

# El conocimiento comunitario como herramienta para difundir la aplicación de la Computación de Alto Desempeño en el entorno productivo local

Martínez Spessot, César - Martínez Spessot, Gustavo - Medel, Ricardo  
Quispe, Marcelo – Serrano, Diego

*Departamento de Ingeniería en Sistemas de Información  
Universidad Tecnológica Nacional – Facultad Regional Córdoba*

## Abstract

*Los clústeres de computadoras para Computación de Alto Desempeño (HPC, por sus siglas en inglés) permiten a diversas industrias acceder a relativamente bajo costo a poderosas herramientas computacionales que reducen los tiempos de exploración, diseño y testeo de nuevos productos y así aceleran su introducción en el mercado, mejorando la competitividad de las empresas. Sin embargo, la falta de conocimiento técnico clave ha limitado la rápida adopción de dichas herramientas en la industria argentina. Por otro lado, el uso de las herramientas de trabajo colaborativo ha crecido exponencialmente debido al mayor acceso a internet. En este trabajo se presenta un enfoque basado en herramientas colaborativas a través de internet para generar y difundir el conocimiento técnico requerido para adquirir, instalar, mantener y programar clústeres de HPC.*

## Palabras Clave

HPC, Herramientas colaborativas, Computación de Alto Desempeño, Sistemas de aprovisionamiento, Clúster, Programación paralela

## Introducción

Con el término Computación de Alto Desempeño (HPC, por sus siglas en inglés: *High Performance Computing*) se define al uso de computadoras en problemas que requieren una gran capacidad de procesamiento de información. Generalmente esto 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 [1, 2, 3]. Esta área se encuentra vinculada con la investigación científica, con aplicaciones en Física, Química y Medicina, pero también con el desarrollo de productos innovadores

tanto en industrias tradicionales, tales como la automotriz y la aeronáutica, como en las nuevas industrias basadas en la biotecnología y la nanotecnología [4, 5].

Un clúster de HPC es un conjunto de computadoras conectadas por una red local de alta velocidad, que se comporta como si fuera una única computadora paralela, que coordina el trabajo de los “nodos” o computadoras individuales a través de software especializado. Un clúster bien configurado tiene un rendimiento similar al de una supercomputadora mucho más cara [1, 4, 6]. Actualmente los clústeres representan más del 90% de las computadoras listadas en el ranking de las 500 computadoras con mejor performance del mundo [7]. Dado que los nodos de un clúster pueden ser computadoras personales (PC) o servidores de bajo costo, 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, por estar formado por componentes de diverso origen, el proceso de adquirir, instalar y mantener un clúster involucra altos costo de planificación, configuración, infraestructura, mantenimiento de sistema operativos y aplicaciones, administración de redes y unidades de procesamiento, entre otras tareas necesarias [8, 9]. 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 [10], justamente aquellos que podrían aprovechar el bajo costo inicial para potenciar su desarrollo industrial.

El objetivo de nuestro grupo de investigación es colaborar en la difusión de la tecnología de HPC en el entorno productivo local, desarrollando una estrategia para obtener, generar y difundir el conocimiento requerido para utilizar clústeres de HPC. Para lograrlo, hemos realizado estudios a fin de responder a las principales preguntas respecto de las barreras de adopción de clústeres de HPC, comparar diferentes sistemas de clústeres, analizando cuáles son los más simples de aprender para el mercado, y cómo afecta el costo de estos ambientes (en cuanto a material y a conocimiento necesario) a su adopción en la industria, y hemos definido estrategias de adquisición y difusión de conocimientos técnicos requeridos para la compra, instalación y mantenimiento de clústeres de HPC.

En este trabajo presentamos algunos resultados de estos estudios y la experimentación realizada con respecto al uso de sistemas informáticos de construcción cooperativa de conocimiento o “groupware” [11] en la creación y mantenimiento comunitario de documentación sobre la instalación y mantenimiento de clústeres.

