## Cloud Computing: Servicios y Aplicaciones



# UNIVERSIDAD DE GRANADA

# PRÁCTICA 2 DESPLIEGUE DE UN SERVICIO CLOUD NATIVE

**Autor** Juan Manuel Castillo Nievas



MÁSTER PROFESIONAL EN INGENIERÍA INFORMÁTICA 2020-2021

Granada, 7 de mayo de 2021

## Índice

1.	. Introducción	2
2.	. Configuración del entorno	3
3.	. Resolución de las tareas	3
	3.1. Preparación del entorno	4
	3.2. Descarga de datos A y B	4
	3.3. Captura de datos	5
	3.4. Almacenamiento de datos	5
	3.5. Creación de los modelos de predicción	5
	3.5.1. Modelo de predicción ARIMA	
	3.5.2. Modelo de predicción Prophet	6
	3.5.3. Problemas encontrados	6
	3.6. Clonación del repositorio	7
	3.7. Unit test	7
	3.8. Build	7
	3.8.1. Creación de la API	8
	3.8.2. Problemas encontrados	8
	3.9. Despliegue	9
	3.9.1. Problemas encontrados	9
4.	. Demostración	10
	4.1. Acceso a la API v1	11
	4.2. Acceso a la API v2	13
<b>5</b> .	. Bibliografía	<b>15</b>

#### 1. Introducción

El objetivo de esta práctica es llevar a cabo un despliegue completo de un servicio **Cloud Native**. Para esto, es necesario entender los aspectos claves para Cloud Native:

- **DevOps:** colaboración entre los desarrolladores de software y operaciones TI para la entrega constante, realizando la construcción, prueba y liberación de software.
- CI/CD: integración continua y entrega continua.
- Microservicios: un pequeño servicio que ejecuta su propio proceso y se comunica a través de API en algunos casos. Cada microservicio se despliega, actualiza, escala y reinicia de manera independiente a los demás.
- Contenedores: utilizados para el despliegue de los microservicios.

Para esta práctica se va a desarrollar un sistema de predicción de temperatura y humedad. Se debe desplegar un servicio que proporcione una API de tipo HTTP RESTful con las predicciones. Esta API consta de 2 versiones, utilizando en cada versión un modelo de predicción diferente, y consta de los siguientes endpoints:

#### ■ Version 1

- /servicio/v1/prediccion/24horas: obtiene la predicción de las siguientes 24 horas usando el modelo de predicción de la versión 1
- /servicio/v1/prediccion/48horas: obtiene la predicción de las siguientes 48 horas usando el modelo de predicción de la versión 1
- /servicio/v1/prediccion/72horas: obtiene la predicción de las siguientes 72 horas usando el modelo de predicción de la versión 1

#### ■ Version 2

- /servicio/v2/prediccion/24horas: obtiene la predicción de las siguientes 24 horas usando el modelo de predicción de la versión 2
- /servicio/v2/prediccion/48horas: obtiene la predicción de las siguientes 48 horas usando el modelo de predicción de la versión 2
- /servicio/v2/prediccion/72horas: obtiene la predicción de las siguientes **72 horas** usando el modelo de predicción de la versión 2

El conjunto de salida debe ser de tipo JSON y debe tener la siguiente estructura:

```
[{''hour'':13:05,''temp'':32.20,''hum'':85.90}, ...]
```

### 2. Configuración del entorno

Para el desarrollo de la práctica se han utilizado las siguientes herramientas:

- Python 3 [1] como lenguaje de programación
- Flask [2] para el desarrollo de la API
- Apache Airflow [3] para el diseño del flujo de trabajo del servicio Cloud Native
- MongoDB [4] como base de datos para el almacenamiento de los datos
- pymongo [5] como biblioteca para el uso de MongoDB en Python
- unittest [6] para la batería de pruebas con TDD de la API
- pandas [7] como biblioteca para la manipulación de datos en Python
- Statsmodels [8] para el uso del modelo ARIMA [9] para la creación del modelo de predicción de la versión 1 de la API
- Prophet [10] para la creación del modelo de predicción de la versión 2 de la API
- Pickle [11] para el almacenamiento de los modelos de predicción creados
- gunicorn [12] para desplegar la API en un servidor HTTP

#### 3. Resolución de las tareas

Todo el código desarrollado se encuentra en mi repositorio de Github:

https://github.com/Jumacasni/MUII-CCSA/tree/main/practica2

A modo de introducción, en la Figura 1 se muestra el diagrama de flujo de trabajo que se ha utilizado para el despliegue del servicio Cloud Native usando Apache Airflow.

