

Aplicación de una herramienta colaborativa en el desarrollo comunitario de conocimiento para Computación de Alto Rendimiento

Gustavo Martínez Spessot, Ricardo Medel, Marcelo Quispe, Diego Serrano,

Departamento de Ingeniería en Sistemas de Información,
Facultad Regional Córdoba – Universidad Tecnológica Nacional,
Maestro López esq. Av. Cruz Roja Argentina, Springer-Verlag, Computer Science Editorial,
Córdoba, Argentina
{gdrake84, ricardo.h.medel, marcelo.quispe, diegojserrano}@gmail.com

Resumen. La Computación de Alto Rendimiento (HPC) puede ser una herramienta clave en la competitividad de un país al acelerar los tiempos de investigación, diseño, ensayo y producción de nuevos productos y servicios. Sin embargo, la falta de RRHH con conocimientos en la instalación, mantenimiento y utilización de clústeres de computadoras para HPC es un obstáculo que debe ser superado a fin de obtener todas las ventajas de esta tecnología. En este trabajo reportamos el uso de una herramienta de creación comunitaria de documentos para el desarrollo de “recetas” que indican paso a paso como instalar un clúster bajo tecnología Linux. Esta primera experiencia muestra promisorios resultados que pueden ser expandidos a la comunidad argentina de HPC, a fin de paliar la falta de conocimientos específicos en el tema.

Palabras clave: Computación de Alto Rendimiento, diseños de referencia, recetas, creación colaborativa de conocimiento, clústeres.

1 Introducción

Con el término Computación de Alto Rendimiento (CAR, o más conocido como HPC, por sus siglas en inglés) se describe al uso de computadoras en problemas que requieren una gran capacidad de procesamiento de información. Actualmente esa capacidad de procesamiento se obtiene utilizando clústeres de computadoras (o “nodos”) conectadas por una red local de alta velocidad y trabajando en paralelo [1]. Dado que los nodos de un clúster pueden ser computadoras personales (PC) o servidores de bajo costo [2,3], 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 o servicios. 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].

Sin embargo, a pesar de su bajo costo inicial, y por estar formado por componentes de hardware y software de diverso origen, el proceso de adquirir, instalar y mantener un clúster HPC involucra altos costo de planificación, configuración, infraestructura, mantenimiento de sistema operativos y aplicaciones, administración de redes y unidades de procesamiento, entre otras tareas necesarias. 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 [6], justamente aquellos que podrían aprovechar el bajo costo inicial para potenciar su desarrollo industrial.

El objetivo del proyecto “Desarrollo de un Plan de Negocios basado en una Solución de Aprovisionamiento y Mantenimiento para Clústeres de Computación de Altas Prestaciones”, homologado por la Secretaría de Ciencia y Tecnología de la Universidad Tecnológica Nacional bajo el código UTN1230, 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 [7]. Como parte de dicho proyecto hemos utilizado en forma experimental una herramienta para la creación y mantenimiento de documentos, denominada Fresita [8] y desarrollada en un proyecto anterior, a fin de crear de manera comunitaria documentos que estipulen los pasos necesarios para la instalación de clústeres para HPC.

Consideramos que la estrategia de creación comunitaria de conocimientos para el área de la instalación y mantenimiento de clústeres HPC puede ser una solución creativa para superar la falta de dicho conocimiento en la región y expandir el uso de HPC tanto en el sistema nacional de I+D como en el sector industrial. El alcance de nuestro experimento fue solamente interno al grupo de I+D, pero esperamos poder expandirlo a una comunidad más amplia en el futuro cercano.

2 Metodología de la experiencia

El grupo de I+D para esta experiencia estuvo compuesto por cuatro investigadores, dos con experiencia relevante en la instalación y mantenimiento de clústeres para HPC y dos sin experiencia en el tema. Para obtener los conocimientos mínimos en el tema se realizó la instalación de un clúster con cuatro computadoras en desuso, basado en el ambiente Windows HPC Server. La instalación de este ambiente se realiza mediante una serie de pasos fijos indicados por el software de instalación. Sin embargo, debido a su flexibilidad y potencia, más del 90% de los clústeres HPC funcionan bajo entorno Linux [2,3], por lo que el siguiente paso fue instalar el mismo clúster pero bajo ambiente Linux y registrando los pasos realizados por medio de la herramienta de creación comunitaria Fresita [8].

La instalación de un clúster Linux requiere primero establecer el hardware que se utilizará, instalarlo de acuerdo a sus características, instalar el sistema operativo, utilizar un sistema de aprovisionamiento para crear la “imagen” del software que se instalará en cada “nodo” del cluster y luego replicar dicha imagen en cada nodo. Cada imagen contiene una serie de componentes de software que interactúan entre sí. El

proceso completo de instalación puede documentarse incluyendo la lista de componentes de hardware y software a utilizar (o BOM, por Bill of Materials) y la lista de pasos a seguir para realizar la instalación, que pueden incluir la ejecución de comandos o instrucciones del sistema operativo o incluso pequeños programas denominados *scripts*. Los documentos que siguen esta estructura (BOM + pasos) se denominan “recetas” o manuales *how-to*.

