

APACHE AIRFLOW



Instalación y configuración

HENRY

A continuación, te ofrecemos un tutorial de instalación y configuración de **Apache Airflow**, una plataforma de código abierto que permite crear, programar y monitorear flujos de trabajo (workflows) de manera programática.



¿Qué necesitas para su instalación? ¡Simple!:

Tener Python 3.8 o superior instalado.

Espacio en disco: al menos 2 GB libres para entorno y dependencias.

Memoria RAM: mínimo 4 GB (recomendado 8 GB).



DESCARGA (instalación paso a paso)

En Linux y MACOS

- 1 Crea entorno virtual: `python3 -m venv airflow_env`
`source airflow_env/bin/activate`
- 2 Instala Airflow (versión estable): `pip install apache-airflow`
- 3 Verifica la instalación: `airflow version`

En Windows

Instalación Directa

- 1** Abre el menú Inicio y busca "**cmd**" o "**Windows PowerShell**".
- 2** Haz clic derecho y selecciona **Ejecutar como administrador**.
- 3** **Crea y activa un entorno virtual de Python** (recomendado):

```
python -m venv airflow_venv
airflow_venv\Scripts\activate
```

- 4** Asegúrate de tener pip actualizado:

```
python -m pip install --upgrade pip
```

- 5** **Instala Apache Airflow** con el siguiente comando (reemplaza X.Y.Z por la versión de Python que tengas, por ejemplo 3.9):

```
pip install "apache-airflow==2.9.3" --constraint "https://raw.githubusercontent.com/apache/airflow/constraints-2.9.3/constraints-3.9.txt"
```

- 6** Espera unos minutos y verás mensajes de descarga e instalación de varios paquetes adicionales.
- 7** Inicializa la base de datos de Airflow:

```
airflow db init
```

- 8** Crea un usuario administrador:

```
airflow users create \
--username admin \
--firstname Nombre \
--lastname Apellido \
--role Admin \
--email admin@example.com
```

- 9** Inicia el **servidor web de Airflow**:

```
airflow webserver
```

- 10** Abre otro terminal (sin cerrar el anterior) y ejecuta:

```
airflow scheduler
```

- 11** Ingresa en tu navegador a **http://localhost:8080**

- 12** **Inicia sesión** con el usuario y contraseña creados.



Con Ubuntu

1 Instala WSL2 y Ubuntu:

- Abre PowerShell como administrador.
- Escribe el comando: `wsl --install -d Ubuntu`
- Reinicia tu computadora si lo pide.
- Al iniciar Ubuntu, crea un usuario y contraseña.

2 Prepara Ubuntu:

- Actualiza los paquetes:

```
sudo apt update && sudo apt upgrade -y
```

- Verifica que tengas Python 3 `python3 --version` (Si no lo tienes, instala con `sudo apt install python3 python3-pip -y`)

3 Instala Airflow:

- Crea un entorno virtual:

```
python3 -m venv airflow_venv source airflow_venv/bin/activate
```

- Actualiza pip:

```
pip install --upgrade pip
```

- Instala Airflow (ejemplo con Python 3.9):

```
pip install "apache-airflow==2.9.3"--constraint "https://raw.githubusercontent.com/apache/airflow/constraints-2.9.3/constraints-3.9.txt"
```

- Inicializa la base de datos:

```
airflow db init
```

- Crea usuario admin:

```
airflow users create \
--username admin \
--firstname Nombre \
--lastname Apellido \
--role Admin \
--email admin@example.com
```

- Inicia Airflow:

```
airflow webserver (en otra terminal Ubuntu)
airflow scheduler
```

- Abre en tu navegador:

```
http://localhost:8080
```



PRIMERAS CONFIGURACIONES

- 1 Inicializa la base de datos interna de Airflow:

```
airflow db init
```

Esto crea las tablas necesarias en SQLite (por defecto).

- 2 Crea un usuario administrador:

```
airflow users create \
--username admin \
--firstname Nombre \
--lastname Apellido \
--role Admin \
--email admin@example.com
```

- 3 Inicia los servicios principales:

- Servidor web (UI): **airflow webserver -p 8080**
- Planificador (scheduler): **airflow scheduler**

- 4 Accede a la interfaz web:

Abre el navegador en **http://localhost:8080**

