

Diseño y desarrollo de un microservicio para la gestión de información de monitorización y predicciones de tráfico en red

9 de diciembre de 2022

TRABAJO FIN DE MÁSTER

Máster Universitario en Ingeniería de
Telecomunicación

Autor: Enrique Fernández Sánchez

Tutor: Pablo Pavón Mariño



Índice general

Índice de figuras	4
Listado de ejemplos	5
1 Introducción	6
1.1 Contexto del trabajo	6
1.2 Motivación	6
1.3 Descripción Global	6
1.4 Objetivos	6
1.5 Resumen capítulos de la memoria	6
2 Tecnologías empleadas	7
2.1 Arquitectura y microservicios	7
2.2 Bases de datos	9
2.2.1 Base de datos tipo relacional/SQL	9
2.2.2 Base de datos tipo “time series”/InfluxDB	10
2.3 Lenguajes y frameworks	11
2.3.1 Python	11
2.3.2 FastAPI	11
2.3.3 Prophet	12
2.4 Despliegue en Producción	13
3 Diseño e implementación del sistema	18
3.1 Descripción REST API	18
3.2 Estructura de la aplicación	20
3.3 Modelos de datos	20
3.3.1 Base de datos SQL	20
3.3.2 Base de datos InfluxDB	21
3.4 Endpoints	21
3.5 Implementación del sistema	21
4 Pruebas y validación del sistema	22
5 Conclusiones	23
5.1 Propuestas futuras	23
6 Bibliografía	24
Enlaces y referencias	24
Imágenes	24

Anexos	26
Anexo I. Generación dataset sintético	26

Índice de figuras

2.1	Comparativa entre arquitectura de microservicios y arquitectura “monolítica”. [1]	8
2.2	Ejemplo de relaciones dentro de una base de datos SQL. [2]	9
2.3	Logotipo base de datos InfluxDB [3].	10
2.4	Logotipo Python [4].	11
2.5	Logotipo FastAPI [5].	12
2.6	Logotipo Prophet [6].	12
2.7	Docker	13
2.8	Captura portal de gestión de Traefik. Detalle de enrutado hacia un servicio.	15
3.1	Modelo de datos para las redes a monitorizar. Equivale con la tabla “networks”.	20
3.2	Modelos de datos para las interfaces a monitorizar. Equivale con la tabla “interfaces”	20

Listado de ejemplos

2.1	Archivo configuración de contenedores para entorno de desarrollo.	14
2.2	Archivo configuración de contenedores para entorno de producción.	16

Capítulo 1

Introducción

- 1.1 Contexto del trabajo
- 1.2 Motivación
- 1.3 Descripción Global
- 1.4 Objetivos
- 1.5 Resumen capítulos de la memoria

Capítulo 2

Tecnologías empleadas

En este capítulo, se van a presentar las diferentes tecnologías utilizadas para la implementación de la aplicación.

2.1 Arquitectura y microservicios

En primer lugar, se va a comentar acerca de la arquitectura escogida. En este caso, se decide realizar una implementación basada en microservicios utilizando una REST API.

Arquitectura basada en microservicios

Lo primero, es entender en que consiste un microservicio. Para ello, podemos definirlo como los sistemas que cumplen las siguientes premisas: [2]

- Los microservicios son sistemas pequeños, independientes y poco “acoplados” (ver figura 2.1).
- Cada servicio tiene su propio código fuente, que esta separado del resto de códigos de los servicios.
- Cada servicio se puede desplegar de manera independiente.
- Cada servicio es responsable de la persistencia de sus datos.
- Los servicios se comunican entre sí utilizando APIs
- Además, como ventaja, los servicios no tienen por qué estar implementados todos en el mismo lenguaje de programación.

Por lo tanto, dado los objetivos presentados en este trabajo, se llegó a la conclusión de que tratar el sistema propuesto como un microservicio podría aportar numerosas ventajas, ya que permitiría ser utilizado por otros servicios, extendiendo la funcionalidad de estos y añadiendo un valor extra. Para ello, será necesario definir la API que utilizaremos para comunicarnos con el sistema.

