



¿QUÉ ES
Airflow?



¿Qué es?

Es una herramienta de código abierto para gestionar, monitorizar y planificar flujos de trabajo mediante programación en Python





Es el nuevo director de orquesta de Google Cloud



Apache
Airflow

**Orquestador
de servicios para automatizar procesos
y llevar a cabo ingesta de datos de forma
programática, dividiéndolos en subtareas**



1
La base son los DAGs (Grafos Acíclicos Dirigidos), que se ejecutan en un flujo de trabajo mediante relaciones de dependencias entre ellos. Cada uno debe tener una entrada y una salida posible para continuar. Cada una de las tareas que componen un DAG se denomina Operadores.

2
Para generar los DAGs se deben unificar todos los operadores necesarios y establecer las relaciones de dependencia entre ellos, para así poder ejecutarlos de la manera en la que sea necesario.



4
Esta herramienta está escrita y como más se utiliza es en Python. Para generar un DAG, toda la estructura se crea en Python, aunque lo que se ejecute puede estar en otros lenguajes.

3
La herramienta cuenta con operadores ya predefinidos como bash, Python, Email, http, SQL. Estos, se configuran para que se ejecute un comando dentro de cada operador.





Apache
Airflow

Características destacables

El potencial de la herramienta se puede extender mediante los plugins de la propia interfaz

Se utiliza sobre todo para el tema de gobierno de los datos

Expone una interfaz web y evita tener que realizar el trabajo de automatización para poder dárselo a un usuario final y, que pueda ejecutar de una manera sencilla

Se puede desplegar mediante contenedores docker

Airflow

Esta información ha sido desarrollada por el equipo de Ayscom:



Alex Méndez Fuentes
Ingeniero de Datos



Darcy Zerpa
Coordinadora DYA