

Implementación Banco de Pruebas GRID

Bertogna, Leandro M - Del Castillo, Rodolfo
Departamento de Informática y Estadística
Universidad Nacional del Comahue
Buenos Aires 1400, Neuquén
Argentina
{mlbertog, rolo}@uncoma.edu.ar

Aggio, L. Santiago - Matrangolo, Carlos
Centro Regional de Investigaciones
Básicas y Aplicadas de Bahía Blanca
Cam. La Carrindanga Km.7, Bahía Blanca
Argentina
{slaggio, symatran}@criba.edu.ar

Resumen: La computación Grid permite a grupos de recursos distribuidos y heterogéneos trabajar en forma coordinada para solucionar problemas complejos, que sería costoso o imposible de solucionar con las máquinas disponibles en forma aislada. En este trabajo desarrollaremos el proyecto de extensión presentado en la Universidad Nacional del Comahue junto con el CRIBABB para la interconexión y puesta en marcha de laboratorios para la investigación.

1. Introducción

La computación Grid [1][2] permite a grupos de recursos distribuidos heterogéneos formar parte de una gran computadora virtual y puede ser usada por grupos de usuarios y organizaciones en colaboración para solucionar problemas complejos

Compartir a través de Grid va mas allá de compartir datos y recursos de red. Permite a los usuarios mejorar la performance, reducir tiempos en las aplicaciones, incrementar la disponibilidad y mejorar la tolerancia a fallos distribuyendo la ejecución y la carga de trabajo intensiva a través de los recursos. La computación Grid también permite la integración efectiva y eficiente de las aplicaciones, que están en plataformas heterogéneas, sistemas operativos, implementados bajo diferentes paradigmas y lenguajes de programación. Además de esto, compartir recursos de lugares aislados formando un grupo mayor, permite el uso de recursos no disponibles o subutilizados, esto no solo optimiza la disposición de los recursos sino que también aumenta la performance de las aplicaciones.

Este tipo de soluciones se encuentra poco aplicada en entornos de producción o en áreas de investigación de nuestro país siendo esto un objetivo para alcanzar con el fin de lograr un incremento en la productividad de los distintos grupos de investigación.

Existen varios ejemplos de aplicaciones de esta tecnología en otras partes del mundo como por ejemplo Teleprecencia Distribuida en el proyecto de NEESgrid[3], Distribución de datos médicos en una red de investigación biomédica[4], integración de conocimiento en el proyecto Silico[5], etc.

El proyecto que presentamos en este trabajo tiene como destinatarios el Centro Regional de Investigaciones Básicas y Aplicadas de Bahía Blanca (CRIBABB) con sus institutos asociados, Planta piloto de Química (PLAPIQUI), Instituto Argentino de Oceanografía (IADO), Investigaciones Bioquímicas de Bahía Blanca (INIBIBB) y Centro de Recursos Renovables de la Zona Semiárida (CERZOS), además del departamento de Ciencias de la Computación de la Universidad Nacional del Comahue.

El proyecto surgió con la necesidad de agrupar recursos tecnológicos heterogéneos dentro y fuera de las organizaciones involucradas, para lograr combinarlos y mediante una computadora virtual lograr un poder de cálculo mayor que el obtenido individualmente. Este proyecto también se encuentra relacionado con una propuesta internacional AlfaGrid[6] donde participan las instituciones involucradas en este proyecto y tiene como fin lograr una cooperación en el ámbito científico sobre la

tecnología Grid entre Europa y Latinoamérica y el proyecto RedCLARA[7] que es una iniciativa de las redes académica avanzadas para Latinoamérica.

En este trabajo se pretende optimizar el uso de los recursos tecnológicos disponibles y analizar los costos de implementación y beneficio de este tipo de soluciones. Una vez implementado se usará como una herramienta de divulgación para otros grupos de investigación e instituciones relacionadas. Actualmente se dispone de capacidad de computación y datos pero estos están dispersos geográficamente y generalmente sobre plataformas incompatibles, haciendo inviable el uso de estos recursos en forma remota y coordinada.