Comunicación basada en API

Una API permite a dos componentes comunicarse entre sí mediante una serie de reglas. Además, supone un “contrato” en el que se establecen las solicitudes y respuestas esperadas en la comunicación. [1]

Dependiendo de la implementación de la API que se realice, distinguimos cuatro tipos de API:

- API de SOAP. Utilizan un protocolo de acceso a objetos. Los interlocutores intercambian mensajes XML. En general, es una solución poco flexible.
- API de RPC. Basado en llamadas de procedimientos remotos. El cliente ejecuta una función en el servidor, y este responde con la salida de la función.
- API de WebSocket. Solución moderna de desarrollo de API, que utiliza objetos JSON y un canal bidireccional para realizar la comunicación entre el cliente y el servidor.
- API de REST. Solución más popular. El cliente envía solicitudes al servidor como datos, utilizando métodos HTTP. Es una opción muy flexible.

En el caso de nuestra aplicación, se decidió utilizar el tipo REST API, ya que permite una sencilla implementación de cara al cliente que quiera utilizar dicha interfaz.

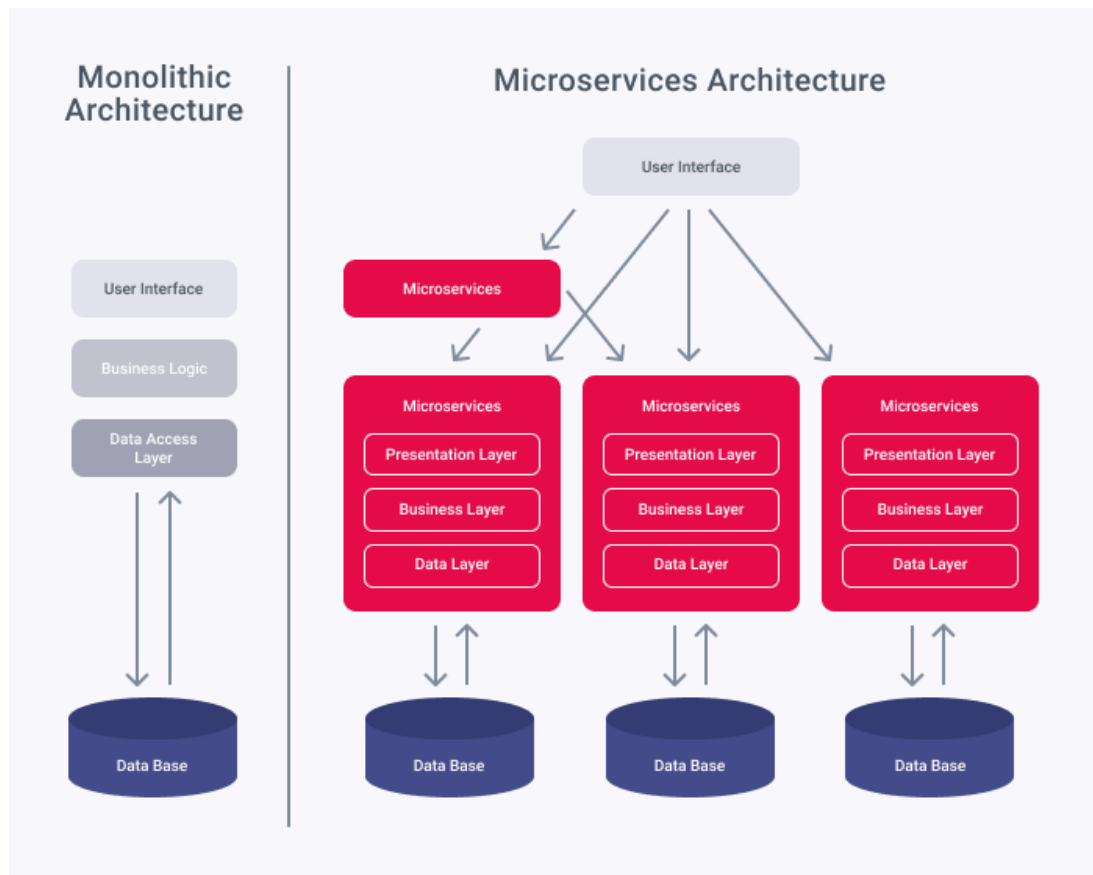


Figura 2.1: Comparativa entre arquitectura de microservicios y arquitectura “monolítica”. [1]

2.2 Bases de datos

Para asegurar la persistencia de los datos en nuestra aplicación, es necesario utilizar una base de datos. En dicha base de datos, guardaremos información relevante para el correcto funcionamiento del sistema, en nuestro caso, redes y/o interfaces a monitorizar, o los datos monitorizados.

