

Hacia la Aplicación de la Computación de Alto Desempeño al Entorno Productivo Local

Ricardo Medel, César Martínez Spessot, Julio Castillo

Laboratorio de Investigación de Software
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Córdoba
Universidad Tecnológica Nacional

Maestro M. Lopez esq. Cruz Roja Argentina
Ciudad Universitaria – Córdoba
{ricardo.h.medel,cspessot,jotacastillo}@gmail.com

Resumen

Los clústeres de computadoras para Computación de Alto Desempeño (HPC) permiten acceder a relativamente bajo costo a herramientas que aceleran los tiempos de exploración, diseño, testeo e introducción al mercado de nuevos productos. Sin embargo, la falta de conocimiento técnico clave está limitando la difusión de dichas herramientas en nuestra industria. El objetivo de nuestra línea de I+D es desarrollar una estrategia para obtener y difundir el conocimiento requerido para adquirir, instalar, administrar, mantener y programar un clúster de HPC.

Palabras clave: Programación de Alto Desempeño, Programación Paralela, Programación Distribuida.

Contexto

La línea de I+D presentada aquí se desarrolla actualmente en el marco del proyecto

de investigación “*Desarrollo de un Plan de Negocios basado en una Solución de Aprovechamiento y Mantenimiento para Clústeres de Computación de Altas Prestaciones*”, dirigido por el Dr. Ricardo Medel y desarrollado en el Laboratorio de Investigación de Software del Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional. El proyecto ha sido aprobado para ser desarrollado entre los meses de Mayo de 2010 y Abril de 2012, y es financiado por la Secretaría de Ciencia y Tecnología de la UTN bajo el código UTN1230.

Un proyecto anterior también es considerado parte de esta línea de I+D, al atacar el problema de la generación y distribución de conocimiento en una comunidad, aplicable a la línea 3 (ver sección *Líneas de investigación y desarrollo*) para el caso de compartir conocimiento sobre la instalación de clústeres de HPC. Dicho proyecto, denominado “*Definición de una estructura estándar para recetas y desarrollo de herramientas para automatizar la creación y*

administración colaborativa de recetas”, dirigido por el Dr. Ricardo Medel, fue desarrollado en la misma institución entre Enero de 2009 y Diciembre de 2010, y financiado por la Sec. de Ciencia y Tecnología de la UTN bajo el código EIUTICO969. En el marco de dicho proyecto, la empresa Intel Software de Argentina S.A. (filial argentina de Intel Corporation) financió una pasantía de 18 meses de duración.

Introducción

Con el término Computación de Alto Desempeño (HPC, por sus siglas en inglés) se describe al uso de computadoras en problemas que requieren una gran capacidad de procesamiento de información. Generalmente involucra usar un conjunto de recursos de computación paralelos que cooperan para solucionar un problema en el cual intervienen gran cantidad de datos y cálculos [L04, LW95, B99]. Esta área se encuentra vinculada con la investigación científica, pero también con el desarrollo de productos innovadores [CWG09, GD09].

Un clúster de HPC es un conjunto de computadoras conectadas por una red local, que se comporta como si fuera una única computadora a través de software especializado. Un clúster bien configurado tiene un rendimiento similar al de una supercomputadora mucho más cara [C09, GD09, L04]. Los clústeres representan el 82% de las computadoras listadas en el ranking de las 500 computadoras con mejor performance del mundo¹. Gracias a la accesibilidad a HPC provista por los clústeres, pequeños centros de I+D y empresas pueden ahora utilizar HPC para acelerar los tiempos de

exploración, diseño, testeo e introducción al mercado de nuevos productos.

Sin embargo, a pesar de su bajo costo inicial y por estar formado por componentes distribuidos y de diverso origen, instalar y mantener un clúster involucra altos costo de planificación, instalación, configuración, infraestructura, mantenimiento de sistema operativos y aplicaciones, administración de redes y unidades de procesamiento, entre otras [MMM08, C+08]. En particular, la escasez de conocimiento específico para la instalación y mantenimiento de clústeres para HPC ha sido un obstáculo para que el crecimiento mundial de uso de clústeres se viera reflejado en igual medida en países en desarrollo, lo que podrían aprovechar el bajo costo inicial para potenciar su desarrollo industrial.

Es por eso que nuestro proyecto apunta a aumentar la automatización del software de instalación y mantenimiento de clústeres para HPC con miras al mercado local, dado que la adaptación y mantenimiento de un software es una actividad netamente local, dependiente de la idiosincracia del mercado regional y de los factores locales que lo afectan.

Nuestra línea de I+D pretende desarrollar una estrategia de adquisición y difusión de los conocimientos requeridos para instalar y mantener clústeres para HPC. En particular, los resultados esperados son:

- Un sistema de aprovisionamiento (instalación y mantenimiento) de clústeres adaptado al mercado local.
- Un estudio de mercado que responda a las principales preguntas respecto del mercado local de provisión de servicios para clústeres.