2. Desarrollo del proyecto

Para llevar adelante el proyecto en estos momentos se cuenta con varios tipos de recursos informáticos, en el CRIBABB existe un numero importante de equipos como son SUN, Silicon Graphics, Macintosh, formando un parque heterogéneo de diferentes fabricantes, la gran mayoría de estos de uso exclusivo de los proyectos de investigación, también se cuenta en experimentación con un cluster de uso dedicado y herramientas electrónicas avanzadas como es el equipamiento de microscopio electrónico. Por su parte en la UNC, el departamento de ciencias de la computación dispone de tres aulas con 20 máquinas cada una todas ellas de P-IV con un mínimo de 256Mb por equipo.

En cuanto a las comunicaciones, se cuenta con conectividad a Internet y recientemente con enlace a redes avanzadas (I2), a través de RETINA, formando parte de la red CLARA, permitiendo de esta manera vincular centros de investigación de Latinoamérica con calidad de servicios en las comunicaciones.

En el aspecto humano contamos con expertos en el área de redes e ínter conectividad y en el área de sistemas distribuidos y seguridad. Este último es un aspecto fundamental para cualquier aplicación que funcione a través de Internet. Salvo el equipamiento dedicado el resto de las máquinas se encuentran libres en horario nocturno.

De no mediar ningún tipo de intervención, los recursos tecnológicos permanecerán aislados, acarreado su subutilización e incrementando el costo de los proyectos con necesidad de alta performance o alto volumen de trabajos, ya que el personal involucrado deberá movilizarse para poder realizar sus labores o se deberá contratar estos servicios en el exterior.

Para la concreción del proyecto se plantearon dos objetivos específicos: el primero es lograr la interconexión de los laboratorios y luego la extensión de estos conocimientos a otras instituciones, que tengan las mismas problemática.

Las metas concretas para lograr cumplir los objetivos propuestos son cuatro, dos por año:

En el primer año, lograr una capacitación de los equipos de trabajo y manejo de la potencialidad de la tecnología e implementar la instalación, puesta en funcionamiento de laboratorios interconectados ofreciendo servicios básicos; y en el segundo año las metas a cumplir son: elevar el nivel de complejidad de los servicios ofrecidos además de agregar nuevas herramientas y lograr una calidad de utilización aceptable de los mismos para luego poder extenderlos a otras instituciones con la misma problemática.

La metodología a emplear en el desarrollo del proyecto incluirá:

a) En lo operativo:

- Establecimiento de tareas para cada uno de los integrantes.
- El trabajo básicamente será en forma remota con un planteo de objetivos a corto plazo. Al final de cada etapa se evaluará el avance del proyecto y se realizarán los ajustes necesarios para poder cumplir con el cronograma propuesto.

- Elaboración de informes bimestrales con el avance del proyecto.
- b) En lo formativo:
- Lectura, análisis, instalación y desarrollo de software referente al proyecto.
 - Obtención de bibliografía a través de la búsqueda de material en Internet, la suscripción a listas de correo, la adquisición de libros, etc.
 - Planificación, dictado y asistencia a cursos sobre la tecnología involucrada.
- c) En lo institucional:
- Establecimiento de contactos con las autoridades y personal del CRIBABB y de la UNC con el fin de lograr ambientes propicios y cooperación para la aplicación de la tecnología surgida de los proyectos de investigación realizados en ambas instituciones.

3. Actividades

Las actividades propuestas a lo largo del desarrollo del proyecto son las siguientes:

1. Conformación de los laboratorios: Se realizarán reuniones con el fin de establecer las aplicaciones en entorno de prueba, el equipamiento necesario y analizar las alternativas en cuanto a software que se requerirá para correr las aplicaciones.
2. Instalación software, para entorno de producción: Una vez realizadas las etapas de prueba en laboratorio de software, se realizara una elaboración de conclusiones, se realizará el diseño de la arquitectura y se instalará las versiones definitivas del software que proveerá los servicios.
3. Extensión de los servicios de Grid: Cuando la tecnología este funcionando y en régimen en su entorno de producción, se iniciará la fase de publicación de nuevas herramientas en el entorno Grid. Estas nuevas herramientas pueden ser microscopios electrónicos, computadoras de alta capacidad, etc.
4. Pruebas de calidad de servicio: Generalmente las aplicaciones que se encuentran disponibles en el Grid, son de alta disponibilidad o de tiempo real, esto requiere un uso intensivo de los equipamientos y de los recursos donde estas son ejecutadas, por ejemplo la visualización por el microscopio electrónico en forma remota. El objetivo de esta actividad es poder realizar las optimizaciones para calidad de servicios y la aplicación de políticas para el acceso y publicación de los servicios en la Grid.
5. Extensión a otras Instituciones: Cuando se complete la etapa anterior se habrán cumplimentado la instalación y optimización contando con una variedad de alternativas en el entorno Grid . El paso siguiente es extender este tipo de soluciones a otras instituciones y así enriquecer los servicios ofrecidos.

Actualmente nos encontramos entre los puntos primero y segundo. Se han instalado 3 clusters dedicados con el software Oscar 3 y Oscar 4[8], con distribuciones Red Hat 9 y Fedora Core 2 respectivamente, las librerías por defecto instaladas son mpich 1.2.5[9] y schedulers OpenPBS[10]. Estos clusters se encuentran monitoreados con el software Ganglia[11] y este se puede acceder desde internet en <http://grid.uncoma.edu.ar:8090/ganglia>. Para el uso de estos clusters se están codificando aplicaciones de test y en lo que se refiere a Grid se ha instalado Globus 3.9.5[12] aún en configuración, esta es una versión beta del release GT 4.0 que tiene como plazo estimado para salida a producción fines de Abril del corriente año.

4. Conclusiones

En este documento se ha presentado una de las líneas de trabajo sobre tecnologías Grid en nuestro país y si bien es prematuro la elaboración de conclusiones, se puede observar que algunas de las dificultades con la que nos encontramos es la disponibilidad de aplicaciones que se pueden adaptar a este tipo de entornos y la limitada infraestructura de comunicaciones con garantías en la calidad del servicio.

Como trabajos a cumplir en forma próxima se encuentran la incorporación de equipamiento no dedicado como son la mayoría de la máquinas en el CRIBABB o la aulas de computadoras en la UNC, para las que se está evaluando el uso de Condor[13]. Otra de las tareas pendientes es la publicación de los servicios Grid ofrecidos por este entorno a través de un portal Grid, algunos de los que esta en evaluación son GridSphere Portal[14], Gridport[15] o el de uso general Jetspeed[16].

5. Bibliografía

- [1] I. Foster, C. Kesselman, S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International J. Supercomputer Applications*, 15(3), 2001.
- [2] Foster, I., Kesselman, C., Nick, J. and Tuecke, S. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. 2002.
- [3] Chase Phillips, Andrei Reinhorn, Hatem A Seliem, Benson Shing, Eric Stauffer, Bozidar Stojadinovic, Kesselman, C., Pearlman, L. and Mehta, G. Design for NEESgrid Telepresence Referral and Streaming Data Services. NEESgrid, Technical Report NEESgrid-2003-09, 2003.
- [4] M. Ellisman and S. Peltier, "Medical Data Federation: The Biomedical Informatics Research Network," in *The Grid: Blueprint for a New Computing Infrastructure (2nd Edition)*, I. Foster and C. Kesselman, Eds.: Morgan Kaufmann, 2004.
- [5] Carole Goble, Stephen Pettifer, Robert Stevens, and Chris Greenhalgh. Knowledge Integration: *In Silico* Experiments in Bioinformatics, 2003.
- [6] Proyecto Alfagrid: <http://grid.ifca.unican.es/alfagrid/>
- [7] Proyecto Red Clara: <http://www.redclara.net/>
- [8] <http://oscar.openclustergroup.org/tiki-index.php>
- [9] A Portable Implementation of MPI: <http://www-unix.mcs.anl.gov/mpi/mpich/>
- [10] OpenPBS Project, A Batching Queuing System, Software Project, Altair Grid Technologies, LLC, www.openpbs.org
- [11] Massie, M., B. Chun, and D. Culler. The Ganglia Distributed Monitoring: Design, Implementation, and Experience. in *Parallel Computing*. May 2004.
- [12] The Globus Project: <http://www.globus.org>
- [13] Livny, M., Tannenbaum T., Thain, D. Condor and the Grid, in Fran Berman, Anthony J.G. Hey, Geoffrey Fox, editors, *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley, 2003.
- [14] GridLab. The gridsphere portal. <http://www.gridsphere.org/gridsphere/gridsphere>.
- [15] Thomas, M. P., Mock, S., Dahan, M., Mueller, K., Sutton, D., The GridPort Toolkit: a System for Building Grid Portals. *Proceedings of the Tenth IEEE International Symposium on High Performance Distributed Computing*, August, 2001.
- [16] Apache Jetspeed open-source portal serve: <http://jakarta.apache.org/jetspeed>