

Integración flexible de *workflows* científicos en entornos de computación dinámicos

S. Hernández de Mesa, P. Álvarez

Grupo de Integración de Sistemas Distribuidos y Heterogéneos (GIDHE)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976761000 ext. 5458, Fax +34-976762043, e-mail: shernandez@unizar.es

Abstract

La aparición de potentes infraestructuras de computación *Grid* ha impulsado un creciente interés por parte de la comunidad científica para automatizar la ejecución de sus experimentos. Para ello, la utilización de *workflows* científicos resulta clave de cara a aprovechar las oportunidades que presentan este tipo de infraestructuras, debido principalmente a las características que presentan: están formados por un elevado número de tareas con un alto coste computacional, que manejan grandes volúmenes de datos complejos y que requieren una gestión adecuada de los recursos tanto *software* como *hardware* disponibles. Sin embargo, la complejidad de los propios *workflows* y de las infraestructuras de computación provoca que el proceso de *scheduling* resulte complejo y difícil. Las propuestas actuales utilizan un *broker* de recursos que facilita esta tarea pero provoca que el diseño del *workflow* esté fuertemente acoplado con el *middleware* que gestiona la infraestructura. Esto imposibilita la utilización conjunta de diferentes *Grids*, la compartición y portabilidad de los *workflows* y la gestión flexible del ciclo de vida de los mismos. Para solucionar estas limitaciones, se plantea una infraestructura flexible a la hora de integrar *workflows* de diferentes lenguajes, diversas infraestructuras de computación (*Grid, Cloud, etc.*) y componentes de gestión que permitan enriquecer y mejorar la ejecución de los *workflows*. La utilización de redes de Petri de alto nivel se plantea como base, dada la amplia experiencia en este campo por parte del Grupo de Integración de Sistemas Distribuidos y Heterogéneos (GIDHE), grupo en el que se enmarca la línea de investigación del autor. Esta propuesta introduce nuevos retos en la decisión de las infraestructuras en las que ejecutar los *workflows* (*meta-scheduling*), siendo la simulación una alternativa muy adecuada al ser capaz de prever el comportamiento del *workflow* en diferentes escenarios.