencia de la Grid, primando la facilidad de llegar a cualquier punto sobre la oportunidad de intervenir cualquier recurso.

Figura 2. Infraestructura Grid que involucra Hardware y Software.



Fuente: Monografía de Adscripción a Sistemas Operativos Grid Computing
Autor: Verónica Vanessa Barrios

3.2 LA EVOLUCIÓN DE LA COMPUTACIÓN GRID

GRID es la informática y la infraestructura de gestión de datos que proporcionará el fundamento electrónico para la conformación de una sociedad global en los negocios, gobierno, investigación, ciencia y entretenimiento. [4].

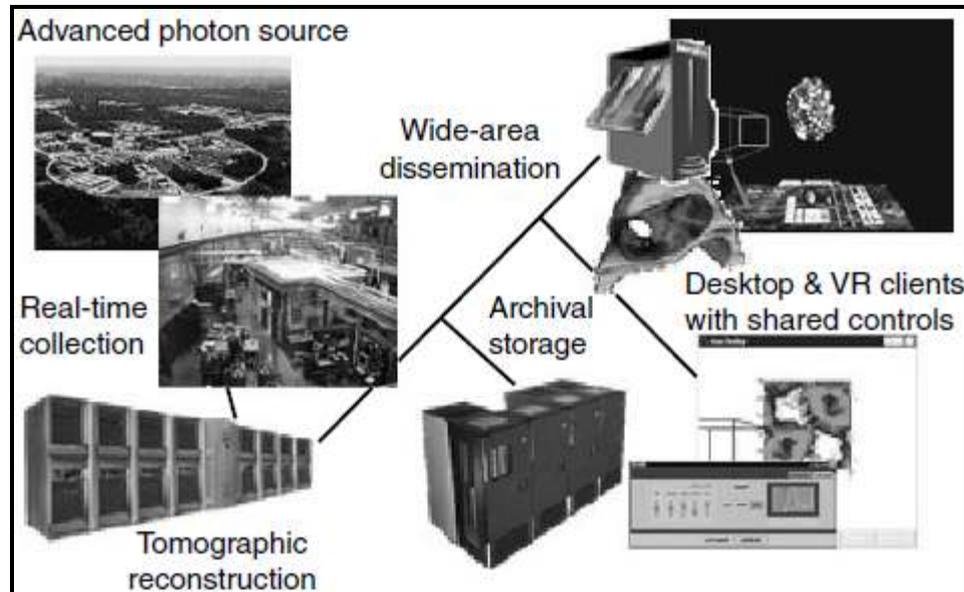
Un esquema de GRID, como se ilustra en la Figura 1, integra redes de comunicación, computación e información proporcionando una plataforma virtual para el cálculo y gestión de datos de la misma manera que Internet integra recursos para formar una plataforma virtual de información.

GRID es la transformación de la ciencia, los negocios, la salud y la sociedad, ya que brinda la posibilidad de enlazar dinámicamente un conjunto de recursos para apoyar la ejecución a gran escala e intensivamente, de recursos y aplicaciones distribuidas.

La Figura 3 muestra una aplicación típica de principios de éxito con la información canalizada a través de los sistemas distribuidos [5].

La realidad es que para alcanzar esta promesa, los sistemas complejos de software y servicios se deben desarrollar permitiendo el acceso de fácil y su eficaz utilización conjunta, adicionando la aplicación de políticas que permitan a las comunidades de usuarios coordinar los recursos de manera estable.

Figura 3. Entorno computacional para el análisis de datos en tiempo real tomadas en la fuente de fotones avanzada.



Fuente: <http://epics.aps.anl.gov/welcome.html>

3.2.1 Comienzos de la Red

Es necesario comprender las influencias que se reunieron para intervenir en el desarrollo de la Computación Grid. Tal vez el mejor momento para empezar es la década de 1980, una década de intensa investigación, desarrollo y despliegue de hardware, software y aplicaciones para computadoras paralelas. La computación paralela en la década de 1980 se centró en el desarrollo de algoritmos, programas y arquitecturas que apoyaron la simultaneidad.

Como desarrolladores de aplicaciones se comenzó con códigos de gran escala lo cual llevó al límite de los recursos, incluso a las computadoras más rápidas. Algunos grupos comenzaron a buscar una distribución más allá de los límites de la máquina como una forma de lograr resultados para los problemas de mayor tamaño.

Durante los años 1980 y 1990, los software para computadoras paralelas se centraron en la prestación de poderosos mecanismos para la gestión de la comunicación entre procesadores, desarrollo y entornos de ejecución para máquinas paralelas. Parallel Virtual Machine (PVM), Message Passing Interface (MPI), High Performance Fortran (HPF) y OpenMP se han desarrollado para apoyar la comunicación de aplicaciones escalables [6].

Los Paradigmas de éxito de aplicación se han desarrollado para aprovechar el inmenso potencial de las arquitecturas de memoria compartida y distribuida. Inicialmente se pensó que la red sería más útil en la ampliación de los paradigmas de computación paralela de grupos fuertemente acoplados a los sistemas distribuidos geográficamente. Sin embargo, en la práctica, la Red se ha utilizado más como una plataforma para la integración de aplicaciones de acoplamiento flexible algunos componentes de los cuales podrían estar en ejecución en paralelo de una máquina de baja latencia y para vincular los recursos dispares (cómputo, almacenamiento, visualización, instrumentos).

Dos conceptos fundamentales en la Computación Grid son la coordinación y distribución de datos. En la década de 1990, el programa de banco de pruebas Gigabit EE.UU. [7] incluye un enfoque en la distribución de área metropolitana y las aplicaciones de área amplia. Cada uno de los bancos de pruebas, Aurora, Blanca, Casa, Néctar y VISTAnet fue diseñado con dos objetivos: investigar posibles arquitecturas de red de banco de pruebas y explorar su utilidad para los usuarios finales. En el segundo objetivo, cada banco de pruebas es siempre un lugar para experimentar con las aplicaciones distribuidas.

La primera Grid moderna fue desarrollada como un proyecto de demostración experimental de SC95. En 1995, durante la conferencia de Supercomputación de una semana de duración, los investigadores pioneros se reunieron para agregar un banco de pruebas nacionales distribuidas con más de 17 sitios conectados en red por el vBNS. Más de 60 aplicaciones fueron desarrolladas

para la conferencia y desplegadas en el I-WAY, así como una rudimentaria infraestructura Grid de software para facilitar el acceso, reforzar la seguridad, coordinar los recursos y otras actividades. Esto era importante, pero el desarrollo de la investigación Grid requiere un enfoque muy diferente a la investigación de computación distribuida.

Considerando que la investigación de computación distribuida en general, se centra en abordar los problemas de la separación geográfica, la investigación Grid se centra en abordar los problemas de integración y gestión de software.

A finales de 1990, los investigadores de Grid se reunieron en el Foro de GRID, posteriormente ampliaron el Global Grid Forum (GGF) [8], donde gran parte de las primeras investigaciones está convirtiéndose en la base de estándares para Redes presentes y futuras. Recientemente, el GGF ha sido instrumento en el desarrollo del Open Grid Services Architecture (OGSA), que integra Globus y enfoques de los servicios web. OGSA está siendo desarrollado tanto por los Estados Unidos como por iniciativas europeas con el objetivo de definir los servicios básicos para una amplia variedad de áreas, incluyendo:

- Sistemas de Gestión y Automatización
- Carga de trabajo / Gestión del rendimiento
- Seguridad
- Disponibilidad y Gestión de Servicios
- Gestión de recursos lógicos
- Servicios por Agrupamiento
- Gestión de Conectividad
- Gestión de Recursos Físicos.

En la actualidad, GRID se ha vuelto global, con muchas colaboraciones a nivel mundial entre los Estados Unidos, Europa y los investigadores de Asia y el Pacífico (Especialmente Suramérica). Los organismos de financiación, los proveedores comerciales, investigadores académicos, y centros y laboratorios

nacionales se han unido para formar una comunidad de amplia experiencia con una enorme dedicación a la construcción de la Computación Grid. Por otra parte, la investigación en las áreas relacionadas con la creación de redes, bibliotecas digitales, Peer-To-Peer, colaboratorios están proporcionando ideas adicionales pertinentes a la red.

3.2.2 Un modelo de comunidad Grid

En la última década, la comunidad Grid ha comenzado a converger en un modelo de capas que permite el desarrollo de complejos sistemas de servicios y software necesario para integrar los recursos de las diferentes estructuras en cuanto a redes se refiere. Este modelo, proporciona una abstracción de capas de la cuadrícula. La Figura 4 ilustra el modelo de Community Grid que se está desarrollando de una manera poco coordinada en todo el mundo académico y el sector comercial.

La capa inferior horizontal del modelo de Community Grid consiste en el hardware los recursos que subyacen a la cuadrícula. Esos recursos incluyen computadoras, redes, archivos de datos, instrumentos, dispositivos de visualización y así sucesivamente. Están distribuidos, heterogéneos y tienen perfiles de rendimiento muy diferentes (como el rendimiento de contraste se mide en FLOPS o ancho de banda de memoria con el rendimiento, medido en bytes y el tiempo de acceso a datos).

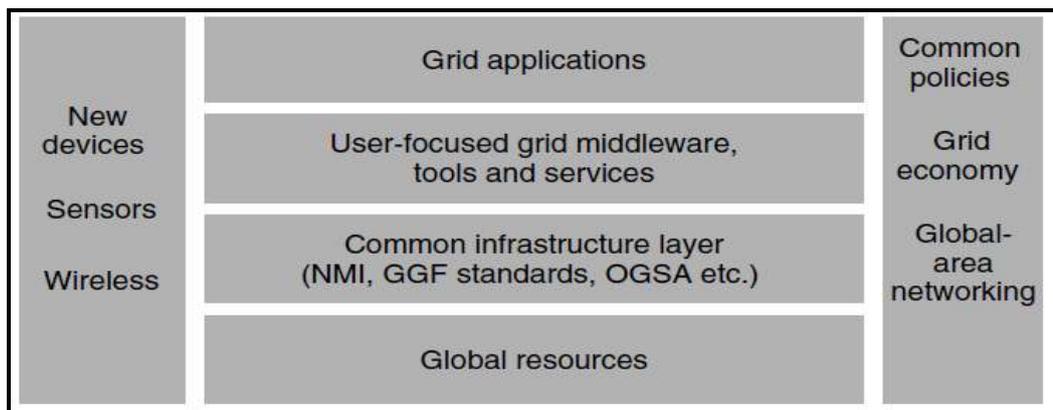
Por otra parte, el fondo de recursos que representa esta capa es muy dinámico, tanto como resultado de los recursos que se agregan a la mezcla y de diferentes resultados observables de los recursos en el entorno multiusuario para compartir, de la Grid.

La siguiente capa horizontal (infraestructura común) se compone de los servicios de software y sistemas que virtualiza la cuadrícula. Comunidad de

apoyo como Middleware NSF Initiative (NMI), OGSA, así como las nuevas normas como Globus proporcionan un común acuerdo sobre las capas en la que además de heterogéneos pueden ser dinámicos los recursos de acceso a la red.

La siguiente capa horizontal (usuario y middleware Grid aplicación enfocada, herramientas y servicios) contiene paquetes de software construido sobre la infraestructura común. Este software sirve para que las aplicaciones puedan utilizar de forma más productiva los recursos Grid y enmascarar algunas de las complejidades implicadas en las actividades del sistema, tales como la autenticación y la transferencia de archivos.