Este flujo está disponible en el archivo del repositorio llamado **practica2\_dag.py**, que es el que se debe incluir en la carpeta **dags** en la instalación de **Airflow** para que pueda ejecutarse. El enlace del archivo en el repositorio es el siguiente:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/practica2\_dag.py

Cada tarea realizada en el flujo de trabajo se va a explicar a continuación en las siguientes subsecciones, indicando en cada caso el fichero del repositorio que se ha utilizado para ello.

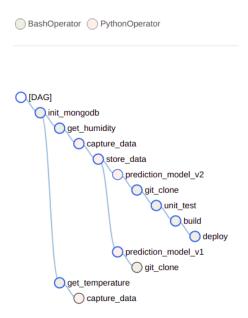


Figura 1: Diagrama de flujo de trabajo

#### 3.1. Preparación del entorno

La preparación del entorno está formada por la tarea init\_mongodb. Esta tarea se encarga de desplegar el servicio de la base de datos de MongoDB, usando para ello una operación BashOperator con el siguiente comando bash:

docker run -d --rm -p 27017:27017 --name mongodb mongo:latest

#### 3.2. Descarga de datos A y B

Aquí se incluyen dos tareas que están al mismo nivel: get\_humidity y get\_temperature, que realizan la descarga de datos de la humedad y la temperatura que se usarán para la creación del modelo de predicción. En el diagrama de flujo de trabajo son dos operaciones **BashOperator**. El comando que se usa en cada operación es curl para descargar un fichero .zip que se guarda en el directorio /tmp/:

curl -o /tmp/humidity.csv.zip
https://raw.githubusercontent.com/manuparra/MaterialCC2020/master/humidity.csv.zip

#### 3.3. Captura de datos

La captura de datos se encuentra en la tarea capture\_data, y es una operación del tipo **Pytho-nOperator**, es decir, se usa una función en Python para su resolución. En este paso se deben **descomprimir**, unificar y limpiar los datos para adminitar sólamente la columna *date* (fecha) y San Francisco (ciudad), de modo que los datos se almacenen en un nuevo fichero .csv con estructura DATE; TEMP; HUM.

Este **PythonOperator** llama al método capture\_data que se encuentra en el archivo **practi-**ca2\_dag.py del repositorio. El siguiente enlace lleva a la línea exacta en el código:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/practica2\_dag.py#L20

#### 3.4. Almacenamiento de datos

Se corresponde con la tarea store\_data. Una vez que se han capturado los datos en el archivo forecast.csv, esta tarea se encarga de almacenar dichos datos en la base de datos MongoDB. De nuevo, se trata de un PythonOperator que llama al método store\_data disponible en el fichero practica2\_dag.py del repositorio. El siguiente enlace lleva a la línea exacta en el código:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/practica2\_dag.py#L46

#### 3.5. Creación de los modelos de predicción

La creación de los modelos de predicción se hacen en dos tareas que están al mismo nivel. Son dos tareas de tipo **PythonOperator**. La idea general de estas tareas es crear el modelo correspondiente y almacenarlo en un archivo .**pkl** para que se pueda hacer la predicción solamente llamando a dicho archivo.

La creación de los modelos suele tardar alrededor de **10 minutos** debido a la gran cantidad de datos recopilados de humedad y temperatura, con lo cual se ha decidido entrenar el modelo con sólamente las **1.000 primeras filas** para reducir el tiempo de ejecución de forma considerable.

#### 3.5.1. Modelo de predicción ARIMA

Este modelo de predicción corresponde a la versión 1 de la API. El método asociado al **Pytho-nOperator** de esta tarea se llama models\_arima y está definido en el archivo practica2\_dag.py del repositorio. Se puede acceder mediante el siguiente enlace:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/practica2\_dag.py#L60

Los archivos creados se guardan en /tmp/practica2/v1.

#### 3.5.2. Modelo de predicción Prophet

Este modelo de predicción corresponde a la versión 2 de la API. El método asociado al **Pytho-nOperator** de esta tarea se llama models\_prophet y está definido en el archivo practica2\_dag.py del repositorio. Se puede acceder mediante el siguiente enlace:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/practica2\_dag.py#L113

Para el entrenamiento de los modelos creados en **Prophet** se ha seguido un pequeño tutorial propuesto por **Caner Dabakoglu** [13].

Los archivos creados se guardan en /tmp/practica2/v2.

