

INVESTIGACIÓN ACADÉMICA Y TRANSFERENCIA DE CONOCIMIENTO SOBRE COMPUTACIÓN GRID EN LA "UNIVERSIDAD DISTRITAL - FRANCISCO JOSÉ DE CALDAS"

Manuel Ernesto Bolaños González

Ingeniero de Sistemas, MS(c) en Ciencias de la Información y las Comunicaciones, aspirante a Doctor en Ingeniería Informática, docente de la Universidad Antonio Nariño.
mbolanosg@udenar.edu.co

Alexis Rojas Cordero

Ingeniero de Sistemas, MSc (c) en Ciencias de la Información y las Comunicaciones, docente de la Universidad la Salle.
rojas.cordero.alexis@ieee.org

José Nelson Pérez Castillo

Ingeniero de Sistemas, MSc. en Teleinformática, Doctorado en Informática, docente de planta de la Universidad Distrital "Francisco José de Caldas".
jnperezc@gmail.com

Tipo: Artículo de investigación

Fecha de Recepción: Agosto 30 de 2010

Fecha de Aceptación: Nov. 5 de 2010

KNOWLEDGE TRANSFER AND ACADEMIC RESEARCH OVER GRID COMPUTING AT UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

Abstract

This article presents innovative experiences, at the national level, resulting from the preparation and delivery of a course in Advanced Information Systems for e-Science-based computing grid (computational meshes). To this end, a specific-purpose computational grid was conceived, designed and implemented. Such an implementation was based on the following: Globus Toolkit 4, an ad hoc network, a virtual infrastructure and a teaching model supported by advanced constructivist concepts and virtual collaborative learning.

Keywords: Grid Computing, ad hoc network, middleware, globus toolkit 4, gLite.

Resumen

Este artículo presenta experiencias innovadoras en el ámbito nacional, resultado de la preparación e realización de un curso en Sistemas de Información Avanzados para Ciencia con base en computación grid (mallas computacionales). A tal efecto, se concibió, diseñó e implementó una malla computacional de propósito específico, con base en GlobusToolkit 4, una red ad hoc, una infraestructura virtual y un modelo pedagógico avanzado apoyado en conceptos constructivistas y de aprendizaje virtual cooperativo.

Palabras clave: Computación grid, red ad hoc, middleware, globus toolkit 4, gLite.

1. INTRODUCCIÓN

En los últimos años, la computación Grid ha revolucionado la manera de investigar y socio-construir el conocimiento, gracias a

las redes sociales que logran el aprovechamiento óptimo de recursos, permitiendo que el trabajo colaborativo entre diferentes actores y organizaciones virtuales, sea más productivo, a partir de elementos geo-

gráficamente dispersos [1], [2], [3], [4].

Es importante que los investigadores colombianos tengan excelentes conocimientos en computación grid, de modo que les permitan aprovechar sus ventajas en beneficio del desarrollo nacional. En el ámbito colombiano, se han impartido cursos similares, pero apoyándose en infraestructuras internacionales, gracias a que las instituciones pioneras cuentan con los recursos suficientes [5]. Bajo este contexto, el curso mencionado anteriormente, constituye una importante innovación que se ha plasmado en varias versiones del mismo a nivel de maestría. Posteriormente se convertirá en un curso dirigido a estudiantes de pregrado [6], que les facilitará adquirir y desarrollar plenamente las competencias cognitivas y laborales para abordar trabajos que necesite el desarrollo científico y técnico nacional, acorde con el reciente Plan Nacional de Tecnologías de la Información y las Comunicaciones, aprobado por el gobierno nacional [7]. Cabe anotar, que estas mallas computacionales, facilitan el trabajo en equipo y la creación de comunidades de conocimiento, situación que marca un avance estratégico en el modo de hacer investigación, desarrollo e innovación, que resultan claves para la inserción de la nación colombiana en la sociedad del conocimiento.

El documento se estructura de la siguiente manera. En la sección titulada experiencia, se describe el proceso llevado a cabo, para el desarrollo del curso. Luego se evalúan los resultados obtenidos y finalmente se presenta el trabajo futuro de investigación y las conclusiones.