Figura 4. Capas de la arquitectura del Modelo de Comunidad Grid

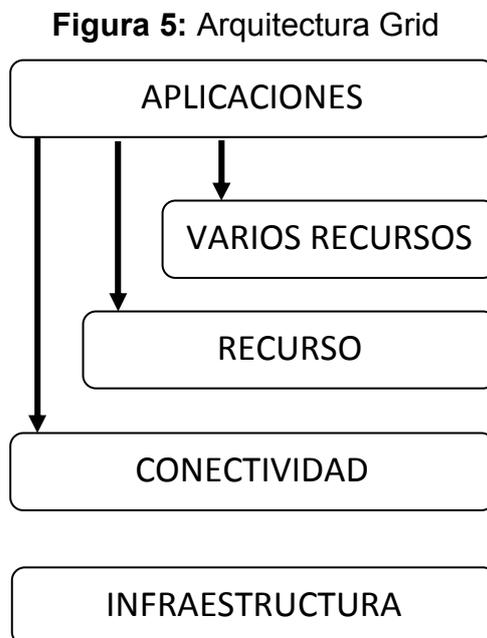


Fuente: Fran Berman and Anthony J.G.Hey. Grid Computing. Making the Global Infraestructure a Reality.

La capa superior horizontal (aplicaciones Grid) representa a las aplicaciones y usuarios. La cuadrícula en última instancia, debe ser tan exitosa como su comunidad de usuarios y todas las otras capas horizontales deben asegurarse de que la cuadrícula presenta una plataforma de gestión sólida, estable, utilizable y útil de cómputo y de datos para el usuario. En el más amplio sentido, las aplicaciones que utilizan sólo un único recurso en la red son aplicaciones Grid.

Las capas verticales representan los pasos a seguir para el desarrollo de la red. La capa vertical de la izquierda representa la influencia de los nuevos dispositivos, PDAs y dispositivos inalámbricos. Durante los próximos 10 años, estos y otros nuevos dispositivos tendrán que ser integrados con la cuadrícula y se agravan los problemas de la gestión de la heterogeneidad y la promoción de rendimiento. Al mismo tiempo, la creciente globalización de la cuadrícula requiere una consideración seria de las políticas para el intercambio y el uso de los recursos, la creación de redes mundiales, de la zona y el desarrollo de las economías. A medida que enlazan redes conjuntas nacionales para formar una red global, será cada vez más importante desarrollar políticas de cuadrícula social y económica que garanticen la estabilidad del sistema, promover la implementación de los usuarios e integrar con éxito diferentes culturas, política y tecnologías en su aplicación.

De una manera más generalizada se puede analizar la distribución de la arquitectura de la *Computación Grid* a través de niveles:



Fuente: Monografía de Adscripción a Sistemas Operativos Grid Computing
Autor: Verónica Vanessa Barrios

Esta arquitectura permite que los recursos físicos existentes se sigan utilizando, sin necesidad de hacer una inversión para su reemplazo, lo cual brinda la posibilidad de utilizar los recursos a medida que las necesidades lo vayan exigiendo.

Gracias a esta arquitectura, es que la Computación Grid ha sido utilizada, para grandes tareas en el área científica, de investigación, con simulaciones, análisis estadísticos, que a diferencia de la Grid empresarial ha sido para actividades en tiempo real y optimización informática y de recursos; en la actualidad Oracle tiene como objetivo lo últimamente mencionado.

Esta arquitectura está basada en protocolos, lo cual quiere decir que, entre los niveles, los usuarios pueden negociar y gestionar condiciones para explotar al máximo el uso compartido de recursos.

En el nivel de infraestructura, se encuentran los recursos computacionales, tales como computadoras, en este caso servidores, bases de datos, etc. También está incluida lo que a estructura de red se refiere, mecanismos de control y gestión.

El nivel de conectividad, concierne a los protocolos de comunicación y seguridad, utilizados para permitir que los recursos físicos puedan comunicarse. Dentro de estos protocolos se encuentran, la pila del protocolo TCP/IP, protocolo SSL, certificados X.509.

La seguridad en la Grid, es un tópico importante, ya que, con esta tecnología se pueden compartir recursos en diferentes ubicaciones y que pueden tener diferentes mecanismos de seguridad; es por ello que una condición favorable de esta arquitectura basada en protocolos, es como se menciono anteriormente, la posibilidad de negociar condiciones.

El nivel recursos, es donde se lleva a cabo la negociación, control y monitoreo de un recurso específico, al igual que toda la información del mismo. En él se encuentran los protocolos que permiten obtener información técnica del recurso, tal como capacidad, carga actual, ubicación, etc. Además se encuentran y ejecutan los protocolos de control y monitoreo del recurso, los cuales permiten acceso al recurso, arranques de procesos, gestión, paradas, monitoreo.

Varios recursos, es el nivel donde se involucran todos los servicios que pueden gestionar a través de un conjunto de recursos. Dentro de estos servicios, se encuentran, el servicio de directorios, el cual permite al usuario encontrar recursos de su interés; los planificadores distribuidos, es el servicio que permite asignar tareas a cada recurso, es decir, programa la actividad y el momento de actuar de cada recurso, según las necesidades del usuario, la monitorización y diagnóstico de la ejecución de distintas tareas en una aplicación, la contabilidad, la cual posibilita el cálculo del costo de la utilización de varios recursos heterogéneos; y el acceso a datos distribuidos, quien gestiona el uso compartido de recursos para la replicación de datos.

El nivel último de aplicación, es donde los protocolos de esta capa permiten a las aplicaciones acceder a la infraestructura de la Grid, por medio de las distintas capas, aunque, es posible que se acceda simplemente la capa necesitada de la infraestructura.

3.2.3 Grid Service (Servicio Grid)

Grid Service es una tecnología que se basa en Arquitectura Orientada a Servicios, donde una aplicación se constituye por componentes independientes y cooperadores llamados servicios. Estos servicios construyen bloques que utiliza un modelo de objeto para crear sistemas distribuidos abiertos y habilitar

a los usuarios para que puedan crear rápida y globalmente sus aplicaciones disponibles para la red. [9]

Existen mecanismos para crear y administrar servicios Grid, los cuales son habilitados al momento de desarrollar un nuevo servicio que será desplegado dentro de un sistema OGSA (Arquitectura de Servicios Grid Estándar). Estos mecanismos son:

Fábrica: Es una clase especial donde se crea dinámicamente instancias de servicios Grid, códigos de servicios Grid ejecutables y se está por requerimientos.

Otros mecanismo es el Registro, que es la interfaz que habilita una serie de momentos en los servicios Grid que permiten registrar el Grid Service Handle (Administrador de Servicios Grid), dentro de un servicio de registro que permite la identificación en ese conjunto.

También existe un mecanismo llamado descubrimientos, el cual facilita al usuario conocer a través de una interfaz los servicios proporcionados.

Ciclo de vida, es referente a todos los instantes por los cuales atraviesa un servicio Grid, desde su creación hasta su destrucción.

Datos del servicio, es un mecanismo que asocia toda la información necesaria de un momento del servicio Grid.

Notificación, es el mecanismo que cumple en la instancia en que una parte solicita una información y parte correspondiente envía dicha notificación.

Invocación Fiable, es el mecanismo que se encarga de garantizar la invocación de métodos en caso que hayan sido creados múltiples instancias con servicios Grid redundantes en el espacio.

Es importante tener en cuenta que el único contacto directo que tiene un usuario con un Grid Service es la interfaz de servicios. Estas interfaces de servicios son definidas por el lenguaje de Descripción de Servicios Web. [10]

3.3 LOS CONCEPTOS Y COMPONENTES DEL GRID

A continuación se introducen varios conceptos del Grid, componentes, y términos más detallados.

3.3.1 Los Tipos de Recursos

Un Grid es una colección de máquinas, a veces llamado “nodos”, “recursos”, “miembros”, “servidores”, “clientes”, “organizadores”, y muchos otros.

Todos ellos aportan grupos de recursos al Grid. Algunos recursos pueden ser usados por todos los usuarios del Grid, mientras que otros pueden tener restricciones específicas.

3.3.2 Computación

El recurso más común lo constituyen los ciclos de cómputo proporcionados por los procesadores de las máquinas en el Grid.

Los procesadores pueden variar en velocidad, arquitectura, plataforma de software, y otros factores asociados, como la memoria, almacenamiento, y conexión.

Hay tres formas primarias para aprovechar los recursos de cómputo de un Grid: El primero y más simple es usarlo para ejecutar una aplicación existente en una máquina disponible en el Grid, en lugar de localmente. El segundo es para usar una aplicación diseñada para dividir su trabajo de tal manera que las partes separadas pueden ejecutarse en paralelo en los diferentes procesadores. El tercero es ejecutar una aplicación que precise ser ejecutada muchas veces en muchas máquinas diferentes en el Grid. “Escalabilidad” es una medida de cómo se usan eficazmente los múltiples procesadores en un Grid.

Si los procesadores realizan dos veces una aplicación completa en la mitad del tiempo, entonces se dice que es absolutamente escalable.

Sin embargo, puede haber límites en la escalabilidad cuando las aplicaciones sólo pueden ser divididas en un número limitado de partes ejecutables separadamente, o si esas partes experimentan alguna otra contención para recursos de algún tipo de almacenamiento.

3.3.3 Almacenamiento

El segundo recurso más común en un Grid es el almacenamiento de los datos. Un Grid que proporciona una visión integrada de almacenamiento de datos es a veces llamado un “Grid de datos”.

Cada máquina en el Grid normalmente mantiene alguna cantidad de almacenamiento para uso del Grid, aun temporalmente. El almacenamiento puede ser memoria unida al procesador o almacenamiento secundario, usando el disco rígido u otros medios de almacenamiento permanente.

La memoria unida a un procesador normalmente tiene un muy rápido acceso pero es volátil. Sería mejor usarlo para guardar datos y servir como un almacenamiento temporal para las aplicaciones ejecutadas.

El almacenamiento secundario en un Grid puede usarse de manera interesante para aumentar la capacidad, la calidad, el compartir, y la fiabilidad de los datos.

Muchos sistemas de Grid usan sistemas de archivo de red montables tales como, Andrew File System (AFS), Network File System (NFS), Distributed File System (DFS), o el General Parallel File System (GPFS). Estos ofrecen diferentes grados de calidad, seguridad y fiabilidad.

La capacidad puede ser incrementada usando el almacenamiento en múltiples máquinas con un sistema de archivo unificado.

Cualquier archivo individual o base de datos pueden expandirse a varios dispositivos de almacenamiento y máquinas, eliminando las restricciones de máxima capacidad, a menudo impuestas por los sistemas de archivo de los sistemas operativos.

Un sistema de archivo unificado también puede mantener un sólo espacio uniforme para el almacenamiento del Grid. Esto facilita a los usuarios referenciar los datos que residen en el Grid, sin considerar su ubicación exacta.

De manera similar, el software especial de base de datos puede “federar” un conjunto de bases de datos individuales y archivos para formar bases de datos más amplias y más comprensivas.

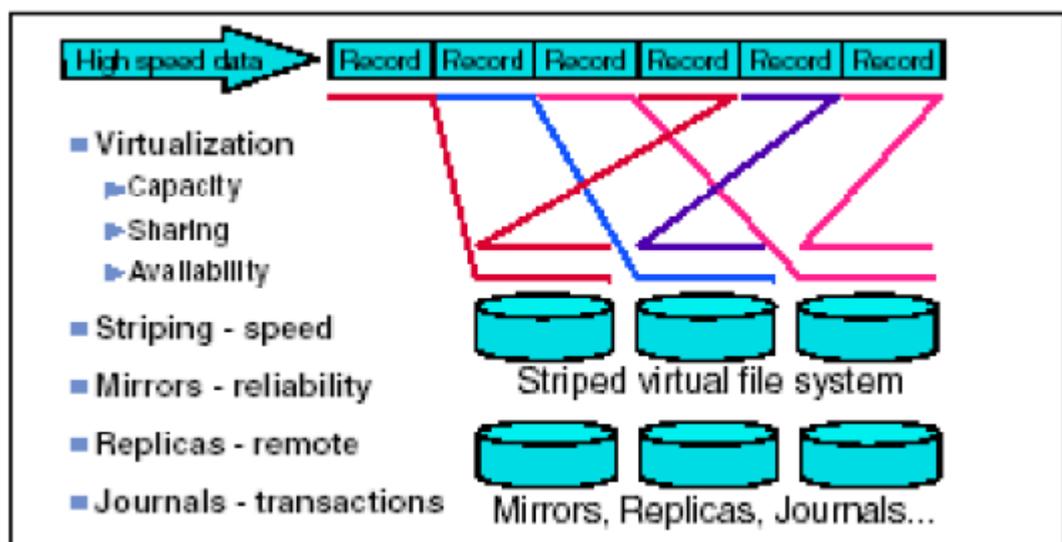
Como se observa en la figura 6, el Data Striping significa escribir o leer sucesivos datos desde o para diferentes dispositivos físicos, solapando el acceso para un mejor rendimiento; asimismo, con técnicas adicionales se aumenta la fiabilidad.