### **Elementos del trabajo y metodología**

A fin de validar nuestra hipótesis de que la falta de conocimientos técnicos es una importante barrera para la adopción de HPC en la industria, se utilizaron los resultados obtenidos de la realización de una encuesta cuantitativa a través del sitio web *The Register* [12], el cual permite utilizar su motor de generación de encuestas y su base de datos para realizar estudios de este tipo. La encuesta, realizada durante octubre y noviembre de 2009 a 254 profesionales de Tecnologías de la Información (TI) de todo el mundo, que se encontraban trabajando con tecnologías de HPC, fue respondida en

forma anónima y cubrió varios aspectos de la instalación y uso de sistemas de HPC, más allá de los utilizados en nuestro proyecto y reportados en este trabajo. Más detalles de la encuesta, cómo se seleccionó la muestra, cuáles fueron las preguntas y qué resultados se obtuvieron pueden verse en la Tesis de Maestría en Administración de Negocios de C. Martínez Spessot [13].

En cuanto a la comparación de los diversos ambientes para clústeres de HPC, se utilizaron computadoras personales en desuso y nuevas para construir un clúster HPC y se documentaron los pasos requeridos para la instalación tanto del hardware como del software en dos ambientes distintos: Windows HPC Server y Linux. Se seleccionaron estos ambientes por ser los dos ambientes que corresponden a prácticamente el total de los sistemas de HPC en funcionamiento en la actualidad, con Linux comprendiendo un 80% de los sistemas y Windows el resto [14]. Los documentos generados, con los pasos necesarios para instalar los clústeres fueron luego volcados a una herramienta de trabajo en grupo, denominada Fresita [11], que permite administrar comunitariamente dichos documentos, permitiéndoles crecer y adaptarse según los conocimientos y necesidades de una comunidad de usuarios.

### **Resultados**

Dos resultados de la encuesta realizada a los 254 profesionales de TI nos indican el costo que el conocimiento de los sistemas de HPC representa para las empresas que quieren adoptar esta tecnología. La Figura 1 muestra las respuestas a la pregunta: “*Más allá de las necesidades de servidores. ¿Cuáles de estos factores cree que requiere de mayor inversión cuando se implementa HPC?*” El color más a la izquierda de cada barra indica una mayor inversión requerida. Puede verse que los “servicios profesionales” (consultoría de expertos, etc.) requieren menos inversión y que “experiencia y conocimiento de bajo nivel” queda en una tercera posición, superadas

ambas por los sistemas de almacenamiento y las redes. Sin embargo, sumando ambos factores (servicios profesionales y experiencia, ambos basados en el conocimiento de los profesionales) se obtiene un costo de inversión similar al mayor de los costos (redes).

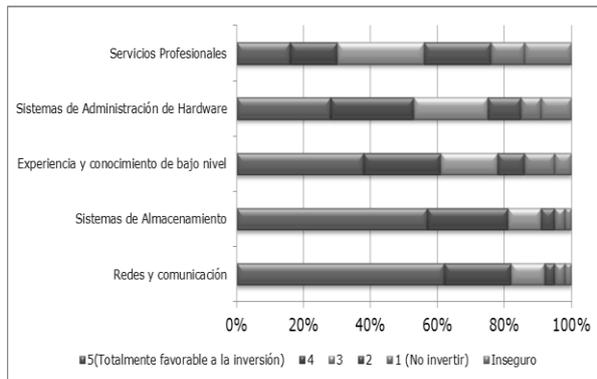


Fig. 1. Respuestas a la pregunta sobre factores que requieren mayor inversión

Asimismo, las respuestas a la pregunta “Basada en su experiencia y conocimiento actual. ¿Cree usted que las organizaciones del futuro podrán usar el conocimiento genérico de sus profesionales TI para implementar una solución de HPC?” se muestran en la Figura 2 e indican que se requiere conocimiento y entrenamiento específico para instalar soluciones HPC.

