

## Resumen tesis doctoral “High Performance Scientific Computing over Hybrid Cloud Platforms”

Las aplicaciones científicas generalmente precisan grandes requisitos de cómputo, memoria y gestión de datos para su ejecución. Este tipo de aplicaciones tradicionalmente ha empleado recursos de altas prestaciones, como supercomputadores de memoria compartida, clústers de PCs de memoria distribuida, o recursos provenientes de infraestructuras Grid, sobre los que se adaptaba la aplicación para que se ejecutara satisfactoriamente. El auge que han tenido las técnicas de virtualización en los últimos años, propiciando la aparición de la computación en la nube (Cloud Computing), ha provocado un importante cambio en la forma de ejecutar este tipo de aplicaciones. Sin embargo, la gestión de la ejecución de aplicaciones científicas sobre plataformas de computación elásticas de altas prestaciones no es una tarea trivial.

En esta tesis doctoral se ha desarrollado Elastic Cloud Computing Cluster (EC3), una herramienta de código abierto capaz de llevar a cabo la ejecución de aplicaciones científicas de altas prestaciones creando para ello clústers virtuales, híbridos y elásticos, autogestionados y eficientes en cuanto a costes, sobre plataformas Cloud de tipo Infraestructura como Servicio (IaaS). Estos clústers autogestionados tienen la capacidad de adaptar su tamaño, es decir, el número de nodos, a la carga de trabajo, creando así la ilusión de un clúster real sin requerir una inversión por encima del uso actual. Además, son completamente configurables y pueden ser migrados de un proveedor a otro de manera automática y transparente a los usuarios y trabajos en ejecución en el cluster.

EC3 también permite desplegar clústers híbridos sobre recursos Cloud públicos y privados, donde los recursos privados son complementados con recursos Cloud públicos para acelerar el proceso de ejecución. Otras configuraciones híbridas, como el empleo de diferentes tipos de instancias y el uso de instancias puntuales combinado con instancias bajo demanda son también soportadas por EC3. Además, el uso de instancias puntuales junto con técnicas de checkpointing permite a EC3 reducir significativamente el coste total de las ejecuciones a la vez que proporciona tolerancia a fallos. EC3 está concebido para facilitar el uso de clústers virtuales a los usuarios, que, aunque no tengan un conocimiento extenso sobre este tipo de tecnologías, pueden beneficiarse fácilmente de ellas. Por ello, la herramienta ofrece dos interfaces diferentes a sus usuarios, una interfaz web donde se expone EC3 como servicio para usuarios no experimentados y una potente interfaz de línea de comandos.

Además, esta tesis doctoral se adentra en el campo de la virtualización ligera, mediante el uso de contenedores como alternativa a la solución tradicional de virtualización basada en máquinas virtuales. Este estudio analiza el escenario propicio para el uso de contenedores y propone una arquitectura para el despliegue de clusters virtuales elásticos basados en esta tecnología. Finalmente, para demostrar la funcionalidad y ventajas de las herramientas desarrolladas durante este proyecto de tesis, esta memoria recoge varios casos de uso que abarcan diferentes escenarios y campos de conocimiento, como estudios estructurales de edificios, astrofísica o biodiversidad.