

Proyecto 1, Entrega 5 - Arquitectura, conclusiones y consideraciones

Integrantes:

- Daniel Santamaría Álvarez - 201720222
- Nestor Gonzalez – 201912670
- Álvaro Plata – 201820098
- Rafael Humberto Rodriguez Rodriguez – 202214371
- Juan Nicolás Bolaños - 201911676

Para migrar la aplicación al servicio de App Engine se crearon los archivos de configuración para el despliegue.

En el servicio default, que corresponde a la aplicación de Flask, se configuró con Python3.9 y se ejecuta la aplicación con el servidor http de gunicorn. Finalmente, se usó una variable de entorno con el id del proyecto.

```
runtime: python39
entrypoint: gunicorn -b :8080 app:app

env_variables:
  PROJECT_ID: swnube
```

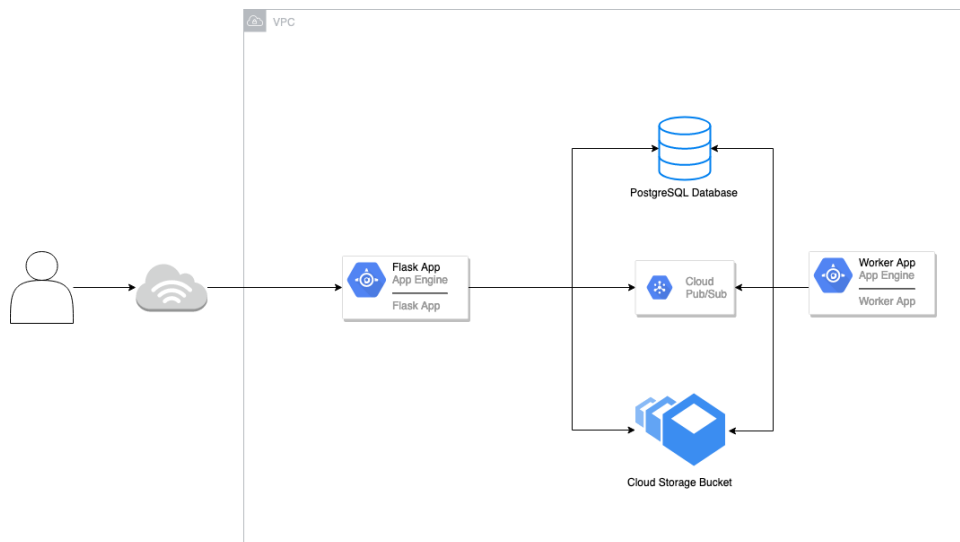
Así mismo, para el servicio del worker, se usó Python3.9, sin embargo, en este caso no hay un comando activo de entrypoint dado que solo se requiere ejecutar el script que está constantemente escuchando a los mensajes de pub/sub. Igualmente, se usó una variable de entorno con la URL donde está desplegada la aplicación de flask.

```
runtime: python39
#entrypoint: python3 -m batch.py
service: worker
env_variables:
  API_INSTANCE_IP: swnube.ue.r.appspot.com
```

Por último, se cambió la dirección a la que se envía la petición informando que ya se procesó el archivo correctamente.