¹ <http://www.top500.org>

- Definición de estrategias de adquisición y difusión de conocimientos técnicos en adquisición, instalación, mantenimiento y programación de clústeres de HPC.

- Un sistema informático de construcción cooperativa de conocimiento, aplicable a la creación de manuales de instalación de clústers.

- Materiales educativos para la difusión del conocimiento técnico adquirido.

Líneas de investigación y desarrollo

Actualmente se siguen activamente dos líneas de investigación y se espera comenzar una tercera en el futuro cercano.

1. **Factibilidad Económica:** Estudio de la factibilidad económica de desarrollos de Computación de Alto Desempeño en el contexto local. En este tema se ha desarrollado un plan de negocios para un sistema de aprovisionamiento para clústeres de HPC.

2. **Programación de Sistemas Paralelos:** Estudio de las complejidades de la programación de sistemas paralelos y distribuidos, análisis de las herramientas de desarrollo y propuestas de enseñanza y desarrollo en el área.

3. **Instalación y Administración:** Análisis de los conocimientos requeridos para la instalación y administración de clústeres de HPC en PyMES locales y desarrollo de estrategias para facilitar la adquisición de dichos conocimientos.

Resultados y Objetivos

Además de los importantes resultados obtenidos en formación de recursos humanos

(ver siguiente sección), los proyectos que pertenecen a esta línea de I+D han producido los siguientes resultados:

- Se diseñó un estándar para “recetas”, documentos que permiten compartir información técnica y procedimientos entre pares de una comunidad.

- Se desarrolló el sistema Fresita, que implementa el estándar mencionado anteriormente como documentos XML y permite crear, modificar y distribuir recetas en una comunidad a través de la web. Disponible bajo licencia GNU GPL v3.

<http://code.google.com/p/fresita/>

- Se publicaron un Reporte Técnico [MM09], un artículo en una conferencia internacional [M+08] y se realizaron dos presentaciones en universidades nacionales [S07, M07].

- Se desarrolló un plan de negocios para la creación de una empresa basada en el servicio de instalación y mantenimiento de clústeres para Computación de Alto Desempeño [M11].

- Se instaló un clúster pequeño basado en Windows HPC Server y se realizaron estudios comparativos de performance entre sistemas basados en MPI y en OpenMP en un ambiente .Net. Como resultados de estos experimentos se publicó un Reporte Técnico [CSM11].

Los objetivos en curso y a futuro son adquirir el material necesario e instalar un clúster de un tamaño mayor (al menos 4 nodos multi-núcleo) y realizar experimentos comparativos de instalación bajo entornos Linux y Windows. El objetivo de estos experimentos será analizar los conocimientos requeridos para la instalación y administración de clústeres y plantear una estrategia de adquisición y

desarrollo de dichos requerimientos en el ámbito local.

Asimismo, se profundizarán los experimentos comparativos entre distintos ambientes de programación paralela y distribuida, a fin de definir una estrategia de entrenamiento de desarrolladores.

Formación de Recursos Humanos

- Formación de RR.HH. en el marco del proyecto **EIUTICO969 – Definición de una estructura estándar para recetas y desarrollo de herramientas para automatizar la creación y administración colaborativa de recetas.**

Integrantes: El proyecto fue dirigido por el Dr. Ricardo Medel y sus integrantes fueron: Mag. César Martínez Spessot, Ing. Martín Farías, Ing. Luis Mariano Guerra, Ing. Javier Dall'Amore, Sr. Diego Sarmentero y Sr. Mateo Bengualid.

Tesina de Grado: Los Ingenieros Farías, Guerra y Dall'Amore realizaron su Proyecto Final de carrera de Ingeniería en Sistemas de Información en el marco de este proyecto, desarrollando "Fresita", un sistema informático para la creación y manipulación comunitaria de recetas, en sus versiones stand-alone y web, que fuera aprobado en 2010.

Pasantías y Becas: El Ing. Farías, en calidad de alumno, realizó una pasantía de 18 meses en la empresa Intel Software de Argentina S.A. (filial de Intel Corporation) estudiando la factibilidad de utilizar el sistema Fresita para el desarrollo de Diseños de Referencia, los cuales permiten a administradores de clústeres de HPC instalar un cluster que pueda ser certificado com

"Intel Cluster Ready" con mínimos conocimientos del tema.

El Ing. Farías recibió una Beca de Investigación SAE (Secretaría de Asuntos Estudiantiles) para participar del proyecto durante el año 2009. El Sr. Bengualid recibió una beca SAE durante 2010, mientras que el mismo año el Sr. Sarmentero recibió una beca de investigación con fondos del Rectorado UTN y el Ing. Dall'Amore recibió una beca BINID de Iniciación en Investigación y Desarrollo para Jóvenes Graduados Universitarios.