En la aplicación de monitorización, distinguimos entre datos de dos tipos:

- Datos clásicos. Como por ejemplo, la información asociada a una red a monitorizar.
- Datos de tipo "time series". Como por ejemplo, las muestras de monitorización de una red.

En primer lugar, se diseña una base de datos tipo SQL para almacenar los "datos clásicos". Y por otro lado, se diseña una base de datos diferente, especializada para el almacenamiento de datos tipo "time series", en este caso, se elige una base de datos llamada InfluxDB.

2.2.1 Base de datos tipo relacional/SQL

SQL es una base de datos de tipo relacional. Dichas bases de datos, suponen una colección de información que organizan los datos en una serie de "relaciones" cuando la información es almacenada en una o varias "tablas". Por lo tanto, las relaciones suponen conexiones entre diferentes tablas, permitiendo así una asociación entre información diferente. [3]

Por ejemplo, si vemos la figura 2.2, podemos comprobar como se realizan las relaciones entre las diferentes tablas (Ratings, Users, Movies o Tags), se realiza mediante uno de los campos definidos en la propia tabla. Por ejemplo, el campo "user_id" de la tabla Ratings, permite una relación con la tabla Users, con el campo "id".

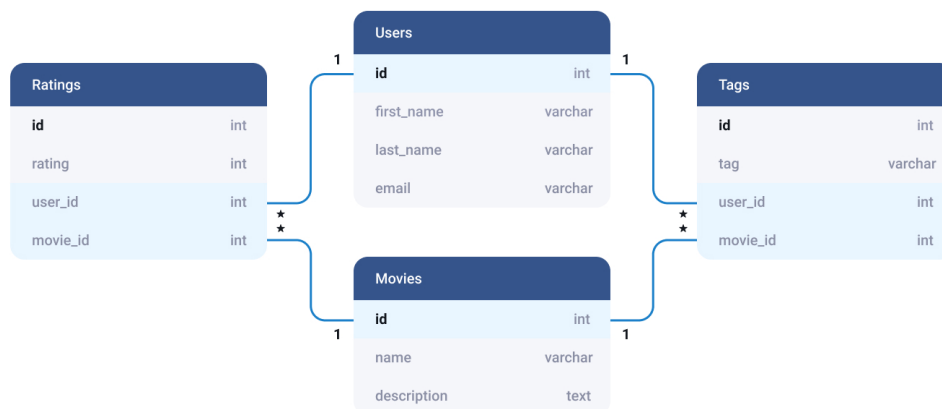


Figura 2.2: Ejemplo de relaciones dentro de una base de datos SQL. [2]

Para el caso de nuestra aplicación de monitorización de tráfico en red, modelamos la base de datos según la información que vamos a almacenar. Por lo que tenemos que definir la estructura de tablas, los campos que van a tener cada una de las tablas, y los campos por los que se van a relacionar entre sí. A esta información la llamamos modelo de datos.

2.2.2 Base de datos tipo “time series”/InfluxDB

InfluxDB es una base de datos diseñada para trabajar con datos tipo time-series. Si bien es cierto que SQL puede gestionar este tipo de datos, no fue creado estrictamente para este objetivo. En este caso, InfluxDB está diseñado para almacenar grandes volúmenes de datos, y además realizar análisis en tiempo real sobre esos datos.

En comparación con SQL, en InfluxDB un “timestamp” identifica un punto en cualquier serie de datos. Esto sería equivalente a SQL, si la clave primaria de una tabla es establecida por el sistema y siempre es equivalente al tiempo. Además, InfluxDB permite reconocer el “schema” de manera dinámica, además de que no estás obligado a definir el “schema” y seguirlo, es decir, se permiten cambios dentro de la misma serie de datos. [6]

Algunas de las razones destacadas para elegir InfluxDB son: [5]

- Perfecto para almacenar datos de telemetría, como métricas de aplicaciones o sensores IoT.
- Los datos son comprimidos automáticamente para ser eficientes con el espacio disponible.
- Se realizan tareas automáticas de “downsampling” para reducir el uso de disco.
- Lenguaje para hacer consultas que permite analizar en profundidad los datos almacenados.
- Disponible una aplicación web para realizar consultas y comprobar los datos disponibles en la base de datos.

