



¿Qué es Airflow?



Un poco de historia

- Fue creado por Maxime Beauchemin dentro de AirBnB en octubre 2014, como proyecto de Open Source.
- En junio 2015 fue puesto a disposición del público en GitHub.
- En marzo de 2016, la fundación Apache Software lo acogió en su programa de incubación.
- En enero de 2019 pasó a ser un proyecto Top Level.



Definición oficial

Airflow is a platform created by community to programmatically author, schedule and monitor workflows.

Airflow es una plataforma creada por la comunidad para programáticamente crear, programar y monitorizar flujos de trabajo.



Componentes principales de Airflow

- **Webserver**
 - Un server Flask con Gunicorn para acceder a la interfaz gráfica
- **Scheduler**
 - Responsable de programar las tareas
- **Metadata database**
 - Base de datos donde se guardan los metadatos de Airflow: información sobre DAGs, conexiones, variables, usuarios...
- **Executor**
 - Define cómo son ejecutadas las tareas



Conceptos clave

- DAG
 - Una representación gráfica de varias tareas
- Operator
 - Representa una tarea de Airflow
- Task
 - Una instancia de un Operator
- Task Instances
 - Una ejecución específica de una tarea (tiene timestamp)
- Workflow
 - Combinación de todos los anteriores