

Escuela Superior Politécnica del Litoral
Facultad de Ingeniería en Electricidad y Computación



Grid Computing Computación Distribuida

Kevin Marlon Calderón Barrera

Investigación referente al tema de **Grid Computing** o también conocido como computación grid o distribuida.

La división de los temas y subtemas han sido realizados combinando diversos conceptos de “Computación grid” de wikipedia y de “Computación Distribuida:Grid Computing” de Benjamín Domínguez Hernández

Índice

1. Computación Distribuida	3
1.1. Objetivos de la investigación	4
2. Conceptos Básicos de Grid Computing	5
2.1. Definición	5
2.2. Middleware	5
2.3. Organización Virtual OV (VO por sus siglas en inglés)	6
2.4. Actualidad	6
2.5. Ventajas y Desventajas	7
2.5.1. Ventajas	7
2.5.2. Desventajas	8
2.6. Funcionamiento	9
3. Aplicaciones	10
3.1. La Supercomputación	10
3.2. Sistemas Distribuidos en Tiempo Real	10
3.3. Servicios Puntuales	10
3.4. Proceso Intensivo de datos	10
3.5. Entornos Virtuales de colaboración	10
4. Diferencia entre otros sistemas de computacion distribuida	12
4.1. Cloud Computing	12
4.1.1. Definición de Cloud Computing	12
4.1.2. La Diferencia	12
4.2. Peer-to-Peer (P2P)	14
4.2.1. Definición de Peer-to-Peer	14
4.2.2. La Diferencia	14
5. Conclusiones	16
6. Referencias	17

1. Computación Distribuida

Resumen

Alguna vez alguien habrá imaginado lograr que los computadores puedan trabajar en equipo, más de una vez se debió encontrar problemas muy complejos para ser resueltos por un solo computador, de allí vemos la necesidad de compartir no solo datos o información sino también recursos, viendo también el nacimiento de un nuevo concepto para la computación que es Grid Computing.

El nombre mismo nos indica a lo que esto se refiere, Grid Computing es una perspectiva diferente en el mundo de la computación, mirando a las computadoras no como varios equipos diferentes que pueden resolver problemas de poca complejidad pero separados, sino que ve a los computadores como verdaderos equipos que trabajan en unidad, logrando que las computadoras puedan trabajar en conjunto, como será observado mas adelante en el presente trabajo de investigación.

Para hacer una imagen sencilla de lo que es grid, lo podemos ver también como una malla de algunos computadores que se unirn para que se vean como un solo computador, con recursos de varios computadores a la vez.

Grid Computing es entonces parte importante de la computación distribuida que nos conviene bastante revisar y conocer acerca de estos conceptos para encontrar sus aplicaciones que sin duda alguna podrán ayudarnos a resolver problemas de mayor complejidad que los que pudieramos resolver sin utilizar la computación distribuida.

Todo nuevo conocimiento nos mostrará mayor cantidad de soluciones a problemas actuales, nos armará para enfrentar mejor problemas futuros y podremos ver que aún hay mucho mas por descubrir y conocer.

1.1. Objetivos de la investigación

Objetivo General.-

Se busca dar a conocer conceptos claves acerca de Grid Computing, mostrar sus ventajas y sus aplicaciones.

Objetivos específicos.-

- Describir de forma clara y sencilla los conceptos de computación grid.
- Mostrar campos de aplicación de la computacion distribuida
- Diferenciar los distintos tipos de sistemas distribuidos y entender cual es la aplicación específica de cada uno de ellos.

2. Conceptos Básicos de Grid Computing

2.1. Definición

Grid: Sistema paralelo y distribuido que permite la compartición , detección y agregación de recursos “autónomos ” geográficamente distribuidos.

(Tomado de “Computación Distribuida:Grid Computing” de Benjamín Domínguez Hernández)

Este paradigma fue propuesto a mediados de los noventa por Ian Foster y Carl Kesselman, no es un concepto nuevo pero tampoco es perfecto, hace falta protocolos y mejor estandarización sin embargo la idea es bastante buena ,en sencillas palabras Grid es un sistema de computación que busca integrar en un conjunto los recursos de diversos computadores a la misma vez, a traves de “middleware” que no es otra cosa que un software que brinda asistencia a la aplicación para que esta pueda comunicarse con otras aplicaciones, redes hasta hardware, ahorrando la tarea compleja de realizar la conexión de un sistema distribuido. Esto nos permite realizar tareas complejas desde el ordenador utilizando recursos de otros.

Otra sencilla definición que se puede dar a Grid Computing y como un concepto acertado es que este se refiere a la acción de compartir tareas con otros equipos, utilizando cualquiera de sus recursos, sea procesamiento o simple almacenamiento de datos, sin importar el lugar geográfico en el que se encuentren. También Grid suele ser traducido como “Red”.

Como se mencionó anteriormente la idea de Grid Computing fue concebida por Ian Foster y Carl Kesselman, quienes no solo dieron la idea, sino que dieron vida a la misma, ya que ellos mismos en conjunto desarrollaron las herramientas para que se puedan manejar las redes y trabajar en conjunto de una manera eficiente y flexible.

2.2. Middleware

El middleware es quien hace que sea posible la existencia del Grid y estas son sus principales funciones:

- Encontrar el lugar conveniente para ejecutar la tarea solicitada.
- Optimizar el uso de recursos.

- Ejecutar las tareas.
- Monitorear el progreso de los trabajos en ejecución.
- Gestionar la recuperación frente a fallos.
- Dar aviso cuando se haya completado una tarea y entregar los resultados.

*División tomada de “gridcomputing” de Ramon Millan.

El Middleware como se puede observar, no tiene una tarea muy sencilla, por esta razón no es tan solo un simple programa, sino que varios programas funcionan en conjunto como un solo software y se puede decir que ellos son los que “arreglan o negocian” el uso de recursos de los proveedores.

2.3. Organización Virtual OV (VO por sus siglas en inglés)

La existencia de organizaciones virtuales, es de gran importancia en el enfoque de computacion grid.

“Una organizacion virtual es un conjunto de individuos y/o instituciones definida por reglas que controlan el modo en que comparten sus recursos.” (Tomado de “computación grid” de www.textoscientificos.com del apartado de redes)

Y un ejemplo de una de estas organizaciones es un proveedor de algún servicio de aplicación que ayuda a analizar los datos científicos de un usuario en algún lugar. Cabe también recalcar que las organizaciones son diferentes las unas de las otras, porque cada una tiene un objetivo o simplemente busca resolver algo distinto.

2.4. Actualidad

Existen otros sistemas que buscan el compartir recursos, tales como el peer to peer(P2P, del cual se mencionará luego), que teniendo en común la idea de compartir recursos este tiene sus diferencias. También existen programas a los cuales la computación grid ha brindado beneficios, las empresas que han notado la importancia que puede llegar a tener el ofrecer este sistema a

sus clientes desarrollan software que permita hacerlo. En la sección anterior quedo definido lo que era un middleware, pero hay que resaltar que no existe tan solo un único middleware, estos tienen diferentes funcionalidades, para realizar la conexión dejando la posibilidad de escoger la más adecuada para la situación.

2.5. Ventajas y Desventajas

2.5.1. Ventajas

Antes de mostrar las ventajas hay que indicar que es necesario que los usuarios estén enlazados entre sí y los datos sean compartidos, asegurando también que se puede ingresar a ellos en cualquier lugar y momento.

- Se puede acceder a recursos desde nuestro computador.
- Muchos campos científicos resultan beneficiados por grid computing.
- El tiempo de análisis en grandes cantidades de datos quedará reducido, porque la capacidad de cómputo se eleva.
- Si por algún motivo los recursos o el ordenador que se está usando se caen, la tarea que se estaba realizando se reasigna hacia otra máquina que cumpla los requerimientos para realizar esta tarea.
- Se aprovechan ciclos de procesamiento no utilizados de todos los ordenadores conectados en la red.
- Permite al usuario un acceso a recursos de datos y almacenamientos que puedan ser requeridos.
- Ayudaría en la reducción de costos, porque no sería necesario el gasto en provisionarse de supercomputadores cada cierto tiempo.
- Sin computación grid una organización solo podría usar recursos locales sobre los cuales tenga control.

2.5.2. Desventajas

Aun cuando la computación Grid es bastante buena presenta ciertos inconvenientes.

- Necesita ser mas organizado en cuanto a políticas de seguridad.
- Las aplicaciones para el control del manejo de recursos debe desarrollarse más.
- La computación grid debería poder manejar cualquier tipo de recurso.
- Necesidad de tener los servicios permanentemente conectados.
- Dificultad en sincronizar además de la adaptación de problemas a redes paralelas.

2.6. Funcionamiento

Un grid se define en una secuencia de pasos a realizar a traves de cada tarea, los cuales seran detallados en enumerados.

1. El usuario tiene un problema un poco complejo para su sistema computacional individual, así que necesita ayuda de otros sistemas. Implementa su tarea y se conecta al grid por medio de un software que ejecuta desde su equipo.
2. El ususario debe de acceder y ser validado, esto es por seguridad.