2. EXPERIENCIA

Uno de los principios de la Grid es apoyar el trabajo colaborativo, razón por la cual, el curso se proyectó teniendo en cuenta la definición de la teoría del aprendizaje de Vigotsky [8], que hace hincapié en que la ganancia cognitiva individual ocurre primero a través de la interacción con el entorno social y que el entorno social del aprendizaje puede ser empleado en trabajo mancomu-

nado apoyado en la informática. A lo largo del desarrollo del curso fue primordial el trabajo en grupo y se encontró que la co-presencia y la discusión ayudaron a obtener mejores resultados en la comprensión y apropiación de los conceptos, ya que la acción comunicativa permitió la planeación de tareas y su orientación al entendimiento [9], [10], de tal manera que se coordinaron los planes de acción sobre la base de acuerdos motivados a partir de los objetivos y los temas del curso.

Una de las primeras actividades posteriores a la concepción del curso, consistió en la fundamentación de virtualización, con base en el paradigma de máquinas virtuales, el cual es una tecnología de particionamiento lógico que aporta aislamiento de los sistemas operativos. Debido a que las máquinas virtuales son abstracciones de los equipos físicos reales, con sistema operacional completamente instalado, que contiene utilidades de administración del sistema, aplicaciones y redes, se convierten en herramientas apropiadas para la implementación y ejecución de diferentes tipos de pruebas. Para el laboratorio del que trata este artículo, se utilizó el software de VMware [11], que permite instalar cualquier sistema operacional, ya sea Linux, Solaris, Windows, AIX, etc. La eficiencia y rendimiento de la plataforma real, depende de los elementos de hardware como: memoria RAM, espacio en disco, velocidad del procesador, etc.

En la definición y preparación del syllabus del curso, se tuvo como base el orden y las presentaciones del seminario internacional de computación grid [5], con el propósito de trabajar con el middleware gLite. Se iniciaron las pruebas de instalación y configuración de los elementos de la grid, pero se presentaron inconvenientes debido a que no se contaba con IPs públicas necesarias para cumplir los requerimientos de trabajo de gLite, razón por la cual se tomó la decisión de construir una red ad hoc y utilizar el middleware GT4, de manera que se pudieran optimizar los recursos de procesamien-

to, reducir costos y responder con mayor rapidez a las necesidades del proyecto. Con la ejecución y resultado de las diferentes pruebas realizadas se pudo identificar que para tomar el curso, los estudiantes debían estar en condiciones de construir una red de computadores de área local, manejar los conceptos fundamentales de sistemas operativos, máquinas virtuales tales como VMware [11], edición básica de programación shell en Linux, conceptos básicos de computación Grid [3], [4] y de sistemas distribuidos en general, manejo básico de los esquemas de seguridad de clave pública [4], [12], servicios Web y estándares de la OGSA[13], con miras a la configuración de organizaciones virtuales en ambiente Grid, con el propósito de abordar la configuración y disposición de los distintos elementos de una malla computacional, que permitieran alcanzar los objetivos específicos de formación planeados para el curso.

El curso se concibió con una duración de 64 horas presenciales teórico/prácticas, en donde los estudiantes debían asumir, extra clase, algunas lecturas y la preparación de los ejemplos del texto guía [6], así como la elaboración de un proyecto de curso consistente en montar una infraestructura de servicios grid.

Al iniciar el desarrollo del curso se conformaron grupos de estudiantes, con el propósito de que en cada grupo, un estudiante en particular, tuviera a cargo el montaje de un componente de la grid. Se empezó con presentaciones sobre los conceptos generales de la computación grid, Scientific Linux, máquinas virtuales [11], servidores NTP, DNS, Webmin y manejo de certificados, de manera que existiera claridad en los objetivos, en los temas a tratar y en cada uno de los componentes del nodo grid. Cada estudiante trabajó con un computador portátil con las siguientes características mínimas: 1 Gb en RAM y 30 Gb de espacio disponible en disco velocidad de procesador superior a 1.5 Ghz y un concentrador por grupo. En cada grupo de trabajo se desarrolló el si-

guiente orden de trabajo:

- Instalación del servidor de máquina virtual VMWare-Server.
- Instalación del sistema operativo Scientific Linux v.4.4.
- Instalación del servidor NTP (Servidor de sincronización de relojes de los portátiles de la red de cada grupo) con los pasos que se muestran en la Fig. 1.
- En este punto del procedimiento se elaboró una copia de la respectiva máquina virtual con los componentes hasta este punto instalados y se nombró un alumno como el responsable del corazón del sistema, quien instaló los servidores HTTP y DNS, reconfigurando las direcciones IP de los miembros de su grupo de acuerdo al diseño de la Fig. 2.
- En otra copia de la máquina virtual pre configurada en el punto C, se configuró el contenedor de GT4 [14] y todos sus componentes como Java, Ant, etc., la cual se denominó máquina base. Posteriormente se realiza una nueva copia de ésta máquina y se distribuye a cada uno de los integrantes del grupo para que proceda con la instalación de los respectivos componentes del nodo Grid como por ejemplo la entidad certificadora (CA), estaciones de trabajo (WN), etc., reconfigurando su respectiva dirección IP, haciendo énfasis en la importancia de la sincronización de los equipos para la certificación [15].

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS Bogotá - Colombia www.udistrital.edu.co

Pre-requisitos (4)

- **Instale ntp si no ha sido instalado en su sistema:** `apt-get install ntp`
`# apt-get -f install`
- **Adicione su servidor de tiempo:** `# vi /etc/ntp.conf`
En el bloque NTP MULTICASTCLIENT
`restrict 192.168.0.119 mask 255.255.255.0 nomodify notrap noquery`
- **Y en el bloque GENERAL CONFIGURATION**
`server ntp.geox.udistrital.edu.co`
- **Edite /etc/ntpstep-tickers** adicione su(s) server(s) hostname de tiempo
Ejemplo: `ntp.geox.udistrital.edu.co`
- **Si un cortafuego está activado, tendrá que permitir la comunicación de entrada en el puerto del NTP**
Edite /etc/sysconfig/iptables y antes de `-A FORWARD` adicione la línea
`-A INPUT -s 192.168.0.119 -p udp --dport 123 -j ACCEPT`
- **Reiniciar servicios de red**
`# /etc/init.d/network restart`

Sistemas de Información Avanzados para e-Science [Ing. José Nelson Pérez Castillo]
 Maestría en Ciencias de la Información y las Comunicaciones (Universidad Distrital Francisco José de Caldas)

Fig. 1. Pasos para configurar el servidor de control de tiempo - NTP.

La configuración de cada servicio con su software respectivo, se realizó a manera de taller.

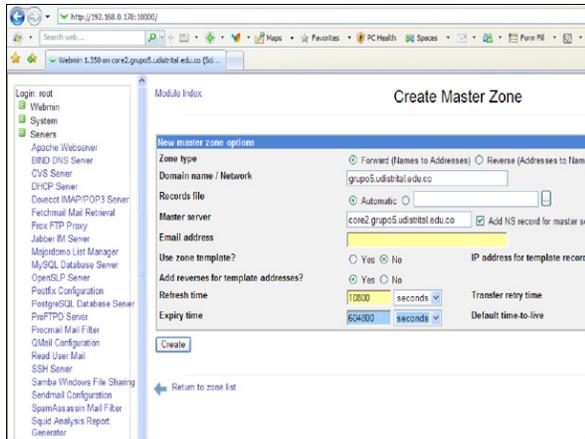


Fig. 4. Configuración de un DNS usando Webmin.

De acuerdo con las necesidades de cada componente de la Grid, se utilizaron diferentes paquetes de software, teniendo en cuenta los requerimientos y pasos especificados en la guía de instalación de GT4 [14]. Cada estudiante realizó la instalación completa en diferentes máquinas virtuales, para adquirir práctica en dicha instalación y configuración.

Cuando se instalaron y probaron los servicios Grid utilizando GT4, los estudiantes implementaron y ejecutaron los ejemplos de los capítulos 6 al 24 del libro de Borja Sotomayor y Lisa Childers. Con estas prácticas se terminó lo correspondiente al middleware GT4 y se procedió a abordar de manera teórica los temas concernientes a gLite de acuerdo con la programación del curso.

3. RESULTADOS

Se configuró una plataforma mínima ad hoc de computación Grid, como la base para un laboratorio que permitió a los estudiantes realizar las diferentes pruebas y demostraciones que exige el desarrollo de una tesis de maestría. Como parte de la evaluación del curso se conformaron dos grupos al

azar, para que cada uno implementara la estructura de la grid ad hoc y los servicios que corren sobre ella.

