MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

**DESARROLLO DE UNA SOLUCIÓN BDaaS (BIG DATA AS A SERVICE) PARA EL ANÁLISIS DE GRAFOS TEMPORALES A GRAN ESCALA**

DAVID PISONERO FUENTES

2022

MÁSTER uNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN

trabajo fin de MÁSTER

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

**Autor:** D. David Pisonero Fuentes

**Tutor:** D. Félix Cuadrado Latasa

**Ponente:** D. Juan Carlos Dueñas López

**Departamento:** Departamento de Ingeniería se Sistemas Telemáticos (DIT)

Miembros del Tribunal

**Presidente:** D. ……………

**Vocal:** D. …………..

**Secretario:** D. …………..

**Suplente:** D. ……………..

Los miembros del tribunal arriba nombrados acuerdan otorgar la calificación de: ………

Madrid, a de de 20…

**UNIVERSIDAD POLITÉCNICA DE MADRID**

**ESCUELA TÉCNICA SUPERIOR**

**DE INGENIEROS DE TELECOMUNICACIÓN**

****

MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

**DESARROLLO DE UNA SOLUCIÓN BDaaS (BIG DATA AS A SERVICE) PARA EL ANÁLISIS DE GRAFOS TEMPORALES A GRAN ESCALA**

DAVID PISONERO FUENTES

2022

RESUMEN

Históricamente, los grafos han sido una herramienta de gran utilidad en diferentes disciplinas, desde las Matemáticas hasta las Telecomunicaciones, pasando por las Ciencias Sociales.

En aras de aprovechar la información que estos nos ofrecen, especialmente en el mundo actual, donde convivimos con la realidad del Big Data, surge la necesidad de procesar dichos grafos de la forma más rápida y eficiente posible, hecho que supone un reto de considerable dificultad, debido a la naturaleza interactiva de los vértices y aristas que componen estas estructuras.

Así surge Raphtory, una herramienta capaz de realizar el procesado **distribuido** de grafos mediante la implementación del paradigma teórico “**Thinking like a vertex**”, desarrollado por Google, que sí permite la distribución de operaciones de procesado de grafos entre diferentes actores. De hecho, Raphtory va más allá, pues también incluye la posibilidad de llevar a cabo análisis temporales, es decir, estudiar la evolución de un grafo a lo largo del tiempo, mediante el uso de ventanas.

Así pues, este Trabajo Fin de Máster realiza un estudio teórico sobre el procesado distribuido de grafos a través del paradigma antes mencionado, analiza el funcionamiento de la herramienta Raphtory, y experimenta con sus diferentes formas de despliegue, desde la opción monolítica, hasta la distribuida basada en servicios alojados en contenedores y pods.

A continuación, se implementa una solución **BDaaS** (Big Data as a Service), que permite el despliegue distribuido de Raphtory en Google Cloud haciendo uso de Kubernetes, aplicando cierto nivel de automatización, de manera que un potencial usuario pueda utilizar esta herramienta eliminando las costosas barreras de entrada que supone el procesado distribuido de grafos masivos, gracias a las ventajas que ofrece la **Nube**.

Se finaliza el trabajo con las conclusiones extraídas sobre el funcionamiento de Raphtory, y los despliegues desarrollados, y se plantean unas líneas futuras para la evolución de este servicio BDaaS.

summary

Historically, graphs have been a very useful tool among different disciplines, ranging from Maths to Telecommunications, including Social Sciences.

In order to take advantage of the information they offer, specially nowadays, when we have to coexist with the reality of Big Data, the necessity of processing said graphs as fast and efficiently possible arises, which creates a considerably difficult challenge, due to the interactive nature of the vertices and edges composing these structures.

This is why Raphtory appears, a tool able to carry out the **distributed** processing of graphs thanks to the implementation of the theoretical paradigm “**Thinking like a vertex**”, developed by Google, which does enable the distribution of the graph processing operations between different actors. In fact, Raphtory goes beyond that, including the possibility of conducting temporal analysis, that is, studying the graph evolution through time, making use of windows.

That being the case, this Master’s Thesis carries out a theoretical study about the distributed graph processing by means of said paradigm, it analyses Raphtory’s functioning, and it experiments with its different ways of deployment, from the monolithic option, to the distributed one based on services hosted inside containers and pods.

Following that, a **BDaaS** (Big Data as a Service) solution is implemented, which offers a distributed deployment for Raphtory on Google Cloud using Kubernetes, applying a certain level of automatization, so a potential user could use this tool without the expensive barrier entries brought up by the massive graphs distributed processing, thanks to the possibilities offered by the **Cloud**.

