

Serverless Strategies and Tools in the Cloud Computing Continuum

En los últimos años, la popularidad de la computación en nube ha permitido a los usuarios acceder a recursos de cómputo, red y almacenamiento sin precedentes bajo un modelo de pago por uso. Esta popularidad ha propiciado la aparición de nuevos servicios para resolver determinados problemas informáticos a gran escala y simplificar el desarrollo y el despliegue de aplicaciones. Entre los servicios más destacados en los últimos años se encuentran las plataformas FaaS (Función como Servicio), cuyo principal atractivo es la facilidad de despliegue de pequeños fragmentos de código en determinados lenguajes de programación para realizar tareas específicas en respuesta a eventos. Estas funciones son ejecutadas en los servidores del proveedor Cloud sin que los usuarios se preocupen de su mantenimiento ni de la gestión de su elasticidad, manteniendo siempre un modelo de pago por uso de grano fino.

Las plataformas FaaS pertenecen al paradigma informático conocido como Serverless, cuyo propósito es abstraer la gestión de servidores por parte de los usuarios, permitiéndoles centrar sus esfuerzos únicamente en el desarrollo de aplicaciones. El problema del modelo FaaS es que está enfocado principalmente en microservicios y tiende a tener limitaciones en el tiempo de ejecución y en las capacidades de computación (por ejemplo, carece de soporte para hardware de aceleración como GPUs). Sin embargo, se ha demostrado que la capacidad de autoaprovisionamiento y el alto grado de paralelismo de estos servicios pueden ser muy adecuados para una mayor variedad de aplicaciones. Además, su inherente ejecución dirigida por eventos hace que las funciones sean perfectamente adecuadas para ser definidas como pasos en flujos de trabajo de procesamiento de archivos (por ejemplo, flujos de trabajo de computación científica).

Por otra parte, el auge de los dispositivos inteligentes e integrados (IoT), las innovaciones en las redes de comunicación y la necesidad de reducir la latencia en casos de uso complejos han dado lugar al concepto de Edge computing, o computación en el borde. El Edge computing consiste en el procesamiento en dispositivos cercanos a las fuentes de datos para mejorar los tiempos de respuesta. La combinación de este paradigma con la computación en nube, formando arquitecturas con dispositivos a distintos niveles en función de su proximidad a la fuente y su capacidad de cómputo, se ha acuñado como continuo de la computación en la nube (o continuo computacional).

Esta tesis doctoral pretende, por lo tanto, aplicar diferentes estrategias Serverless para permitir el despliegue de aplicaciones generalistas, empaquetadas en contenedores de software, a través de los diferentes niveles del continuo computacional. Para ello, se han desarrollado múltiples herramientas con el fin de: i) adaptar servicios FaaS de proveedores Cloud públicos; ii) integrar diferentes componentes software para definir una plataforma Serverless en infraestructuras privadas y en el borde; iii) aprovechar dispositivos de aceleración en plataformas Serverless; y iv) facilitar el despliegue de aplicaciones y flujos de trabajo a través de interfaces de usuario. Además, se han creado y adaptado varios casos de uso para evaluar los desarrollos conseguidos.