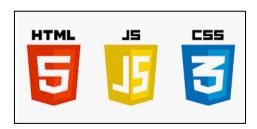
ETL Project II: Docker

Aplicación web con Flask y MongoDB









Profesor: Jorge Peralta Dolz

Escuela: La Salle Bonanova

Máster: Data Science

Asignatura: Infraestructuras de computación



Índice

Introducción	3
Arquitectura	3
Dockerfile	
Docker-Compose	
Desarrollo – Flujo	
Dificultades encontradas	18
Bibliografía	19

Introducción

En este proyecto se explicará el funcionamiento de una aplicación llamada, "Supergestor Marvel", en Flask conectada con el sistema gestor de base de datos MongoDB, sobre el lenguaje de programación Python, además esta aplicación se automatizará mediante Docker Compose aplicando conceptos de microservicios.

El esqueleto de este proyecto ha sido obtenido a través de la plataforma Crashell, donde siguiendo este pequeño tutorial se ha ido entendiendo, paso a paso, el funcionamiento de la aplicación inicial.

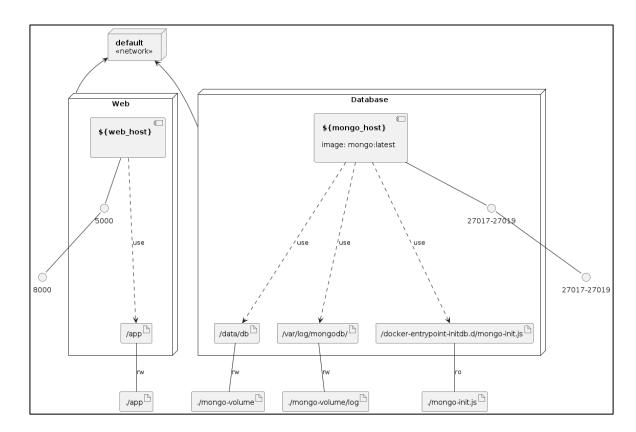
Una vez entendido el proyecto base, se han realizado una serie de cambios que han permitido crear "Supergestor Marvel":

"Supergestor Marvel" es una aplicación web que permite al usuario buscar cualquier super héroe de Marvel que se encuentre dentro de la Web Api de Marvel, y después en base al número de comics, series, historias y eventos el usuario decide si desea guardarlo en su base de datos o no.

Arquitectura

En la *llustración 1* se puede observar la arquitectura creada mediante Docker-Compose. Ambos contenedores se encuentran en la misma red "default". El contenedor web internamente usa el puerto 5000 y se accede a él a través del puerto 8000, es decir, el tráfico que llega al puerto 8000 de la máquina host se está redirigiendo al puerto 5000 del contenedor Docker. Por lo tanto, si desde la máquina virtual se quiere acceder a la web se tendrá que usar la notación "localhost:8000/".

Por otro lado, la base de datos abre los puertos en el rango 2701[7-9] con el rango homónimo 2701[7-9] que se encuentra en el contenedor, por lo tanto, serán expuestos los mismos puertos. Además, se crean diferentes volúmenes para persistir datos y además con el archivo "mongo-init.js" se crea un usuario con sus respectivas credenciales además de una base de datos.



Dockerfile

El Dockerfile original es el que se muestra en la *llustración 2*, se basa en Python:3.7-alpine que es una distribución Linux ligera que permite el funcionamiento del aplicativo. Se probó de cambiar a la última versión estable reducida, Python:3.12-slim, pero esta no funcionaba.

En la línea 5 se crea dentro del contenedor un directorio /app y acto seguido se declara este como directorio de trabajo. A continuación, en la línea 7 se copia todo el directorio desde donde se esta ejecutando el Dockerfile al directorio destino /app.

El paso de la línea numero 9 ha sido eliminado para la "Supergestor Marvel", ya que se trata de la instalación del compilador gcc y de las cabeceras de Linux que para el proyecto no son necesarias:

- No hay dependencias de Python con extensiones C.
- No se usan extensiones de Flask que requieran compilación.
- No se compilan recursos Frontend.
- En el caso de las cabeceras de Linux, no se interactúa directamente con el kernel.

Por último, se instalan los requerimientos necesarios con la línea 10 (*Ilustración* 3) y se expone el contendor para que escuche en el puerto 5000. Como requerimientos necesarios básicos, se tiene *Flask* ya que esté si que se necesita para el desarrollo de la aplicación, *requests* que permite hacer las llamadas a la api de Marvel y *pymongo* que permite comunicar MongoDB con Python. La librería *xmltodict* no se ha usado, pero se ha considerado dejarla ya que a futuro podría ser de gran utilidad manejar data XML de forma mucho más sencilla.

```
# syntax=docker/dockerfile:1

FROM python:3.7-alpine

RUN mkdir /app
WORKDIR /app
COPY . /app

RUN apk add --no-cache gcc musl-dev linux-headers
RUN pip install -r requirements.txt

EXPOSE 5000
```

Ilustración 2 Ilustración 3

Docker-Compose

Una vez entendido el Dockerfile que construirá la imagen sobre la cual se construirá el contendor encargado de ejecutar la aplicación Flask, se mostrará como a través de Docker-Compose se construyen y se conectan estos dos contenedores. Para ello, vienen declaradas unas variables de entorno predefinidas en un archivo llamado .env (llustración 5).

```
ETL > 👉 docker-compose.yml
     services:
         container name: ${WEB HOST}
         hostname: ${WEB_HOST}
         build: ./app
         environment:
           FLASK APP: ./app.py
           FLASK RUN HOST: 0.0.0.0
           FLASK ENV: development
          - '8000:5000'
20
           - database
         depends on:
           - database
           - ./app:/app
           - default
```

```
ETL > .env

1 WEB_HOST=cs_api
2 |
3 MONGO_HOST=cs_mongodb
4 MONGO_PORT=27017
5 MONGO_USER=root-crashell
6 MONGO_PASS=password-crashell
7 MONGO_DB=db_crashell
```

Ilustración 4

En la *llustración 4* aparece la construcción del contenedor web donde se usan las variables de entorno para declarar los nombres, además en vez de usar la *images* y declarar una imagen, se indica que la imagen que se usara es la que se tiene que construir y que se encuentra en el directorio ./app es decir, se construirá en base al Dockerfile que se ha comentado anteriormente. Se especifican ciertos parámetros para el entorno de Flask que permiten el desarrollo más eficiente y cómodo, se declara la conexión de puertos antes mencionada en la sección y se especifica que esta vinculada al otro servicio que se verá a continuación que es nuestra base de datos MongoDB (database). Por último, se especifica que depende de la base de datos, se crea un volumen para persistir los datos y que estos estén sincronizados y se especifica la red en la que se encontrará este servicio (default), la misma que la base de datos.

Por otro lado, en la *llustración 5* aparece la construcción del contenedor de la base de datos. En este caso la imagen si que es la de *mongo:latest* para poder obtener la última versión estable. De igual forma que en el contenedor web, también se declaran los nombres y ciertas variables a través de las variables de entorno, también se crean volúmenes para persistir datos y se declaran los puertos como se ha comentado al principio de la sección y se establece la red *default*.

```
database:
    image: mongo:latest
    env_file:
        - .env
    container_name: ${MONGO_HOST}
    hostname: ${MONGO_HOST}
    environment:
        MONGO_INITDB_ROOT_USERNAME: ${MONGO_USER}
        MONGO_INITDB_ROOT_PASSWORD: ${MONGO_PASS}
        MONGO_INITDB_DATABASE: ${MONGO_DB}
    volumes:
        - ./mongo-init.js:/docker-entrypoint-initdb.d/mongo-init.js:ro
        - ./mongo-volume:/data/db
        - ./mongo-volume/log:/var/log/mongodb/
    ports:
        - '27017-27019:27017-27019'
    networks:
        - default
    volumes:
    persistent:
```

Ilustración 4

Desarrollo - Flujo

Con tal de hacer una explicación de principio a fin, se explicará paso a paso todo el flujo de funcionamiento de la aplicación y paralelamente se mencionarán por encima fragmentos de código para complementar la explicación.

En primer lugar, suponiendo que el repositorio de la aplicación "Supergestor Marvel" esté disponible para el usuario, éste tendrá que clonar el repositorio. Una vez clonado, deberá de lanzar el comando "docker-compose up -d" para que estos contenedores se levanten y se ejecuten en segundo plano (en la *llustración* 7 se ejecuta el comando con el Dockerfile original, y en la *llustración* 8 se ejecuta el comando con la versión final personalizada, eliminando la instalación necesaria del compilador y de las cabeceras Linux, entre otros, y se observa como disminuye considerablemente el tamaño.

```
✓ Container cs_mongodb Started

 ✓ Container cs_api
root@ETL:/home/didac/Documents/MyLocalRepo/Data-Science-Master/ETL# docker images
                   IMAGE ID
                      IMAGE ID CREATED
fac5269dfc68 41 minutes ago
REPOSITORY
                                                       SIZE
             latest
                                                       222MB
etl-web
            latest
                      2e123a0ccb4b
                                     11 days ago
                                                       757MB
nongo
root@ETL:/home/didac/Documents/MyLocalRepo/Data-Science-Master/ETL#
```

Ilustración 8

Una vez los contenedores están en marcha, ya se puede entrar en el navegador para acceder a la web "localhost:8000/".

La *llustración* 9 muestra la vista de la página web principal con el inspector abierto, la *llustración* 10 muestra el *body* de esta vista y la 11 como ésta interactúa con Flask y devuelve el templete renderizado.

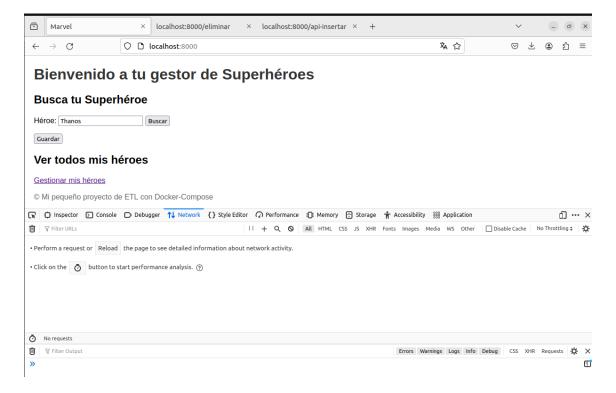


Ilustración 9

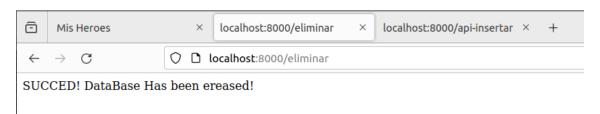
```
<h1>Bienvenido a tu gestor de Superhéroes</h1>
<section id = "FormularioBusqueda">
   <h2>Busca tu Superhéroe</h2>
   <form id="buscarhero">
       <label for="hero">Héroe:</label>
       <input type="text" name="hero" id="hero">
       <button type="button" onclick="sendHero()"> Buscar </button>
<section id = "RespuestasApiMongo">
   <button type="button" onclick="saveHero()"> Guardar </button>
   <div id ="temporal"></div>
<section id = "listarSuperheros">
   <h2>Ver todos mis héroes</h2>
   <a href="{{url_for('list_heros')}}"> Gestionar mis héroes</a>
   © Mi pequeño proyecto de ETL con Docker-Compose
```

```
109
110 @app.route('/', methods=('GET', 'POST'))
111 def index():
112 | return render_template('index.html')
```

Ilustración 11

A parte de la *url* principal, también se disponen de dos *urls* extras que han servido de ayuda durante el desarrollo.

La primera es "localhost:8000/eliminar", esta básicamente cada vez que la llamas elimina todos los elementos de la colección de la base de datos que se está utilizando en la aplicación (*llustración 12 para vista y 13 para código*).



```
def delete_al():
    ok, connection = getMongoClient()
    if not ok:
        return Response(response=json.dumps(connection), mimetype='application/json')
    collection = getCollection(connection, "superheros")
    collection.delete_many({})
    return "SUCCED! DataBase Has been ereased!"
```

En la *llustración 13* se aparecen dos funciones globales, que viene definidas por la *llustración 14*.

```
def getMongoClient():
    return ConnectionMongoDB(DB_SERVER,DB_PORT,DB_USER,DB_PASS,DB_NAME).getDB()

def getCollection(connection,collection_name):
    db = connection[DB_NAME]
    return db[collection_name]
```

Ilustración 14

Así pues, desde nuestra pagina principal (*Ilustración 9*) haciendo clic en el link hacía "Gestionar mis Héroes", éste conduce hacia una página ("localhost:8000/list_heros") donde se consultan y listan todos los superhéroes guardados en la colección de la base de datos (*Ilustración 15 para vista y 16 para código*).



Ilustración 15

Ilustración 16

La segunda *url* que ha ayudado mucho para el desarrollo del proyecto ha sido "localhost:8000/api-insert". Esta *url* inserta unos superhéroes inventados por defecto dentro del código y después lista de forma encadenada y sin espacios los nombres de todos los superhéroes de la colección de la base de datos. De esta manera, permitió testear para el correcto guardado en la base de datos y para el correcto leído (*Ilustración 17 para los superhéroes inventados, 18 para la vista y 19 para el código*).

```
63
64  #Custom Hereos for testing
65  MisSuperheros = [
66  Superhero("Juan", "mortal",23, 3, 6, 6),
67  Superhero("Julio", "lento",254, 2, 41, 2),
68  Superhero("Cesar", "rapido",2355, 3, 49, 1),
69  Superhero("Patroclo", "inutil",234, 9, 43, 8)
70 ]
```

Ilustración 17



Ilustración 18

Ilustración 19

Una vez guardados estos superhéroes en la base de datos, volvemos a "localhost:8000/list_heros" y vemos como ahora si que estos ya se muestran (*Ilustración 20*).



Si ahora, volvemos por un momento a la *llustración 9*, donde se ve que en el texto esta escrito el superhéroe "Thanos" y le damos al botón de buscar, la función de JavaScript de la *llustración 21* recolectara la información la enviará a nuestra aplicación Flask y ésta realizará una llamada a la api de Marvel devolviendo en caso de que el superhéroe exista, el nombre, el número de comics, de series, de historias y de eventos (*llustración 22*). Una vez esta información sea devuelta al Frontend, se rellenará la web con el contenido (*llustración 23*).

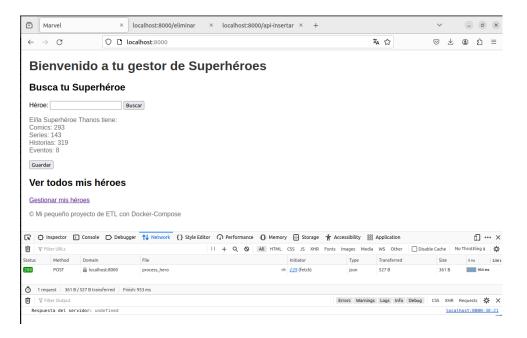


Ilustración 23

Ahora, el usuario es el encargado de escoger si desea guardar este superhéroe en su base de datos o no. En caso afirmativo, deberá pulsar el botón de "Guardar" que activará la función de JavaScript de la *llustración 24* que se conectará con la parte del Backend de Flask tal y como indica la *llustración 25*. El resultado en la vista quedará tal y como indica la *llustración 26*.

Ilustración 25

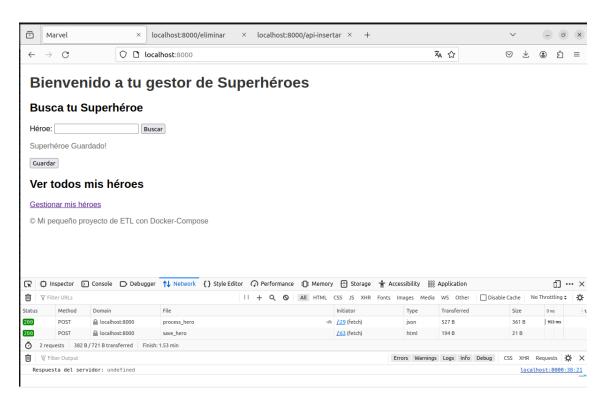


Ilustración 26

Cuando se ha guardado el superhéroe el usuario puede ir a su página de gestión de superhéroes y puede ver que efectivamente Thanos, se ha guardado correctamente (*Ilustración 27*).

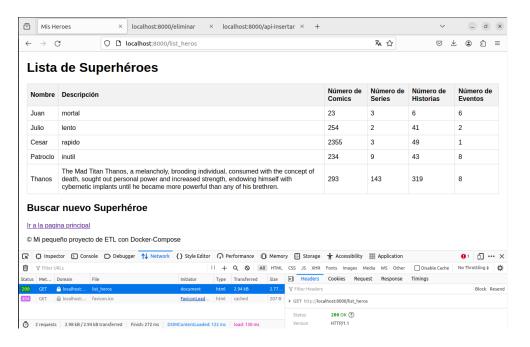


Ilustración 27

El body de la vista que lista los superhéroes queda definido en la Ilustración 28.

```
<h1>Lista de Superhéroes</h1>
          Nombre
          Descripción
          Número de Comics
          Número de Series
          Número de Historias
          Número de Eventos
       {% for hero in allHeros %}
              {{ hero['nombre'] }}
              {{ hero['descripcion']}}
             {{ hero['comics'] }}
{{ hero['series'] }}
             {{ hero['historias'] }}
{{ hero['eventos'] }}
       {% endfor %}
   <h2>Buscar nuevo Superhéroe</h2>
   <a href="{{url_for('index')}}"> Ir a la pagina principal</a>
   © Mi pequeño proyecto de ETL con Docker-Compose
```

Para que todo el desarrollo sea posible y eficiente se han trabajado con dos clases:

- Clase Superhero: permite guardar la información necesaria del superhéroe y manejarla de forma rápida y sencilla para su guardado y listado (Ilustración 29).
- Clase ConnectionMongoDB: permite realizar la conexión con la base de datos o devolver una excepción en caso de error (Ilustración 30)

```
37 @dataclass
38 class Superhero:
39 nombre: str
40 descripcion: str
41 comics: int
42 series: int
43 historias: int
44 eventos: int
45
46 def toDBCollection (self):
47 return {
48 "nombre":self.nombre,
49 "descripcion": self.descripcion,
50 "comics":self.comics,
51 "series": self.series,
52 "historias": self.historias,
53 "eventos": self.eventos
54 }
```

Ilustración 29

Por último, el directorio de archivos y carpetas del proyecto quedaría tal y como se muestra en la *Ilustración 31*.

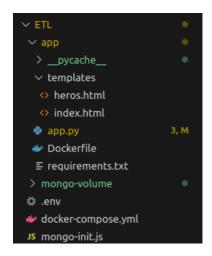


Ilustración 31

Dificultades encontradas

Durante la realización del proyecto se han encontrado múltiples dificultades, pero sin embargo, la que más tiempo consumió fue el no tener la versión correcta de Ubuntu y que esto resultaba altamente conflictivo con Docker a la hora de montar el servidor web Flask. Siempre, daba igual que repositorio se cogiera, resultaba con el error "http: Invalid Host Header". Este error se solvento gracias a que se probo de instalar Docker desde otro ordenador y clonando los mismos repositorios con los que se estaba testeando y viendo que en el Mac si que funcionaban. Así pues, esto llevo a la conclusión de que el problema era la maquina virtual, lo que llevo a más investigaciones hasta encontrar que se trataba de la versión de Ubuntu. Gracias al comando "sudo snap refresh docker --channel=latest/edge" se actualice, y se pudo empezar el proyecto.

Por otro lado, otras dificultades fueron las conexiones con la base de datos y el entendimiento general de la estructura que a medida que se iba avanzando en el proyecto se iban adquiriendo estos conocimientos.

Por último, el depurar ha sido un reto todo y que gracias a poderse conectar a los logs del servidor de Flask se podían ver cuales eran los errores y donde se fallaba.

Añadir, que todo y que ChatGPT ha sido de gran ayuda para poder tener rápido acceso a por ejemplo las plantillas de código html y css, también se han usado los foros ya que muchas veces se tiene que ir más allá en la búsqueda.

Bibliografía

ChatGPT. (s. f.). https://chat.openai.com/

Docker-compose ERROR [internal] booting buildkit, http: invalid host header. (s. f.). Stack Overflow. https://stackoverflow.com/questions/77225539/docker-compose-error-internal-booting-buildkit-http-invalid-host-header

Flores, J. M. &. F. (s. f.). *Python, Flask y MongoDB con Docker compose*. Python, Flask y MongoDB con Docker Compose. https://www.crashell.com/estudio/python_flask_y_mongodb_con_docker_compose

How to use Render_Template in Flask. (2023, 9 marzo). https://pytutorial.com/how-to-use-render_template-in-flask/?expand_article=1

Jarroba, R. [. (2017, 21 octubre). *Python MongoDB Driver, con ejemplos*. Jarroba. https://jarroba.com/python-mongodb-driver-pymongo-con-ejemplos/

Marvel Developer Portal. (s. f.). Marvel.com. https://developer.marvel.com/