#### 3.5.3. Problemas encontrados

Hay dos problemas principales que se encontraron durante esta etapa:

- Al principio no se almacenaban los modelos entrenados en un archivo .pkl, con lo cual cada vez que se hacía una llamada a la API, la API llamaba el método de entrenamiento y luego hacía la predicción. Como es obvio, teniendo en cuenta el volumen de datos de entrenamiento, esto hacía que en cada llamada a la API se tardara más de 10 minutos en realizar una sola predicción. Es por esto que se decidió almacenar los modelos de predicción en archivos .pkl.
- El modelo de predicción **Prophet** parecía sencillo de usar pero se tardó un poco en hacer que el modelo de funcionara debido a que no hay demasiada documentación buena acerca de este. También dio muchos problemas a la hora de desplegarlo como servicio debido a todas las librerías necesarias para su uso, pero eso es algo que se explicará en su sección más adelante.

#### 3.6. Clonación del repositorio

Esta tarea está asociada a la que se llama git\_clone. Se encarga de clonar el repositorio de Github para coger los archivos necesarios para el despliegue del servicio.

Esta tarea es del tipo **BashOperator** y el comando que tiene asociado es el siguiente:

Este comando copia los archivos necesarios en la carpeta /tmp/practica2.

#### 3.7. Unit test

Una vez que se ha clonado el repostorio, se deben lanzar los tests que se han preparado como batería de pruebas con TDD para verificar el funcionamiento de ambas versiones de la API. Esto se realiza en la tarea llamada unit\_test del flujo de trabajo.

Esta tarea es del tipo BashOperator y el comando que se ejecuta es el siguiente:

cd /tmp/practica2 && python3 tests.py

Los tests se encuentran en el archivo del repositorio llamado **tests.py**, al que se puede acceder mediante el siguiente enlace:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/tests.py

#### 3.8. Build

Llegando casi al final de esta práctica, una vez que se han verificado los tests, se debe pasar a la construcción del contenedor que se va a desplegar. Esta tarea se llama build y es de tipo **BashOperator**, y el comando que tiene asociado es el siguiente:

cd /tmp/practica2 && docker build --rm -t api .

El **Dockerfile** usado para ello se encuentra en el repositorio y se puede acceder mediante el siguiente enlace:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/Dockerfile

Este **Dockerfile** ejecuta el comando **gunicorn** que permite desplegar la API realizada en Flask en un servidor. En concreto, se despliega en el puerto 8000. También se instalan muchas librerías debido a que **Prophet** las necesita, tal y como se va a explicar en la Sección 3.8.2.

Este proceso tarda entre **5-10 minutos** aproximadamente debido a la cantidad de librerías que se deben instalar.

#### 3.8.1. Creación de la API

La **API** se ha creado usando **Flask**, tal y como se ha comentado anteriormente. Se encuentra en el archivo **api.py** del repositorio al que se puede acceder mediante el siguiente enlace:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/api.py

Se han creado tres rutas:

- /: se accede a través de http://localhost:8000/ y simplemente devuelve un texto indicando que el microservicio está funcionando.
- /servicio/v1/prediccion/<string:horas>: que permite el acceso a las siguientes rutas que devuelven la predicción en formato JSON usando ARIMA:
  - localhost:8000/servicio/v1/prediccion/24horas
  - localhost:8000/servicio/v1/prediccion/48horas
  - localhost:8000/servicio/v1/prediccion/72horas
- /servicio/v2/prediccion/<string:horas>: que permite el acceso a las siguientes rutas que devuelven la predicción en formato JSON usando Prophet:
  - localhost:8000/servicio/v2/prediccion/24horas
  - localhost:8000/servicio/v2/prediccion/48horas
  - localhost:8000/servicio/v2/prediccion/72horas

#### 3.8.2. Problemas encontrados

Esta etapa ha sido la más problemática debido al modelo de predicción **Prophet**. En mi ordenador funcionaba muy bien, pero al instalar esta librería en docker, empezaron a aparecer muchos errores de dependencias de otras librerías y demás.

Buscando información acerca de esto, las "soluciones" encontradas no llegaban a ayudarme del todo. Dos soluciones probadas y que no funcionaron:

 Usar conda para instalar fbprophet. Usé imágenes docker de conda pero no llegó a funcionar.  Usar imágenes docker de fbprophet (https://hub.docker.com/r/safakcirag/fbprophet), entre otras. Debido a que son versiones bastante antiguas, había bastantes problemas entre dependencias y demás.