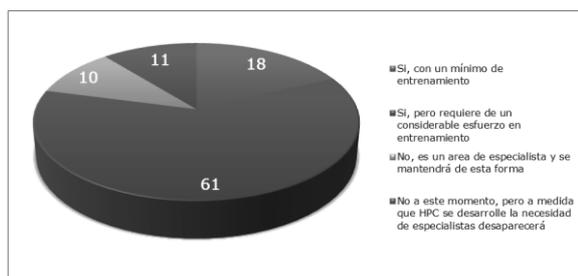


Fig. 2. Respuestas a la pregunta sobre el conocimiento necesario para implementar HPC.

En la Figura 2 se indica que el 61% de los encuestados opinan que se requiere un “considerable esfuerzo de entrenamiento” y que el 10% opina que es un área de especialistas y que ello no va a cambiar.

Solo un 18% opina que con mínimo entrenamiento los profesionales generalistas de TI pueden implementar una solución de HPC.

Con estos resultados avalando nuestra opinión, se realizaron experimentos de instalación de clústeres en los ambientes Windows HPC Server y Linux, en los que dos alumnos realizaban las tareas con mínimo entrenamiento y registraban todos los pasos del proceso en un documento. Esos documentos fueron luego pasados al formato de la herramienta Fresita, que permite trabajar cooperativamente en línea, en una forma distribuida.

<receta>	<ingrediente>
<intro>	<id/>
<id/>	<version/>
<version/>	<titulo/>
<titulo/>	<descripcion>
<autores>	TEXTO
<autor>	</descripcion>
<nombre/>	<autores>
<apellido/>	<autor>
<email/>	<nombre/>
</autor>	<apellido/>
</autores>	<email/>
<revisiones/>	</autor>
<tags>	</authors>
<tag/>	<revisiones/>
</tags>	<pasos>
</intro>	<paso>
<contenido>	<titulo/>
<abstract/>	<contenido>
<indice/>	TEXTO
<capitulo/>	</content>
<orden/>	</paso>
<nombre/>	</pasos>
<ingredientes>	<tags>
<ingrediente/>	<tag/>
</ingredientes>	</tags>
</contenido>	</ingrediente>
</receta>	

Fig. 3. Estructura XML de receta e ingredientes.

Esta herramienta permite crear “recetas”, es decir documentos estructurados de modo de explicar paso a paso cómo realizar alguna

tarea con mínimos conocimientos técnicos. La estructura de los documentos, mostrada en formato XML en la Figura 3, permite definir una serie de *ingredientes* y, para cada ingrediente, una serie de *pasos* para su instalación.

El usuario crea una receta utilizando una interfaz muy simple (ver Figura 4), se le asigna un número de versión y se hace la copia local accesible para todos los usuarios del grupo de colaboración, los cuales, de manera asíncrona pueden ir mejorando la receta o creando nuevas recetas a partir de ésta.

Las recetas, conocidas como documentos *how-to* en inglés, pueden ser utilizadas para compartir conocimiento en diversas disciplinas. En nuestro caso, los documentos generados indicaban paso a paso cómo instalar un clúster de HPC. Dichos documentos fueron luego enviados a un Laboratorio de Física de la Universidad Nacional del Comahue, que estaba interesado en adquirir un clúster y que deseaba evaluar el costo en experiencia y conocimiento que tendría instalarlo con Windows o con Linux. Estos mismos documentos fueron también enviados a la empresa local VENG S.A., proveedora de la CONAE – Comisión Nacional de Actividades Espaciales, para colaborar con la instalación de un clúster recién adquirido.



Fig. 4. Pantalla principal de la herramienta Fresita.

