Entorno de supercomputación Grid para aplicaciones de altas demandas computacionales

A Grid Supercomputing Environment for High Demand Computational Applications

I. Barcena, J. A. Becerra, C. Fernández et al.

Resumen

Este trabajo demuestra que la colaboración entre centros de supercomputación permite resolver problemas más complejos de los que se podrían llevar a cabo con los recursos de cada centro por separado. Se ha demostrado que es posible conseguir una buena escalabilidad de las aplicaciones bajo la premisa de que el cálculo y el tráfico de red puedan ser desacoplados. Un problema real puede ser resuelto utilizando esta infraestructura sin grandes cambios en el código, lo cual resulta una solución más atractiva para el programador de software. Las infraestructuras de red de área extensa de nueva generación han probado su capacidad en términos de fiabilidad y ancho de banda para interconectar de modo seguro y óptimo recursos de computación distribuidos, y en el futuro serán estas infraestructuras los catalizadores para ejecutar nuevas aplicaciones en un entorno informático Grid y distribuido.

Palabras clave: Supercomputación, Grid, altas demandas computacionales.

Summary

This work demonstrates that collaboration between supercomputing facilities can enable more demanding problems to be solved. Scalability has been demonstrated under the premise that computing and network traffic can be decoupled. A real problem can be solved using this infrastructure without major changes in the code, which results in a more attractive solution for the software programmer. New generation wide area network infrastructures have proven their adequacy in terms of reliability and bandwidth to safely and optimally interconnect spread computing resources, and in the future will be the catalyst to execute new applications in a distributed and grid computing environment.

Keywords: Supercomputing, Grid, high demand computacional applications.

La mayor parte de los proyectos relativos al Grid se centran en la ejecución de tareas secuenciales

1.- Introducción

Los centros de supercomputación ofrecen recursos informáticos de alto rendimiento enfocados hacia la resolución de los problemas informáticos más complejos (los llamados "grandes retos computacionales"). Sin embargo, muchos problemas de cálculo en diferentes áreas no han podido aún ser resueltos de una manera práctica, debido a que requieren tiempos computacionales extremadamente largos o por falta de recursos computacionales como la memoria disponible o el almacenamiento permanente de datos.

Como en otras muchas áreas, la cooperación y la unión de recursos entre diferentes instituciones constituye la única solución real para proporcionar capacidades suficientes que permitan resolver estos problemas en la práctica. La computación Grid establece algunos de los mecanismos necesarios para conseguir este tipo de unión de potencia computacional.

En los diferentes proyectos Grid han aparecido muchos testbeds diferentes, cuya función es experimentar y explotar esta tecnología emergente. Sin embargo, la mayor parte de los proyectos relativos al Grid se centran en la ejecución de tareas secuenciales, lo que ha sido denominado como "alta productividad" (o High Throughput Computing, HTC), como es el caso del proyecto europeo Datagrid, mientras que existen pocos proyectos enfocados hacia la ejecución de tareas paralelas utilizando tecnologías Grid. Algunos ejemplos de este tipo de experiencias son el entorno de

Seleccionamos un problema con comunicaciones poco frecuentes o en el cual la latencia de red no fuera un problema real

simulación Cactus y el proyecto europeo Crossgrid. Estas aplicaciones paralelas pueden beneficiarse especialmente de los últimos desarrollos en redes de área extensa en cuanto a ancho de banda y fiabilidad, así como de la reciente introducción de políticas de calidad de servicio en estas infraestructuras.

2.- Objetivos

Los principales objetivos de esta experiencia son:

1.- Establecer un entorno de computación Grid entre dos centros de supercomputación independientes.
2.- Demostrar cómo las aplicaciones secuenciales pueden ser utilizadas dentro de este entorno Grid.
3.- Demostrar que los códigos paralelos pueden también ser ejecutados dentro de este entorno Grid.
4.- Determinar de qué manera la heterogeneidad de sistemas puede afectar a la eficiencia de la computación paralela.
5.- Analizar de qué manera la heterogeneidad de la red y de los equipos de comunicación pueden influir en este tipo de problemas y determinar la fiabilidad, disponibilidad y capacidad para prestar servicios (RAS) de la nueva generación de redes de investigación en áreas extensas.

Tras todos estos importantes objetivos subyace un interés común por proporcionar recursos de cálculo dedicados y altamente eficientes que ayuden a resolver problemas científicos que requieran las mayores capacidades computacionales disponibles.

3.- Problema y aplicaciones

Para esta experiencia queríamos realizar una simulación con altas necesidades computacionales y de memoria y que no se pudiese realizar por separado por ninguno de los dos centros.

Seleccionamos un problema con comunicaciones poco frecuentes o en el cual la latencia de red no fuera un problema real. Debemos tener en cuenta que a pesar de las crecientes capacidades de las nuevas redes de área extensa en términos de ancho de banda, la latencia se mantiene como un obstáculo ya que los enlaces ópticos tienen, hasta ahora, el límite impuesto por la velocidad de la luz. En nuestro caso, hay más de 1.200 kilómetros de fibra óptica separando el CESGA y el CESCA, y las señales luminosas necesitan más de 8 milisegundos para un viaje de ida y vuelta completo. Debido a la conversión electro-óptica y al enrrutado de los paquetes, el tiempo medido de un viaje completo de ida y vuelta es casi dos veces mayor, del orden de 20 a 30 ms. Esta latencia es más de 1.000 veces mayor que las redes de interconexión de sistemas paralelos de alta velocidad y baja latencia como Quadrics, Myrinet, o Memory-Channel.

Una vez analizadas las características de las aplicaciones de los usuarios de ambos centros de supercomputación, una aplicación propia de Inteligencia Artificial, SEVEN, fue seleccionada como el código más adecuado de acuerdo con las características que se requerían para este experimento:

- Se trata de una aplicación paralela ampliamente testeada y de la cual los usuarios finales tienen un profundo conocimiento acerca de su estructura interna.
- El problema resuelto por esta aplicación precisa recursos de computación de altas prestaciones en términos de potencia de procesamiento y memoria, y puede ser fácilmente redimensionada para solventar problemas de diferente envergadura.

- A pesar de que en cada paso de la simulación se deben transmitir decenas de megabytes de datos, estas transmisiones sólo tienen lugar después de una significativa cantidad de tiempo de procesamiento.

El objetivo de la simulación es obtener un controlador para un robot autónomo Pioneer 2-DX. El propósito de este controlador es dotar al robot de un comportamiento dado

4.- Objetivo de la simulación

El objetivo de la simulación es obtener un controlador para un robot autónomo Pioneer 2-DX. El propósito de este controlador es dotar al robot de un comportamiento dado: este robot (R) está situado en una habitación cuadrada con otros dos robots. Uno de esos dos robots es un perseguidor y el otro es una presa. El robot R debe escapar del perseguidor y seguir a la presa. El perseguidor y la presa no tienen ninguna diferencia especial salvo su velocidad de movimiento, así que la información temporal es necesaria para distinguir a los dos robots, porque sólo pueden ser diferenciados controlando el modo en el que se mueven. El controlador debe ayudar a este comportamiento autónomo con la información facilitada por los sensores del sónar del robot R.

5.- Configuración Grid: sistemas y redes

Para esta prueba, se utilizó el superordenador más potente disponible en cada centro: 2 Hewlett Packard HPC320. En la tabla adjunta aparecen reflejadas las principales características de ambos servidores. Estos dos superordenadores unidos ofrecen un pico de potencia de 117,31 Gflops, 85 Gflops sostenidos, 64 procesadores, 100 Gbytes de memoria y 4 Terabytes de almacenamiento en disco.

	Sistema Operativo	Nodos	Procesadores	Memoria	GFlops	Conexion Internet
CESGA	Tru64 v5.1A	8	32 Alpha EV68@1GHZ	80 Gbytes	64	Gb y Fast-Ethernet
CESCA	Tru64 v5.1A	8	32 Alpha EV68@833MHz	20 Gbytes	53	Fast-Ethernet

El CESGA se ubica en el noroeste de España y el CESCA en el noreste. La red nacional de investigación, RedIRIS2, ha mejorado recientemente su infraestructura incorporando velocidades Multigigabit y alta disponibilidad en todos sus puntos presenciales en el país. En esta red mallada hay más de diez rutas diferentes para la conexión entre el CESGA y el CESCA. Dependiendo de la carga y la disponibilidad de los enlaces, el tráfico en la red puede discurrir a través de diferentes saltos intermedios. Sin embargo, dada la carga actual y la potencia requerida para esta prueba, se determinó que ir a través de Madrid implicaba un menor número de saltos y por tanto la conexión con menor latencia. Esta línea tiene 1.200 kms. de largo y la media observada de los tiempos de ida y vuelta de paquetes es de 20 ms. (paquetes ICMP de 64 bytes). Los enlaces de alto ancho de banda podrían proporcionar un total de más de 5 Gbps o más de 600Mbytes/s, equivalente o incluso superior a cualquier otra interconexión local.

6.- Componentes software

Dos componentes fundamentales fueron utilizados para esta simulación: Globus (versión 2.2.4) y MPICH-G2 (versión 1.2.5-1a). Además, teniendo en cuenta que ambos sistemas estaban en producción mientras se desarrollaban los primeros tests, fue necesario instalar los correspondientes job managers de Globus para cada centro: la versión "Pro" del Portable Batch System (PBS Pro v. 5.2) en el caso del CESGA, y el Load Sharing Facility (LSF Base version 5.0) en el CESCA. También se utilizó la Autoridad de Certificación de la Redegrid para los certificados de usuarios y sistemas necesarios en Globus.

El diagrama muestra que los nodos del CESGA tienen mejor eficiencia CPU

7.- Ejecución del código

El código fue enviado con un script RSL desde el sistema del CESGA, distribuyendo 32 trabajos en cada centro. Una vez iniciados todos los procesos, la simulación comenzó y transcurrió durante más de 9 horas. A continuación describimos y representamos algunas de las mediciones recogidas.

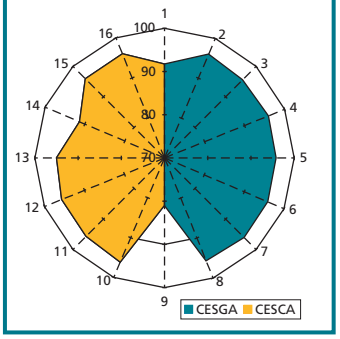
7.1.- Rendimiento observado

Definimos la eficiencia como el ratio del tiempo de cálculo empleado por la aplicación y el tiempo total. La eficiencia global conseguida durante el proceso fue de 94.3%. El diagrama kiviatt representa la eficiencia lograda por cada nodo individual de los sistemas HPC320. Del 1 al 8 pertenecen al HPC320 del CESGA y del 9 al 16 pertenecen al HPC320 del CESCA. El diagrama muestra que los nodos del CESGA tienen mejor eficiencia CPU (95.3% *vs.* 93.2%). Los menores rendimientos del CESCA son principalmente debidos a la falta de memoria y a la posible falta de páginas de memoria disponibles.

Cada paso o generación empleó 17 minutos en el CESGA y 22 en el CESCA. Esta proporción en la cual los nodos del CESCA son 1.29 veces más lentos que los del CESGA es equivalente al ratio entre las velocidades de los procesadores (833Mhz vs. 1000MHz) y a las diferentes eficiencias conseguidas en cada sistema.

El tráfico total transferido entre los dos superordenadores fue de 8.3 Gbytes, dividido en 4.6 Gbytes en el camino desde el CESGA al CESCA y 3.7 Gbytes en el camino contrario. Cada proceso necesita 963 Mbytes de memoria virtual, y la memoria residente observada por proceso fue de 300 Mbytes en el CESCA y 600 en el CESGA (debemos señalar cómo la falta de memoria pudo afectar negativamente a la potencia de los nodos en el CESCA).

EFICIENCIA POR NODO



Agradecimientos

Nos gustaría agradecer a Hewlett Packard el facilitarnos las plataformas necesarias para las pruebas iniciales y por su continuo apoyo durante esta experiencia. Así como a la Red nacional de investigación española, RedIRIS, por su cooperación y esfuerzos realizados proporcionando una infraestructura de red de alto rendimiento y fiabilidad para desarrollar este trabajo.

Ingrid Barcena, Joan Cambras, Caterina Parals
(ibarcena@cesca.es), (jcambras@cesca.es), (cparals@cesca.es)
Centre de Supercomputació de Catalunya (CESCA)
José Antonio Becerra, Richard Duro
(ronin@udc.es), (richard@udc.es)
Grupo de Sistemas Autónomos (GSA), (UDC)
Carlos Fernández, Javier Fontán, Juan Villasuso
(carlosf@cesga.es), (jfontan@cesga.es), (jvilla@cesga.es)
Andrés Gómez, Ignacio López, José Carlos Pérez
(agomez@cesga.es), (nlopez@cesga.es), (jcarlos@cesga.es)
Centro de Supercomputación de Galicia (CESGA)