# Taller 1 – Sistema Conversión Cloud

Gina Eveling Posada , Martin Daniel Rincón, Juan Camilo Muñoz, Felipe Serrano MINE semestre 202410

Universidad de los Andes, Bogotá, Colombia {g.posadas, md.rincon, jc.munozc12, ff.serrano42}@uniandes.edu.co Fecha de presentación: abril 09 de 2024

github, video sustentacion

# 1) Arquitectura de referencia

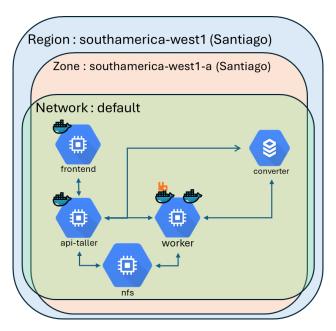


Ilustración 1 Arquitectura general de la aplicación

La aplicación ha sido desarrollada implementando 4 componentes de Docker, un servicio CLOUD SQL, al igual que UN servidor (MV) que sirve de servidor NFS para el almacenamiento de archivos. Estos componentes interactúan en la nube de GCP, dentro de la misma región/zona y red para poder desplegar y cumplir con la funcionalidad requerida para convertir archivos de diferentes (docx, pptx, xlsx, odt) formatos a PDF:

#### 1. Frontend:

- Desarrollado utilizando Streamlit Web API para garantizar una presentación ágil y eficiente para el usuario.
- Facilita la interacción con los servicios del API Rest del Backend a través del protocolo HTTP.

#### 2. Api-taller:

- Implementado en el framework FastApi, proporcionando todas las funcionalidades de negocio.
- Incluye características como registro y login de usuarios mediante JSON Web Tokens (JWT).

 Gestiona la carga y descarga de archivos, así como la gestión de mensajes en la cola de conversión de archivos.

### 3. Converter (Base de Datos):

 Para esta entrega se está utilizando el servicio SQL ofrecido por la plataforma GCP (CloudSQL), como motor de db se utiliza PostgreSQL para alojar las tablas relacionadas con usuarios y las tareas de conversión de documentos.

#### 4. Worker (Cola de Mensajes rabbitMQ):

 Cada solicitud de conversión genera un mensaje en la cola de rabbitMQ la cual se encuentra sobre su propio contenedor dentro de la maquina denominada Worker

### 5. Worker (Consumer):

 Programa en pyhton, ejecutado en su propio contenedor, el cual Procesa los mensajes puestos en la cola rabbit para la conversión de archivos de manera asíncrona. Recupera la ruta desde la que debe leer el archivo desde la base de datos y hace llamado al comando en Linux para el libreoffice, y así convertir utilizando este componente gratuito un documento a formato pdf

subprocess.call(['soffice', '--headless', '--convert-to', 'pdf', '--outdir', upload\_folder, source\_file\_without\_first\_slash])

- Por último guarda la ruta en la que se almaceno el archivo en formato pdf
- Este contenedor tiene como base una imagen sobre la cual se instaló Libreoffice lo que lo hace mas escalable, ya que libreoffice hace parte de la imagen y no de la máquina.

```
RUN apt-get update && apt-get install -y --no-install-recommends \
libreoffice \
&& rm -rf /var/lib/apt/lists/*
```

Ilustración 2 Comando para instalar libreoffice en la imagen

#### 6. NFS (máquina virtual):

• Máquina virtual que sirve como servidor NFS para que la maquina api y worker puedan intercambiar los archivos. Para la configuración de esta máquina se siguieron los pasos del VIDEO del siguiente enlace <a href="https://jpadillaa.hashnode.dev/tutorial-nfs-network-file-system">https://jpadillaa.hashnode.dev/tutorial-nfs-network-file-system</a>

A continuación se colocan imágenes de referencia¹ de la arquitectura desplegada en la nube publica GCP



Ilustración 3 Lista de máquinas virtuales.

