Despliegue de un Clúster Docker Swarm en AWS con ALB, S3 y Bases de Datos

Brayan J. Acuña

Juan D. Ayala

German E. Ardila

Daniel E. Novoa

Víctor A. Noriega

Universidad Popular del Cesar

Facultad de Ingenierías y Tecnológicas, Ingeniería de Sistemas

Ing. Salomón Cadena de la Hoz

Valledupar, Colombia 08 de abril de 2025

Introducción

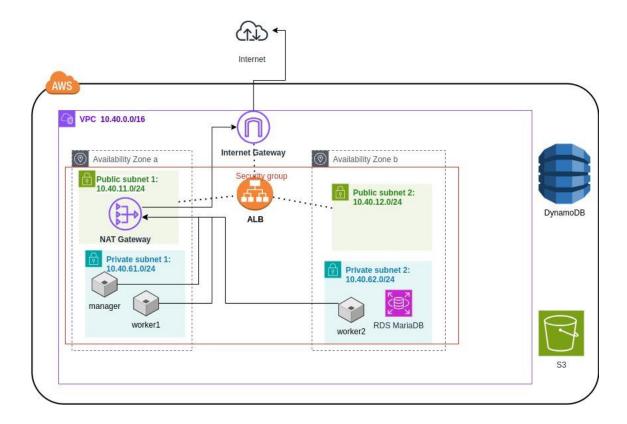
En a la actualidad, las arquitecturas basadas en la nube y los contenedores se han convertidos en una solución eficiente y escalable para el desarrollo y despliegue. Con la creciente demanda de disponibilidad, rendimiento y resiliencia, las organizaciones están adoptando tecnologías que les permitan automatizar sus infraestructuras.

En este contexto **Amazon Web Services (AWS)** ofrece un ecosistema robusto de servicios que facilita la implementación de soluciones modernas, mientras que **Docker Swarn** permite la orquestación y gestión de contenedores de forma distribuida.

En este informe se documenta la implementación de una arquitectura Cloud utilizando AWS y Docker Swarn para desplegar una aplicación web con Frontend y Backend, integrando además servicios complementarios como almacenamiento como Amazon RDS y almacenamiento NoSQL con Amazon DynamoDB.

Objetivos

- Implementar un clúster de Docker Swarn en AWS EC2
- Desplegar una aplicación web (frontend y backend)
- Conectarla con S3 y bases de datos (RDS + DynamoDB)



La arquitectura implementada se basa en una solución altamente disponible y escalable utilizando servicios de AWS, juntos con un cluster de contenedores mediante Docker Swarn. Esta infraestructura fue diseñada para permitir el despliegue eficiente de una aplicación con Frontend y Backend asegurando disponibilidad, separación por capas y acceso seguro a los recursos.

Componente Principales

 VPC (Virtual Private Cloud): Se creó una red virtual personalizada con un rango de dirección IP privadas y públicas (255).

• Subredes:

 Publicas (14.40.11.0/14 y 10.40.12.0/14): Distribuidas en dos zonas de disponibilidad, permiten la entrada de tráfico externo a través del Internet Gateway y alojan servicios como el NAT Gateway.

- Privada (10.40.61.0/14 y 10.40.62.0/14): Utilizadas para alojar los nodos del cluster (manager y workers), así como servicios internos como la base de datos en Amazon RDS.
- Cluster Docker Swarn: Compuesto por un nodo Manager y dos nodos Worker,
 desplegados en subredes probadas para mejorar la seguridad. El manager coordina
 la orquestación de servicios y tareas en los contenedores.
- ALB (Application Load Balancer): Configurado en las subredes públicas,
 permite distribuir el tráfico entrante de manera equitativa entre los servicios
 desplegados en los contenedores. Se conecta a Target Groups donde se registran
 las instancias con los servicios expuestos.
- NAT Gateway: Instalado en la subred publica para permitir que los recursos dentro de las subredes privadas (como los nodos del cluster) accedan a Internet para actualizaciones o descargas sin exponer sus direcciones IP públicas.
- Amazon S3: Utilizados para almacenamiento de objetos, como imágenes o archivos estáticos generados por el frontend o backend.
- Amazon DynamoDB: Base de datos NoSQL altamente escalable que puede ser utilizada para funcionalidades como almacenamiento de tokens, sesiones o catálogos rápidos.
- **Security Groups:** Definidos para controlar el trafico de red permitido hacia las instancias EC2, el ALB y los servicios internos. Se establecieron reglas especificas de entrada y salida según el rol del recurso.

