

# Grid Computing

Javier Ferrés

Facultad de Ciencias y Tecnología, Universidad Católica de Asunción

Asunción, Paraguay

[javierferres@telesurf.com.ay](mailto:javierferres@telesurf.com.ay)

## Resumen

*Grid Computing o su traducción al español computación en malla, consiste en redes de computadoras, almacenamiento u otros dispositivos que comparten y otorgan recursos. La arquitectura de malla provee de dichos recursos a usuarios autorizados. Dichos recursos son utilizados virtualmente y en demanda.*

### 1. Introducción

Un sistema distribuido es aquel en el que los componentes localizados en computadores, conectados en red, comunican y coordinan sus acciones únicamente mediante el paso de mensajes.[1]

Al dar una definición de sistema distribuido, podemos ir adentrándonos en que significa grid computing, cuales son sus alcances, y sus posibles aplicaciones.

La arquitectura de computación en malla posee millones de definiciones, algunas de estas son muy extensas y consisten en clusters interconectados que comparten una piletta común de datos. Otras definen a esta arquitectura como ambientes de uso masivo interconectados, que utilizan gran cantidad de sistemas de información y diversifican el almacenamiento en diferentes subsistemas como mallas.

Triangulando la definición y ajustándonos al ambiente en el cual estos sistemas trabajan, podemos decir que:

1. Estos realizan un balanceo inteligente de carga sobre redes de gran ancho de banda
2. Poseen una arquitectura de red para incrementar la potencia de computo o almacenamiento
3. Standards que hacen que sistemas y aplicaciones heterogéneas compartan computabilidad y almacenamiento.

Con lo anteriormente nombrado podemos concluir que Grid es una arquitectura Standard para compartir ya sea una aplicación o almacenamiento, que hace posible que sistemas heterogéneos y aplicaciones, compartan recursos de cómputo y almacenamiento transparentemente.

Generalmente hay una confusión a la hora de hablar de clusters y grids. La diferencia entre estos, es que, en los clusters, la distribución de los recursos esta hecha por un manejador centralizado, y todos los nodos que forman parte del cluster cooperan y trabajan juntos como una única fuerza. En el caso de las grids, cada nodo tiene su propio manejador de recursos, y no se apunta a dar una ilustración de un sistema centralizado.

La función principal de una malla es proveer recursos, para esto, hay que tener en cuenta primero:

### 1.1 Aprovisionamiento

El aprovisionamiento involucra encontrar recursos de almacenamiento y computo, y hacer disponibles dichos recursos a los demandantes.

Existen diferentes maneras de realizar dicha tarea, podemos nombrar algunos ejemplos de diseño para alcanzar esto:

- ❖ Reconfigurar recursos para poder tener computo/almacenamiento adicional
- ❖ Componentes instalados, pero inactivos, de manera a poseer siempre recursos disponibles
- ❖ Acceder a recursos adicionales que pueden ser encontrados una LAN(Local Area Network) o una WAN(Wide Area Network)
- ❖ Encontrar y explotar recursos computacionales/almacenamiento usando una malla

Además del aprovisionamiento, la Virtualización de dichos recursos es muy importante,

### 1.2 Virtualización:

Existen dos formas de realizar la virtualización. Una de ellas es la que ocurre en un Servidor Local, en este los recursos pueden ser manualmente configurados. Para configurar los recursos locales, podemos configurar el servidor de dos formas:

- ❖ Partición Lógica: Permite que las unidades de almacenamiento sean virtualmente particionadas para hacer lugar y de este modo poder tener mayor capacidad de computo.
- ❖ Reconfiguración dinámica: Algunos sistemas altamente escalables permiten ser reconfigurados on-the-fly.
- ❖ Y finalmente, sobredimensionar la capacidad de cómputo, y en caso de necesidad, utilizar los recursos excedentes disponibles.

La otra forma de virtualización ocurre cuando los recursos se encuentran disponibles en una LAN o WAN. Podemos nombrar:

- ❖ Sistemas organizados en clusters ofrecen acceso a CPU adicionales
- ❖ El software de que maneja la grid puede encontrar recursos disponibles en las LAN's o WAN's
- ❖ Servicios virtuales externos pueden ser utilizados

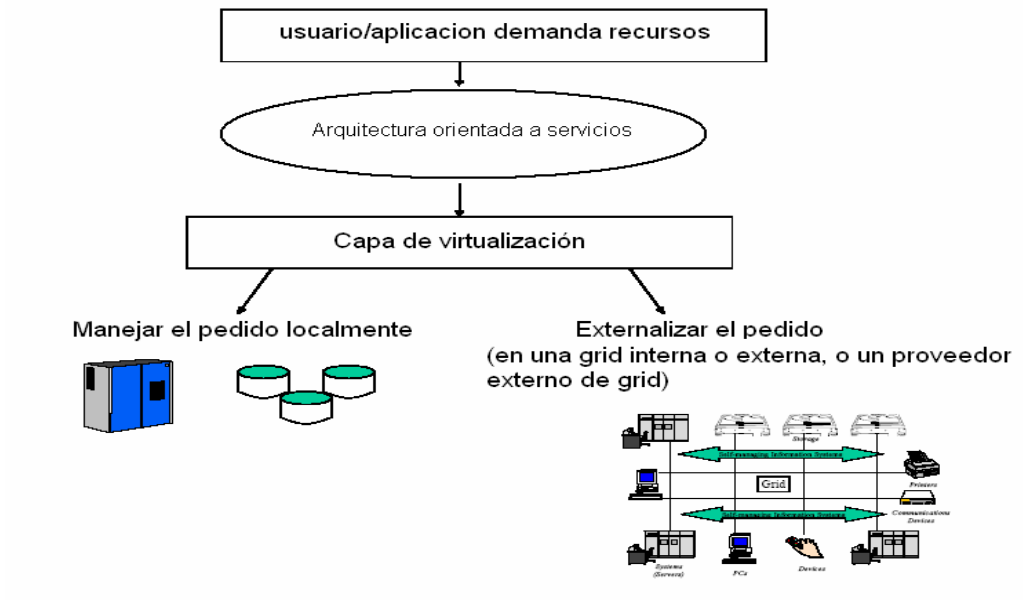


Figura 1.1

### 1.3 Configuración

Las grids pueden ser de gran tamaño (miles de nodos) o pequeñas, pocos procesadores y nodos de almacenamiento. Pero, como son desplegadas dichas mallas, podemos nombrar tres formas de categorizarlas en su configuración:

1. A nivel departamental(departamental grids)
2. A lo largo de una organización (intergrids)
3. Entre múltiples organizaciones (extragrids)

#### 1.3.1 Extra Grids

Las aplicaciones que requieren mayor poder de cómputo, son generalmente aquellas que son llevadas a cabo por comunidades de investigación científica, o cálculos de ingeniería. La arquitectura de malla posee la posibilidad de agregar el poder miles de PC'S o servidores vía Internet, creando un ambiente que alcance a veces el desempeño de súper computadoras. Alguno de los proyectos de mayor tamaño en este campo son SETI@home , Mersenne, The Nacional Technology Grid, y the NASA Information Power Grid.

#### 1.3.2 Intergrids

Grids íter departamentales son aquellas que facilitan recursos a través de una organización. La gran mayoría de dichas grids, se encuentran detrás de los firewalls empresariales, brindando soporte íter departamentales .

#### 1.3.3 Departamental

Hasta la fecha, el uso de las grids ha sido en el terreno científico o de investigación. En dichos campos, las grids funcionan a nivel de departamentos de investigación, y su implementación y configuración se realiza en los diferentes departamentos de investigación.

### 1.4 Campos de utilización de las grids

La utilización de las grids puede ser encontrada en las siguientes industrias:

- ❖ Automovilística y Aeroespacial (diseño y modelado)
- ❖ Arquitectura (Construcción e ingeniería)
- ❖ Electrónica (diseño y testing)
- ❖ Energía (Exploración de gas y petróleo)
- ❖ Finanzas/Seguros/Bienes Raíces (corredores de bolsas, especiales para control de stock, análisis de riesgo de inversión)
- ❖ Producción de Fármacos
- ❖ Producción (administración de procesos industriales)
- ❖ Media/entretenimiento (producción de animación digital)}

#### 1.4.1 Utilización Comercial

- ❖ Contabilidad
- ❖ Análisis (estadística, análisis de mercados y negocios)
- ❖ Animación
- ❖ Bancos
- ❖ Visualización de imágenes
- ❖ Soporte de decisión/Minería de datos
- ❖ Juegos Virtuales
- ❖ Testing (testeo de software intensivo)

## **2. Tipos de Grids**

Existen muchos tipos de mallas, desde las de uso científico hasta las que se ofrecen comercialmente en demanda.