Terminología

Comparando con los conceptos ya existentes en bases de datos de tipo SQL, se definen los siguientes conceptos en InfluxDB:

- “measurement”: equivalente a una tabla.
- “tags”: equivalente a columnas indexadas dentro de una tabla.
- “fields”: equivalente a columnas no indexadas dentro de una tabla.
- “points”: similar a las filas en una tabla.



Figura 2.3: Logotipo base de datos InfluxDB [3].

2.3 Lenguajes de programación y frameworks

En resumen, para el desarrollo de esta aplicación se ha utilizado el lenguaje de programación Python, con el framework de desarrollo para APIs llamado FastAPI.

2.3.1 Python

Python [7] es un lenguaje de programación orientado a objetos, interpretado y de alto nivel con tipado dinámico. Es muy atractivo ya que permite un desarrollo rápido de aplicaciones, además de ser muy adecuado para realizar tareas de “scripting”. Python es un lenguaje simple y sencillo de aprender. Por otro lado, dispone de multitud de “librerías” o “módulos” publicados por usuarios, dando lugar a una gran comunidad y una gran variedad de alternativas para implementar soluciones.

Actualmente, Python destaca como lenguaje de programación en los siguientes ámbitos:

- Desarrollo de aplicaciones web, utilizando los frameworks Django, Flask o FastAPI.
- Tareas asociadas a “data science”, utilizando librerías como Pandas o NumPy.
- Inteligencia artificial, utilizando frameworks como TensorFlow o scikit-learn.



Figura 2.4: Logotipo Python [4].

2.3.2 FastAPI

FastAPI [8] es un framework moderno y rápido para construir APIs utilizando la versión de Python 3.7+. Algunas de las características más destacadas son:

- Rápido: rendimiento muy alto, prácticamente a la par con otros lenguajes de programación destinados al desarrollo de backend (como NodeJS o Go).
- Intuitivo: soporta auto completado en el código.
- Robusto: código pensado para entornos de producción, además de incluir documentación automática (usando Swagger o ReDoc).
- Basado en estándares: al utilizar estándares de tipo de datos, permite ser totalmente compatible con los estándares de [OpenAPI](#) [9] y [JSON Schema](#) [10].



Figura 2.5: Logotipo FastAPI [5].

2.3.3 Prophet

Prophet [11] es un framework del lenguaje de programación Python, desarrollado por [Meta](#), que recoge una serie de procedimientos que permiten realizar predicciones de un dataset de series temporales, en el que se pueden encontrar diferentes efectos no lineales, llamados tendencias (como puede ser una tendencia anual, semanal, o mensual), además de otros efectos causados por fechas concretas.

Es un framework de predicción basado en inferencia estadística, lo que permite tener un rendimiento mayor que si utilizamos técnicas de Machine Learning para solucionar el mismo problema de predicción de datos. Además, es robusto a datos no disponibles y a modificaciones aleatorias sobre las tendencias.



Figura 2.6: Logotipo Prophet [6].

2.4 Tecnologías utilizadas en un despliegue en producción

Otro de los aspectos importantes para la realización de este proyecto, es el hecho de que la aplicación desarrollada debe ser apta para desplegarse en un entorno de producción y funcionar correctamente para que sea implementada como microservicio por otras aplicaciones. Es por este motivo por el que nos decantamos por implementar este microservicio dentro de un contenedor.

Docker

Utilizamos Docker como la tecnología para realizar un contenedor de nuestra aplicación, implementando dentro del contenedor todas las librerías y código necesario para hacer funcionar la aplicación.

Por otro lado, también será necesario desplegar diferentes contenedores que tendrán alojadas las bases de datos que utilizaremos en el proyecto. En este caso, al utilizar dos bases de datos, tendremos que hacer uso de un contenedor para desplegar una base de datos SQL, y otro contenedor para desplegar una base de datos InfluxDB.

Los contenedores Docker necesarios se desplegarán sobre una máquina host que tenga en funcionamiento el “daemon” de Docker.

