

Francia Silvana Sánchez Vásquez
Tecnóloga en Sistemas Empresariales de
Información
Egresada Ingeniería de Sistemas
Institución Universitaria Tecnológica de
Comfacauca
email: silvana.sanchezv@hotmail.com

Ingrid Paola Solano Benítez
Docente Investigadora
Institución Universitaria Tecnológica de
Comfacauca
email: isolano@unicomfacauca.edu.co
Sede Popayán - Cauca

Diego Andrés Alegría Ruiz
Egresado Ingeniería de Sistemas
Institución Universitaria Tecnológica de
Comfacauca
email: daar_1684@yahoo.es

Usos y Aplicaciones de la Grid en la Ciencia y la Investigación.

Resumen: la computación en malla ha tenido grandes repercusiones en los adelantos de la ciencia, apoyando los proyectos de investigación que requieren alto rendimiento frente a sus recursos computacionales y facilitando así la labor de la e-ciencia en cada uno de los aspectos del ser humano, gracias a la obtención de productos en una menor cantidad de tiempo. De esta manera, este artículo da a conocer algunos aspectos de la tecnología Grid, sus usos y aplicaciones en la actualidad, con el fin de implementar en la Institución Universitaria Tecnológica de Comfacauca un nodo de Grid Computing. Éste estará basado en un cluster de CPUs para su conexión a Grid Colombia, buscando promover el interés, incentivar la investigación y desarrollar las capacidades investigativas en la Institución Universitaria.

Palabras clave: Grid Computing, e-ciencia, cluster, aplicaciones.

36

INTRODUCCIÓN

Con el crecimiento de la ciencia, los investigadores de múltiples disciplinas han encontrado gran cantidad de obstáculos en el almacenamiento y procesamiento de los datos, cada vez son más robustos, considerando la necesidad de hacer investigaciones paralelas, lo que implica la obtención de los recursos computacionales necesarios para ejecutar dichos procesos o grandes inversiones en la adquisición de servidores y/o supercomputadoras para estas tareas.

Pero de la misma forma en que crece la cantidad de datos a almacenar o procesar, han sido encontradas soluciones a estos inconvenientes. Una de ellas es la Grid Computing o computación en malla, la cual permite a los investigadores, de diversos campos, centrarse en sus temas de dominio sin tener que preocuparse por los recursos computacionales que serán necesarios para el desarrollo de la investigación.

Así, lo importante no es entonces la información que un determinado colectivo de investigación posea sino a la que éste puede acceder. Bajo esa premisa, el compartir recursos computacionales entre distintos grupos de investigación, aún en competencia, resulta de altísimo interés en la ciencia moderna [1].

Para los investigadores que están interesados en la computación en malla, es de vital importancia conocer que tipo de investigaciones se están llevando a cabo mediante la Grid, considerándola como estrategia de agilización de procesos y almacenamiento de datos.

PLANTEAMIENTO DEL PROBLEMA

La computación de alto desempeño o Grid Computing, también conocida como computación en paralelo, está asociada a desarrollar aplicaciones complejas que puedan ser procesadas por más de un computador a la vez. Incluye soluciones que modelan escenarios o situaciones que son muy costosas de implementar

en la vida real. Estos costos han motivado el interés de crear modelos que permitan recrear la vida real en todas sus dimensiones y variables.

La computación en malla también se ha implementado para generar un eficiente y simultáneo uso de un gran número de computadoras, interconectadas en una red de telecomunicaciones de alta velocidad. Esto permite que esta última no se convierta en un cuello de botella, especialmente en el momento de procesar grandes volúmenes de información.

El mercado de usuarios de las redes de computación, de alto desempeño, está compuesto por grandes corporaciones o centros de investigación, los que necesitan disponer de la capacidad de computación de varios equipos de manera simultánea, buscando procesar grandes volúmenes de información por medio de sistemas altamente complejos.