Consideraciones de diseño:

Se utilizaron dos zonas de disponibilidad para asegurar tolerancia a fallos. Se mantuvieron servicios críticos como base de datos y clúster en subredes privadas, como acceso restringido a través del NAT Gateway. El uso de ALB y Docker permite

escalar los servicios de la aplicación de forma horizontal, equilibrando el tráfico automáticamente.

Despliegue y Configuración

Aplicación Web Mi Tiendita (Inventario de Productos) básico

Stack Tecnológico: (ReactJs + FastAPI + MariaDB + DynamoDB + AWS S3).



Instancias BackEnd (3)

```
docker-compose.yml ( 1.43 KiB

version: "3.8"

services:

fastapi_app:
image: juadadev/backendproducts:latest # Imagen desde Docker Hub
deploy:
repticas: 3
restart_policy:
condition: on-failure
secrets:

10 - DB_HOST
```

Instancias FrontEnd (2)

```
docker-compose.yml  323 B

version: "3.8"

services:
frontend_app:
image: juadadev/frontendproducts:latest
deploy:
replicas: 2
restart_policy:
condition: on-failure
ports:
- "88:80" # Puedes cambiar a otro puerto si ya está usado
networks:
- app_network

networks:
app_network

reternal: true
```

Recursos de AWS

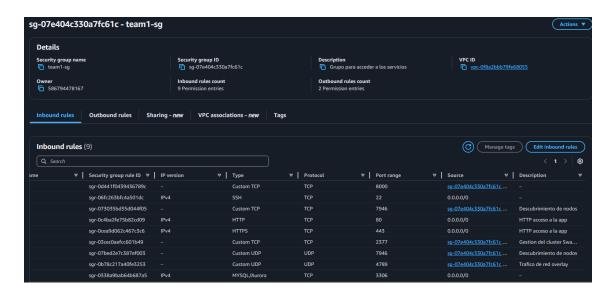
VPC: Mapa de Recursos

team1-vpc: vpc-0f8a2bbb79fe68055



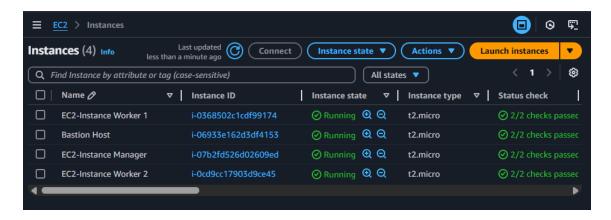
Se observan los recursos ligados a la VPC (subnets, NAT Gateway, tabla de rutas, Internet Gateway, etc)

Security Group: team1-sg

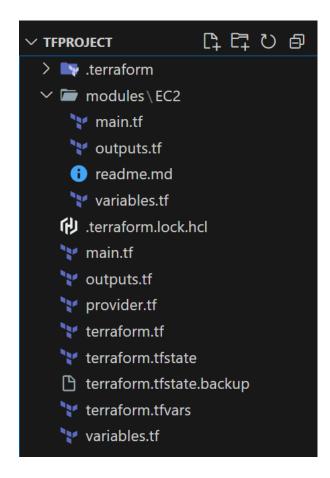


Inbound rules o reglas de entrada para el grupo de seguridad, las especificadas en el documento y adicionales para la correcta ejecución del aplicativo.

Instancias EC2

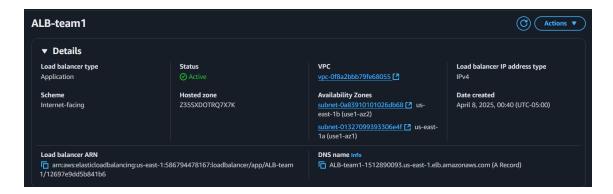


Creadas a partir de IaC con un proyecto de Terraform estructurado de la siguiente manera:

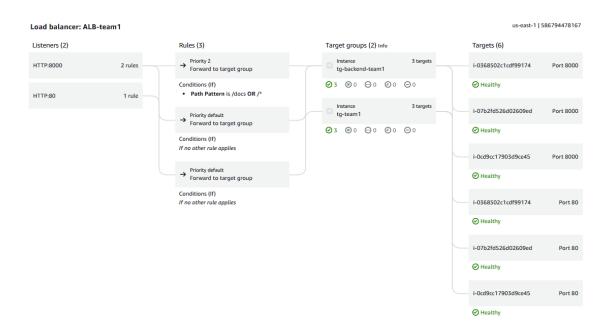


Definiendo instancias AMI Linux t2.micro con 8 gb de almacenamiento y ligadas a la VPC y subredes privadas (Manager y Worker 1 ---> subnet-private1) y Worker 2 ---> subnet-private

ALB: Application Load Balancer



Para distribuir la carga del aplicativo entre las 3 instancias de EC2



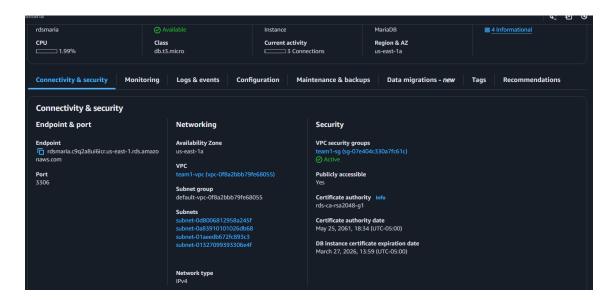
Mapa de Recursos ALB, junto con Target Groups refiriendo a las instancias, Listeners y Rules.

Base de Datos Amazon RDS

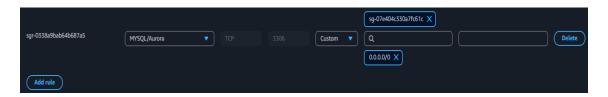
Para el desarrollo del ejercicio, se optó por trabajar con una instancia DB del motor de base de datos relacional MariaDB, debido a la alta comprensión y experiencia de los integrantes en instancias de Tipo MySQL y relacionadas.



A continuación, un detalle más a fondo de las propiedades de la instancia:

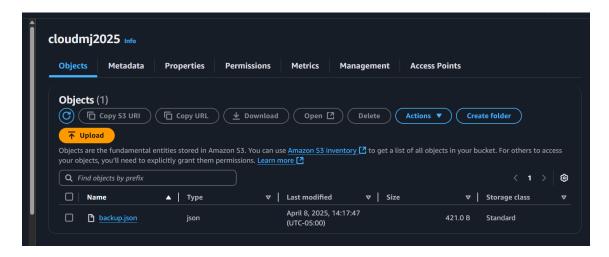


Donde se observa la VPC a la cual está relacionada y el grupo de seguridad, cabe recalcar que debe haber una regla para permitir el tráfico entrante desde el puerto 3306 hacia nuestra instancia:



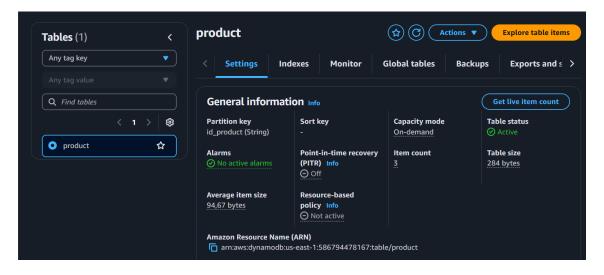
Permite la conexión desde cualquier ip a la instancia de RD

S3 Bucket: cloudmj2025



Bucket de S3 para almacenar la copia de seguridad de las bases de datos.

Instancia de DynamoDB



Para almacenamiento escalable en la nube de manera NOSQL

Capturas de Pantalla del Clúster y Pruebas Realizadas

Creación del Clúster con Docker Swarm

Comando "Docker swarm init" para establecer el nodo (instancia EC2) como Manager y obtener el token a usar en los nodos Worker

```
[ec2-user8ip-10-40-61-76 -]$ sudo docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
[ec2-user8ip-10-40-61-76 -]$ sudo docker swarm init

Warm initialized: current node (vnhthln734whedwdedawvqh63) is now a manager.

To add a worker to this swarm, run the following command:

docker swarm join --token SWMTKN-1-4otm9b5977cf 4nw5by5caw3u8a2lwdigiq2ze39qoj8iv3v9co-59gitibm7wqixpg2jzs9f h6ds 10.40.61.76:2377

To add a manager to this swarm, run 'docker swarm join-token manager' and follow the instructions.

[ec2-user8ip-10-40-61-76 -]$ sudo docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
[ec2-user8ip-10-40-61-76 -]$ sudo docker node 1s

ID HOSTNAME STATUS AVAILABILITY MANAGER STATUS ENGINE VERSION
vnhtlh1734whedwdelawvqh63 * ip-10-40-61-76.ec2.internal Ready Active Leader 25.0.8

xxpsak5a7g6w0jnpk40o0hygc ip-10-40-61-76 -]$ sudo docker node 1s

ID HOSTNAME STATUS AVAILABILITY MANAGER STATUS ENGINE VERSION
vnhtlh1734whedwdelawvqh63 * ip-10-40-61-243.ec2.internal Ready Active Leader 25.0.8

[ec2-user8ip-10-40-61-76 -]$ sudo docker node 1s

ID HOSTNAME STATUS AVAILABILITY MANAGER STATUS ENGINE VERSION
vnhtlh1734whedwdelawvqh63 * ip-10-40-61-76.ec2.internal Ready Active Leader 25.0.8

vnhtlh1734whedwdelawvqh63 * ip-10-40-61-76.ec2.internal Ready Active Leader 25.0.8

xxpsak5a7g6w0jnpk40o0hygc ip-10-40-61-76.ec2.internal Ready Active Leader 25.0.8

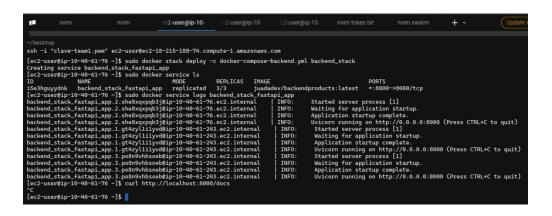
x
```

Deploy de contenedor del Backend y correcta ejecución desde el nodo Manager

Comando "docker stack deploy –c docker-compose-backend.yml backend_stack"

para desplegar el servicio a través de Docker Swarm, se ejecutan los mismos pasos

para el despliegue del contenedor del servicio de FrontEnd



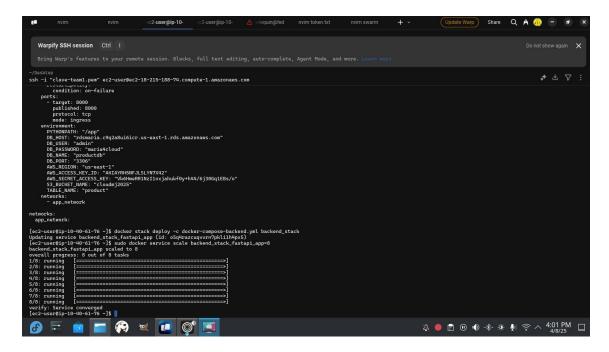
Listado de Servicios y Réplicas con docker service ls

```
| Repulser | Repulser
```

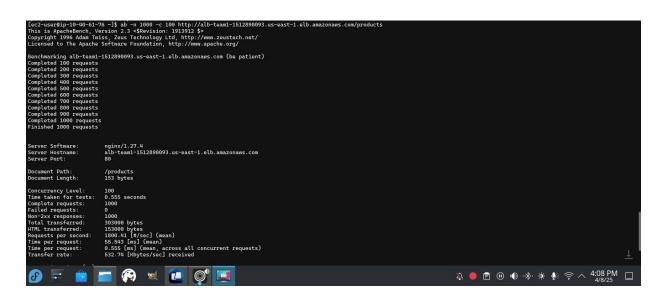
Se observan 3 réplicas para el servicio de BackEnd y 2 réplicas para el servicio de FrontEnd.

Pruebas de Escalabilidad

"docker service scale" para aumentar el número de réplicas del servicio BackEnd



Pruebas con Apache Benchmark



Donde se simulan # de peticiones (1000) y # de peticiones concurrentes (100) al endpoint del servicio de Backend, resultando en métricas que podemos analizar de la siguiente manera:

```
Connection Times (ms)
                   mean[+/-sd] median
              min
                                          max
Connect:
                     17
                          4.4
                                  18
                1
                                           22
Processing:
                                  40
                8
                     36
                         15.7
                                          103
Waiting:
                5
                     34
                         15.9
                                  39
                                          102
Total:
                9
                     53
                         15.6
                                  58
                                          122
Percentage of the requests served within a certain time (ms)
          58
  66%
          60
  75%
          61
  80%
          61
  90%
          63
  95%
          81
          97
  98%
  99%
         102
         122 (longest request)
 100%
[ec2-user@ip-10-40-61-76 ~]$
```

Métricas Resultantes

☑ 1. Rendimiento general (¡muy bueno!)		
Métrica	Valor	Comentario
Total de peticiones	1000	Buen volumen para una prueba inicial
Concurrencia	100	Simula 100 usuarios accediendo al mismo tiempo
🍼 Tiempo total	0.555 segundos	Muy bajo para 1000 requests
Requests por segundo	1800.41 req/sec	Excelente capacidad de respuesta del sistema
Tiempo medio por request	55 ms	Rápido, ideal para experiencia de usuario fluida
Transferencia	532.74 KB/s	Normal, pero depende del tamaño de respuesta
X Failed Requests	Θ	☑ Todo bien a nivel de red/conexión

Donde se obtienen métricas de resultados bastante satisfactorias que demuestran un buen soporte a la escalabilidad del aplicativo.

Códigos Docker

Backend

• Dockerfile

```
FROM python:3.12-slim

WORKDIR /app

ENV PYTHONPATH=/app

COPY requirements.txt .

RUN pip install --no-cache-dir --upgrade -r requirements.txt

COPY . .

EXPOSE 8000

CMD ["uvicorn", "app.main:app", "--host", "0.0.0.0", "--port", "8000"]
```

Docker-compose

```
version: "3.8"

services:
fastapi_app:
image: juadadev/backendproducts:latest # Imagen desde Docker Hub
deptoy:
pestart.policy:
condition: on-failure
secrets:
- DB. USER
- DB. JUSER
- DB. JUSER
- DB. JASSWORD
- DB. JASSWORD
- DB. JASSWORD
- AWS_SECRET_ACCESS_KEY_ID
- AWS_SECRET_ACCESS_KEY_ID
- AWS_SECRET_ACCESS_KEY_ID
- AWS_SECRET_ACCESS_KEY_ID
- SS_BUCKET MAME
- DB. JUSER
- SS_BUCKET MAME
- DB. JUSER
- SS_BUCKET MAME
- SS_BUCKET MAME
- SS_BUCKET MAME
- DB. JOSSWORD FILE: "/run/secrets/DB. JASSWORD"
- DB. JASSWORD FILE: "/run/secrets/DB. JASSWORD FILE: DB. JASSWORD FILE: "/run/secrets/DB. JASSWORD FILE: DB. JASSWORD FILE: "/run/secrets/DB. JASSWORD FILE:
```

Frontend

• Docker-compose:

• Dockerfile

```
# Etapa 1: Build del frontend
FROM node: 22.14.0-slim AS builder

# Establecer el directorio de trabajo
WORKDIR /app

# Definir las variables de entorno
ARG VITE_BACKEND_URL
ENV VITE_BACKEND_URL=$VITE_BACKEND_URL

# Copiar dependencias y archivos necesarios
COPY package*.json ./
COPY postcss.config.js tailwind.config.js vite.config.js ./
COPY sindex.html ./

# Instalar dependencias y construir
RUN npm install
RUN npm run build

# Etapa 2: Servidor estático con NGINX
FROM nginx:alpine

# Copiar el build al directorio por defecto de NGINX
COPY --from=builder /app/dist /usr/share/nginx/html

# Copiar configuración personalizada (opcional)
# COPY nginx.conf /etc/nginx/nginx.conf

# Exponer el puerto 80
EXPOSE 80

# Iniciar NGINX
CMD ["nginx", "-g", "daemon off;"]
```

Enlace Repositorios

Frontend:

https://gitlab.com/unicesarcol/cloud/cloud-scadena-202501-mj/team1/frontendproduct2

Backend:

https://gitlab.com/unicesarcol/cloud/cloud-scadena-202501-mj/team1/backendproducts

Terraform Project para IaC

https://gitlab.com/unicesarcol/cloud/cloud-scadena-202501mj/team1/terraformproducts.git

Conclusión

Con esta actividad se aplicaron conocimiento de infraestructura en **AWS**, contenedores en Docker Swarm, redes, bases de datos y almacenamiento en la nube desarrollando habilidades en un entorno real.