## Discusión

La empresa Intel ha sido pionera en compartir recetas para la instalación de clústeres de HPC. Su programa Intel®

Cluster Ready [15] provee a las empresas asociadas con recetas (documentos en formato PDF y algunos *scripts* en lenguaje Python) que, de ser seguidas paso a paso, les permiten instalar, sin demasiados conocimientos técnicos, un clúster de HPC basado en arquitectura Intel y utilizando ciertas distribuciones Linux específicas. Esto permite a las empresas socias de Intel minimizar las barreras de entrada al campo de HPC y acelerar la llegada al mercado de nuevos clústeres basados en nuevos procesadores de Intel.

Nuestro enfoque ha sido similar, pero con diferencias clave. Por una parte, en vez de tener un “centro de conocimiento” desde donde se provee la información y conocimiento necesarios a todos los otros actores, confiamos en una “red de conocimiento” donde cada actor aporta el conocimiento específico que tiene y entre todos se va creando el conocimiento necesario y que puede ser reutilizado por todos los miembros de la comunidad.

Por otra parte, no se limita la producción de recetas ni a una arquitectura en particular ni a un ambiente. Todo responde a las necesidades de la comunidad.

En nuestros experimentos utilizamos con éxito la herramienta Fresita para documentar y compartir el conocimiento adquirido sobre la instalación de clústeres de computadoras para HPC, tanto bajo entorno Windows como Linux. Los encargados de esta tarea pudieron comunicarse efectivamente a través de esta herramienta, incorporando los cambios en sucesivas versiones de las recetas. Sin embargo, el hecho de trabajar localmente, es decir, compartir un laboratorio, hizo que la comunicación no se diera solamente a través de la herramienta. Es probable que si se desea trabajar con equipos de personas geográficamente dispersas sea necesario incorporar otros medios de comunicación (chat, emails, etc.) a la herramienta.

Los documentos producidos fueron remitidos a una empresa y un laboratorio de una universidad nacional a fin de ser

utilizados en los estudios previos a la instalación de sus clústeres. Sin embargo, no se pudo evaluar su importancia ni su uso por falta de comunicaciones posteriores. Es de mencionar, además, que el uso más adecuado de esta herramienta no es el generar recetas que luego sean compartidas con otras personas, sino involucrar a todos los usuarios en su creación y mantenimiento (modificación, mejora, actualización, etc.).

Los resultados obtenidos son alentadores, sin embargo se requieren más experimentos, en particular estableciendo al menos una comunidad geográficamente distribuida, cuyas comunicaciones sean totalmente en-línea, a fin de establecer conclusiones más firmes sobre la utilidad de herramientas de creación y administración cooperativa de recetas en el ámbito de la Computación de Alto Desempeño.

## Conclusión

Los clústeres de computadoras para HPC son un conjunto de computadoras (PCs o servidores) conectadas por una red local de alta velocidad, que se comporta como si fuera una única computadora paralela. Su relativo bajo costo de hardware y software (generalmente software libre [16]) y su rendimiento similar a una supercomputadora paralela de un costo mucho mayor, permite a pequeños centros de I+D y empresas utilizarlos para acelerar los tiempos de exploración, diseño, testeo e introducción al mercado de nuevos productos, mejorando su competitividad. Sin embargo, la falta de conocimiento técnico clave, tal como lo muestran los resultados de la encuesta realizada, ha limitado la rápida adopción de estas herramientas en la industria. Para subsanar dicho inconveniente proponemos utilizar Fresita, una herramienta de trabajo colaborativo a través de internet que permite generar y difundir conocimiento técnico sobre instalación y mantenimiento de clústeres de HPC.

La herramienta Fresita permite crear “recetas” (documentos XML estructurados que permiten definir y/o seguir un proceso paso a paso) y compartirlas a fin de que una comunidad pueda usarlas y crear versiones mejoradas o adaptadas a nuevas situaciones. La utilización de dicha herramienta en el seno de nuestro grupo de I+D ha producido resultados positivos, y el marcado interés con que se recibieron fuera de nuestro grupo, indican que, aunque se requieren más experimentos para confirmar totalmente nuestra hipótesis, es provechoso utilizar una herramienta colaborativa para suplir la falta de conocimiento en la instalación y mantenimiento de clústeres para HPC.

Como trabajos futuros nos proponemos expandir la comunidad de usuarios de Fresita para evaluar su impacto en la generación de conocimiento y, por otro lado, mejorar algunos aspectos aún no muy desarrollados de la herramienta, como el control de versiones.

## Agradecimientos

A la Dra. Adriana Zambon, de la Universidad Nacional del Comahue, por colaborar con nuestro proyecto y contactarnos con interesados en utilizar nuestras “recetas”. A Nicolás Cugat de la empresa Veng S.A. por su interés en nuestro trabajo.

## Referencias

- [1] Lucke, R.W., *Building Clustered Linux Systems*, Prentice Hall, 2004.
- [2] Lenoski, D. E., Weber, W. D., *Scalable Shared-Memory Multiprocessing*, Morgan Kaufmann Publisher, San Francisco, 1995.
- [3] Buyya, R. (Ed.), *High Performance Cluster Computing - Vol.1: Architectures and Systems, Vol.2: Programming and Applications*, Prentice Hall, 1999.
- [4] Chorley, M. J., Walker, D. W., Guest, M. F., *Hybrid Message-Passing and Shared-Memory Programming in a Molecular Dynamics Application On Multicore Clusters*, 2009.
- [5] Hernández, E. B., *Bioinformática: una oportunidad y un desafío*, en Revista

- Colombiana de Biotecnología, 10(1), pág.132-138, 2008.
- [6] Conway, S., *Energy Profiling and Analysis of the HPC Challenge Benchmarks*, International Journal of High Performance Computing Applications, 2006.
- [7] Top500, [www.top500.org](http://www.top500.org) (visitado por última vez el 13/08/13)
- [8] Medel, R., Martínez Spessot, C., McMillan, S., *Development of a Verification Tool for HPC Clusters*, en HPC Symposium 2008 – 37 Jornadas Argentinas de Informática (JAIIO), Santa Fe, Argentina, 2008.
- [9] Arellano, G. E., Hoet, L. J., *Cluster modular autocontenido*, en XI Workshop de Investigadores en Ciencias de la Computación, San Juan, Argentina, 2009.
- [10] Kitchens, F. L., *High performance computing as an educational experience well suited to developing nations*, en Fourth IEEE International Workshop on Technology for Education in Developing Countries, pág. 38-43, 2006.
- [11] Farías, M., Guerra, M., Sarmentero, D., Martínez Spessot, C., Medel, R., *Fresita: una herramienta distribuida para la creación comunitaria de recetas*, en 34<sup>a</sup> Conferencia Latinoamericana de Informática – CLEI 2008, Santa Fe, Argentina, 2008.
- [12] The Register®, <http://www.theregister.co.uk/> (visitado por última vez el 13/08/13)
- [13] Martínez Spessot, C., *Desarrollo de un plan estragógico de negocios sobre la solución de instalación de clústeres de computación de alta performance*, Tesis de Maestría en Administración de Negocios, Escuela de Posgrado, Facultad Regional Córdoba, Universidad Tecnológica Nacional, Córdoba, 2011.
- [14] Crimson Consulting Group, *Choosing between Windows and Linux for High Performance Computing: A Crimson Perspective*, White Paper, January 2010.
- [15] Intel® Cluster Ready Program, [www.intel.com/go/cluster](http://www.intel.com/go/cluster) (visitado por última vez el 13/08/13)
- [16] Sloan, J. D., *High Performance Linux Clusters with OSCAR, Rocks, OpenMosix, and MPI*, O'Reilly, 2004.

**Datos de Contacto:**

Ricardo Medel. Universidad Tecnológica Nacional – Facultad Regional Córdoba. Edificio Maders, 2º Piso, Oficina 4, Maestro M. López esq. Cruz Roja Argentina. Córdoba, 5000. Argentina.  
[ricardo.h.medel@gmail.com](mailto:ricardo.h.medel@gmail.com)