Para esta práctica se puso un tiempo límite de 6 horas, con el propósito de que se delegaran funciones y se trabajara de manera colaborativa e integral. Terminado este lapso de tiempo el primer grupo logró implementar en un 90% de los requerimientos, el segundo grupo logró implementar un 75% de los requerimientos, debido a que tuvieron dificultades en el momento de integrar su trabajo. Se utilizó eficiente y efectivamente, el software GT4, como middleware de computación grid, para el desarrollo y avance de los procesos de investigación.

4. CONCLUSIONES Y TRABAJOS FUTUROS

Con el propósito de socializar el conocimiento, se está desarrollando el proyecto de investigación “Estrategia de e-Learning para Capacitación en Computación Grid” [6], con el cual se pretende tener una herramienta computacional que permita el entendimiento claro de los pasos a seguir en la configuración de una grid, mediante la implementación de un laboratorio de práctica que emule la realidad de una grid. Además, se pretende a través de Access Grid [24], enlazar espacios de trabajo activos y distribuidos para soportar trabajo colaborativo que permita compartir, difundir y desarrollar el conocimiento, proporcionando acceso a los recursos de la tecnología grid.

La red ad hoc es una idea innovadora que se usó en una segunda versión del curso con excelentes resultados, y se pretende usarla en otros cursos ya que permite implementar, configurar y simular el comportamiento de una grid en sitios en los que no se cuenta con los elementos computacionales necesarios para implementar una red real.

La experiencia adquirida en la preparación y desarrollo del curso, da el conocimiento necesario para concluir que es factible la

realización de un nuevo ejercicio con red Ad hoc usando los middlewares Globus toolkit 4.2 y gLite 3.1, como se muestra en la Fig. 5. El reto consiste en que los elementos de la red gLite se autentiquen en la CA de Globus. Se recomienda bajar el repositorio de la página de Gilda [25], [26] (con el siguiente comando: `wget -mp -nH -np http://grid018.ct.infn.it/rep`), para tener a la mano la funcionalidad de todos los paquetes de los componentes del middleware gLite [23]. Este repositorio se configura en el directorio html de Linux en una máquina alterna de la respectiva red Ad hoc.

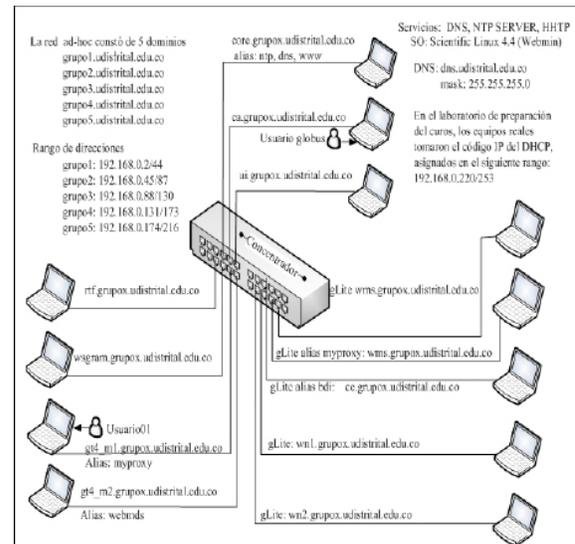


Fig. 5. Configuración propuesta con dos middlewares diferentes sobre una red Ad hoc.