This work ends with the conclusions obtained about Raphtory’s functioning, and the implemented deployments, as well as some future proposals relating to the evolution of this BDaaS service.

PALABRAS CLAVE

Raphtory, procesado de grafos, análisis temporal, sistemas distribuidos, Big Data, Nube, Kubernetes, Docker, Docker Compose, GKE, Cloud Storing, BdaaS.

KEYWORDS

Raphtory, graph processing, temporal analysis, distributed systems, Big Data, Cloud, Kubernetes, Docker, Docker Compose, GKE, Cloud Storing, BDaaS.

índice del contenido

[1. Glosario 7](#_Toc106032598)

[2. introducción y objetivos 8](#_Toc106032599)

[2.1. Introducción 8](#_Toc106032600)

[2.2. Estado del arte 8](#_Toc106032601)

[2.3. Objetivos 8](#_Toc106032602)

[3. desarrollo 2](#_Toc106032603)

[4. resultados 3](#_Toc106032604)

[5. conclusiones y líneas futuras 4](#_Toc106032605)

[5.1. ConclusioneS 4](#_Toc106032606)

[5.2. Líneas futuras 4](#_Toc106032607)

[6. bibliografía 5](#_Toc106032608)

[anexo a: aspectos ÉTICOS, económicos, sociales y ambientales 6](#_Toc106032609)

[A.1 iNTRODUCCIÓN 6](#_Toc106032610)

[A.2 DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PROYECTO 6](#_Toc106032611)

[A.3 ANÁLISIS DETALLADO DE ALGUNO DE LOS PRINCIPALES IMPACTOS 6](#_Toc106032612)

[A.4 CONCLUSIONES 6](#_Toc106032613)

[anexo b: presupuesto económico 7](#_Toc106032614)

# Glosario

|  |  |
| --- | --- |
| Término | Significado |
| BDaaS | Big Data as a Service, Big Data como Servicio |
| GKE | Google Kubernetes Engine |
| IAM | Identity and Access Management, Gestion de la Identidad y el Acceso |

# introducción y objetivos

## Introducción

…(A partir de ésta, se escribirán hasta 50 páginas, sin contar los anexos) HABLAR SOBRE TEORÍA DE GRAFOS, LA IMPORTANCIA DE LOS GRAFOS EN LA ACTUALIDAD, CASOS DE USO

## Estado del arte

HABLAR SOBRE EL PROCESADO DISTRIBUIDO DE GRAFOS, CON EL PARADIGMA DE “THINKING LIKE A VERTEX”, LOS GRAFOS TEMPORALES, AMAZON ELASTIC MAP-REDUCE, Y MENCIONAR BREVEMENTE RAPHTORY. Comentar que hay una barrera de entrada muy grande, y que este tipo de despliegues BDaaS permite que cualquier usuario pueda acceder a este tipo de servicios, y centrarse más en la algoritmia. Importancia de la nube, que te facilita un poco el asunto. Ventajas de la nube.

## Objetivos

…ANÁLISIS DE LA HERRAMIENTA, PROBAR VARIOS DESPLIEGUES, QUIZÁS MEDIR TIEMPOS DE CADA UNO, Y HABLAR DEL DESPLIEGUE EN LA NUBE. Objetivo final: hacer un prototipo de analítica como servicio. Llevar a cabo estudio de DevOps, virtualización, almacenamiento de datos en la nube. Analizar/Evaluar este tipo de cosas es objetivo de esta herramienta. Que todo esto tenga un peso importante, que se vea que ha habido un trabajo de aprendizaje.

# desarrollo

Esto es un ejemplo de cita a una referencia bibliográfica [1]… EMPEZAR CON UN ANÁLISIS MÁS DETALLADO DE LA HERRAMIENTA, Y LUEGO SEGUIR CON LOS DIFERENTES MECANISMOS DE DESPLIEGUE (SINGLE, PD, DOCKER, K8S, GOOGLE CLOUD). El tema del rendimiento, comentarlo de manera cualitativa, no hacer medidas. Atacar por escalabilidad vs comunicaciones.

Para la automatización, dejarlo como un despliegue de GitHub Actions, pero que sea algo particularizado, con un ejemplo concreto. No poner demasiado código, poner un enlace a GitHub.

# resultados

…. Hacer una prueba del sistema, explicando poco a poco cómo se va haciendo un despliegue.

# conclusiones y líneas futuras

## ConclusioneS

Enfatizar la dificultad de trabajar un proyecto tan dinámico (que te cambian las cosas de un día para otro).