Finalmente encontré un perfil de Github con un Dockerfile que instalaba **fbprophet**, pero no funcionó en un principio. Sin embargo, estos errores parecían ser de versiones de librerías, así que decidí poner las versiones exactas que yo tenía en mi ordenador personal y funcionó.

Las versiones usadas junto con las librerías se encuentran en el fichero **requirements.txt**, disponible en el siguiente enlace del repositorio:

https://github.com/Jumacasni/MUII-CCSA/blob/main/practica2/requirements.txt

El Dockerfile encontrado en Github pertenece al usuario **stefanproell**, y se puede encontrar en el siguiente enlace:

https://gist.github.com/stefanproell/89a00f3a2c18a7e9549a1de6f82cf0f8

Tal y como se puede ver, al final que da una imagen con bastantes librerías y un poco pesada, pero todo funciona correctamente.

#### 3.9. Despliegue

Este es el último paso del diagrama de flujo. El nombre de esta tarea es deploy y es de tipo **BashOperator**, que contiene el siguiente comando:

docker run --rm --name api -p 8000:8000 api

Este comando lo que hace es desplegar el servicio de la API creada en el puerto **8000**. Desde este momento, se pueden acceder a los distintos **endpoints**.

#### 3.9.1. Problemas encontrados

No fue un problema grave, pero al principio se usaba la orden app.run(debug=True, host='0.0.0.0', port=8000) para desplegar el servicio de Flask en el archivo api.py. Esto se puede ver en el historial de *commits* del repositorio de Github. Concretamente, en el siguiente enlace:

#### https:

//github.com/Jumacasni/MUII-CCSA/commit/10b28f2bbeb9d58c8673cd28e407de1e0260bf97#diff-e13f9fc15b9d1f9c6e9f3c48c0709a47ffbde737be1ebb1137f34b29a2fbe08cR60

El problema de esto es que esta forma de "desplegar" el servicio sólo debe usarse en **desarrollo**, y **nunca en producción**. Es por esta razón por la que, una vez verificado que todo funcionaba correctamente, se decidió usar **gunicorn** para su despliegue.

#### 4. Demostración

En esta sección se va a demostrar el funcionamiento de todo el despliegue y funcionamiento de la API.

Lo primero que se debe hacer es colocar el archivo **practica2\_dag.py** en la carpeta **dags** que se encuentra en la instalación de **Apache Airflow** (ver Figura 2).



Figura 2: Archivo practica2\_dag.py en la carpeta dags

Apache Airflow tarda unos 30 segundos en reconocer el flujo de trabajo. Una vez que se ha reconocido en la versión web, se procede a activar el flujo de trabajo (ver Figura 3).

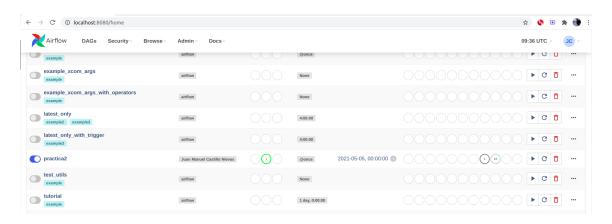


Figura 3: Activación del fujo de trabajo en Apache Airflow

Una vez que hayan terminado todas las tareas, el flujo de trabajo en **Apache Airflow** se verá como se muestra en la Figura 4. Esto significa que la API ya está desplegada y lista para su uso.

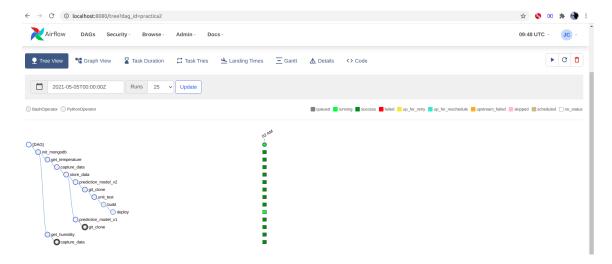


Figura 4: Despliegue final de la API con Apache Airflow

#### 4.1. Acceso a la API v1

Tal y como se comentó anteriormente, la versión 1 de la API está formada por los siguientes **endpoints**:

- localhost:8000/servicio/v1/prediccion/24horas (ver Figura 5)
- localhost:8000/servicio/v1/prediccion/48horas (ver Figura 6)
- localhost:8000/servicio/v1/prediccion/72horas (ver Figura 7)

