

CC II

Práctica # 2: AirFlow

Mario César Rosales Castro

Descripción y objetivo del sistema

En esta práctica el objetivo es utilizar una herramienta de orquestación que permita llevar a cabo un conjunto de tareas de distinta índole y complejidad, las cuales permitan simular un sistema real de predicción de humedad y temperatura. Si bien existen varias herramientas de orquestación ([Luigi](https://github.com/spotify/luigi), [Rundeck](https://www.rundeck.com/open-source), [Microsoft Flow](https://flow.microsoft.com/en-us/)), en este caso se utilizará [Airflow](https://airflow.apache.org/),

Las tareas a realizar concretamente son las siguientes:

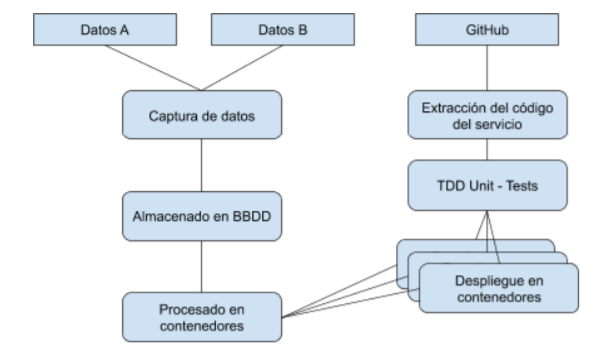


Figura 1 Descripción del sistema (https://pradoposgrado1920.ugr.es/pluginfile.php/136804/mod\_resource/content/1/Pr%C3%A1ctica%202%20CC%20AirFlow.pdf)

A continuación, se analizarán cada una de ellas en detalle

Captura de datos

En esta etapa se hace necesario capturar los datos que servirán de entrada a nuestros servicios de predicción (al menos uno de ellos). A pesar de estar disponibles mediciones de temperature y humedad diarias desde el 2012 en el repositorio <https://github.com/manuparra/MaterialCC2020>, para acelerar el procesamiento del flojo y demostrar las capacidades de Airflow sin optimizar la precisión de la predicción se decidió limitar el número de mediciones a 1000 y colocar los ficheros con extensión .csv en el siguiente repositorio público <https://github.com/mcrosales/cc_ii/tree/master/data>.

A modo de ejemplo, para obtener los datos podemos crear una tarea bash en Airflow con la siguiente información:

CapturaDatosB = BashOperator(

                 task\_id='CapturarDatosB',

                 depends\_on\_past=False,

                 bash\_command='curl -o /tmp/workflow/humidity\_min.csv https://raw.githubusercontent.com/mcrosales/cc\_ii/master/data/humidity\_min.csv',

                 dag=dag

                )