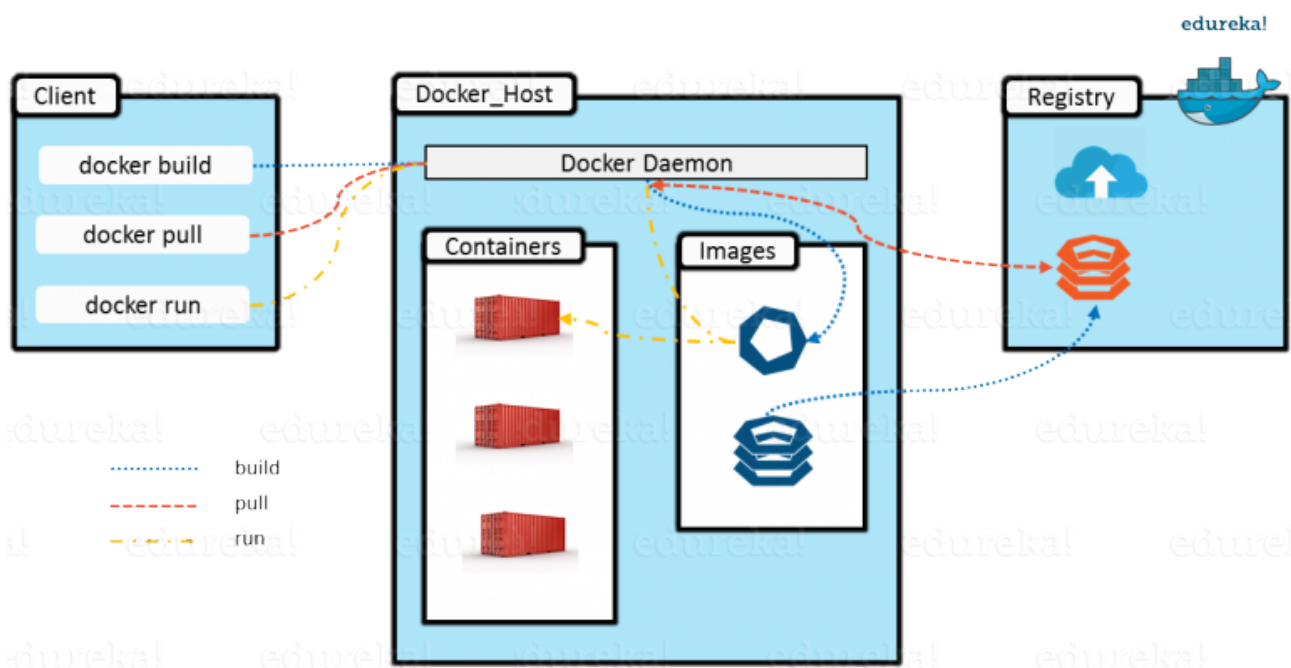


Figura 2.7: Docker

Docker Compose

Compose es una herramienta de Docker que permite definir y ejecutar instancias multi contenedores de Docker [12]. Con la herramienta Compose, podemos utilizar un archivo YAML para describir y configurar los contenedores que vamos a utilizar, dando lugar a que con un único comando puedas crear y ejecutar todos los servicios de la aplicación, y con la misma configuración.

Compose funciona para los diferentes entornos: producción, desarrollo, y o pruebas. Además, tiene disponibles comandos para hacer más sencilla la tarea de iniciar o parar servicios, o ver la salida de consola producida por los contenedores.

En el caso de nuestra aplicación, el archivo “*docker-compose.yml*” utilizado para el entorno de desarrollo sería el siguiente:

Ejemplo 2.1: Archivo configuración de contenedores para entorno de desarrollo.

```
1 version: '3.8'
2
3 services:
4   traffic_forecast:
5     build: ./backend
6     ports:
7       - 5000:5000
8     environment:
9       - POSTGRES_USER=monitor
10      - POSTGRES_PASSWORD=forecast2022
11      - POSTGRES_DB=traffic-forecast
12      - INFLUX_TOKEN=*****
13      - INFLUX_ORG=e-lighthouse
14      - INFLUX_BUCKET=traffic-forecast
15      - SECRET_KEY=upct2022_sk
16      - FASTAPI_CONFIG=development
17   volumes:
18     - ./backend:/app
19   depends_on:
20     - db
21     - influxdb
22
23 db:
24   image: postgres:13
25   expose:
26     - 5432
27   environment:
28     - POSTGRES_USER=monitor
29     - POSTGRES_PASSWORD=forecast2022
30     - POSTGRES_DB=traffic-forecast
31   volumes:
32     - ./data/postgres_db:/var/lib/postgresql/data
33
34 influxdb:
35   image: influxdb:latest
36   volumes:
37     - ./data/influxdb/data:/var/lib/influxdb2:rw
38   ports:
39     - 8086:8086
```

Para desplegar los contenedores, según el archivo YAML definido, tenemos que ejecutar el siguiente comando:

```
docker compose up -f <filename> traffic_forecast
```

Traefik

Traefik es una herramienta que permite hacer de “reverse proxy” y “load balancer” en contenedores Docker, permitiendo el despliegue sencillo de microservicios en un entorno de producción.