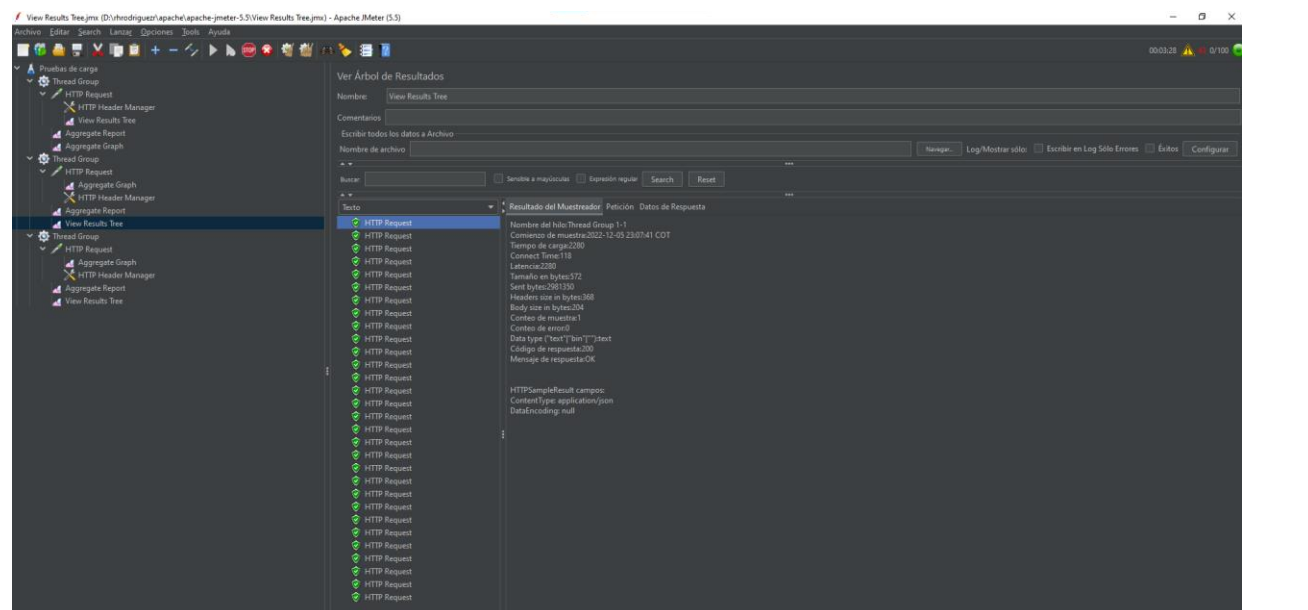
```
r = requests.post(f'https://{os.environ.get("API_INSTANCE_IP")}/api/tasks/{task_id}/processed')
```