Para Colombia, es indispensable fortalecer las capacidades de colaboración y cooperación entre las instituciones de educación superior y sus centros de investigación. En este contexto, un componente importante es el desarrollo de un servicio de computación de alto desempeño, distribuido a escala nacional y el cual aproveche las ventajas de la Red Nacional de Tecnología Avanzada RENATA. Todo ello en busca de un bien común y considerando el fortalecimiento de la investigación y la educación, específicamente en áreas que requieran un alto nivel de cómputo y capacidad de almacenamiento [2].

La evolución en el uso de estas infraestructuras de cómputo distribuido ha dado lugar al surgimiento de tecnologías de alta complejidad. En el caso de Colombia, ya se cuenta con una Grid nacional, donde uno de sus principales frentes es la formación en tecnologías de Grilla, capacitándose a técnicos, docentes y personal de diferentes universidades para que logren apropiarse de estas tecnologías, labor para la cual se está trabajando en asociación con RENATA.

Para la Institución Universitaria Tecnológica de Comfacauca, es de vital importancia iniciar un proceso de apropiación de estas nuevas tecnologías, de tal manera que puedan incluirse en sus actividades académicas y de investigación, logrando hacer uso de la alta capacidad de procesamiento de que se

dispondría al contar con una Grid local con acceso a una Grilla nacional. De igual modo, se deberá contar con un grupo de personas para colocar en operación y mantener un nodo del prototipo Grid Computing, el cual pertenezca a Grid Colombia.

Así, en la investigación es de vital importancia contar con herramientas que agilicen los procesos, ya que los retos planteados por la ciencia han crecido a un ritmo acelerado, para el cual la tecnología no estaba preparada. De un trabajo empírico y teórico la ciencia ha pasado hoy a un proceso de simulaciones y de manipulación de datos, para los cuales los centros de cómputo, aun los más potentes, no ofrecen una respuesta satisfactoria. Los retos actuales implican hablar de experimentos que requieren miles de horas de CPU, terabytes y hasta pentabytes de almacenamiento, al igual que gigabytes por segundo en capacidad de comunicación [3]. Esto genera gigantescos gastos de recursos económicos al inicio de cada proyecto. Pero como ventaja, las grillas poseen la reducción del "costo" de cómputo, porque proveen bajo demanda accesos confiables y económicos [4]. Pero para los investigadores que no conocen los alcances de la Grid, es necesario identificar las áreas de investigación en las que ha sido un apoyo en la agilización de procesos, logrando identificar cómo podría contribuir a los proyectos que se esperan iniciar.

Conceptualización

Existen diferentes definiciones para computación en Grid de las cuales se retomarán las siguientes:

"La computación en grillas es un tipo de cluster heterogéneo y distribuido que comparte recursos de servidores y computadores ubicados en diferentes dominios de red, con su propio administrador de recursos y políticas de asignación. Cada uno de estos nodos puede tener procesadores, hardware y un sistema operativo totalmente diferente. La computación Grid cada vez toma más relevancia en los problemas que requieren soluciones de multiprocesamiento y grandes cargas de trabajo" [5].

"Grid Computing (computación en malla) es la respuesta de la tecnología a las necesidades de la e-ciencia. Permite compartir recursos entre grupos que no pertenecen a una misma organización y

poder hacerlo de manera que todos los participantes obtengan una experiencia satisfactoria (en términos de rendimiento, seguridad, usabilidad, etc.), para lo cual es necesario tener muchas capas de software, pero también de política” [3].

“Una infraestructura Grid ofrece una capa común para poder integrar plataformas computacionales no compatibles (silos verticales), por medio de la definición de un conjunto consistente de interfaces para acceder y gestionar recursos compartidos. Los servicios Grid incluyen, entre otros, descubrimiento y monitorización de recursos, asignación y gestión de recursos, infraestructura de seguridad y transferencia de ficheros” [6].

En su artículo “What is the Grid? A Three Point Checklist” Ian Foster define Grid en tres aspectos:

“Coordina recursos que no están sujetos a un control centralizado”.