## Líneas futuras

…

# bibliografía

|  |  |
| --- | --- |
| [1] | «Portal web de la ETSIT UPM,» [En línea]. Available: http://www.etsit.upm.es/. |
| [2] | P. Taylor, Text-to-Speech Synthesis, Cambridge University Press, 2009. |
| [3] | Zen, H., Tokuda, K., & Black, A. W., «Statistical parametric speech synthesis,» *Speech Communication, 51(11),* pp. 1039-1064, 2009. |
| [4] | Pardo, J.M., Giménez de los Galanes, F.M., Vallejo, J.A., Berrojo, M.A., Montero, J.M., Enríquez, E., Romero, A., «Spanish text-to-speech, from prosody to acoustics,» de *International Congress of Acoustics*, 1995. |

# anexo a: aspectos ÉTICOS, económicos, sociales y ambientales

El apartado “Requisitos de las acreditaciones internacionales EUR-ACE y ABET” de la Normativa de TFT de la ETSIT-UPM establece que “*La memoria del TFT del GITST, GIB y MUIT, y en general la de aquellas titulaciones que hayan obtenido o para las que se desee solicitar una acreditación internacional EUR-ACE o ABET, debe mostrar conciencia de la responsabilidad de la aplicación práctica de la ingeniería, el impacto social y ambiental, el compromiso con la ética profesional, la responsabilidad y las normas de la aplicación práctica de la ingeniería, así como sobre las prácticas de gestión de proyectos, gestión, y control de riesgos, entendiendo sus limitaciones*”.

Este anexo obligatorio del TFT tendrá un carácter sintético con los siguientes apartados:

## A.1 iNTRODUCCIÓN

Breve descripción del contexto del proyecto, objetivos, necesidades que pretende cubrir o problemas que pretende resolver, centrándose en su relación con los temas sociales, económicos, éticos, legales y/o ambientales que se hayan identificado.

## A.2 DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PROYECTO

Síntesis del trabajo realizado en la fase 2, de selección y descripción de impactos. Presentar y justificar las conclusiones a las que se haya llegado sobre cuáles son los asuntos más relevantes relacionados con la sostenibilidad social, económica o ambiental, así como los principales grupos de interés identificados y que se han considerado en los análisis posteriores.

## A.3 ANÁLISIS DETALLADO DE ALGUNO DE LOS PRINCIPALES IMPACTOS

Síntesis del trabajo de análisis realizado.

## A.4 CONCLUSIONES

Valorar el proyecto desde un punto de vista ético, social, económico y medioambiental y justificar si el uso de criterios de sostenibilidad ha aportado o puede aportar valor añadido al proyecto.

# anexo b: presupuesto económico

El apartado “Requisitos de las acreditaciones internacionales EUR-ACE y ABET” de la Normativa de TFT de la ETSIT-UPM establece que “*La memoria del TFT del GITST, GIB y MUIT, y en general la de aquellas titulaciones que hayan obtenido o para las que se desee solicitar una acreditación internacional EUR-ACE o ABET, … debe incluir un presupuesto económico*”. A modo de ejemplo, la tabla del presupuesto de un proyecto podría ser la siguiente:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **COSTE DE MANO DE OBRA (coste directo)** | | | | | **Horas** | | **Precio/hora** | **Total** |
|  | | | | | 300 | | 15 € | **4.500 €** |
|  | | | |  |  | |  |  |
| **COSTE DE RECURSOS MATERIALES (coste directo)** | | | | **Precio de compra** | **Uso en meses** | | **Amortización (en años)** | **Total** |
| Ordenador personal (Software incluido)....... | | | | 1.500,00 € | 6 | | 5 | 150,00 € |
| Impresora láser | | | | 500,00 € | 6 | | 5 | 50,00 € |
| Otro equipamiento | | | |  |  | |  |  |
|  | | | |  |  | |  |  |
| **COSTE TOTAL DE RECURSOS MATERIALES** | | | | | | | | **200,00 €** |
|  | |  | |  | |  | |  |
| **GASTOS GENERALES (costes indirectos)** | | 15% | | sobre CD | | | | **705,00 €** |
| **BENEFICIO INDUSTRIAL** | | 6% | | sobre CD+CI | | | | **324,30 €** |
|  | |  | |  | |  | |  |
| **MATERIAL FUNGIBLE** | |  | |  | |  | |  |
| Impresión | | | | | | | | **100,00 €** |
| Encuadernación | | | | | | | | **300,00 €** |
|  |  | |  | | |  | |  |
| **SUBTOTAL PRESUPUESTO** | | | | | | | | **6.129,30 €** |
| **IVA APLICABLE** | | | | | | | 21% | **1.287,15 €** |
|  |  | |  | | | |  |  |
| **TOTAL PRESUPUESTO** | | | | | | | | **7.416,45 €** |

Esta tabla podría ser rellenada mediante una sencilla hoja de cálculo como la siguiente:

