

FBID: APACHE AIRFLOW

- **TODO:** Explain the architecture of Apache Airflow (see the official documentation of Apache Airflow).

Apache Airflow se basa principalmente en los DAGs, que son colecciones de tareas a ejecutar y están conectados mediante relaciones y dependencias. Cada tarea definida en un DAG está representada como un nodo y cada uno de estos DAGs es creado a través de código Python configurando la estructura de cada uno.

Existe una interfaz gráfica para monitorizar el estado de la ejecución de cada uno de los DAGs que se alcanza a través del link introducido en el README.txt, pero no permite crear nuevos flujos de trabajo o DAGs.

Airflow está compuesto de los siguientes componentes:

- Servidor web que sirve la API, interfaz de usuario, etc.
- Scheduler con ejecutores encargados de lanzar los nodos y dividir tareas entre ellos.
- Base de datos a modo de backend encargada de almacenar metadatos, usuarios y ejecuciones.

Los ejecutores contenidos en el planificador pueden ser de dos tipos principalmente:

Ejecutores Monomodo: Permiten la ejecución de tareas en un nodo, no siendo escalables y siendo un único punto de fallo. Pueden ser locales, que ejecutan las tareas en paralelo, o secuenciales, usados para debuggear DAGs que ejecutan una tarea en cada momento.

Ejecutores Cluster: Realizan una asignación fija de nodos para distribuir las cargas de trabajo. Pueden ser Celery, tareas en paralelo en varios nodos separados, Dask, cada tarea en varios nodos dividiendo la carga de trabajo, o Kubernetes, que ejecuta cada tarea en un pod de Kubernetes, desplegando pods en función de los recursos necesarios.

- **TODO:** Analyzing the setup.py: What happens if the task fails?, What is the periodicity of the task?

La tarea no tiene periodicidad, por lo que se ejecutará una vez inicialmente cuando se lance y ejecute el dag en Apache Airflow no volviendo a ejecutarse más.

En el caso de que dicha tarea falle, no se podrá entrenar el modelo corriendo el dag ni se logrará añadir el mismo a los dags disponibles para su posterior visualización y lanzamiento, por lo que se podría cambiar dicha periodicidad para lograr correr el dag desde la interfaz web de Airflow.