“Un Grid integra y coordina recursos y usuarios que pertenecen a diferentes dominios de control - por ejemplo, un usuario con un ordenador de sobremesa frente a un sistema centralizado; diferentes unidades administrativas de una misma compañía o diferentes compañías - y se encarga de los problemas de seguridad, las políticas, el pago, la pertenencia y de todos los demás asuntos que se dan en este tipo de configuraciones. En caso contrario, estaríamos operando con un sistema de gestión local”.

“Usa protocolos e interfaces estándar, abiertos y de propósito general”.

“Un Grid se construye a partir de protocolos e interfaces con múltiples propósitos que se encargan de problemas fundamentales como la autenticación, autorización, descubrimiento de recursos y acceso a éstos”.

“Es importante que estos protocolos e interfaces sean estándar y abiertos. En caso contrario, estaríamos manejando una aplicación específica del sistema”.

“Para proporcionar calidades de servicio no triviales”.

“Un Grid permite que los recursos que lo forman sean utilizados de una forma coordinada para

proporcionar varias calidades de servicio, relativos por ejemplo al tiempo de respuesta, a la productividad, la disponibilidad, la seguridad y/o a la asignación de diferentes tipos de recursos, para dar soporte a demandas de usuario complejas, de tal forma que la utilidad del sistema combinado es significativamente mayor que la de la suma de sus partes”.

Conociendo la definición de Grid se genera la pregunta de investigación ¿En qué campos de la ciencia y la investigación es utilizada la Grid?

GRID COMPUTING Y LA INVESTIGACIÓN

Para toda organización es importante realizar un plan que permita economizar en los gastos generados por las constantes investigaciones, las cuales son necesarias para actualizar o generar nuevos productos. De igual forma, se hace necesario minimizar las dificultades que puede traer el no poseer los recursos computacionales adecuados, para procesar o almacenar la información o para realizar investigaciones en convenio con otros grupos e ingresar a procesos colaborativos con diferentes colectivos de investigación, locales, nacionales o internacionales, expertos en el tema de preferencia.

Para ello se deberá crear un cluster en la organización, ya sea universitaria o empresarial, que esté conectado con otros grupos, ya que un Grid permite que los recursos que lo forman sean utilizados de una forma coordinada para proporcionar varias calidades de servicio, relativas por ejemplo al tiempo de respuesta, a la productividad, la disponibilidad, la seguridad, y/o a la asignación de diferentes tipos de elementos para dar soporte a demandas de usuario complejas; considerando que la utilidad del sistema combinado es significativamente mayor que la de la suma de sus partes [7], un ejemplo de esto son las Enterprise Grids que se describen a continuación.

Enterprise Grids

El objetivo de este tipo de Grid es permitir que diferentes colectivos puedan compartir diversidad de recursos, con el fin de que logren mejorar la colaboración interna y alcanzar un mayor retorno de la inversión en TIC's (Tecnologías de la Información y las comunicaciones).

Las principales ventajas de esta infraestructura Grid son:

- Minimizar costes
- Maximizar prestaciones [8]

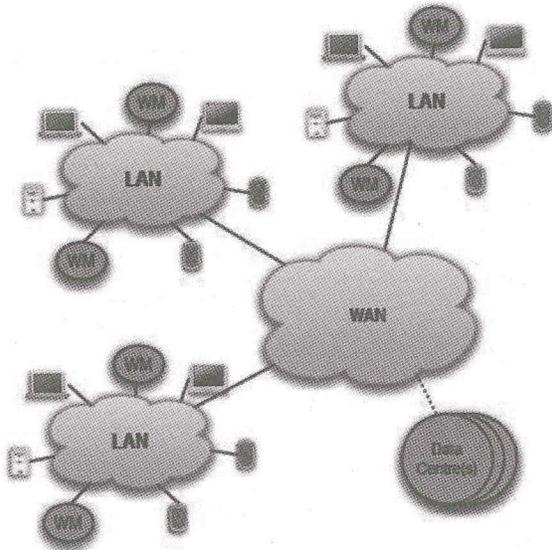


Figura 1. Grid de empresa [8]

Grid asociado

El objetivo de esta infraestructura es proporcionar compartición fiable y segura de recursos a gran escala, entre socios o participantes en la cadena de valor.