Traefik está diseñado para ser simple de operar, pero con una gran capacidad de gestión en entornos complejos. Además, se integra perfectamente con la infraestructura ya existente y configurar el sistema de manera dinámica. Algunas de las tareas que realiza Traefik son: [13]

- Gestionar middlewares necesarios para la aplicación (forzar protocolos específicos, configurar contraseña para el servicio...).
- Funcionar con API Gateway, gestionando los dominios y certificados para cada uno de los microservicios desplegados.
- Orquestador de nodo de entrada, gestionando la comunicación de tráfico de un dominio hacia el servicio asociado.

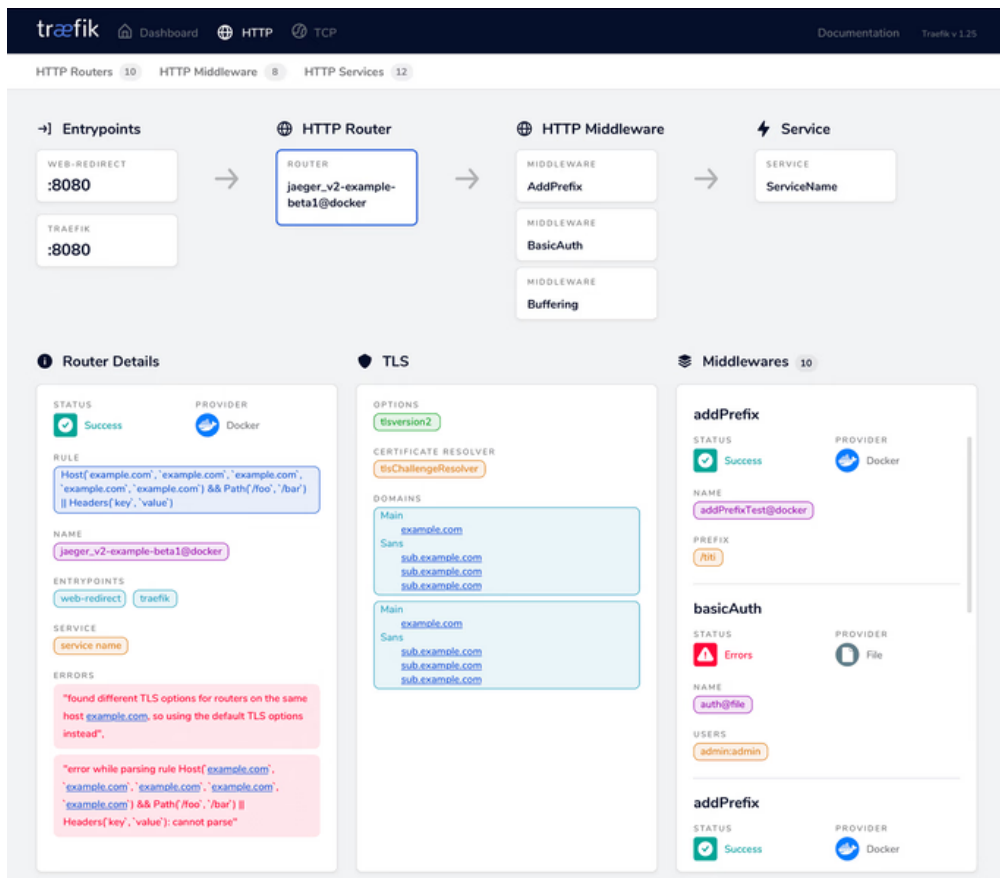


Figura 2.8: Captura portal de gestión de Traefik. Detalle de enrutado hacia un servicio.

En el caso de nuestra aplicación, para el entorno de producción usamos el siguiente archivo que permite a la herramienta docker compose ejecutar los servicios con sus configuraciones asociadas. Como podemos ver, en la sección de “labels” se encuentra la configuración asociada a Traefik. Dicha configuración permite a Traefik asignar un servicio a un dominio, crear un certificado SSL y configurar los puertos de acceso para dicho servicio. [14]

Ejemplo 2.2: Archivo configuración de contenedores para entorno de producción.

```
1 version: '3.8'
2
3 services:
4   traffic_forecast:
5     build: ./backend
6     container_name: traffic_forecast-api
7     restart: unless-stopped
8     environment:
9       - POSTGRES_USER=monitor
10      - POSTGRES_PASSWORD=forecast2022
11      - POSTGRES_DB=traffic-forecast
12      - INFLUX_URL=https://tfm-influx.ranii.pro:8443/
13      - INFLUX_TOKEN=****
14      - INFLUX_ORG=e-lighthouse
15      - INFLUX_BUCKET=traffic-forecast
16      - SECRET_KEY=upct2022_sk
17   volumes:
18     - ./backend:/app
19   labels:
20     - traefik.enable=true
21     - traefik.http.routers.tf-api.entryPoints=web-secure
22     - traefik.http.routers.tf-api.rule=Host('tfm-api.ranii.pro')
23     - traefik.http.routers.tf-api.tls.certresolver=default
24     - traefik.http.services.tf-api.loadbalancer.server.port=5000
25   depends_on:
26     - db
27     - influxdb
28
29 db:
30   image: postgres:13
31   container_name: traffic_forecast-postgres
32   restart: unless-stopped
33   environment:
34     - POSTGRES_USER=monitor
35     - POSTGRES_PASSWORD=forecast2022
36     - POSTGRES_DB=traffic-forecast
37   volumes:
38     - ./data/postgres_db:/var/lib/postgresql/data
39
40 influxdb:
41   image: influxdb:latest
42   container_name: traffic_forecast-influx
43   restart: unless-stopped
44   volumes:
45     - ./data/influxdb/data:/var/lib/influxdb2:rw
46   labels:
47     - traefik.enable=true
48     - traefik.http.routers.tf-influx.entryPoints=web-secure
49     - traefik.http.routers.tf-influx.rule=Host('tfm-influx.ranii.pro')
50     - traefik.http.routers.tf-influx.tls.certresolver=default
```



```
51     - traefik.http.services.tf-influx.loadbalancer.server.port=8086
52
```

Para desplegar los contenedores, según el archivo YAML definido, tenemos que ejecutar el siguiente comando:

```
docker compose up -f <filename> traffic_forecast
```

En el caso de querer profundizar más en el funcionamiento de Traefik, se puede consultar el siguiente enlace [Traefik: Get Started](#)

Capítulo 3

Diseño e implementación del sistema

En este capítulo se va a comentar más en detalle la implementación del sistema propuesto para la monitorización y predicción de tráfico en una red. Tal y como vimos en el capítulo anterior, se va a utilizar el lenguaje de programación Python, con el framework de desarrollo FastAPI.

3.1 Descripción REST API

Tal y como pudimos ver en la sección 2.1, para la aplicación propuesta hemos elegido que siga la estructura de REST API, ya que es la más sencilla de implementar y la más flexible.

Para nuestro caso particular de monitorización del tráfico de una red, se han detectado tres “agentes” involucrados dentro del sistema. Dichos agentes son los siguientes:

- Redes (desde ahora `networks`). Corresponde con una red en la que queremos monitorizar el tráfico de ciertas interfaces.
- Interfaces de red (desde ahora `interfaces`). Corresponde con las interfaces de red de las que queremos monitorizar su tráfico.
- Muestras de monitorización (desde ahora `samples`). Corresponden con los diferentes datos de monitorización de tráfico asociados a una interfaz.

Definidos los agentes, se procede a diseñar el almacenamiento de la información de cada uno de ellos. Para ello, en primer lugar, nos decantamos por aplicar el concepto de CRUD en los agentes `networks` y `interfaces`.

CRUD: `networks`

Tal y como se refiere la definición de CRUD (**C**reate **R**ead **U**pdate **D**elete), en el caso del agente de `networks`, queremos definir el funcionamiento que tiene que seguir la aplicación cuando queramos referirnos a una red a monitorizar. Además, asumimos que dentro de una misma red podemos tener diferentes interfaces, que también serán dadas de altas en la aplicación. Por lo tanto, una única red puede tener muchas interfaces de red monitorizadas.

