

CloudButton: Plataforma Serverless para Analítica de Datos

El proyecto CloudButton ‘democratiza’ el big data simplificando el modelo de programación y el ciclo de vida del software gracias a las tecnologías serverless.

La idea de desarrollar CloudButton germina a partir de una frase de un profesor de Berkley: “¿Por qué la nube no tiene un botón?” Con esta sencilla frase sintetizaba el deseo de sus alumnos de simplificar el proceso de despliegue y ejecución de sus desarrollos en la nube.

Basándose en esta premisa, que muestra el trasfondo de una necesidad real de cualquier desarrollador, CloudButton diseña la arquitectura de su plataforma en torno a tres objetivos:

1. **Desarrollar el primer entorno FaaS** (Function as a Service) para el análisis de datos usando Apache OpenWhisk. Dicha plataforma FaaS da servicio al paradigma de arquitecturas serverless mediante el cual el complejo proceso de abstracción de recursos ya no recae en el usuario final. De este modo, es el proveedor de recursos en la nube quien se encarga de gestionar el despliegue de las aplicaciones y su ejecución bajo demanda.
2. **Desarrollar estructuras distribuidas de datos mutables** utilizando Infinispan. Para ello se ha desarrollado un middleware capaz de proveer las funcionalidades necesarias para asegurar la persistencia, confiabilidad y control de concurrencia de datos en funciones serverless.
3. **Desarrollar el CloudButton Toolkit**, que provee la capa de abstracción necesaria para el manejo de aplicaciones intensivas de datos con mínimos cambios, basándose en los puntos anteriores. Para poder migrar fácilmente estas aplicaciones a la plataforma provista por CloudButton, se han desarrollado nuevas herramientas y métodos utilizando tecnologías HPC, machine learning y analítica de datos.

Como sucede habitualmente con aquellas soluciones más horizontales, no resulta sencillo para el usuario final comprender su funcionamiento o visualizar su posible aplicación, o los beneficios resultantes de la misma. Por ello, CloudButton desarrolla tres casos de uso distintos, todos ellos con una gran necesidad de computación para realizar el análisis de grandes volúmenes de datos, donde se demuestra el potencial de la plataforma:

1. *Análisis de datos genómicos*: las limitaciones en cuanto a escalabilidad en los recursos computacionales de los centros de investigación pueden reducirse usando tecnologías serverless, de este modo se mejora no sólo la escalabilidad sino la productividad en el procesamiento de grandes conjuntos de datos.
2. *Análisis de datos metabolómicos*: la solución propuesta mejora la capacidad analítica de datos metabolómicos sin procesar, impulsando el acceso externo y la reutilización eficiente de datos abiertos.
3. *Análisis de datos geoespaciales*: para el análisis de datos geoespaciales se utilizan datos satelitales y LiDAR de acceso abierto, mejorando la productividad, escalabilidad y rendimiento de las aplicaciones ambientales.

Para estar al tanto de las últimas novedades sigue el desarrollo del proyecto en:



<https://cloudbutton.eu/>



H2020 CloudButton



@cloudbutton2020



<https://github.com/cloudbutton>