Las principales ventajas de esta infraestructura Grid son:

- Acceder a más recursos para satisfacer picos de demanda.
- Proporcionar soporte para hacer frente a proyectos colaborativos [8].

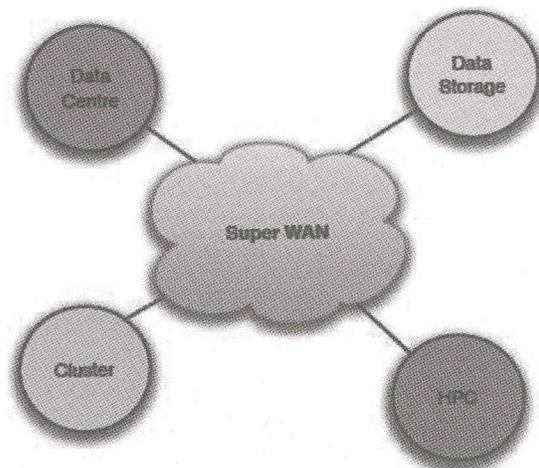


Figura 2. Grid asociado a nivel nacional o internacional [8]

Grid público

El objetivo de este tipo de Grid es proporcionar recursos bajo demanda. Los beneficios que pueden obtenerse por utilizar este tipo de infraestructura son los siguientes:

- Flexibilidad para ajustar la capacidad
- Acceder a capacidad ilimitada
- Transformar costes TIC fijos en variables [8]

APLICACIONES DE LA GRID

Existen múltiples tipos de aplicaciones que se pueden beneficiar de esta arquitectura Grid [9].

Supercomputación distribuida

Se trata de aplicaciones cuyas necesidades no se pueden satisfacer por una sola organización. Este tipo de aplicaciones se caracterizan por demandas puntuales e intensas de computación. Dentro de éstas se engloban las simulaciones de fenómenos físicos o de cálculos numéricos [10].

Sistemas distribuidos en tiempo real

Estas aplicaciones generan un flujo continuo de datos, que deben ser analizados en tiempo real. E-medicine (experimentos de física de alta energía) y el control remoto de recursos complejos son ejemplo de éstos sistemas [11].

Entornos virtuales (Teleinmersión): son aplicaciones que aprovechan la potencia computacional y la naturaleza distribuida del Grid, con el fin de crear entornos virtuales distribuidos en tres dimensiones [12].

Procesamiento intensivo de datos: este tipo de aplicaciones operan con grandes volúmenes de datos que son imposibles de almacenar en un único nodo. En su lugar, los datos se distribuyen por todo el Grid [13].

Servicios puntuales: son aquellas aplicaciones que realizan accesos puntuales a determinados recursos. Éstas se diferencian de las anteriores en que los servicios que utilizan no son para satisfacer potencia computacional y tampoco son servicios en tiempo real.

Como ejemplo se tienen aplicaciones que realizan análisis químico o biológico, para lo cual necesita acceder a un hardware específico [14].

En este contexto, es importante resaltar que la Grid ha sido utilizada en diferentes disciplinas de la e-ciencia, dando como resultado una mayor agilidad en los procesos. Esto ha llevado a los investigadores, que tienen acceso a ella, a tener resultados en menor tiempo del esperado y poder realizar sus publicaciones antes que sus competidores. Así, la gran demanda de computación, espacio y gestión de almacenamiento, requeridos por un gran número de aplicaciones (que gestionan grandes cantidades de datos de forma eficiente), exige cada vez más el uso de nuevas

tecnologías, como es el caso de la computación grid o Grid Computing [15].

CONCLUSIÓN

La computación Grid juega un papel de gran importancia en la ciencia moderna, ya que agiliza la obtención de resultados de investigación y ha ayudado a los científicos e investigadores en áreas como física, medicina, meteorología, biología, química, electrónica, matemáticas y busines intelligence, entre otras. Logrando así integrar esfuerzos incluso cuando están compitiendo por un mismo producto.