Nuestro enfoque no es completamente original, ya que la empresa Intel sigue la misma estrategia para su programa Intel Cluster Ready [9], produciendo y distribuyendo gratuitamente Diseños de Referencia (o “recetas”) que permiten instalar clústeres HPC que cumplan con las estrictas especificaciones del programa, orientado a reducir las barreras de entrada al mercado de HPC para empresas proveedoras de clústeres basados en tecnología Intel. Nuestro enfoque se diferencia del de la empresa Intel en que el desarrollo y mantenimiento de la documentación se realiza en forma comunitaria, no en una relación jerárquica proveedor-consumidores sino en una relación plana, formando una red entre *prosumidores* (productores a la vez que consumidores). Es por eso que la herramienta Fresita fue diseñada para crear y mantener comunitariamente documentos con la mencionada estructura de “recetas” y fue liberada como software libre [10].

En Fresita, cada receta está compuesta por una serie de “ingredientes”, donde cada ingrediente explica los pasos requeridos para completar una parte de la receta. Estas estructuras son almacenadas como archivos XML con estructuras especializadas (ver Fig. 1).

<pre> <recipe> <intro> <id/> <reciperevision/> <title/> <authors> <author> <name/> <surname/> <email/> </author> </authors> <revisions/> <tags> <tag/> </tags> </intro> <contenido> <abstract/> <index/> <chapter/> <order/> <name/> <ingredients> </ingredients> </contenido> </recipe> </pre>	<pre> <ingredient> <id/> <ingredientrevision/> <title/> <description> TEXT </description> <authors> <author> <name/> <surname/> <email/> </author> </authors> <revisions/> <steps> <step> <title/> <content> TEXT </content> </step> </steps> <tags> <tag/> </tags> </ingredient> </pre>
---	--

Figura 1. (a) Esquema XML para recetas. (b) Esquema XML para ingredientes.

Las recetas pueden ser creadas a partir de ingredientes ya existentes (ver Fig. 2) o creando ingredientes específicos para dicha receta utilizando una interfaz amigable con el usuario. Además provee un sistema de control de revisiones que permite al usuario trabajar con la versión más reciente o volver a una versión previa, en el estado en que estaba la receta antes de las últimas modificaciones.

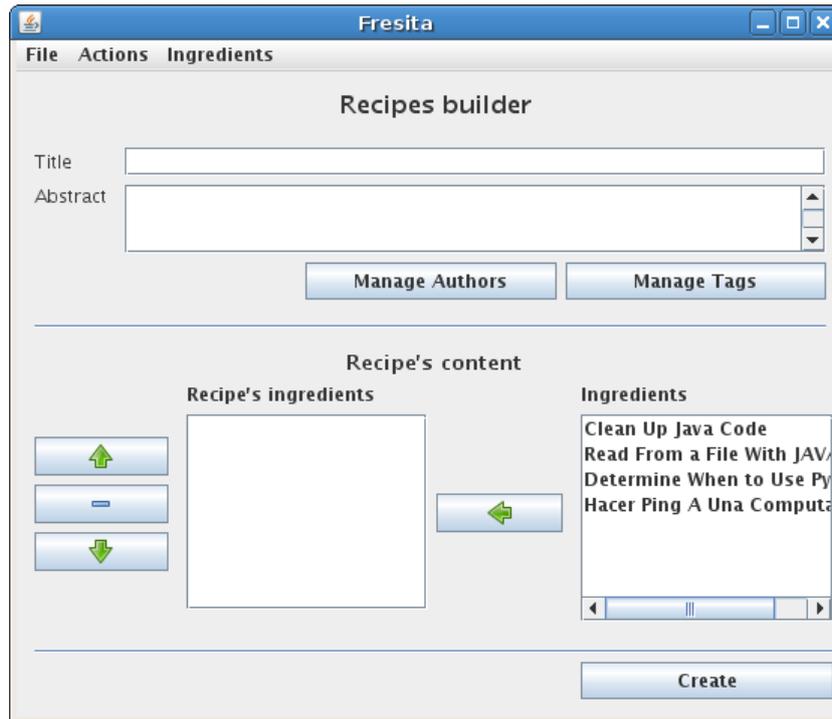


Figura 2. “Recipes builder” permite crear una receta a partir de ingredientes ya existentes.

La característica más relevante de Fresita para nuestro experimento es su capacidad para trabajar en red. Una vez que el usuario ingresa la dirección IP de las otras máquinas que forman la red, el sistema local puede acceder a sus repositorios y realizar búsquedas de recetas en las otras máquinas de la red, hacer copias locales y realizar modificaciones. El sistema de control de versiones se encarga de mantener la coherencia en la historia de modificaciones de las recetas.