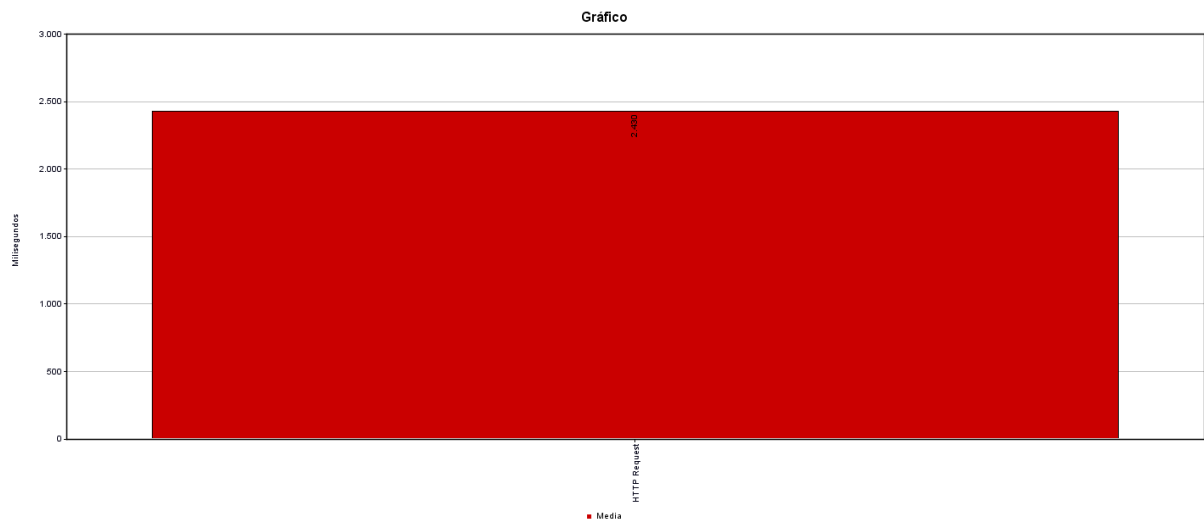
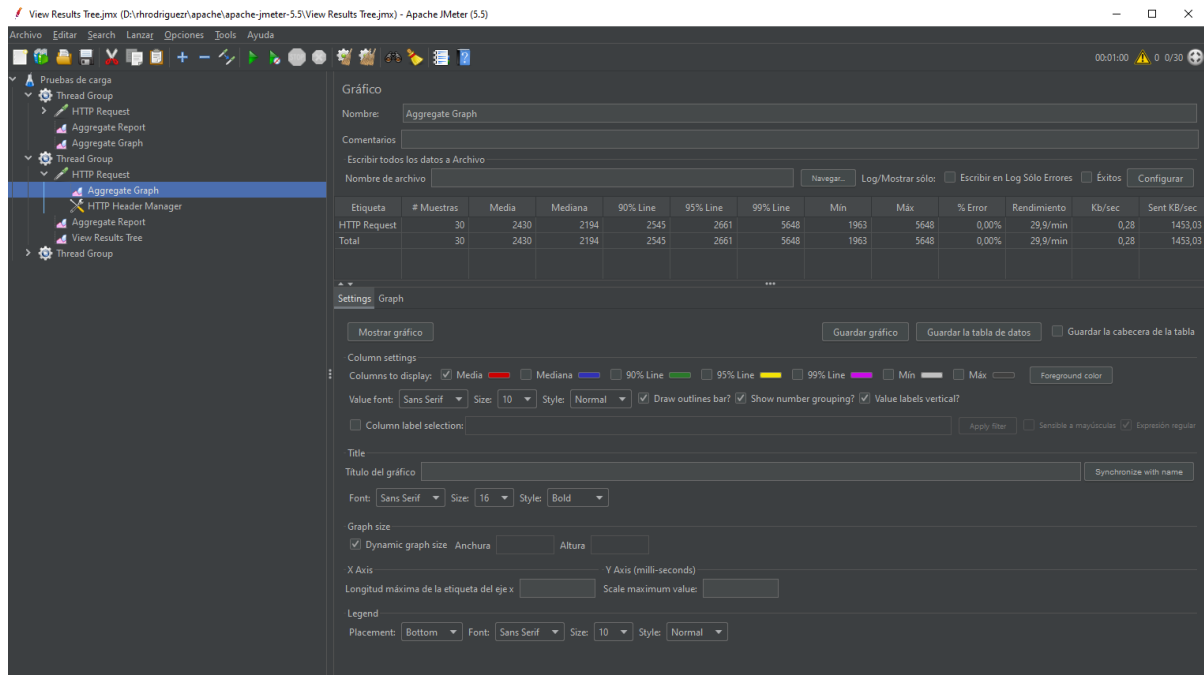
Descripción de la arquitectura desplegada en Google Cloud



A continuación, se describen los principales cambios realizados a la arquitectura de la aplicación desde la última entrega:

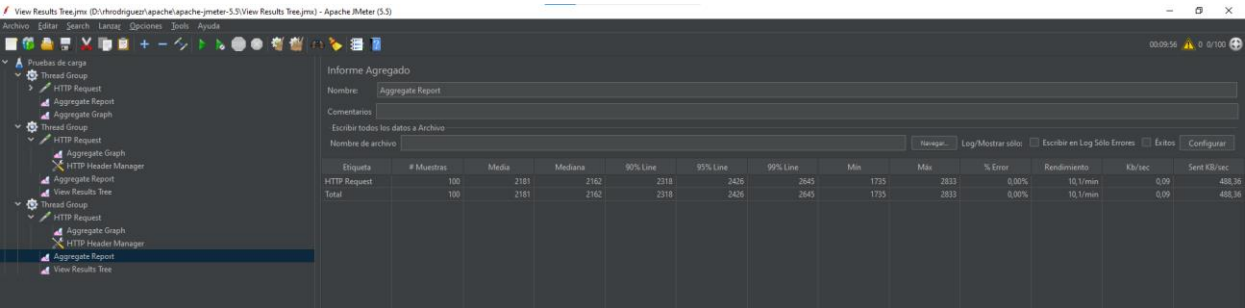
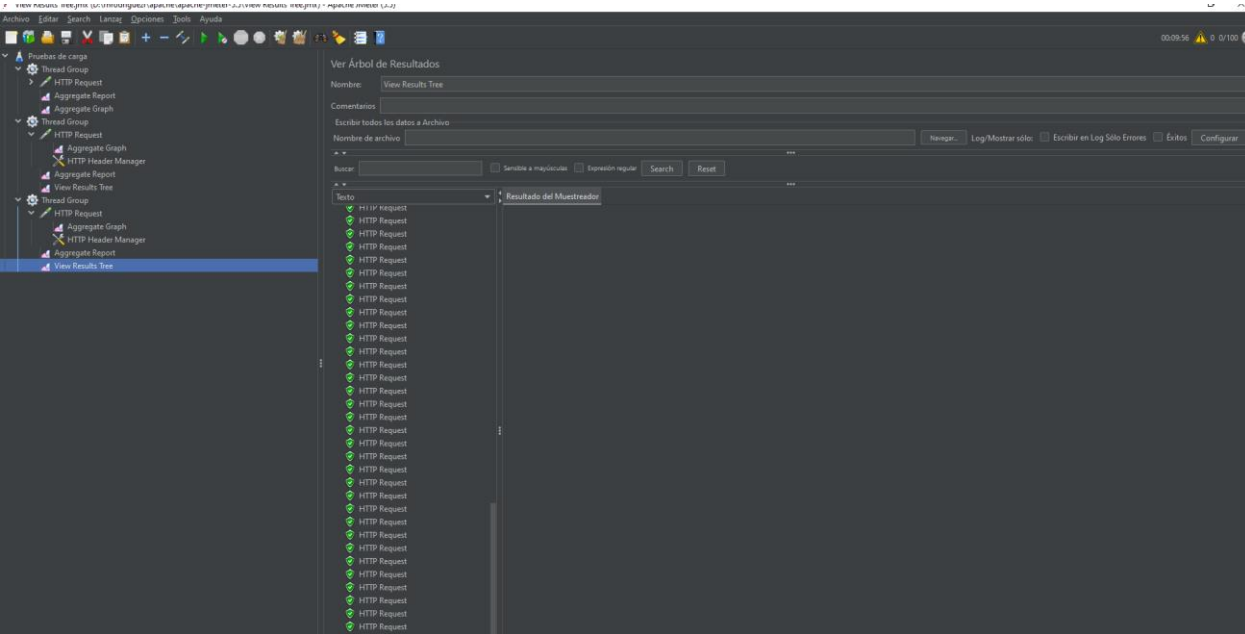
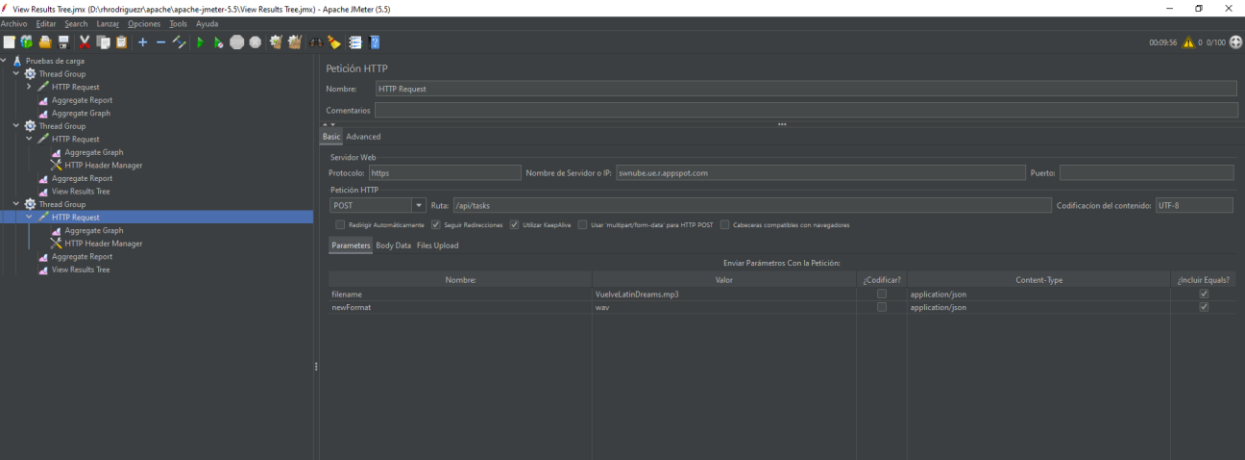
- Se desplegó la capa web (aplicación Flask) en el servicio Google App Engine, que de forma administrada se encarga del escalamiento de la infraestructura necesaria. Únicamente debemos realizar un archivo de configuración y proveer el código a ejecutar.
- Se eliminó el balanceador de carga que recibía las peticiones de la capa web, ya que esta función es realizada de forma automática por Google App Engine.
- Se mantienen los componentes de la base de datos PostgreSQL para almacenamiento de datos de las tareas realizadas y credenciales de los usuarios, el bucket en Cloud Storage para almacenamiento de los archivos de audio y Pub/Sub para la creación de una cola de mensajes que almacena las tareas a realizar de forma asincrónica.
- Al igual que con la capa web, se desplegó la capa de la aplicación Worker en el servicio de Google App Engine, que maneja el escalamiento de esta capa de forma automática, al proveerle un archivo de configuración y el código a ejecutar.