<sup>&</sup>lt;sup>1</sup> Algunas de las imágenes pueden no coincidir con las IPs del video, ya que se habían acabado los créditos de la cuenta donde se hizo el despliegue

Nombre de maquina	Características
api-taller; frontend; worker	Tipo de maquina: e2-medium (1-2 vcpu ; 4 Gig RAM; OS debian-12-bookworm-v20240312; SSD 30 GB)
server-nfs	Tipo de maquina: e2-standard-4 (4 vcpu ; 16 Gig RAM; OS ubuntu-
	2204-jammy-v20240319; SSD 250 GB)



Ilustración 4 SQL como servicio GCP2

Nombre de instancia	Características
converter	PostgresSQL15; Enterprise; Sandbox (2Vcpus;8GB RAM; Storage 10GB)

### Aspectos transversales para la aplicación:

#### Seguridad:

 Se implementan prácticas de seguridad, como la gestión adecuada de tokens JWT y la validación de datos de entrada para prevenir vulnerabilidades.

#### Despliegue:

- Se utiliza Docker Compose para la creación de los contenedores de la aplicación, facilitando el despliegue y la escalabilidad (Se elimino el contenedor de Postgress, ya que se esta utilizando la base de datos como un servicio en GCP)
- La infraestructura puede ser fácilmente replicada y gestionada, garantizando consistencia en diferentes entornos.
- No se deben crear variables ni conexiones de "red" a nivel de contenedores, lo que hace más fácil su despliegue, ya que reduce la carga de dependencia de comandos en el contenedor.
- Es importante que al momento de crear los servicios en GCP, se:
  - Creen en la misma zona y región
  - Se modifiquen las reglas de firewall, en la capa de seguridad de la red para abrir los puertos
    - 8501;5001;5432;5672

### • Desarrollo y documentación:

- Manejo de versiones de aplicación con repositorio de desarrollo unificado en GITHUB.
- Set de Pruebas de consumo de servicios apoyados con la aplicación de POSTMAN y JMETER

#### • Escalabilidad:

\_

<sup>&</sup>lt;sup>2</sup> Esta imagen puede ser diferente de la del video ya que se habían acabado los créditos en la cuenta donde se hizo el despliegue

- La arquitectura de contenedores permite escalar los distintos componentes de manera independiente, asegurando la capacidad de manejar cargas variables.
- O NO se instalaron componentes dentro de las maquinas virtuales, y todo lo necesario esta dentro de la imagen de cada componente, lo que hace que sea más fácil escalar ya que no requiere prerrequisitos ni actividades manuales para el óptimo funcionamiento del sistema

# 2) Paso a paso Despliegue en una nube pública (GCP-AWS):

a) Descargar las siguientes imágenes desde docker hub

docker pull rabbitmq --> el rabbit

docker pull ffserrano42/fastapi:2203204 --> imagen de fastapi

docker pull ffserrano42/t1\_streamlit:02192024--> Imagen del streamlit

docker pull ffserrano42/consumer:24032024--> Consumidor de la cola

b) Comandos de docker útiles

docker rm [id del contenedor] --> elimina el contenedor

docker rmi [id de la imagen] --> elimina la imagen

-v /home/[nombreusuariogcp]/remote\_folder:/remote\_folder → hace la adición de una ruta local a docker para que puedan hacer llamados los programas a una ruta en específico.

- c) Pasos para levantar la app por primera vez en la máquina virtual
  - (1) docker run --name rabbitmq -p 5672:5672 -p 15672:15672 [id\_imagen]
  - (2) docker run --name fast\_api -p 5001:5001 -e RABBIT=10.194.0.5 -e DATABASE\_URL='postgresql://api:Uniandes2025!@10.66.192.9:5432/converter' -e TOCONVERT=/remote\_folder -v /home/[nombreusuariogcp]/remote\_folder:/remote\_folder [id\_imagen]
    - (a) En las variables de DATABASE\_URL debe ir el valor entre ' ya que se tiene el caracter de ! o de lo contrario Linux lo detecta como un carácter de salida
    - (b) Se debe obtener la ip privada del servicio de SQL de GCP y configurar de manera correcta
  - (3) docker run --name streamlit -p 8501:8501 -e API\_URL=http://10.194.0.2:5001 [id\_imagen]
  - (4) docker run --name consumer --privileged -e RABBITMQ\_URL=10.194.0.5 -e DATABASE='postgresql://api:Uniandes2025!@10.66.192.9:5432/converter' -e TOCONVERT=remote\_folder/ -v /home/[nombreusuariogcp]/remote\_folder:/remote\_folder [id imagen]
    - (a) En las variables de DATABASE\_URL debe ir el valor entre ' ya que se tiene el carácter de !
    - (b) Se debe obtener la ip privada del servicio de SQL de GCP y configurar de manera correcta
    - (c) Es recomendable que este contenedor sea el Ultimo en iniciar, y después de haber depositado un mensaje en la cola (que de rabbit). Esto debido a que por el momento si no hay mensaje en la cola, el consumidor se apagara.
    - (d) Para depositar un mensaje en la cola, favor ingresar al portal WEB y dejar cargado un documento
- d) Pasos para levanta la app después de haber corrido los contenedores y que la maquina se haya apagado
  - (1) Obtener el listado de todos los contenedores con docker ps -a
  - (2) Utilizar el comando docker start [id contenedor] para iniciar el contenedor
  - (3) Se recomienda la siguiente secuencia si se desea levantar, en orden de independencia
  - (4) Levantar la pagina WEB (Streamlit)
  - (5) Levantar el contenedor de rabbit
  - (6) Levantar el API (fast\_api)