De esta forma, teniendo en cuenta la gran aplicabilidad de las grillas mencionadas anteriormente, se

<i>Física</i>	<i>Física de alta energía</i> <i>Aplicaciones de partículas físicas</i> <i>Astrofísica</i> <i>Física Experimental de Partículas</i>	<i>Meteorología</i>	<i>Observaciones de la tierra</i> <i>Modelos atmosféricos</i> <i>Ciencias atmosféricas</i> <i>Observatorios</i> <i>Aplicaciones climáticas</i>
<i>Medicina</i>	<i>Procesamiento de imágenes médicas</i> <i>Problemas médicos en Latinoamérica</i> <i>Área Temática de Salud</i>	<i>Química</i>	<i>Erupciones volcánicas</i> <i>Procesos bioquímicos computacionales</i> <i>Modelos bioquímicos</i>
<i>Biología</i>	<i>Procesamiento de imágenes biológicas</i> <i>Aplicaciones biológicas</i> <i>Biología computacional</i> <i>Bioinformática</i> <i>Biorreducción</i>	<i>Electrónica</i> <i>Matemáticas</i>	<i>Química Computacional</i> <i>Aplicaciones e-Learning</i> <i>Sistemas Complejos</i>

Tabla 1. Disciplinas científicas usando Grid [15]

<i>Busines Intelligence</i>	<i>Data warehouse</i> <i>Datamart</i> <i>Gestión del conocimiento</i>
-----------------------------	---

Tabla 2. Complemento a tabla 1. Disciplinas científicas usando Grid

pretende implementar un nodo de Grid Computing en la Institución Universitaria Tecnológica de Comfacauca - Unicomfacauca, con el fin de promover el interés, incentivar la investigación y desarrollar las capacidades investigativas en la Institución a través del diseño, montaje y puesta en operación de servicios de procesamiento distribuido tipo grilla, utilizando la Red Académica de Alta Tecnología Avanzada – RENATA.

[15] *Computación Grid*. Disponible en: <http://is.ls.fi.upm.es/doctorado/Curso2007-2008/Asignaturas/CompGrid.html>

REFERENCIAS

- [1] H. Castro, J. Chacón, C. Díaz, E. González y J. Zuluaga. *Grid Colombia: Soporte para Investigaciones de Avanzada*. 2011.
- [2] Disponible en: <https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/GridColombiaWorkshop2010>
- [3] I. Foster y C. Kesselman. *The Grid 2: Blueprint for a New Computing*. (2nd. Ed.), Morgan Kaufmann, 2003.
- [4] G. P. Ledesma, M. A. Aguirre, M. F. Zuleta y N. L. Payares. *Simulación de Flujos De Fluidos Utilizando Grillas De Cómputo*. 2008.
- [5] G. A. Isaza, N. D. Duque. *Arquitecturas y Modelos de Programación en Computación Grid*
- [6] K. L. Algara. *Tesis Doctoral. Estrategias De Planificación En Infraestructuras Grid.Federadas*.
- [7] I. Foster. *What Is The Grid? A Three Point Checklist*. 2002
- [8] B. Sotomayor. *Un Nuevo Paradigma de Computación Distribuida*. 2003.
- [9] *Nug30*. Disponible en: <http://www-unix.mcs.anl.gov/metaneos/nug30>, 2009.
- [10] Sensima Research. *The End of EAI As We Knew It*. GRID today, December 2003.
- [11] Red Académica de Centros de Investigación y Universidades Nacionales (Reacciun2). Disponible en <http://www.reacciun2.edu.ve>, 2009.
- [12] GriPhyN. *Grid Physics Network*. Disponible en: <http://www.griphyn.org>, 2009.
- [13] S. R. Amendolia, F. Estrella, W. Hassan, T. Hauer, D. Manset, R. McClatchey, D. Rogulin y T. Solomonides. *Grid and Cooperative Computing – GCC 2004*. Chapter *MammoGrid: A Service Oriented Architecture based Medical Grid Application*. Lecture Notes in Computer Science, 2004, pp 939–942.
- [14] J.C. Martínez. *Hacia una Iniciativa Grid Nacional en Colombia: Infraestructura y Aplicaciones*. Tesis De Maestría.