Referencias Bibliográficas

- [1] YAHYAPOUR, R. Attributes for Communication between Scheduling Instances. Copyright (c) Global Grid Forum (2001). All rights reserved.
- [2] Foster, I. What is the Grid? A The Three Point Checklist. Argonne National Laboratory & University of Chicago, foster@mcs.anl.gov, 2002. Consultado Julio 2008.
- [3] Ferreira Luis y Otros, Introduction to Grid Computing with Globus. IBM Red Books. 2005.
- [4] Verstis Viktors y Otros. Fundamentals of Grid Computing. IMB RedPapers. Austing Texas. U.S.A. 2002.
- [5] Seminario Internacional de Computación Grid. Universidad de los Andes. Bogotá Colombia, Marzo 2007. <http://notauniandina.edu.co/html/nota6/070221SeminarioInternacionalGridComputing.php>, consultada: Abril del 2007 y Julio del 2008.
- [6] Bolaños Manuel. Estrategia de e-Learning para formación en computación Grid. Tesis de grado. Universidad Distrital. Bogotá – 2008.
- [7] Plan Nacional de TIC 2008 – 2019. Ministerio de Comunicaciones de Colombia. Bogotá - 2008 <http://www.colombiaplantic.org>, consultada: Mayo de 2008.
- [8] James Britton. Vygotsky's Contribution to Pedagogical Theory. English in Education 21 (3), 22–26 doi:10.1111/j.1754-8845.1987.tb00951.x. 1987
- [9] Rita Radl Philipp. La teoría del actuar comunicativo de Jürgen Habermas: un marco para el análisis de las condiciones socializadoras en las sociedades modernas. 15076 Santiago de Compostela. Spain, 1998
- [10] HABERMAS, "Teoría de la Acción Comunicativa", I, Taurus, Madrid: España 1987, pp.122-146
- [11] Hewlett-Packard Company. Instalación, configuración y administración de HP Integrity Virtual Machines. Fort Collins, Colorado. USA. 2005, pp 10-12.
- [12] Sotomayor Borja and Childers Lisa. Globus Toolkit 4. Programing Java

- Services. University of Chicago. The Elsevier inc. San Francisco CA. U.S.A. 2006. Pp 3-39, 305-223, 431-491.
- [13] Verónica Vanessa Barrios y otros. Estándares de Grid Computing. Facultad de Ciencias Exactas, Naturales y Agrimensura. Universidad Nacional del Nordeste. Corrientes – Argentina. 2006. Pp 2-10
- [14] Globus Toolkit Development Documentation. Chapter 1.4.0.x quickstart. <http://www.globus.org/toolkit/docs/4.0/admin/docbook/quickstart.html>. Consultada, Abril de 2007.
- [15] Ian Foster, Carl Kesselman, Gene Tsudik, Steven Tuecke. A Security Architecture for Computational Grid.
- [16] Mary R. Thompson, LBNL, Doug Olson, LBNL, Robert Cowles, SLAC, Shawn Mullen, IBM, Mike Helm, ESnet. CA-based Trust Issues for Grid Authentication and Identity Delegation. GFD-I.17. Grid Certificate Policy Working Group. June 2003.
- [17] Randy Butler, NCSA, Tony J. Genovese, ESnet/LBNL. Global Grid Forum Certificate Policy Model. GFD-C.16. CAOPS Working Group. June 1, 2003.
- [18] W. Allcock. GridFTP: Protocol Extensions to FTP for the Grid. GFD-R-P.020. Editor Argonne National Laboratory. April 2003. Revised April 2003
- [19] Igor Mandrichenko, FNAL (editor). GridFTP Protocol Improvements. GWD-E. GridFTP WG. July 3, 2003. Revised July 11, 2003
- [20] W. Allcock, J. Bester, J. Bresnahan, A. Chervenak, L. Liming, S. Tuecke. GridFTP Protocol Specification. GGF GridFTP Working Group Document, September 2002.
- [21] K. Czajkowski, I. Foster, C. Kesselman, S. Martin, W. Smith, and S. “Tuecke. A resource management architecture for metacomputing systems”. Job Scheduling Strategies for Parallel Processing. 1998. pp 62-82
- [22] K. Czajkowski, I. Foster, N. Karonis, C. Kesselman, S. Martin, W. Smith, and S. Tuecke. A Resource Management Architecture for Metacomputing Systems. Proc. IPPS/SPDP '98 Workshop on Job Scheduling Strategies for Parallel Processing, 1998.
- [23] Burke S y otros. gLite 3.0 and 3.1. User Guide Manual Series. CERN-LCG-GDEIS-722398, Abril 7 del 2008. <https://edms.cern.ch/file/722398/gLite-3-UserGuide.pdf>. Consultada. Abril 2007 y Julio 2008.
- [24] <http://www.accessgrid.org>. Consultada Febrero 2008.
- [25] Bruno Ricardo. Preparing to Host a Gilda Tutorial. INFN – Sicilia 2006. Disponible en: <https://grid.ct.infn.it/twiki/bin/view/GILDA/PreparingToHostAGILDATutorial>. Consultado Mayo 2007.
- [26] A brief guide using GILDA gLite middleware. <https://grid.ct.infn.it/twiki/bin/view/GILDA/EUMEDGridSiteInstallation>. Consultada Abril 2008.