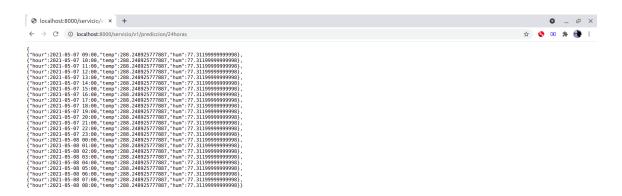


Figura 5: Acceso a localhost:8000/servicio/v1/prediccion/24horas

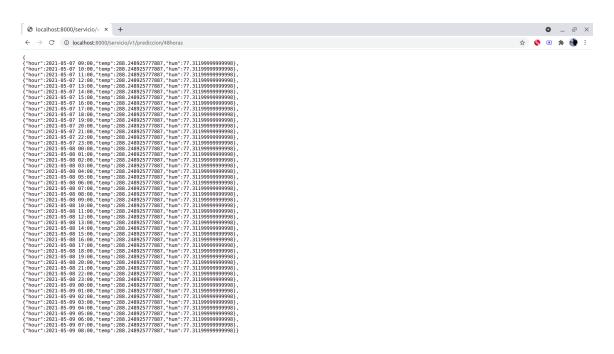


Figura 6: Acceso a localhost:8000/servicio/v1/prediccion/48horas

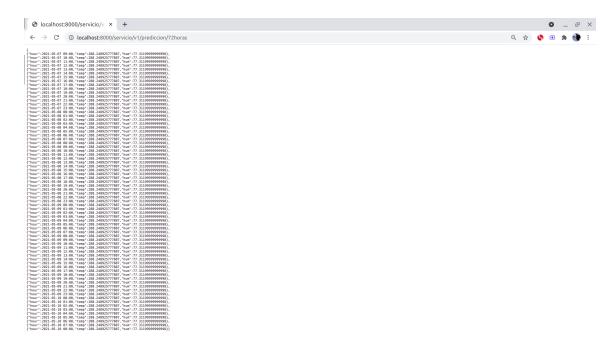


Figura 7: Acceso a localhost:8000/servicio/v1/prediccion/72horas

#### 4.2. Acceso a la API v2

La versión 1 de la API está formada por los siguientes endpoints:

- localhost:8000/servicio/v2/prediccion/24horas (ver Figura 8)
- localhost:8000/servicio/v2/prediccion/48horas (ver Figura 9)
- localhost:8000/servicio/v2/prediccion/72horas (ver Figura 10)

Figura 8: Acceso a localhost:8000/servicio/v2/prediccion/24horas

Figura 9: Acceso a localhost:8000/servicio/v2/prediccion/48horas

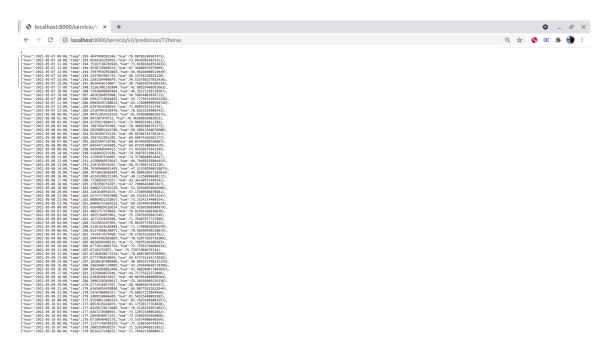


Figura 10: Acceso a localhost:8000/servicio/v2/prediccion/72horas

### 5. Bibliografía

- [1] Python. https://www.python.org/.
- [2] Flask. https://flask.palletsprojects.com/en/1.1.x/.
- [3] Apache Airflow. http://airflow.apache.org/.
- [4] MongoDB. https://www.mongodb.com/es.
- [5] pymongo. https://pypi.org/project/pymongo/.
- [6] unittest. https://docs.python.org/3/library/unittest.html.
- [7] pandas. https://pandas.pydata.org/.
- [8] Statsmodels. https://www.statsmodels.org/stable/index.html.
- [9] ARIMA. https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima\_model.ARIMA.html.
- [10] Prophet. https://facebook.github.io/prophet/docs/quick\_start.html.
- [11] Pickle. https://docs.python.org/3/library/pickle.html.
- [12] gunicorn. https://gunicorn.org/.
- [13] Caner Dabakoglu. Time series forecasting arima, lstm, prophet with python. https://medium.com/@cdabakoglu/time-series-forecasting-arima-lstm-prophet-with-python-e73a750a9887, 2019.