Los sistemas de archivo más avanzados en un Grid pueden duplicar automáticamente conjuntos de datos, para proporcionar redundancia para mayor fiabilidad y aumento del rendimiento.

Un scheduler de Grid inteligente puede ayudar a seleccionar los dispositivos de almacenamiento apropiados para guardar datos, basados en patrones de uso.

Los trabajos pueden ser programados más cerca de los datos, preferentemente en las máquinas directamente conectadas a los dispositivos del almacenamiento que guardan los datos requeridos.

Figura 6: Data Striping



Fuente. Vannesa Barrios V. Licenciatura en Sistemas de Información.
Corrientes - Argentina: Universidad Nacional del Nordeste

3.3.4 Las Comunicaciones

El crecimiento rápido en la capacidad de comunicación entre las máquinas hace que el Grid Computing sea práctico, comparado con el limitado ancho de banda disponible cuando la computación distribuida emergió por primera vez.

Por consiguiente, no debe ser una sorpresa que otro recurso importante de un Grid sea la capacidad de comunicación de datos. Esto incluye las comunicaciones dentro del Grid y fuera de él.

Las comunicaciones dentro del Grid son importantes para enviar trabajos y sus datos requeridos hacia puntos dentro del Grid.

Algunos trabajos exigen procesar una cantidad grande de datos y no siempre los mismos pueden residir en la máquina que ejecuta el trabajo.

El ancho de banda disponible para tales comunicaciones frecuentemente puede ser un recurso crítico que puede limitar la utilización del Grid.

El acceso de comunicación externo a Internet, por ejemplo, puede ser valioso al construir motores de búsqueda.

Las máquinas en el Grid pueden tener conexiones a Internet externas, en adición a las conexiones entre las máquinas del Grid. Cuando estas conexiones no comparten la misma ruta de comunicación, se agregan al ancho de banda disponible, para acceder a Internet.

A veces se necesitan rutas de comunicación redundantes para manejar mejor fallas potenciales en la red y el excesivo tráfico de datos. En algunos casos se deben utilizar redes de altas prestaciones para satisfacer demandas de trabajos que transfieren cantidades muy grandes de datos.

Un sistema de administración de Grid puede mostrar mejor la topología del Grid y puede resaltar los cuellos de botella de comunicación. Esta información puede usarse para planificar las actualizaciones del hardware.

3.3.5 El Software y las Licencias

El Grid puede tener el software instalado que puede ser demasiado caro para instalar en cada máquina de éste.

Usando un Grid, los trabajos que requieren este software son enviados a máquinas particulares en las cuales este software está instalado.

Cuando las tarifas de licencia son significantes, esta aproximación puede ahorrar gastos importantes para una organización.

Algún arreglo de licencia de software permite instalarlo en todas las máquinas de un Grid pero puede limitar el número de instalaciones que pueden usarse simultáneamente en cualquier momento.

El software de administración de licencias registra cuántas copias coexistentes de éste están usándose y previene que se ejecute un número mayor en un tiempo dado.

Los schedulers de trabajo de Grid pueden configurarse para tener en cuenta las licencias de software, opcionalmente balanceándolas contra otras prioridades o políticas.

3.3.6 El Equipo Especial, Capacidades, Arquitecturas, y Políticas

Las plataformas en el Grid tendrán a menudo diferentes arquitecturas, sistemas operativos, dispositivos, capacidades, y equipos.

Cada uno de estos ítems representa un tipo diferente de recurso que el Grid puede usar con criterio para asignar trabajos a las máquinas.

Mientras algún software puede estar disponible en varias arquitecturas, por ejemplo PowerPC y x86, tal software es a menudo diseñado para ejecutar sólo un tipo particular de hardware y sistema operativo. Tales atributos deben ser considerados al asignar recursos en el Grid.

En algunos casos, el administrador de un Grid puede crear un nuevo tipo de recurso artificial que será usado por el schedulers para asignar el trabajo según el tipo de política u otras restricciones. Por ejemplo, algunas máquinas pueden diseñarse para sólo ser usadas para investigación médica y otras para sólo participar en el Grid si no se usan para propósitos militares.

4. COMPUTACIÓN GRID Y OTRAS TECNOLOGÍAS

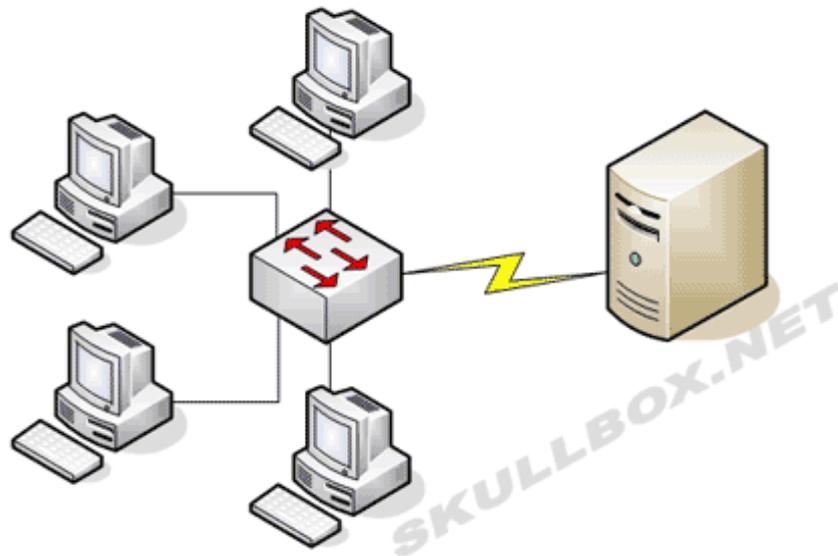
Existen muchas tecnologías que han sido desarrolladas y/o elaboradas con la finalidad de agilizar la transferencia de información, se trata de programas para compartir datos a nivel mundial entre diferentes máquinas. La computación GRID es una de ellas, la cual es muy semejante a la tecnología Peer-To-Peer (P2P), principalmente en la compartición de recursos.

Hoy día la computación GRID está en batalla campal (por denominarlo de alguna manera) con otras tecnologías semejantes, que buscan desbancarse unas a otras a partir de un criterio básico establecido "La satisfacción total del cliente".

La llamada contienda se lleva principalmente con las tecnologías Peer-to-Peer (P2P), Tecnología Clúster (Grupal) y Wireless Sensor Networks (Redes de Sensores sin Cables).

Realizando la comparación entre la computación GRID y la P2P, entre las características diferenciadoras podemos ver la P2P como más anónima y generalizada en ordenadores de usuarios de Internet, mientras que la computación GRID nace de una estructura de nodos más controlada y jerarquizada en centros científicos. Una primera experiencia fue GriPhyN, para unir nodos en el proceso de plantas físicas de altas energías en Estados Unidos.

Figura 7. Diagrama P2P, Topología Estrella



Fuente: Basic Networking Topology (Topología Básica de Redes), Server – Base Network. Pagina Web www.skullbox.net/ntopology.php

Figura 8. Concepto de Computación GRID básica



Fuente: Ramón Jesús Millán Tejedor, Artículo Formativo n°43, acta del 2007. Consultoría Estratégica en Tecnologías de la Información y Comunicaciones.

La comparación entre estas dos tecnologías puede verse ítem por ítem en el siguiente cuadro dándose a conocer las ventajas de la GRID frente a la P2P:

Tabla 1. Comparación entre tecnologías GRID y P2P

ITEM	Computación GRID	Peer-To-Peer (P2P)
Descentralización	Pueden agregarse recursos sin importar su localización geográfica.	Estas redes por definición son descentralizadas y todos los nodos son iguales. No existen nodos con funciones especiales.
Heterogeneidad	<p>Todo recurso puede ser integrado. Pueden ser dos tipos:</p> <ul style="list-style-type: none"> - Hardware: Ordenadores o dispositivos electrónicos. - Software: Los recursos agregados pueden ser de todo tipo (Servicios, documentos, imágenes, bases de datos...) y estar presentes en cualquier sistema operativo (Windows, UNIX, MacOSX...) 	Se comparten o donan recursos a cambio de recursos. Según la aplicación de la red, los recursos pueden ser archivos, ancho de banda, ciclos de proceso o almacenamiento de disco.
Escalabilidad	La infraestructura puede ser aumentada continuamente sin alterar procesos y sin que se resienta su eficiencia.	Son diferentes a una arquitectura del modo servidor-cliente con un sistema fijo de servidores, en los cuales la adición de clientes podría significar una transferencia de datos más lenta para todos los usuarios.
Multipropósito	La infraestructura puede	No cumple para todas las

	utilizarse para todo tipo de aplicaciones.	aplicaciones debido a la repartición de costos entre los usuarios.
Seguridad	Es la tecnología más eficiente que existe hoy día en este ámbito de información.	Es una de las características deseables de las redes P2P menos implementada.

Fuente: Borja Sotomayor, Introducción a la computación GRID

Se puede considerar también una tercera tecnología en esta misma gamma y es la tecnología “Clúster”, que se aplica a los conjuntos y/o conglomerados de computadoras construidos mediante la utilización de componentes de hardware comunes y que se comportan como si fuesen una única computadora.

La computación GRID no es un “Clúster” de órdenes (puesto que en un GRID puede haber integrados muchos o ningún “Clúster”).

Se puede considerar a la tecnología GRID como una versión mejorada de P2P; por lo que permite compartir todos los recursos de un equipo (no solo ficheros) y garantiza calidades de servicio no triviales. Además es multipropósito. [11]

A continuación una tabla comparativa donde se muestra la ventaja de la computación GRID, frente a las tecnologías semejantes, “Clúster” y P2P.

Tabla 2. Comparación entre tecnologías GRID, Clúster y P2P (Supercomputador)

	Grid Computing	Cluster	Supercomputador
Descentralización	✓	✗	✗
Heterogeneidad de hardware	✓	✗	✗
Heterogeneidad de software	✓	✗	✗
Escalabilidad	✓	✓	✗
Multipropósito	✓	✗	✗

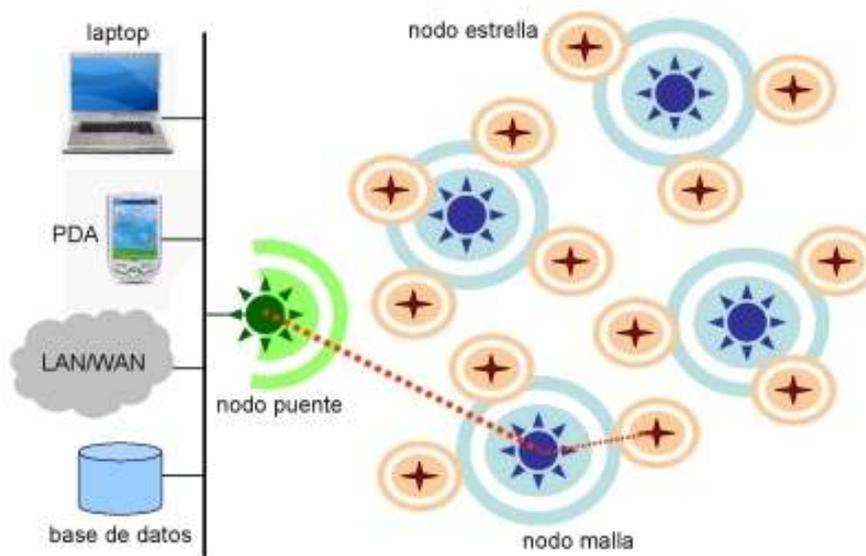
Fuente: Borja Sotomayor, Introducción a la computación GRID

Por otra parte existen otras tecnologías que pueden ser homologables con la tecnología GRID, pero sin embargo no contemplan las ventajas de esta, como es el caso de la tecnología “Redes de sensores sin cable (Wireless Sensor Networks)”, las cuales son redes de nano aparatos autónomos capaces de una comunicación sin cable.

A través de redes de sensores, se puede integrar funcionalidades que antes eran independientes unas de otras, con el fin de lograr máxima eficiencia.

Pero la computación GRID supera a la de Redes de sensores sin Cable debido a que esta última no es nueva y además, sus funciones incluyen medir niveles de temperatura, líquido, humedad, sensibilidad, etc. Aunque las redes de sensores pueden utilizar distintas tecnologías sin cable, incluyendo IEEE 802.11, LANS sin cable, Bluetooth e identificación de la frecuencia de radio.

Figura 9. Esquema de la Tecnología Redes de sensores sin cable (WSN)



Fuente: Tecnologías que cambiarán al mundo, Redes de sensores sin cable (Wireless Sensor Networks). Pagina web: www.taringa.net

El cambio está siendo progresivo, lo que conduce a la mayoría de los grupos de investigación a buscar soluciones a la medida ya que los métodos en uso se iban quedando cortos (Como por ejemplo la tecnología P2P – Peer to Peer). La primera fase de contención ha consistido en la instauración de sistema “Clusters” y adaptación de algoritmos y procedimientos para su ejecución paralela a pequeña escala. Estas soluciones, si bien están permitiendo desarrollar y probar nuevos algoritmos y soluciones, resultan insuficientes para el tratamiento de datos experimentales masivos [12].

Por ello se puede expresar las necesidades básicas como una mayor demanda de capacidad de almacenamiento y tratamiento de información, y un crecimiento desmesurado de la capacidad de cálculo.

Los costes asociados al tratamiento de información derivada de las nuevas técnicas experimentales superan con creces la capacidad de prácticamente cualquier grupo o entidad aislada, aún recurriendo a soluciones paralelas de bajo costo como las granjas de computadores personales.

La computación GRID tiene, como toda tecnología, ventajas sobre otras tecnologías de su misma área; e igual manera desventajas, que se dan a saber a continuación de forma generalizada, a partir de las modificaciones buscando una mejor calidad, para ofrecer mejor servicio a diferencias de otras tecnologías:

Ventajas:

En definitiva, GRID supone un avance respecto a la World Wide Web (WWW): El World Wide Web proporciona un acceso transparente a información que está almacenada en millones de ordenadores repartidos por todo el mundo. Frente a ello, el GRID es una infraestructura nueva que proporciona acceso transparente a potencia de cálculo y capacidad de almacenamiento distribuida por una organización o por todo el mundo. Claramente cumpliendo ciertos requisitos a saber:

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

Y los beneficios que se obtienen:

- Proporciona un mecanismo de colaboración transparente entre grupos dispersos, tanto científicos como comerciales.
- Posibilita el funcionamiento de aplicaciones a gran escala.
- Facilita el acceso a recursos distribuidos desde nuestros PC.

- Todos estos objetivos y beneficios se engloban en la idea de "e-Ciencia".

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

Desventajas

No obstante, la computación GRID presenta algunos inconvenientes que deben solucionarse. Estos problemas son:

- Recursos heterogéneos: la computación GRID debe ser capaz de poder manejar cualquier tipo de recurso que maneje el sistema, si no resultará totalmente inútil.
- Descubrimiento, selección, reserva, asignación, gestión y monitorización de recursos son procesos que deben controlarse externamente y que influyen en el funcionamiento del GRID.
- Necesidad de desarrollo de aplicaciones para manejar el GRID, así como desarrollo de modelos eficientes de uso.
- Comunicación lenta y no uniforme.
- Organizativos: dominios de administración, modelo de explotación y costes, política de seguridad.
- Económicos: precio de los recursos, oferta/demanda.

5. APLICACIONES DE LA COMPUTACION GRID

5.1 ACTUALIDAD EN EL AMBITO INTERNACIONAL

Hoy en día existen un gran número de aplicaciones reales que usan computación GRID, mayoritariamente en el ámbito científico-académico. Aunque existe muchos Grids para tipos específicos de aplicaciones en todo el mundo, el objetivo es la construcción de una Computación Grid a nivel mundial. Los usuarios se conectarían a la "GRID" y accederían a la capacidad de cómputo y de almacenamiento que precisen sin preocuparse de dónde se genera.

Las instituciones y organismos más interesados en el desarrollo de la Computación Grid son, principalmente, los que comparten un objetivo común y que, para poder alcanzarlo, lo más efectivo es compartir sus recursos: Gobiernos y organizaciones internacionales (respuesta a desastres, planificación urbana, etc.); sanidad (análisis rápido de imágenes médicas complejas, etc.); educación (creación de aulas virtuales, teleconferencias, etc.), empresas y grandes corporaciones (cálculos complejos, reuniones virtuales, etc.).

Los beneficios de la Computación Grid, gracias a la integración de recursos distribuidos, están teniendo repercusión en muchísimos campos, de entre los que cabe destacar: medicina (imágenes, diagnosis y tratamiento), ingeniería genética y biotecnología (estudios en genómica y proteómica), nanotecnología (diseño de nuevos materiales a escala molecular), ingeniería (diseño, simulación, análisis de fallos y acceso remoto a instrumentos de control), y recursos naturales y medio ambiente (previsión meteorológica, observación del planeta, modelos y predicción de sistemas complejos).

Las aplicaciones en las cuales se puede destacar la Computación Grid se establecen en tres planos y/o áreas principales, las cuales se detallan consecuentemente:

- 1- Aplicaciones con grandes necesidades computacionales.
- 2- Aplicaciones con grandes necesidades de almacenamiento o proceso de datos.
- 3- Aplicaciones colaborativas.

1- APLICACIONES CON GRANDES NECESIDADES COMPUTACIONALES.

- Simulación, predicción, monitorización.
- Proyecto Cross Grid:
 - ✓ Inundaciones: Modelado y simulación de regiones susceptibles de inundaciones para predecir inundaciones y/o proporcionar datos a equipos de crisis en caso de inundación.
 - ✓ Medicina.
 - ✓ Partículas Físicas.
 - ✓ Aplicaciones meteorológicas: Modelado de la atmósfera y el océano para avisar y diagnosticar el estado del mar.
 - ✓ Contaminación: Simulación de la contaminación aérea.

2- APLICACIONES CON GRANDES NECESIDADES DE ALMACENAMIENTO O PROCESO DE DATOS:

- Aplicaciones que generan un flujo grande y constante de datos.
 - ✓ Por ejemplo el LHC (Gran Colisionador de Hadrones) de 27km de circunferencia desarrollado por el CERN (Organización Europea para la Investigación Nuclear).
 - ✓ El LHC genera datos que se corresponden con 56 millones de cedes cada año. El análisis de datos del LHC requiere un poder

de computación equivalente a 100000 de los procesadores más rápidos que existen en el mercado hoy en día.

- ✓ The Grid: Red de transmisión de fibra óptica que conectara centros de investigación de todo el mundo para almacenar los datos producidos por el LHC. (Aproximadamente 55000 servidores)
- Aplicaciones que se benefician del acceso a datos similares situados en distintas organizaciones.
 - ✓ Por ejemplo, el proyecto “eDiaMoND” para el análisis distribuido de mamografías.

3- APLICACIONES COLABORATIVAS:

- Aplicaciones que, por su naturaleza, son multiorganización y se benefician de una tecnología que facilita la comunicación entre distintas organizaciones.
- Tele conferencias, reuniones virtuales...
 - ✓ AccessGrid: uVigo TV es una de las tecnologías pioneras del uso de esta tecnología en el mundo.
- Colaboración entre sismólogos.
 - ✓ NEESGrid (Network for Earthquake Engineering Simulation): provee software y servicios que permiten a los sismólogos compartir datos y participar en experimentos remotos y simulaciones.
 - ✓ World Comunity Grid
 - Organización no lucrativa
 - Mayor Grid de computación para abordar proyectos que beneficien a la humanidad (catástrofes, enfermedades...)
 - Aprovechar el tiempo de los PC's estén encendidos pero desocupados para ser útiles a la ciencia.

Actualmente la implementación Computación Grid lleva consigo grandes proyectos en desarrollo en el continente Europeo (específicamente en la Europa Occidental) y los Estados Unidos (Al norte del continente americano); de los cuales se conocen los nombres a continuación:

Tabla 3: Proyectos utilizando la Computación Grid en Estados Unidos y Europa.

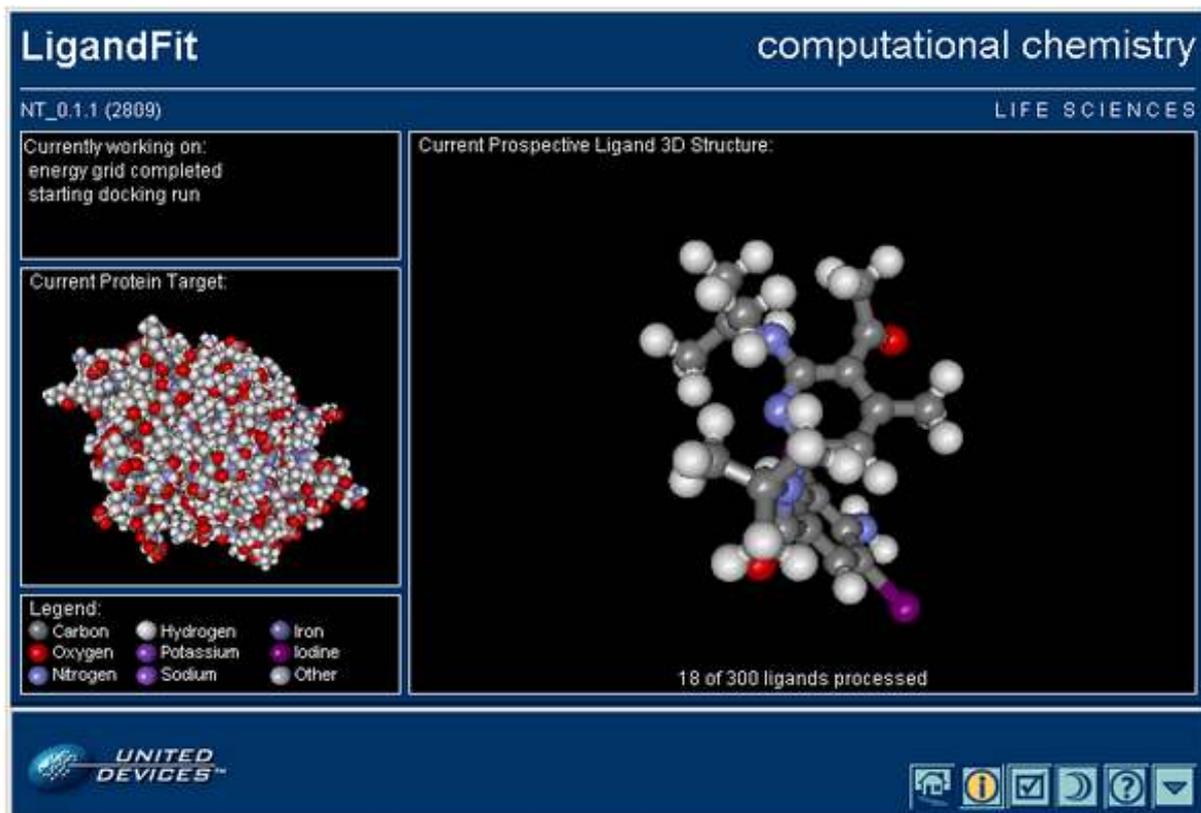
USA	EUROPA
<i>NASA Information Power Grid</i>	<i>EGEE (CERN)</i>
<i>DOE Science Grid</i>	<i>DataGrid (CERN)</i>
<i>NSF National Virtual Observatory</i>	<i>EuroGrid (Unicore)</i>
<i>NSF GriPhyN</i>	<i>DataTag (CERN)</i>
<i>DOE Particle Physics Data Grid</i>	<i>Astrophysical Virtual</i>
<i>NSF TeraGrid</i>	<i>Observatory</i>
<i>DOE ASCI Grid</i>	<i>GRIP (Globus/Unicore)</i>
<i>DOE Earth System Grid</i>	<i>GRIA (Aplicaciones Industriales)</i>
<i>DARPA CoABS Grid</i>	<i>GridLab (Cactus Toolkit)</i>
<i>NEESGrid</i>	<i>CrossGrid (Componentes de infraestructura)</i>
<i>DOH BIRN</i>	<i>EGSO (Física solar)</i>
<i>NSF iVDGL</i>	<i>NorduGrid (Noruega y Suecia)</i>

Fuente: World Community Grid. Technology Solving problems

<http://worldcommunitygrid.org/>

La siguiente Figura nos muestra una de las tantas aplicaciones que tiene la Computación Grid hoy en día, que para este caso se destaca en el ámbito de la bioinformática.