Todas estas tienen una cosa en común, que todas proveen de recursos. Pero cada una está preparada y pensada de acuerdo a las necesidades del demandante.

En lo que resta de este punto, hablaremos seis tipos de Grids (poder de cómputo, datos, gubernamentales, utilidades, empresariales), mostrando como son utilizadas en el mundo real y como dichas organizaciones las utilizan, para incrementar la utilización de recursos y de esta forma adquirir mayores ventajas competitivas.

### 2.1 Mallas de cómputo

Las actividades de investigación y desarrollo son las que demandan mayor computabilidad y almacenamiento.

Algunos de los más famosos proyectos incluyen a:

- ❖ El proyecto Seti@home – miles de pc's usadas para la búsqueda de vida extraterrestre;
- ❖ The Mersenne Project - The Great Internet Mersenne Prime Search (GIMPS) es un proyecto de investigación matemática;
- ❖ The NASA Information Power Grid – Este proyecto une las supercomputadoras y el almacenamiento disponible entre las diferentes organizaciones que componen a la NASA, para tener así un único súper recurso informático;
- ❖ The Oxford e-Science Grid - Es una colaboración internacional distribuida con científicos alrededor del mundo.

❖ The Intel-United Devices Cáncer Research Project – Este proyecto tiene como fin encontrar nuevas fármacos para el tratamiento contra el cáncer. Los usuarios pueden donar tiempo de cómputo que no es utilizado por las personas. El mayor desafío en este momento es el proyecto “TeraGrid”. Dicho proyecto fue lanzado por la United State’s National Foundation en Agosto del 2001. Dicho proyecto tiene como meta el poder armar la mayor arquitectura de malla para desarrollos científicos. En el 2004, es posee 20 teraflops de poder de computo, un petabyte de almacenamiento, y la posibilidad de realizar ambientes simulados para modelado con excelente calidad de visualización. La malla se espera que alcance una velocidad de transmisión de 40 giga bits por segundo.

#### 2.1.2 Mallas utilizadas en el campo médico y de desarrollo empresarial

La preponderancia de las mallas puede ser encontrada en el ambiente científico, pero en los últimos años, el crecimiento de éste campo de investigación fue expandido a ambientes comerciales (particularmente a los mercados de combustibles y medicinales).

Nombremos algunos ejemplos:

Medicinales:

Estos ámbitos de estudio han tenido una demanda creciente de poder de cómputo en la última década. (Por ejemplo, el mapeo del genoma humano para realizar descubrimientos de nuevos fármacos). Hoy en día, casi todas las empresas top del mundo de los medicamentos, utilizan grid computing para producir nuevas drogas.

- ❖ BASF, una de las mayores compañías en este ramo, utiliza una arquitectura de malla, para darle mayor poder de cómputo a un proyecto que examina el efecto de los catalizadores en la aceleración de reacciones químicas. El problema que tenía BASF, era que usando arquitecturas normales de computadoras, los análisis requerían de análisis súper complejos, y las simulaciones tomaban tiempo excesivo. Además, los investigadores debían de intervenir siempre en el proceso de cómputo, perdiendo así valioso tiempo de investigación. BASF necesitaba una solución que aumente el poder de cómputo y disminuya el tiempo de simulación. Con esta arquitectura, BASF logro incrementar la utilización de poder de cómputo en un 90 %;
- ❖ Monsanto, líder en productos agrícolas, se enfoca en mejorar la producción agrícola, mejorando de este modo la calidad de los alimentos. Monsanto posee un staff de más de 200 científicos que utilizan el manejo de genética para desarrollar híbridos resistentes a diferentes agentes externos y con mayor potencial productivo. El mainframe de dicha compañía no podía soportar la carga de computo que le era demandada, para solucionar dicho problema, Monsanto apostó por grid computing para poder tener un mejor balanceo de carga;
- ❖ ISIS Pharmaceuticals utiliza dicha arquitectura para el estudio de fármacos contra el sida, cáncer y el síndrome de Crohn’s.