Seguridad.- La seguridad es muy importante, porque los recursos que son compartidos perenecen a persosnas distintas, entre los aspectos que se tienen en cuenta son las políticas que nos dirá que y a quién se permitirá compartir, autenticaciones de usuarios y la autorización.

3. Luego de su verificación de seguridad el usuario se podrá comunicar y conectarse a los recursos.
4. Se buscará el servicio que sea mas adecuado para cada caso y una vez localizados se enviará la tarea para que se ejecute a través del mismo.
5. El ususario podrá revisar el estado de su proceso, una vez terminado deben ser devueltos los resultados.

3. Aplicaciones

Segun “Computación Distribuida:Grid Computing de Benjamín Domínguez Hernández” se tienen 5 campos de aplicación para la tecnología grid.

3.1. La Supercomputación

El sistema grid funciona con nodos en paralelo de modo que al unirse son como si fueran un solo nodo pero con mayor potencia par resolver problemas mas complejos y puntuales que no pueden ser resultados por un solo nodo apartado de la red. Entre sus principales usos en este campo están las simulaciones, el calculo y análisis de grandes cantidades de datos.

3.2. Sistemas Distribuidos en Tiempo Real

Estos son sistemas que requieren de una respuesta inmediata y que la gran cantidad de datos generada pueda ser analizada al instante así mismo como poder recibir actualizaciones del análisis en tiempo real.

3.3. Servicios Puntuales

Son esos servicios que no queremos tener en nuestra organización porque los vemos como de baja prioridad, pero aún así es requerido que podamos acceder a ellos.

3.4. Proceso Intensivo de datos

Los datos viajan a través de la red y son distribuidos para que sean manejados por diferentes nodos, el uso principal de esto es que estos datos no pueden ser almacenados en un único nodo.

3.5. Entornos Virtuales de colaboración

En base al uso de la potencia computacional que nos permite conseguir la malla del grid computing se crean los entornos virtuales 3D.

Las instituciones y organismos que son mas atraídos para utilizar grid son aquellas que comparten un objetivo común y encuentran mas efectivo para

alcanzarlos el compartir y utilizar recursos compartidos.

Se ve repercusión de los beneficios de usar grid en campos importantes para el desarrollo como lo son la medicina (para diagnosticos y análisis), la ingeniería que realiza simulaciones y análisis de fallos, entre otras ramas más que se ven beneficiadas. Estas aplicaciones que se tiene de Grid Computing son aprovechadas actualmente por los científicos. Se espera que se pueda utilizar en el futuro en escala mundial, para que todos seamos usuarios a la vez que proveedores de los servicios o parte de estas organizaciones virtuales.

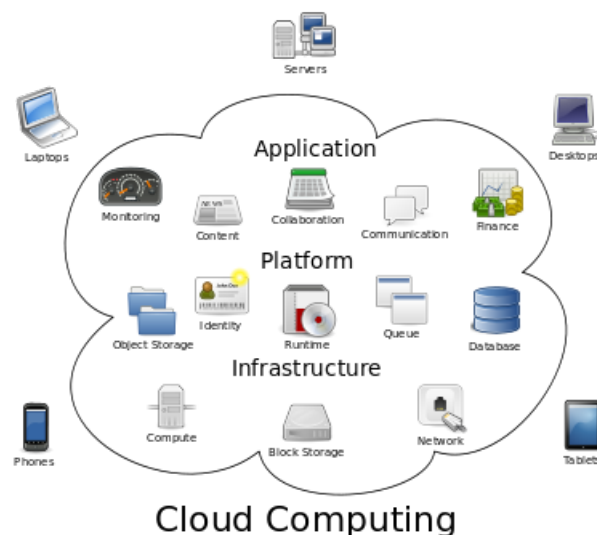
Así es como vemos que este sistema es utilizado por investigadores, científicos hasta médicos, pero aún no ha llegado a manos de todos.

4. Diferencia entre otros sistemas de computacion distribuida

4.1. Cloud Computing

4.1.1. Definición de Cloud Computing

A este concepto se le conoce como computación en la nube o servicios en la nube y este enfoque de sistema distribuido es acerca de los servicios, que estos se encuentren en internet y no los tenga que poseer el usuario. La siguiente imagen tomada de wikipedia permitirá entender mejor lo que es cloud y a lo que se refiere que los servicios se encuentren en Internet o en una “nube”.



4.1.2. La Diferencia

Grid Computing suele ser confundido con Cloud Computing, por esta razón es necesario expandirse para tener mejor conocimiento y poder diferenciar el uno del otro, además de saber cual será nuestra mejor opción frente a un requerimiento.