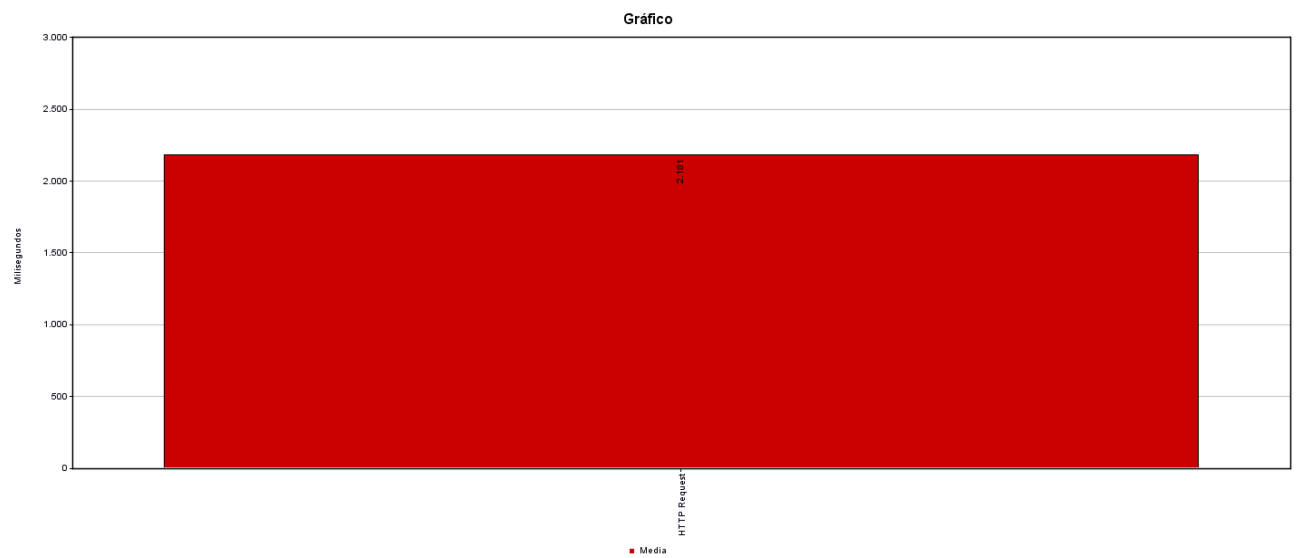




Realizando la revisión de la entrega número 4 los tiempos de respuesta mejoraron considerablemente, también no se tiene porcentaje de error y el tiempo de la media es mucho mejor.

Escenario 2





Para el escenario 2 mejoran también de manera significativa los tiempos de respuesta, no se tiene porcentaje de error. El rendimiento es de 10,1 archivos por minuto.