Dado que estamos planteando un servicio HTTP, debemos asignar una ruta (desde ahora `endpoint`) asociada a la colección de métodos de la CRUD de `networks`. En este caso, el `endpoint` de la colección será: “/networks”

A modo de resumen, recogemos en la siguiente tabla el funcionamiento esperado de la aplicación, en función del método HTTP recibido y el recurso sobre el que se ejecute la petición HTTP.

Recurso	GET	POST	PATCH	DELETE
Colección de redes: <code>/networks</code>	Lista de redes dadas de alta	Añade una nueva red	*	*
Red en particular: <code>/networks/<id_net></code>	Información de una red en particular	*	Modificamos la informacion de una red	Eliminamos una red

Tabla 3.1: CRUD `networks` (* equivalente a “acceso denegado”)

3.2 Estructura de la aplicación

3.3 Modelos de datos

3.3.1 Base de datos SQL

Modelos de datos

El primer modelo de datos de la aplicación, es el referido a la información de una red a monitorizar. La descripción del modelo la podemos ver en la figura 3.1.

Networks
<ul style="list-style-type: none"> • “id_net”: Int, PubK, Unique (lo pasa user) • “name”: String() • “descripcion”: String() • “ip_red”: String() • “influx_net”: String()

Figura 3.1: Modelo de datos para las redes a monitorizar. Equivale con la tabla “networks”.

El segundo modelo de datos de la aplicación, es el referido a la información de una interfaz a monitorizar, que esta dentro de una red monitorizada. La descripción del modelo la podemos ver en la figura 3.2.

Interfaces
<ul style="list-style-type: none"> • “id_if”: Int, PubK, Unique (lo pasa user) • “id_net”: Int, ForK • “name”: String() • “description”: String() • “influx_if_rx”: String() • “influx_if_tx”: String()

Figura 3.2: Modelos de datos para las interfaces a monitorizar. Equivale con la tabla “interfaces”

Se establece una relación del modo que una interfaz solo puede pertenecer a una red monitorizada, y que además, una red monitorizada puede tener muchas interfaces. Dicha relación se realiza mediante el campo “id_net” de la tabla Interfaces.

Por último, la base de datos SQL utilizada para esta aplicación es PostgreSQL [4], ya que además de ser Open Source, permite una gran escalabilidad, amoldándose a los recursos de la máquina en la que esté funcionando.

3.3.2 Base de datos InfluxDB

Modelo de datos

Para la aplicación a diseñar, se plantea la premisa de que en la base de datos InfluxDB solo se van a guardar los datos de cada muestra de monitorización de una interfaz. Además, se pueden tener tantas redes como sean necesarias, y dentro de cada red, tantas interfaces como creamos necesarias.

En resumen, definimos el modelo de datos que tiene que seguir nuestra aplicación:

- “measurement”: equivalente a una red a monitorizar, valor almacenado en `Networks::influx_net`.
- “fields”: nombre del valor a monitorizado, en este caso el default es *link_count*.
- “tags”: solo tenemos un tag llamado *interface*, su objetivo es identificar a que interfaz pertenece el punto de monitorización. El valor puede estar ser `Interfaces::influx_if_rx` o `Interfaces::influx_if_tx`.
- “points”: corresponde con el valor numérico del “field”. En este caso, corresponde con el valor numérico de *link_count* en ese periodo de 5 minutos.

3.4 Endpoints

3.5 Implementación del sistema

Capítulo 4

Pruebas y validación del sistema

Capítulo 5

Conclusiones

5.1 Propuestas futuras

Capítulo 6

Bibliografía

Enlaces y referencias

1. [¿Qué es una API?](#)
2. [Microservice architecture style](#)
3. [What is a relational database?](#)
4. [PostgreSQL](#)
5. [InfluxDB: Introduction](#)
6. [InfluxDB: Comparison to SQL](#)
7. [Python](#)
8. [FastAPI framework](#)
9. [GitHub: OpenAPI-Specification](#)
10. [JSON Schema](#)
11. [Prophet](#)
12. [Docker Compose Overview](#)
13. [Traefik Webpage](#)
14. [Traefik Wiki](#)

Figuras

1. [Monolithic vs Microservices Architecture](#)
2. [Relational Databases](#)
3. [InfluxDB](#)
4. [Python](#)

5. [FastAPI](#)
6. [Prophet](#)

Anexos

Anexo I. Generación dataset sintético