#### 2.1.3 Mallas utilizadas en los mercados financieros

Los mercados financieros encontraron la forma de utilizar a las mallas para acortar tiempos en cómputo y de esta manera poder tener mayores oportunidades de negocios.

- ❖ Una de las mayores compañías americanas financieras, Wachovia, que se dedica al manejo de fondos en inversiones, utiliza dicha arquitectura para obtener derivados de la renta fija, y la obtención del riesgo de las inversiones.

#### 2.1.4 Mercados energéticos

Estos utilizan la arquitectura de malla para procesar los datos sísmicos y potenciar las aplicaciones de imágenes. Estas aplicaciones manejan flujos intensivos de datos, en el orden de terabytes de datos.

- ❖ Amerada-Hess Corporation, una compañía petrolera de 6.6 mil millones de dólares. Esta utiliza aplicaciones asistidas por computadoras en 3-D, para el análisis geofísico de las imágenes de profundidades en la búsqueda de petróleo y gas.

### 2.2 Grids de datos

Estas proveen datos para realizar análisis en profundidad con el apoyo de grandes bases de datos distribuidas.

Uno de los mayores problemas que las mallas de datos ayudan a paliar, son la integridad de los datos. Las empresas, generalmente deben realizar análisis en tiempo real, con datos internos o/y externos.

#### 2.2.1 Mercados Financieros

En el área de análisis de riesgo de inversión.

Algunas empresas que apostaron por compartir una grid de datos son, el Deutsche Bank que utiliza un grid de datos para nutrirse en el cálculo de riesgos de inversión. Pacific Life Insurance, la quinceava compañía en el ranking de compañías de seguro, utilizo la arquitectura de grid para unir sus desktops para el cálculo de modelado y simulación de los mercados financieros.

### 2.3 Grids de Colaboración

Estas son utilizadas para procesar y senderear datos. Estos datos son presentados en forma gráfica, generalmente a localidades geográficamente dispersas.

Podemos encontrar un basto número de ejemplos, en el ámbito aeroespacial, automovilístico, gubernamental y energético.

- ❖ Aeroespacial: Se utilizan para minimizar el tiempo de diseño de los productos, su desarrollo y testeo. Por ejemplo, el desarrollo de alas que tengan mayor índice de planeo y un mayor poder de carga, por parte de estas industrias, hace que el proceso de testeo, desarrollo, y diseño se realice en lugares geográficamente distantes. Se utiliza la arquitectura de malla para aumentar el poder de cómputo y storage.
- ❖ Automovilística, idem aeroespacial, además del área de simulación de choques, como por ejemplo, el estudio de modelos estocásticos de simulación de choques. En estos no se prueba un modelo de vehículo en particular, sino que se analiza el comportamiento de la población en los vehículos ante los choques.

### 2.4 Grids Gubernamentales

Tal como empresas, los gobiernos se benefician de la escalabilidad y la mejor utilización de los sistemas que los sistemas de mallas ofrecen. Estas aceleran los

descubrimientos científicos, nutren a las empresas de herramientas que reducen el tiempo de cómputo, e incrementan el almacenamiento. Además, estas son utilizadas en seguridad nacional, en los centros de inteligencia, para tener resultados casi en tiempo real y manejando inmensos volúmenes de datos.

### 3 Grids utilitarias y de optimización

Las grids utilitarias están empezando a surgir en el ambiente de los negocios. Lo trascendental de estas, es que están disponibles a los usuarios o empresas, cuando los recursos de sus propias grids hayan sido acabados. Entonces, en demanda uno contrata los servicios de estas grids utilitarias para paliar la necesidad ya sea en poder de cómputo o en capacidad de almacenamiento, sin tener que gastar grandes sumas en nuevo hardware, comparativamente al costo de utilización de estas.