En nuestro experimento instalamos Fresita en dos computadoras portátiles que eran utilizadas por los investigadores de menor conocimiento para documentar cada paso en la instalación del clúster Linux, con el apoyo de los investigadores de mayor experiencia. Se obtuvo así un documento que indica, además de los componentes de hardware y software utilizados, qué pasos se deben realizar para instalar el clúster. El documento es almacenado en ambas máquinas como una serie de archivos XML (receta e ingredientes) y puede ser accedido para consulta o modificación desde cualquiera de las dos computadoras. Dado que el formato XML permite el

procesamiento de la información por distintos sistemas, es posible obtener un impreso de la receta en un formato legible para el usuario, tal como .doc de MS Word.

3 Resultados

Nuestra experiencia resultó ser la primera vez que el sistema Fresita fue utilizado en una operación completa. Se descubrieron muchos inconvenientes en la instalación del software, que demoraron en resolverse debido a que los principales desarrolladores (Ingenieros en Sistemas de Información recientemente recibidos de la Universidad Tecnológica Nacional – Facultad Regional Córdoba, que realizaron el sistema como parte de su trabajo final, en el marco de un proyecto de investigación [8]) se encontraban desarrollando tareas profesionales en el extranjero.

Además, se formaron dos recursos humanos en el importante tema de instalación y mantenimiento de clústeres para HPC. Cabe acotar que los dos investigadores con menos experiencia recibieron becas SAE (Secretaría de Asuntos Estudiantiles) de iniciación a la investigación científica.

En cuanto a la documentación, se obtuvieron dos “recetas”, una en MS Word para un clúster basado en un entorno Windows y otra en formato Fresita para un clúster basado en un entorno Linux. Ambas recetas fueron enviadas a un grupo de investigación de la Universidad Nacional del Comahue, que estaba evaluando los conocimientos requeridos para instalar un clúster HPC, y a la empresa local VENG, proveedora de la CONAE, que evaluaba en ese entonces la adquisición de un clúster similar.

4 Discusión

Nuestro enfoque es generar conocimiento sobre instalación y mantenimiento de clústeres para HPC en forma comunitaria utilizando una herramienta diseñada específicamente para compartir “recetas” (documentos conteniendo una lista de ingredientes y los pasos requeridos para obtener el resultado final).

El uso de la herramienta Fresita permitió al grupo trabajar en forma distribuida (temporalmente, ya que no geográficamente) y crear conocimiento que ha sido comunicado a la comunidad y que está disponible para ser utilizado, perfeccionado y adaptado a las necesidades de cada integrante de la comunidad.

Los inconvenientes encontrados en la fase de instalación de la herramienta hablan de la necesidad imperiosa de mantener documentados los proyectos distribuidos como software libre (en este caso como proyecto de Google Code) y buscar formas efectivas de realizar la apropiación de la tecnología por parte de los nuevos integrantes del grupo. Sin embargo, cabe acotar que los problemas fueron resueltos por parte del equipo original y el *feedback* provisto por los nuevos usuarios ha servido para simplificar la instalación del sistema.

Los pasos futuros de este trabajo serán extender el grupo de trabajo sobre las recetas, invitando a la incipiente comunidad HPC de Argentina a sumarse al esfuerzo de crear documentación y compartir información a través de este sistema.

Referencias

1. Buyya, R. (Ed.): High Performance Cluster Computing - Vol.1: Architectures and Systems, Vol.2: Programming and Applications. Prentice Hall (1999)
2. Lucke, R.W.: Building Clustered Linux Systems. Prentice Hall (2004)
3. Sloan, J. D.: High Performance Linux Clusters with OSCAR, Rocks, OpenMosix, and MPI. O'Reilly (2004)
4. Grandinetti L., Joubert W. Gentsch, G. R.: High Speed and Large Scale Scientific Computing - Volume 18: Advances in Parallel Computing. IOS (2010)
5. Hager, G., Wellein, G.: Introduction to High Performance Computing for Scientists and Engineers. Chapman & Hall (2010)
6. Martínez Spessot, C.: 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, Facultad Regional Córdoba, Universidad Tecnológica Nacional. Córdoba (2011)
7. Castillo, J., Martínez Spessot, C., Medel, R.: Hacia la Aplicación de la Computación de Alto Desempeño al Entorno Productivo Local. En: XIII Workshop de Investigadores en Ciencias de la Computación – WICC 2012. Misiones, Argentina (2012)
8. Farías, M., Guerra, M., Sarmentero, D., Martínez Spessot, C., Medel, R.: Fresita: una herramienta distribuida para la creación comunitaria de recetas. En: 34ª Conferencia Latinoamericana de Informática – CLEI 2008. Santa Fe, Argentina (2008)
9. Intel Cluster Ready Program. www.intel.com/go/cluster
10. <http://code.google.com/p/gpi-xrecipes-maker/>