Figura 10: Aplicación del GRID en la bioinformática.



Fuente: www.ramonmillan.com/tutoriales/gridcomputing

LA COMPUTACIÓN GRID EN LOS CENTROS ACADÉMICOS Y DE INVESTIGACIÓN.

La computación distribuida empezó a ser escalada a niveles globales con la madurez de Internet en los años 90. Dos proyectos en particular han demostrado que el concepto es totalmente viable y eficiente, incluso más de lo que los expertos auguraban en un principio: Distributed.net y SETI@home.

Distributed.net: Emplea miles de ordenadores distintos para crackear códigos de encriptación (RC5-64, CSC, DES-III, DES-II-1, DES-II-1, RC5-56, etc.). Fundado en 1997, el proyecto ha crecido hasta abarcar en el año 2009 más de 60.000 usuarios alrededor de todo el mundo. El poder de cómputo de Distributed.net ha ido creciendo hasta llegar a ser el equivalente a más de

160.000 computadoras PII 266MHz trabajando 24 horas al día, 7 días a la semana, y 365 días al año.

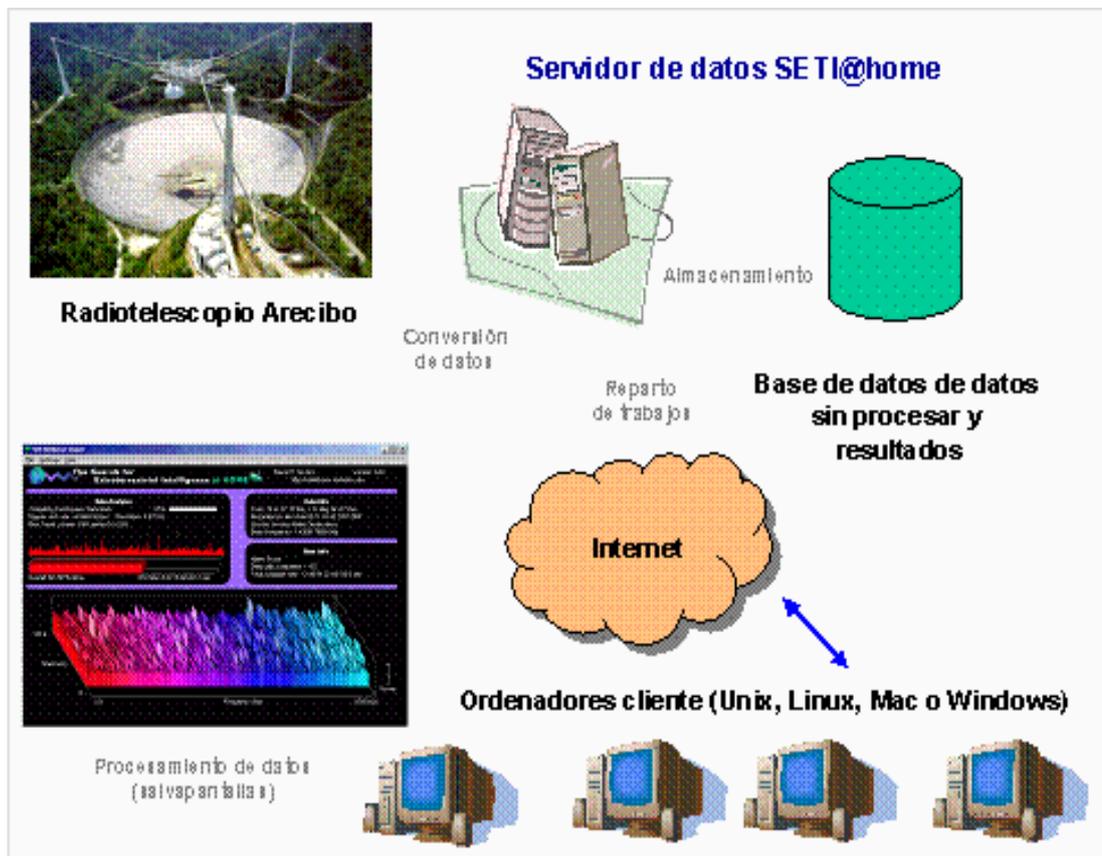
SETI@home ha sido el proyecto de computación distribuida más popular de la historia, cuyo objetivo era la búsqueda de vida extraterrestre mediante la detección de su tecnología de comunicaciones, buscando patrones que demuestren inteligencia en las ondas de radio procedentes del espacio. Para ello, cualquier persona que quisiera colaborar podía descargarse un salvapantallas gratuito (de este modo aprovechaba los ciclos del ordenador sólo cuando éste no estaba en uso), que instalado en su ordenador analiza señales del espacio captadas con el radiotelescopio de Arecibo en Puerto Rico. El software se hizo público el 17 de mayo del 1999 y, desde entonces hasta su finalización el 15 de diciembre de 2006, más de 5 millones de voluntarios han instalado el programa en su ordenador, se han conseguido un total acumulado de dos millones de años de tiempo de CPU y se han analizado alrededor de 50 TB de datos, convirtiendo al proyecto en el mayor computador virtual de la historia de la humanidad por análisis realizados. En la actualidad, SETI@Home sigue en funcionamiento pero integrado en BOINC (Berkeley Open Infrastructure for Network Computing), un proyecto basado en recursos de redes abiertas con los mismos principios que el proyecto original. La nueva infraestructura continuará con la búsqueda de señales de radio extraterrestres, pero ahora además la potencia de CPU de los participantes se dedicará también a investigar sobre otras tareas, como el cambio climático, la astronomía y la cura de enfermedades.

El sector más involucrado en todo el mundo en la puesta a punto de plataformas Grid en el desarrollo de aplicaciones adaptadas a esta nueva tecnología es, sin lugar a dudas, el de la investigación.

En España las principales universidades e instituciones de investigación han participado intensamente en diversos proyectos sobre Grid, tanto nacionales

como internacionales, entre otros: “IRISGrid”, “Damien”, “HealthGrid”, “CrossGrid”, “Enabling Grids for e-Science in Europe”, etc. Existen ya varios agentes Grid disponibles para colaborar activamente en este tipo de proyectos, por ejemplo, en “Grid.org” de United Devices, cualquiera puede bajarse uno, instalarlo y configurarlo, para colaborar activamente en el desarrollo de esta fascinante tecnología y, a la vez, desempeñar una labor altamente solidaria (por ejemplo, ayudando al descubrimiento de nuevos fármacos y vacunas contra algunas de las enfermedades más terribles de nuestra Era).

Figura 11: Proyecto SETI@home.



Fuente: www.ramonmillan.com/tutoriales/gridcomputing

Por otra parte, dentro de las aplicaciones de la Grid a nivel internacional, se destacan los avances en el área de la física, sobre todo en reacciones moleculares con diferentes propósitos. Dentro de los más destacados se

encuentra PPDG; que investiga soluciones de infraestructura a corto plazo para las necesidades específicas y críticas en los experimentos sobre física de partículas que tienen o poseen un desarrollo activo (como CMS y ATLAS). También para el trabajo combinado con otras ciencias como la astronomía y con Detectores de Onda de Gravedad. El Laboratorio internacional Grid de Datos Virtuales (iVDGL) proporciona bloques de prueba globales y recursos de cálculos para experimentos que busquen las comprobaciones en estos campos.

El DataGrid de la Unión Europea tiene objetivos similares a GriPhyN e iVDGL, y es financiado por la Unión Europea.

LCG es una colaboración basada en CERN que se soporta en la infraestructura de la Computación Grid y en aplicaciones para los experimentos de LHC. Finalmente, CrossGrid es otra iniciativa financiada por la Unión Europea que amplía el trabajo de la Computación Grid a once países no incluidos en la DataGrid de la Unión Europea (Como se había indicado anteriormente en la tabla 3).

Otras aplicaciones a nivel internacional como proyectos de investigación en las escuelas académicas se nombran la PPDG, GriPhyN, DataGrid y CrossGrid.

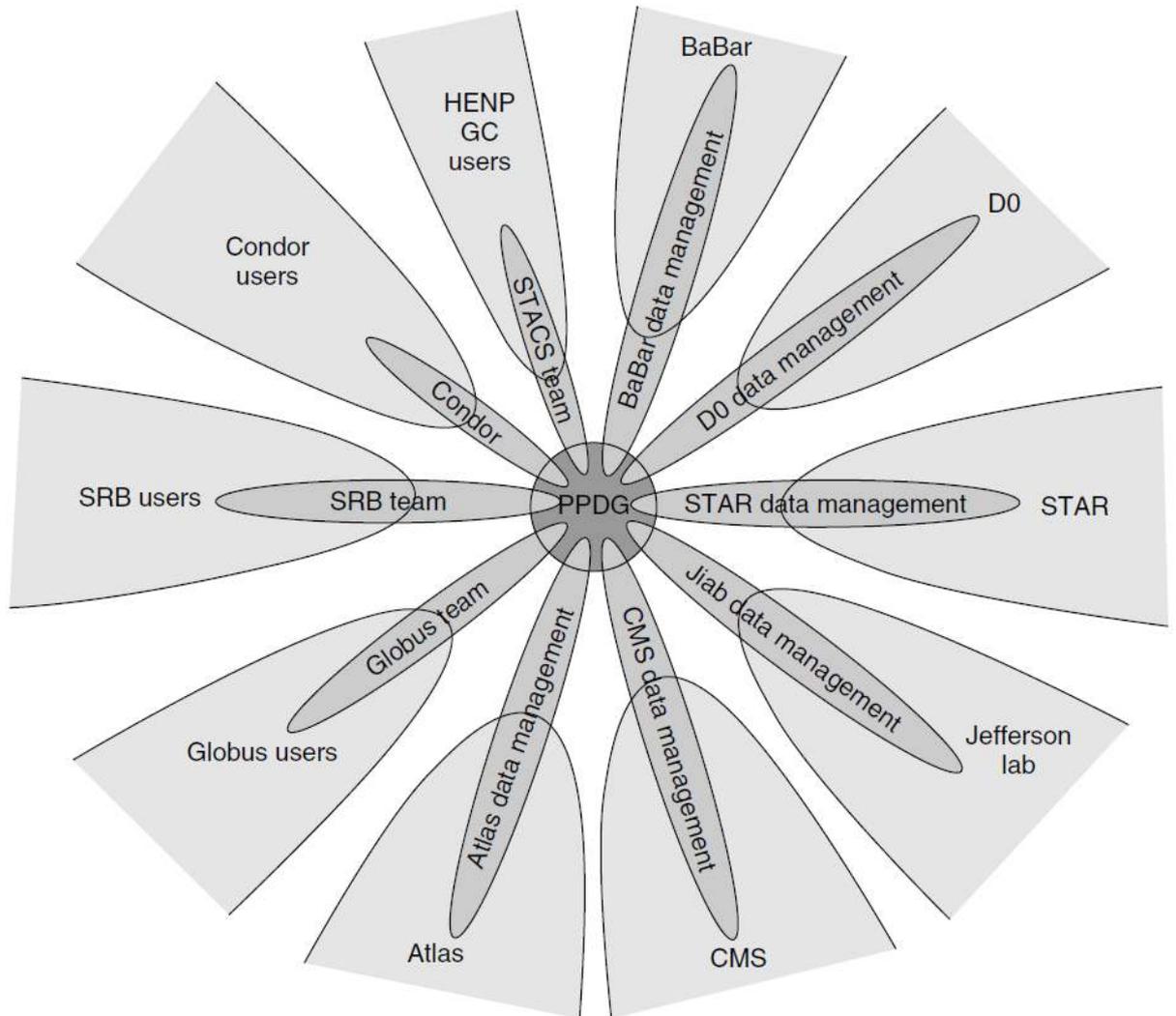
PPDG, La Grid de Datos de Física de Partícula (www.ppdg.net) fue fundada en 1999, con la finalidad de dirigirse a la necesidad de servicios de Datos en base a la Computación Grid para permitir el modelo de cálculo distribuido por todo el mundo de la física de gran energía, nuclear y experimentos futuros (HENP). Al principio financiado por el programa NGI [13] del Ministerio de Energía y más tarde del MICS [13] y programas HENP [14], esto ha proporcionado una oportunidad de desarrollo temprano de la arquitectura Grid y además para la evaluación de algún prototipo de middleware de prototipo Grid.