- Cloud Computing tiene sus servicios en la red y son mas abiertos a los usuarios que Grid.

- La computación en Grid une los recursos de varios ordenadores en una sola red, Cloud presta los servicios en internet pero con un comportamiento diferente a Grid, puesto que no busca el procesamiento de datos.
- El sistema Grid es mas estable porque si se cae la organizacion virtual que esta prestando el servicio, se busca otra organización que puede entrar en su lugar, en cambio Cloud si se cae el servidor principal, se cae el servicio.
- Básicamente el uso y objetivos de Grid son diferentes a los de computing, si bien ambos buscan la distribucion de un servicio utilizando la red, ambos tienen diferente aplicación.
- Grid requiere de un software que haga posible el compartir recursos dentro de la red, mientras que Cloud solo utiliza un software como máscara, para accesar al servicio que se usara en internet.

4.2. Peer-to-Peer (P2P)

4.2.1. Definición de Peer-to-Peer

Una red peer-to-peer o red de pares, también llamada red entre iguales o por sus siglas en inglés P2P, es una red en la que todos son clientes y servidores a la vez, los nodos son interconectados y son tratados como iguales para cada nodo de la red. La mayor utilidad que se aprovecha de este tipo de redes es el de compartir archivos de cualquier tipo y lo hacen aprovechando y optimizando el espacio de banda de los demás usuarios de la red, también en el momento de transmitir los datos, los nodos y el enlace tendrán mayor eficiencia dependiendo de la capacidad de almacenamiento y de la cantidad de usuarios que tengan el mismo recurso ya que se comportarán también como servidores.

En general este tipo de redes es descentralizada, pero estas redes pueden ser clasificadas por su grado de centralización. Donde lo ideal o una red peer-to-peer pura es una totalmente descentralizada.

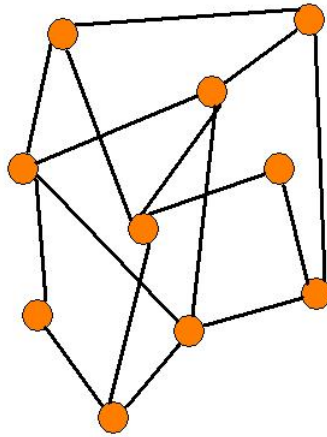


Ilustración de como se vería una red peer-to-peer

4.2.2. La Diferencia

Grid Computing y peer-to-peer tienen similitudes, puesto que son bastante descentralizadas y tienen el principio de formar una malla, pero también existen diferencias que es necesario conocer para escoger el sistema adecuado cuando es necesario.

- Grid Computing tiene un grupo de usuarios mas selecto que peer-to-peer, porque en Grid, se disponen a compartir por un objetivo similar, mientras que P2P comparte archivos a todos los de la red.
- El peer-to-peer se ha popularizado porque cualquiera puede pertenecer a esta red, se mantiene un nivel menor de seguridad dando posibilidad al anonimato y poca motivación al trabajo en conjunto.
- Los recursos de Grid son mas potentes y mejor conectados para trabajar de manera distribuida, con equipos de diversas clases, como una base de datos o equipos científicos, mientras que los recursos peer-to-peer son en su mayoría típicos computadores.

5. Conclusiones

Esta investigación ha llevado al mejor comprendimiento de diversas definiciones y a llegar a un mayor conocimiento en cuanto a computación distribuida.

Fue necesario realizar la investigación completa y no solo del tema escogido porque muchos conceptos eran similares, por esta razón necesitaba conocer mejor todo lo referente a computación distribuida para poder diferenciar a Grid Computing de los demás sistemas de distribución.

Aún así toda esta investigación giro en torno a Grid Computing, solo que con una visión más amplia del mismo, viendo no solo lo que es Grid Computing sino también lo que rodea al tema.

Se llegó a la conclusión que gracias al sistema grid es posible conseguir una supercomputadora, esto porque los recursos se comparten para realizar las tareas, todo como si se tratara de un solo computador.

Grid Computing necesita algunos requerimientos para ponerse en funcionamiento, como son las organizaciones virtuales y otro muy importante que es el middleware.

Middleware se trata de un software que hace posible la existencia de este sistema de redes en paralelo.

Finalmente Grid Computing es un sistema utilizado en el mundo científico, por la misma dificultad que representa hacerlo, sin embargo no se pierde la expectativa que algún día pueda ser más conocido y se pueda utilizar por mayor cantidad de personas, todo está en lograr concretar el conocimiento de seguridad del Grid y mostrar a los usuarios el potencial que representa utilizarlo. De hecho ya hay empresas desarrolladoras de software que han puesto la mira en que llegue ese día

6. Referencias