- (7) Depositar un mensaje en la cola de conversión (ingresando a la aplicación y dejando un documento)
- (8) Esto solo se hace la PRIMERA vez
- (9) Levantar el consumer

#### (10) ;;;Recomendación!!!

- (a) Antes de levantar el contenedor del Consumer, haber levantado todos los otros contenedores, loguearse en la app y dejar un archivo, esto hará que la cola tenga algún mensaje y así el consumer quede "activo".
- (b) Crear todos los servicios en la misma zona y región (para facilitar la interoperabilidad de componentes). También habilitar los puertos de las aplicaciones en el firewall "lógico" 8501;5001;5432;5672

# 3) Paso a paso despliegue en visual studio Code:

- 1) Descargar el código fuente del portal de github
- 2) Para ejecutar el consumer, se debería:
  - i) Ejecutar desde una maquina Linux, que tenga instalado libreoffice
  - ii) Ejecutar desde una maquina Windows que tenga UNICAMENTE instalado <u>libreoffice</u> (no recomendado)
- 3) Instalar los módulos correspondientes a cada una de las soluciones, algunos son
  - a. Fast\_api
  - b. Streamlit
  - c. Jose (Jwt)
  - d. Pika (gestión de colas de RabbitMq)
  - e. Chardet (gestión en el frontend)

La estructura del código fuente se vera de la siguiente manera

Ilustración 5 Estructura codigo Visual Studio

#### Carpeta API

- o Contiene el proyecto en FAST\_API el cual es el API de integración.
  - Para ejecutar el proyecto será necesario estar ubicado sobre el proyecto y en una terminal escribir el comando **uvicorn app:app --port 5001 --reload**
- Carpeta WEB
  - o Contiene el proyecto en streamlit el cual es el FRONTEND
  - Para ejecutar el proyecto será necesario estar ubicado sobre el proyecto y en una terminal escribir el comando streamlit run Login.py
- Carpeta Rabbit
  - o Contiene el proyecto en (consola) python el cual es el consumer de mensajes de rabbit

- Su función es leer el mensaje de la cola, que contiene el id del documento. Luego buscar ese registro en la base de datos postgress, obtener del campo <u>source file</u> que es la ruta compartida donde está el archivo físico. Posterior, ejecuta el siguiente llamado al proceso de libreoffice <u>subprocess.call(['soffice', '--headless', '--convert-to', 'pdf', '--outdir', upload\_folder, source\_file\_without\_first\_slash])</u> después de ejecutar el comando, registra en la base de datos, en la columna <u>pdf\_file</u>, la ruta donde se almaceno el archivo convertido
- Para ejecutar el proyecto será necesario estar ubicado sobre la carpeta y ejecutar la clase converter.py
  - Las otras clases dentro del proyectos son referencia a futuras mejoras (ejercicios) que se deseaban implementar, pero por motivo de tiempo no fue posible eliminarlas
- Recomendaciones para construir las imágenes desde visual studio:
  - Existe un archivo llamado docker-compose.yml, el cual es el compose de todas las imágenes dockerFile de cada proyecto.
  - En este docker-compose, no está la imagen de Rabbit, la cual debe ser descargarda y ejecutada por aparte.
    - docker pull rabbitmq
    - docker run --name rabbitmq -p 5672:5672 -p 15672:15672 [El id de la imagen]
  - Para que se genere automáticamente las imágenes, desde visual studio se puede utilizar el comando docker-compose build. y este comando construirá las imágenes de
    - Api (Fast\_Api)
    - Consumer (Pyhton)
    - Frontend (Streamlit)
  - o En caso de ejecutar el comando docker-compose up, pueden presentar fallos si las variables de entorno no se han configurado correctamente en la maquina ó si no están asociadas los contenedores a la misma red. Para esta etapa del taller los contenedores se iban a desplegar en múltiples maquinas, se elimino el uso de una network en común y se recomienda que se mantenga este esquema también en desarrollo.
  - Adicionalmente, la imagen de postgress no se genera de manera automática con el dockerCompose, ya que se desea hacer la prueba directamente conectando el sistema a servicio googleSQL o AWS RDS. Sin embargo se dejó dentro del código fuente una carpeta de postgress (con un dockerfile) por si más adelante los créditos se agotan y se debe utilizar una imagen nuevamente.