La segunda fase del PPDG se denomina la Física de Partículas Data Grid piloto Collaboratory. Esta fase se refiere al desarrollo, adquisición y entrega de herramientas, una necesidad vital de la Grid con capacidad para satisfacer los requisitos de uso intensivo de datos de las partículas y física nuclear. Nuevos mecanismos y políticas se integran verticalmente con el middleware Grid y las aplicaciones específicas de los experimentos y los recursos de computación para desarrollar capacidades efectivas de extremo a extremo.

PPDG es una colaboración de científicos de la computación con un sólido historial en la computación distribuida y la tecnología Grid, físicos con liderazgo en el diseño de software y de infraestructuras de red para las plantas de alta energía y los experimentos nucleares. Un programa de tres años se ha trazado para un proyecto que aprovecha al máximo la fuerza motriz proporcionada por experimentos de física que operan actualmente o que están en curso.

Los objetivos y planes de PPDG son en última instancia, guiados por las necesidades inmediatas, a mediano plazo y largo plazo y las perspectivas de los experimentos de la física, y por el programa de investigación y desarrollo de los proyectos que participan en el CS PPDG y otros esfuerzos orientados a red.

Figura 12: PPDG.



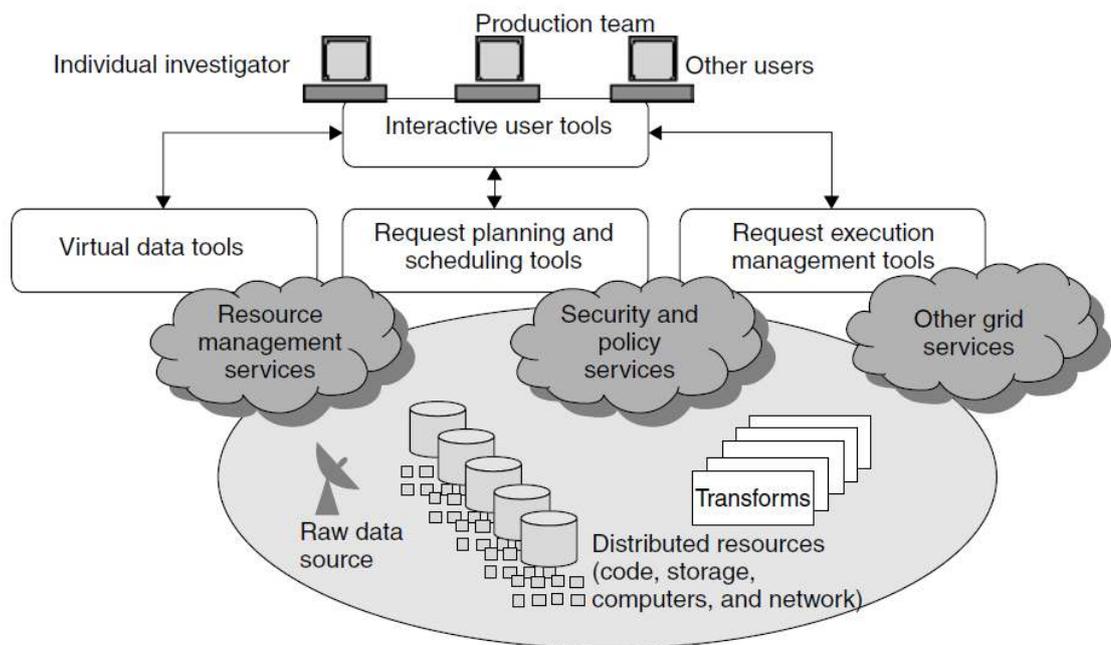
Fuente: Mathematics, information, and computer sciences, through the SCIDAC (Scientific discovery through advanced computing) initiative.

GriPhyN, El proyecto GriPhyN (Grid Física de red - <http://www.griphyn.org>), es una colaboración de la CS y otros investigadores de TI y los físicos de ATLAS, CMS, Laser Interferometer Gravitational-wave Observatory (LIGO) y experimentos SDSS. El proyecto se centra en la creación de una Virtual Data Grid de alto nivel que satisfaga las necesidades de cómputo intensivo de datos de una comunidad diversa en todo el mundo, lo cual abarca la

definición y la entrega de un espacio (potencialmente ilimitado) virtual de productos de datos derivados de los datos experimentales.

En dicho espacio virtual de datos, las solicitudes pueden ser satisfechas a través de acceso directo y / o de computación, con la gestión de los recursos locales y globales y la política de restricciones de seguridad para determinar la estrategia utilizada. La superación de este desafío y la realización del concepto de Datos Virtual requieren avances en tres áreas principales:

Figura 13: Prospecto de GriPhyN.



Fuente: Fuerte integración de la generación de datos, almacenamiento, computación y servicios de red, herramientas de programación, gestión y seguridad. Grid Física de Red. <http://www.griphyn.org>

Tecnologías de datos virtuales: se requieren avances en los modelos de información y de nuevos métodos de catalogación, caracterización, validación y archivo de los componentes de software para implementar las manipulaciones de datos virtuales.

Solicitud de Política impulsada por la planificación y programación de datos en red y los recursos de cómputo: Se requieren mecanismos para representar y hacer cumplir las limitaciones de las políticas locales y globales y las nuevas técnicas de recursos por descubrir.

Gestión de las transacciones y la ejecución de tareas a través de organizaciones virtuales a escala nacional y en el mundo: se requieren nuevos mecanismos para satisfacer las necesidades del usuario en la obtención de un mejor rendimiento, una mayor confiabilidad y bajo costo. Un agente de la informática será importante para permitir que la Red pueda equilibrar las necesidades del usuario y el rendimiento de la Grid, con tolerancia a determinadas fallas.

El proyecto GriPhyN se centra principalmente en la consecución de los avances fundamentales que necesitarán la Virtual Data Grid para crearla a nivel de escala cinco, pero también está trabajando en la creación de sistemas de software para uso comunitario, y la aplicación de la tecnología que permite la distribución y análisis de colaboración de datos. Una multifacética aplicación, independiente del dominio virtual de datos Toolkit está siendo creada y utilizada para el prototipo virtual de Redes de Datos, y para apoyar el CMS, ATLAS, LIGO y las tareas de análisis del SDSS.

DataGrid, El DataGrid Europeo (UE-datagrid.web.cern.ch) es un proyecto financiado por la Unión Europea con el objetivo de establecer un cálculo y los datos de uso intensivo de la Grid de los recursos para el análisis de reseñas provenientes de la exploración científica ya que la ciencia de última generación requiere una acción coordinada en el intercambio de recursos, elaboración, colaboración y análisis de grandes cantidades de datos producidos y almacenados por muchos laboratorios científicos pertenecientes a varias instituciones.

El principal objetivo de la iniciativa DataGrid es desarrollar y poner a prueba la infraestructura tecnológica que permita la participación en los equipos científicos de todos los investigadores que lo requieran sin importar su ubicación geográfica pero con una interacción e intercambio de datos e instrumentos en una escala previamente contemplada.

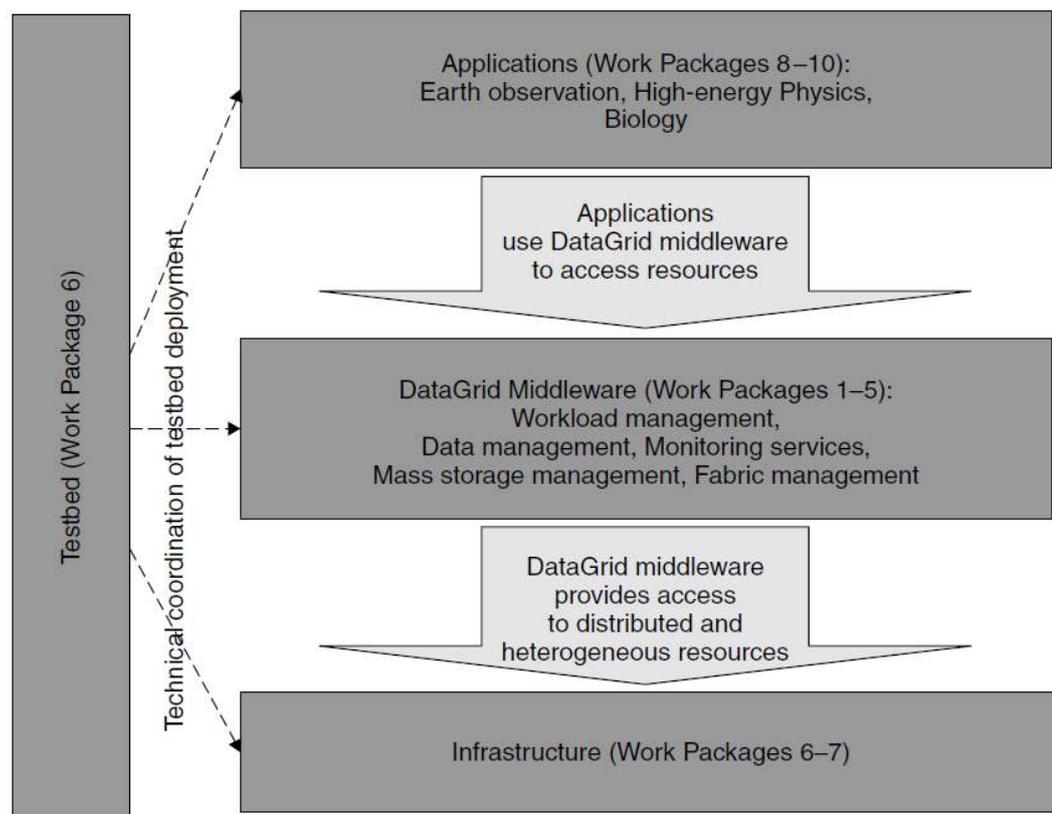
En síntesis, el proyecto consiste en la elaboración y desarrollo de soluciones escalables de software y bancos de pruebas con el fin de manejar muchos petabytes de datos distribuidos, decenas de miles de recursos informáticos (procesadores, discos, etc), y miles de usuarios simultáneos de múltiples instituciones de investigación.

La iniciativa DataGrid es liderado por el CERN, junto con otros cinco socios principales y quince miembros asociados. El proyecto reúne a los siguientes importantes organismos europeos de investigación: la Agencia Espacial Europea (ESA), el Centro de Francia National de la Recherche Scientifique (CNRS), Instituto de Italia Nazionale di Fisica Nucleare (INFN), los holandeses del Instituto Nacional de Física Nuclear y Altas Energía Física (NIKHEF) y Física de Partículas del Reino Unido y el Consejo de Investigación en Astronomía (PPARC). Los quince socios asociados son procedentes de la República Checa, Finlandia, Francia, Alemania, Hungría, Italia, Países Bajos, España, Suecia y el Reino Unido.

DataGrid es un proyecto ambicioso con beneficios para el desarrollo de muchos tipos diferentes de tecnología y conocimientos. El proyecto abarca tres años, de 2001 a 2003, con más de 200 científicos e investigadores involucrados.

El proyecto DataGrid se divide en doce paquetes de trabajo distribuidos en cuatro grupos de trabajo: banco de pruebas y de infraestructura, aplicaciones de cómputo y DataGrid middleware, gestión y difusión.

Figura 14: Estructura del EU DataGrid en sus bloques de componentes de trabajo.



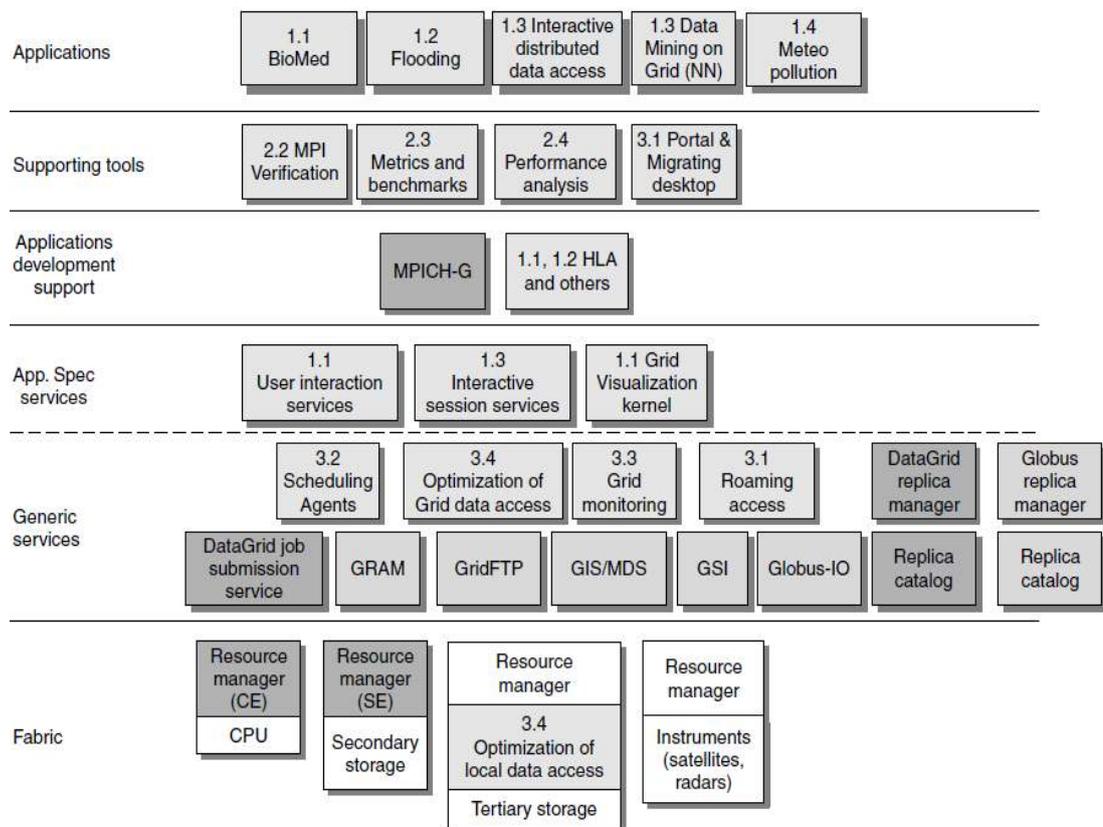
Fuente: UE-datagrid.web.cern.ch

CrossGrid; CrossGrid (<http://www.crossgrid.org>) es un proyecto europeo de desarrollo, ejecución y explotación de nuevos componentes de Grid para aplicaciones interactivas de computación y uso intensivo de datos como la simulación y visualización de los procedimientos quirúrgicos, las inundaciones críticas de los sistemas de equipo de apoyo a las decisiones, análisis de datos distribuidos en la física de alta energía y la contaminación del aire en combinación con las previsiones meteorológicas.

La metodología elaborada, la arquitectura de aplicación genérica, entorno de programación, y los nuevos servicios Grid están siendo validados y probados en el banco de pruebas CrossGrid, con énfasis en un entorno fácil de usar. CrossGrid colabora estrechamente con el Global Grid Forum (GGF) y el proyecto DataGrid con el fin de beneficiarse de sus resultados y la experiencia, y para garantizar la plena interoperabilidad.

El objetivo principal de CrossGrid es ampliar aún más el entorno Grid para una nueva categoría de aplicaciones de gran importancia práctica. Once países europeos se encuentran implicados en este proyecto.

Figura 15: Arquitectura de Cross Grid.



Fuente: <http://www.crossgrid.org>

La novedad esencial del proyecto Cross Grid consiste en extender la red a una categoría completamente nueva e importante socialmente. El rasgo característico de estas aplicaciones es la presencia de una persona en un ciclo de procesamiento, con el requisito de respuesta en tiempo real del sistema informático. Las aplicaciones interactivas son elegidas de computación y uso intensivo de datos.

5.2 A NIVEL DE LATINOAMÉRICA

América latina no podía ser la excepción ante este auge de la tecnología GRID. Pero si en los países desarrollados la GRID Computing esta en formación, en nuestros países (Latinoamérica) apenas está en gestación, por ello se vienen adelantando varios proyectos y alianzas con instituciones que tienen experiencia en la implementación de tecnologías GRID. En la actualidad se cuenta con CLARA Cooperación Latino Americana de Redes Avanzadas que es una organización no gubernamental sin fines de lucro, la cual estimula la cooperación regional en actividades educativas, científicas y culturales, además promueve la integración directa con las comunidades científicas de Europa [CLARA (2006)], esta organización trata de integrar una red regional de telecomunicaciones de la más avanzada tecnología para interconectar a las redes académicas nacionales de la región, hasta el momento tiene los siguientes miembros:

- Argentina – Red Teleinformática Académica-RETINA
- Bolivia – Agencia para el Desarrollo de la Sociedad de la Información en Bolivia-ADSIB
- Brasil – Red Nacional de Enseñanza e Investigación – RNP
- Colombia – Red Nacional Académica de Tecnología Avanzada – RENATA
- Costa Rica – Red Nacional de Investigación – CR2Net
- Cuba – RedUniv
- Chile – Red Universitaria Nacional – REUNA - GREUNA

- Ecuador – Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado – CEDIA
- El Salvador – Red Avanzada de Investigación, Ciencia y Educación Salvadoreña – RAICES
- Guatemala – Red Avanzada Guatemalteca para la Investigación y Educación – RAGIE
- Honduras – Universidad Tecnológica Centroamericana – UNITEC
- México – Corporación Universitaria para el Desarrollo de Internet – CUDI
- Nicaragua – Red Nicaragüense de Internet Avanzada – RENIA
- Panamá – Red Científica y Tecnológica – RedCyT
- Paraguay – Arandu
- Perú – Red Académica Peruana – RAAP
- Uruguay – Red Académica Uruguayaya – RAU
- Venezuela – Red Académica de Centros de Investigación y Universidades Nacionales – REACCIUN

Otro proyecto importante para el trabajo en GRID Computing, que se adelanta en Latino América actualmente es EELA (E-infrastructure Shared between Europe and Latin America), en el cual participan 10 países con sus redes académicas y diferentes instituciones entre universidades y centros de investigación; el objetivo es llevar las e-Infraestructuras de los países latinoamericanos al nivel de explotación de los países Europeos. La EELA se beneficia del estado maduro del proyecto ALICE (América Latina Interconectada Con Europa) y de la red CLARA [EELA].

5.3 A NIVEL NACIONAL

Ahora bien, si en el mundo se está desarrollando esta gran tecnología de la Computación Grid, en Colombia la Universidad Industrial de Santander coordina en el ámbito regional la implantación de la tecnología GRID, y de acuerdo al Director del Grupo de Investigación en Energía y Medio Ambiente -

GIEMA, de la UIS, esa coordinación se reconoce no sólo en Colombia sino incluso en Latinoamérica. La computación Grid es una nueva tecnología, catalogada por los expertos como una revolución similar al Internet que va a cambiar la manera cómo los científicos hacen Ciencia.

"La computación GRID es un nuevo paradigma. Para que la gente nos entienda, GRID Computing es toda una infraestructura tecnológica que le permite a un país como Colombia, participar en un proyecto internacional como es el Proyecto Atlas que es la física de estas energías sin tener todos los recursos tanto materiales como humanos entonces un físico acá de la UIS puede estar participando en el proyecto de física de altas energías, que es tan importante para la humanidad y que se lleva a cabo en cercanías de Ginebra (Suiza)" explicó el también Doctor en Ingeniería, Jorge Luis Chacón. [15]

Es importante anotar, que Colombia lidera la red de computadores más grande de Latinoamérica. Después de muchos años de trabajo dos colombianos lo habían logrado: eran parte del selecto equipo de 2.500 científicos del Laboratorio CERN (Centro Europeo de Investigación Nuclear), que con sus investigaciones en física de altas energías intenta descubrir el origen del universo. Sólo tenían un problema: para mantenerse en el grupo les faltaba crear la red de computadores más grande de Latinoamérica.

Todo había empezado trece años antes, en 1998, cuando ellos, Carlos Ávila y Bernardo Gómez, docentes del Departamento de Física de Los Andes, se unieron a las investigaciones de altas energías del laboratorio Fermilab en Chicago. Allí se experimenta con el segundo acelerador de partículas más grande del mundo, el Tevatrón, que a altas energías hace colisionar partículas (como protones y antiprotones) entre sí, para realizar investigaciones sobre el origen de la masa.

Las cosas marchaban bien, pero los experimentos en Chicago terminaban en el

2011 y Carlos y Bernardo no pensaban detenerse cuando ese proyecto llegara a su fin. Con sus estudios y experiencia en altas energías prefirieron buscar más.

“Queríamos mantenernos en este tipo de estudios y el Laboratorio CERN, que tiene el acelerador de partículas más grande del mundo, el Large Hadron Collider LHC (gran colisionador de hadrones), era nuestro objetivo. Allí se hacen experimentos que pueden llevar a descifrar el origen del universo. Nos contactamos, les ofrecimos mano de obra y aceptaron. Primero nos unimos los profesores Bernardo Gómez y yo (2006). Luego el ingeniero Marcelo Baquero viajó a Suiza y también hizo parte del equipo (2007). El CERN tiene más de 20 años trabajando con los aceleradores que producen partículas como las que existieron al comienzo de la creación del universo. Las podemos estudiar y obtener datos. Esto nos da claves sobre el origen de la masa en relación con la teoría del Big Bang”, dice Ávila.

“El LHC o gran colisionador de hadrones, arroja tantos datos que en un segundo la información que acumula es como almacenar diez enciclopedias británicas”, explica Ávila, pero no sólo es obtener la información. Hay que almacenarla y analizarla. Un computador, con una buena configuración, podría demorarse 13.000 años en procesar tantos datos.

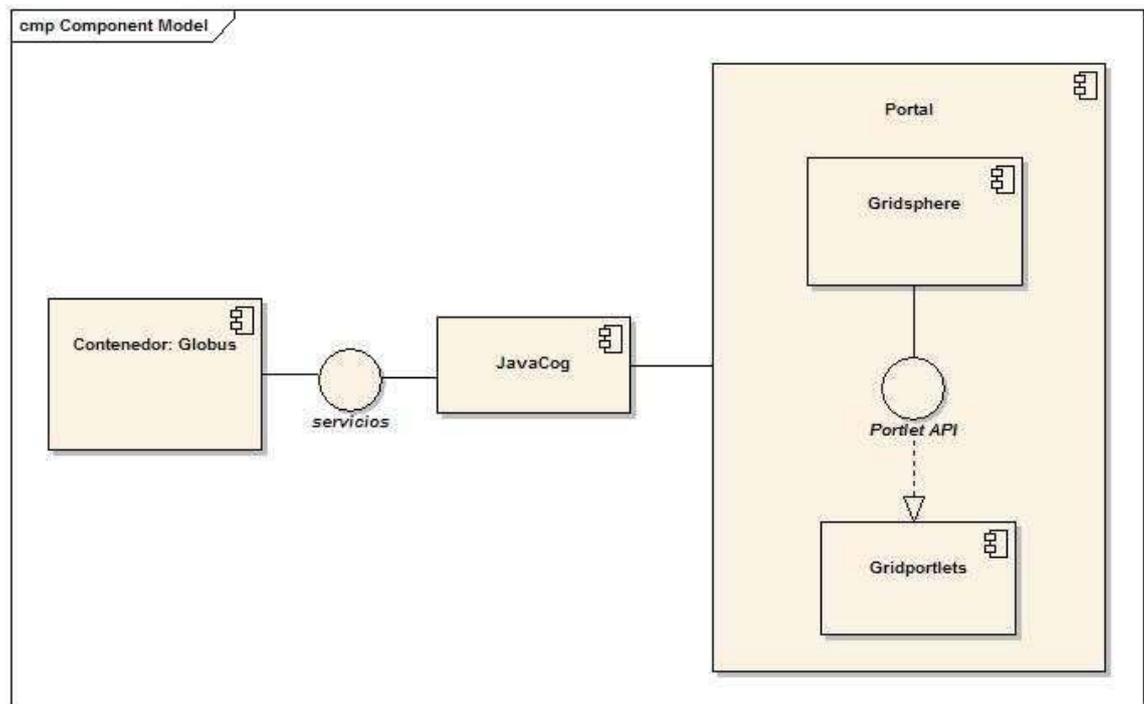
Pero el nivel de computación que se necesitaba no era un problema que no pudieran solucionar 2.500 científicos de 38 países y 183 instituciones que trabajan con el CERN. Si hace 20 años tuvieron que resolver cómo sus científicos podrían comunicarse y compartir sus avances en investigaciones desde cualquier lugar del mundo y para eso crearon la World Wide Web (www), que dio origen a Internet, esta vez no se quedaron atrás: crearon redes que unen varios computadores (sin importar su ubicación geográfica) para que hagan el trabajo de 13.000 años en sólo uno. Esto es la Computación Grid.

El CERN, a su vez, empezó a unir gran cantidad de Grids de las diferentes instituciones que estaban vinculadas con el proyecto y era necesario que Los Andes aportara su propio Grid si quería mantenerse activa en la investigación.

El Grupo de Física de altas energías, entonces, le solicitó a la Dirección de Tecnologías de Información (DTI) de Los Andes un Grid. Petición que llegó en el momento oportuno, pues se sumó a las del Departamento de Sistemas y Computación y del Centro de Computación Avanzada de Ingeniería (MOX), quienes también buscaban una solución para trabajos de gran magnitud en sus investigaciones. La solicitud se aprobó y nació el Grid Uniandes, que se convirtió en la red de computadores más grande del país para trabajos científicos.

El desempeño fue tan bueno, que en ese mismo año Los Andes recibió la certificación necesaria para que el Grupo de Altas Energías del Departamento de Física pudiera procesar los datos obtenidos por sus investigadores en el CERN con el Grid Uniandes. [16]

Recientemente, encontramos en nuestro país otra interesante aplicación de la computación GRID, aplicada a un MODELO DE PORTAL PARA EL ACCESO A UNA GRID DE GEOSENSORES. Esta aplicación fue diseñada e implementada por la ingeniera Sandra Montañez Torres, como tesis de grado en la Maestría en ciencias de la información y comunicaciones con énfasis en teleinformática de la Universidad Distrital Francisco José de Caldas. El modelo general del portal Grid, cuenta con tres componentes bien definidos: contenedor Globus, JavaCog y portal que a su vez contiene dos componentes: Gridsphere y Gridportlets, los cuales tienen una función particular que permite poner en marcha el portal y proporcionar nuevos servicios a los investigadores e ingenieros del Laboratorio en múltiples disciplinas.



La plataforma de conectividad o contenedor de la *Grid* es uno de los componentes del sistema *Grid*, ésta ofrece un conjunto de servicios que hacen posible el funcionamiento y control de aplicaciones distribuidas sobre plataformas heterogéneas. Para el Laboratorio de computación Grid de Gicoge se usó como *Globus Toolkit* versión 4. Las diferentes aplicaciones para el *Portal Grid* requieren usar los servicios de esta plataforma, cada una de ellas implementando una funcionalidad, estos servicios son conceptualizados principalmente como servicios implementados de *WSRF*.

El servicio de gestión y ejecución de trabajos en recursos, es el encargado de la gestión y ejecución de trabajos o *jobs* en recursos. El servicio *Web Service Grid Resource Allocation and Management (WS-GRAM)* proporciona la interfaz de servicios web para iniciar, supervisar, gestionar y coordinar la ejecución de trabajos en recursos remotos. Para el desarrollo de este modelo se implementaron dos servicios básicos: *Managed Job Factory* y *Managed Job*. Este servicio también permite la transferencia de archivos al nodo de ejecución utilizando para este propósito *Reliable File Transfer (RTF)*.

El servicio de gestión de datos, es el encargado de gestionar y transportar datos entre recursos remotos. Para éste modelo se hace uso de GridFTP que provee un conjunto de librerías y herramientas. GridFTP es confiable, seguro y posee un alto rendimiento en la transferencia de datos, especialmente optimizado para grandes cantidades de datos, basado en el protocolo File Transfer Protocol (FTP) y constituye el servicio más importante dentro de la gestión de datos. GridFTP es FTP + Grid Security Infrastructure (GSI). Este posee como una de sus características significativas la posibilidad de realizar transferencias entre dos nodos o servidores remotos desde un tercer nodo, capacidad conocida como *third-party transfer*, también utiliza una base de datos para mantener una lista de transferencias de archivos y sus estados, y es capaz de recuperar una solicitud de transferencia que fue interrumpida.

La gestión de recursos, es el servicio encargado de la gestión estática y dinámica de información sobre el estado de una red computacional y todos sus componentes tales como: las redes, nodos, y sistemas de almacenamiento. En particular, mediante el GRIS, se puede consultar el estado, la configuración, y las prestaciones de cada recurso de la Grid. La información suministrada por cada GRIS se agrupa en el GIS, que ofrece una imagen conjunta y coherente de los recursos de la Grid.

Servicio de Seguridad, este servicio se encarga de afrontar los desafíos de seguridad en el contenedor, para ello usa GSI o también llamados servicios de seguridad, que ofrece un sistema de criptografía de clave pública *PKI*, como la autenticación mutua mediante certificados *x509*, delegación de credenciales y *single sign-on*. La configuración de *GSI* básicamente se reduce a conseguir uno o varios certificados digitales para una máquina. Se pueden tener múltiples certificados para usuarios individuales, como para una máquina o para un servicio en concreto.

JavaCog, es un componente del **kit Java CoG** es usado para utilizar los servicios de **Globus**, éste identifica una serie de utilidades básicas y avanzadas que ayudan en el desarrollo de aplicaciones *Grid* que permiten la interconexión con el portal usando implementaciones de seguridad basadas en **GSI, GridFTP, MyProxy** y **GRAM**. El acceso a los servicios *Grid* se proporcionan a través del **Java CoG** introduciendo flujos de trabajos *Grid* y un modelo de programación basado en tareas. En ésta investigación se han introducido desarrollos hechos por otras organizaciones de esta manera la arquitectura **Java CoG** permite la fácil integración de mejoras en desarrollos hechos por la comunidad.

El portal, es el encargado de mostrar al usuario final la interfaz para usar las tecnologías Grid del Laboratorio de computación Grid de Gicoge.

6. CONCLUSIONES

- La Computación GRID es una infraestructura relativamente nueva y en permanente proceso de mejoramiento, hecho demostrado en los continuos avances que proporciona cada una de las tecnologías que han servido como bases teóricas y experimentales desde las cuales se ha desarrollado.
- Esta tecnología proporciona un práctico mecanismo de colaboración entre grupos dispersos, tanto científicos como comerciales y posibilita claramente el funcionamiento de aplicaciones a gran escala, lo cual la hace más eficaz que otras redes de comunicaciones actualmente en funcionamiento.
- Colombia lidera la red de computación más grande de Latinoamérica y participa en grandes proyectos a nivel internacional, sin tener aún todos los recursos materiales con los que cuentan los científicos de otros países; por lo tanto desde las universidades públicas y privadas se debe incentivar la investigación en este campo y la adquisición de equipos que permitan a nuestra nación tener un impacto aún mayor en los resultados globales.
- El descubrimiento, selección, reserva, asignación, gestión y monitorización de recursos para la investigación en computación GRID son procesos que deben liderarse desde los entes gubernamentales

BIBLIOGRAFIA

- [1] Vannesa Barrios V. Licenciatura en Sistemas de Información. Corrientes - Argentina: Universidad Nacional del Nordeste; 2005. Página 1.
- [2] J. Nick S. Tuecke I. Foster, C. Kesselman. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. Global Grid Forum, USA, 2002.
- [3] Vannesa Barrios V. Licenciatura en Sistemas de Información. Corrientes - Argentina: Universidad Nacional del Nordeste; 2005. Página 4.
- [4] Berman, F., Fox, G. and Hey, T. (2003) *Grid Computing: Making the Global Infrastructure a Reality*. Chichester: John Wiley & Sons.
- [5] Web Site associated with book, *Grid Computing: Making the Global Infrastructure a Reality*, <http://www.grid2002.org>.
- [6] NSF Grand Challenge as part of 1997 HPCC Implementation Plan <http://www.itrd.gov/pubs/imp97/>; their 2002 view at <http://www.itrd.gov/pubs/blue02/national-grand.html>, National Grand Challenge Applications, National Coordination Office for Information Technology Research and Development.
- [7] CNRI, Corporation for National Research Initiatives, Gigabit Testbed Initiative Final Report, December, 1996, <http://www1.cnri.reston.va.us/gigafr/>.
- [8] NSF Middleware Initiative, <http://www.nsf-middleware.org/>.
- [9] L. Joyanes Aguilar. Cibersociedad. Mac Graw-Hill, 1997.
- [10] Vannesa Barrios V. Licenciatura en Sistemas de Información. Corrientes - Argentina: Universidad Nacional del Nordeste; 2005. Página 66
- [11] Leticia Rocha González, María Rosa Varela González, Vanesa Vidal Castro, Diapositivas Grid Computing y Globus Toolkit, 2009.
- [12] Área temática de bioinformática (2003)
- [13] Mathematics, information, and computer sciences, through the SCIDAC (Scientific discovery through advanced computing) initiative.
- [14] High energy and nuclear physics.

[15] Universidad Industrial de Santander (UIS), Curso en computación en Tecnología Grid, 2004.

[16] Universidad De Los Andes. <http://www.fundetec.edu.co/topic.aspx?a=33>