# Título

Desarrollo de una solución BDaaS (Big Data as a Service) para el análisis de grafos temporales a gran escala

# Descripción

Actualmente, los grafos se han convertido en una herramienta muy importante de análisis, visualización, o solución de problemas, con un ámbito de aplicación que se extiende más allá de las matemáticas. Así, es habitual la utilización de grafos para calcular los caminos más cortos en una red de comunicaciones, analizar el comportamiento y las interacciones de los usuarios en una red social, o, más recientemente, estudiar los desplazamientos de los ciudadanos durante la pandemia de COVID-19, entre muchos otros ejemplos.

En este contexto, a fin de poder aprovechar esta información, se hace evidente la necesidad de procesar dichos grafos de la forma más rápida y eficiente posible, lo que, tradicionalmente, ha supuesto un reto considerable, debido a la dificultad de paralelización y distribución de este tipo de cálculos, provocada por la naturaleza interactiva de los elementos que componen un grafo.

Sin embargo, recientemente, han surgido paradigmas de procesado de grafos, como el concepto de “Thinking like a vertex”, que sí permiten la distribución de las operaciones entre diferentes actores.

Una de las herramientas que se basan en este paradigma para llevar a cabo dicho procesado es Raphtory, desarrollada en Scala por una startup inglesa llamada Pometry, la cual, además, implementa un modo de análisis “temporal” o “dinámico”, que permite procesar la información del grafo no solo en el estado actual, sino también durante instantes anteriores, e, incluso, durante un intervalo temporal especificado por el usuario, plasmando de este modo la evolución que se ha producido.

Así, el objetivo de mi Trabajo Fin de Máster es el de analizar dicha herramienta, sobre todo en lo que se refiere a sus diferentes modos de ejecución (en una sola máquina, distribuido, mediante contenedores de Docker, etc.), en relación con el concepto de “Thinking like a vertex”, para, a partir de ello, realizar un despliegue automatizado en la nube, como un servicio de tipo BDaaS, basado en Docker, Google Cloud y Kubernetes.

# Metodología

El proceso a seguir para llevar a cabo este TFM es el siguiente:

* Análisis del procesado distribuido de grafos, en especial el paradigma “thinking like a vertex”, mediante el estudio de diferentes artículos y papers.
* Experimentación y pruebas con la herramienta Raphtory, en sus diferentes modos de ejecución (un host, pseudo-distribuido, en contenedores de Docker).
* Despliegue de Raphtory en Kubernetes, sobre la nube de Google Cloud.
* Automatización del despliegue de Raphtory en la nube, con Helm, y GitHub Actions.