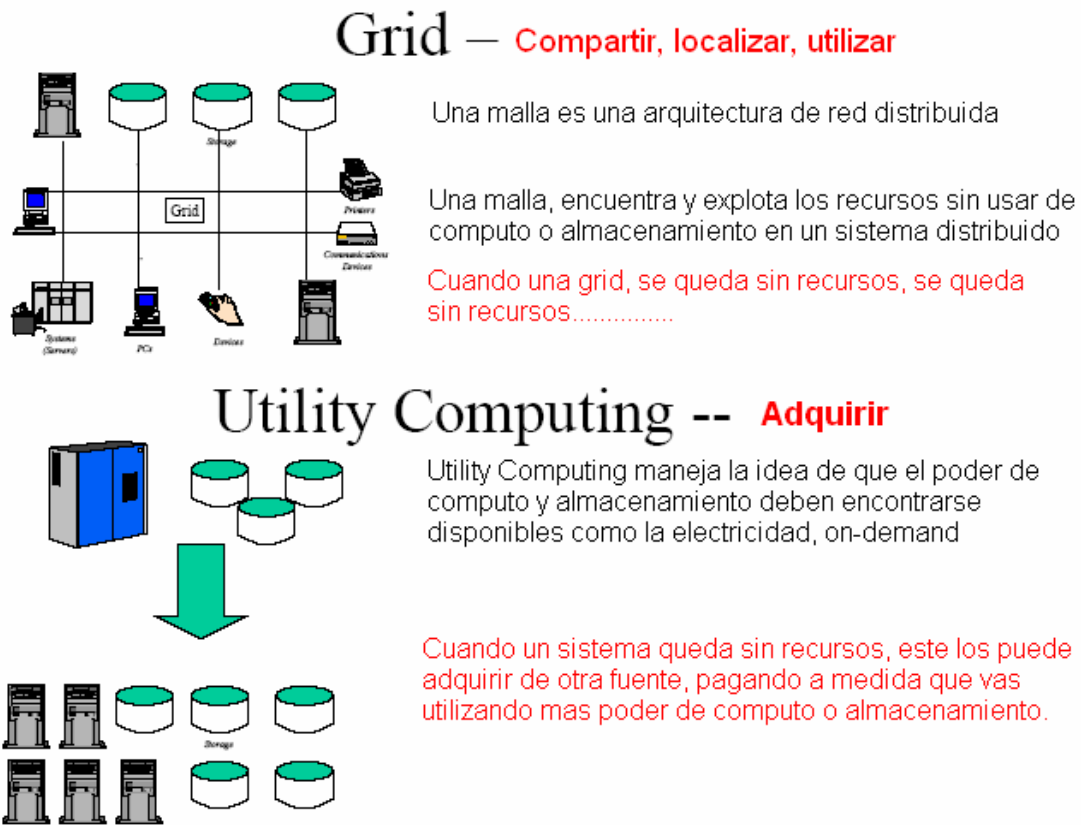


Figura 1.2

#### **4. Conclusión**

El hecho de que el mundo actual evoluciona, hace que los problemas planteados sean cada vez más complejos. Ante esto, se deben buscar nuevos paradigmas de resolución de problemas, y como hacer que los recursos disponibles satisfagan las necesidades de estos cómputos. Esta arquitectura desdobra y provee una solución distribuida para encarar las dificultades o penurias existentes a veces en sistemas centralizados, además de poseer la flexibilidad de realizarla en demanda.

Además, a medida que las redes se hacen más confiables, y la velocidad de transmisión aumenta, se pueden paralelizar los procesos o soluciones, ganando así performance y tiempo, a un menor costo. Siempre y cuando la complejidad de los problemas induzca a esto.

#### **5. Bibliografía**

[1]George Coulouris, Jean Dollimore, Tim Kindberg, Sistemas Distribuidos Conceptos y Diseño, Pearson Education , Tercera Edición, 2001.

[2]Kantardzic, Mehmed , Data Mining:Concepts, Models, and Algorithms, John Wiley and Sons, 2003.

<http://www.gridcomputing.com/gridfaq.html>

<http://www.gridforum.org/>

<http://www.grid.org/home.htm>

<http://www.gridcomputingplanet.com/>

[http://en.wikipedia.org/wiki/Grid\\_computing](http://en.wikipedia.org/wiki/Grid_computing)