# 4) Limitaciones de la solución

- Aún no se ha implementado balanceadores de carga para un escalamiento horizontal de la aplicación sin tener que configurar todos los contenedores. Por lo que uno de los límites es la capacidad máxima de peticiones HTTP que acepta cada una de las maquinas, al igual que su proceso de escalabilidad ( con la arquitectura actual se debe realizar muchos procesos manuales)
- Una vez desplegada la solución en las maquina virtuales de GCP al igual que la base de datos SQL de GCP, se recomienda:
  - 1. Apagar las máquinas virtuales
  - 2. Eliminar la base de datos de GCP
  - Esto con el fin de no consumir los créditos, ya que incluso si el servicio de SQL esta detenido, este sigue consumiendo créditos.
- 3. Por el momento, las maquinas virtuales al igual que el servicio SQL en GCP, no están en redundancias de zonas, y todas se encuentran en la misma zona southamerica-west1-a (esto con el fin de facilitar la comunicación entre componentes). Esto implica que si existe un fallo en esa zona, TODA la aplicación presentaría una falla
- 4. A pesar de haber habilitado los puertos https y realizar la validación de CORS, las peticiones al API en GCP únicamente se pueden realizar utilizando el protocolo http.
- 5. Al contar con una MAQUINA VIRTUAL como servidor NFS, si este servidor no se encuentra activo, la funcionalidad principal (conversión) de todo el sistema fallaría, motivo por el cual la configuración y administración de esta maquina virtual es vital para el correcto funcionamiento del sistema
  - Este es una limitante crítica y de pronto para futuras iteraciones se podía remplazar este servidor, por un servicio como Filestore (en el caso de GCP)
- 6. Al momento de eliminar la base de datos de GCP, la ip que tienen los contenedores del API y Consumer a la que hacen referencia al servicio SQL queda obsoleta (cuando se crea una nueva instancia se asigna otra IP). Esto hace que cada vez que se cree una nueva instancia de base de datos, se deba:
  - 1. Eliminar los contenedores que están corriendo
  - Volver a crear los contenedores con sus correspondientes comando pero modificando la ip PRIVADA de la base de datos.<sup>3</sup>



Ilustración 6 IP privada de la base de datos

<sup>&</sup>lt;sup>3</sup> También es recomendable habilitar la conexión de la base de datos desde fuera, ya que en algunas ocasiones GCP crea la ip privada dentro de otra red, que no están las maquinas y es mas complejo la conectividad internamente.

# 5) Conclusiones

- El uso de contenedores para el despliegue de la solución reduce la carga manual de configuración de componentes uno a uno, además de permitir contar con una arquitectura de solución desacoplada y escalable eficiente ya que solo se deber crecer en el componente necesario (front, back, consumer).
- La curva de aprendizaje en los múltiples servicios dificulta el tiempo de despliegue final en la nube pública. Es indispensable tener un conocimiento claro de la arquitectura "de red" de los distintos componentes para que la interconexión entre distintas maquinas no se vuelva un inconveniente mayor.
- 3. Se debe conocer los distintos modelos de cobro que tienen las nubes públicas, ya que dependiendo del servicio y de la nube publica se pueden utilizar estrategias para optimizar el costo total de la solución.
- 4. A medida que la arquitectura de la solución se va complejizando con máquinas/servicios de nube, el ambiente de desarrollo también se va complejizando para poder hacer una transición sencilla entre desarrollo y despliegue. Es muy importante tener esto definido desde el inicio ya que puede demorar los tiempos de inicio de desarrollo al igual que el despliegue final.