- Formación de RR.HH. en el marco del proyecto **UTN1230 – Desarrollo de un Plan de Negocios basado en una Solución de Aprovisionamiento y Mantenimiento para Clústeres de Computación de Altas Prestaciones IUTICO969.**

Integrantes: El proyecto es dirigido por el Dr. Ricardo Medel y sus integrantes son: Ing. León Shocrón Benmuyal, Mag. César Martínez Spessot, Ing. Emanuel Cuevas, Lic. Julio Castillo, Ing. Diego Serrano y Sr. Vladimir Tula. Integrantes en el pasado han sido: Ing. Martín Farías, Sr. Diego Martínez, Sr. Ignacio Giagante, Sr. Néstor Valentinis, Sr. Leandro Molina y Sr. Rodrigo Martínez.

Tesis de Posgrado: El Mag. César Martínez Spessot, co-dirigido por el Dr. Medel y el Ing. Shocrón Benmuyal, realizó su tesis de Maestría en Administración de Negocios (en la Escuela de Posgrado de la UTN Fac. Reg. Córdoba) en el marco de este proyecto. La tesis se titula "*Desarrollo de un plan estragégico de negocios sobre la solución de instalación de clústeres de computación de alta performance*" y fue defendida con éxito en Diciembre de 2011.

Pasantías y Becas: El Sr. Rodrigo Martínez recibió una Beca de Investigación SAE

(Secretaría de Asuntos Estudiantiles) para participar del proyecto durante el año 2010, mientras que el Sr. Vladimir Tula recibió una beca similar en 2011. El Sr. Leandro Molina realizó actividades de investigación en el proyecto como parte de su Plan de Trabajo de la Beca TICs (Beca de Fin de Carrera para Estudiantes de Grado en Carreras TICs) de la Agencia Nacional de Promoción Científica durante el año 2011.

Referencias

- [B99] Buyya (Ed.), “*High Performance Cluster Computing - Vol.1: Architectures and Systems, Vol.2: Programming and Applications*”, Prentice Hall, 1999.
- [CSM11] Castillo, Serrano, Molina, “*Comparación entre los modelos de programación paralela MPI y OpenMP*”, Reporte Técnico, Laboratorio de Investigación de Software, UTN-FRC, 2011.
- [CWG09] Chorley, Walker, Guest, “*Hybrid Message-Passing and Shared-Memory Programming in a Molecular Dynamics Application On Multicore Clusters*”, 2009.
- [C09] Conway, “*Energy Profiling and Analysis of the HPC Challenge Benchmarks*”, International Journal of High Performance Computing Applications, 2009.
- [C+08] Conway, Walsh, Joseph, Wu, “*Intel Cluster Ready Program Builds HPC Momentum*”, IDC, 2008.
- [GD09] Gai, DeSanti, “*I/O Consolidation in the Data Center: A Complete Guide to Data Center Ethernet and Fibre Channel over Ethernet*”, Cisco Press, 2009.
- [LW95] Lenoski, Weber, “*Scalable Shared-Memory Multiprocessing*”, Morgan Kaufmann Publisher, San Francisco, 1995.
- [L04] Lucke, “*Building Clustered Linux Systems*”, Prentice Hall, 2004.
- [M11] Martínez Spessot, “*Desarrollo de un plan estratégico de negocios sobre la solución de instalación de clústeres de computación de alta performance*”, Tesis de Maestría en Administración de Negocios, Escuela de Posgrado, UTN-FRC, Diciembre 2011.
- [MM09] Martínez Spessot, Medel, “*Una propuesta de formalización del entorno Fresita*”, Reporte Técnico, Laboratorio de Investigación de Software, UTN-FRC, 2009.
- [M07] Medel, “*Creación y Mantenimiento Comunitario y Distribuido de HOWTOs*”, Seminario Técnico, Dpto. de Computación, Universidad Nacional de Río Cuarto, Noviembre de 2007.
- [M+08] Medel, Farías, Guerra, Sarmentero, Martínez Spessot, “*Fresita: una herramienta distribuida para la creación comunitaria de recetas*”, 34ª Conferencia Latinoamericana de Informática, CLEI 2008, Santa Fe (Argentina), 8-12 Sep 2008.
- [MMM08] Medel, Martínez Spessot, McMillan, “*Development of a Verification Tool for HPC Clusters*”, HPC Symposium 2008, 37 JAIIO – Jornadas Argentinas de Informática, Santa Fe, (Argentina), Agosto de 2008.
- [S07] Sarmentero, “*Fresita: A distributed and collaborative documentation management project*”, Seminario Técnico, ASDC-Intel, Córdoba, 2007.