

**UNIVERSIDAD DE GUADALAJARA**  
**CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERIAS**

**Materia:** Computación tolerante a fallas

**Clave:** I7036

**Sección:** D06



**NRC:** 179961

**Airflow**

**Código estudiante:** 221350567

**Alumno:** Rauf Alfonso Hamden Estrada

**Carrera:** Ingeniería en Computación

**Fecha:** 16/10/2023

**Docente:** Michel Emanuel Lopez Franco

2023B

## Introducción

Apache Airflow es una plataforma de código abierto utilizada para orquestar y automatizar flujos de trabajo en una variedad de aplicaciones, desde tareas simples hasta flujos de trabajo complejos y dependientes. Airflow se utiliza comúnmente en la programación de tareas y la administración de flujos de trabajo en entornos de procesamiento de datos, como la ingeniería de datos, la inteligencia artificial, el análisis de big data y la automatización de procesos.

## Desarrollo

### Instalamos virtualenv

```
C:\Users\raufa>pip install virtualenv
Collecting virtualenv
  Obtaining dependency information for virtualenv from https://files.pythonhosted.org/packages/4e/8b/f0d3a468c0186c603217a6656ea4f49259630e8ed99558501d92f6ff7dc3/virtualenv-20.24.5-py3-none-any.whl.metadata
  Downloading virtualenv-20.24.5-py3-none-any.whl.metadata (4.5 kB)
Collecting distlib<1,>=0.3.7 (from virtualenv)
  Obtaining dependency information for distlib<1,>=0.3.7 from https://files.pythonhosted.org/packages/43/a0/9ba967fdbd55293bacfc1507f58e316f740a3b231fc00e3d86dc39bc185a/distlib-0.3.7-py2.py3-none-any.whl.metadata
  Downloading distlib-0.3.7-py2.py3-none-any.whl.metadata (5.1 kB)
Collecting filelock<4,>=3.12.2 (from virtualenv)
  Obtaining dependency information for filelock<4,>=3.12.2 from https://files.pythonhosted.org/packages/5e/5d/97afbaf9d584ff1b45fcb354a479a3609bd97f912f8f1f6c563cb1fae21/filelock-3.12.4-py3-none-any.whl.metadata
  Downloading filelock-3.12.4-py3-none-any.whl.metadata (2.8 kB)
Requirement already satisfied: platformdirs<4,>=3.9.1 in c:\users\raufa\appdata\local\programs\python\python311\lib\site-packages (from virtualenv) (3.10.0)
Downloading virtualenv-20.24.5-py3-none-any.whl (3.7 MB)
----- 3.7/3.7 MB 311.4 kB/s eta 0:00:00
Downloading distlib-0.3.7-py2.py3-none-any.whl (468 kB)
----- 468.9/468.9 kB 264.5 kB/s eta 0:00:00
Downloading filelock-3.12.4-py3-none-any.whl (11 kB)
Installing collected packages: distlib, filelock, virtualenv
Successfully installed distlib-0.3.7 filelock-3.12.4 virtualenv-20.24.5

[notice] A new release of pip is available: 23.2.1 -> 23.3
[notice] To update, run: python.exe -m pip install --upgrade pip
```

### Primero creamos un entorno virtual

```
C:\Users\raufa>virtualenv airflow_env
created virtual environment CPython3.11.4.final.0-64 in 10995ms
creator CPython3Windows(dest=C:\Users\raufa\airflow_env, clear=False, no_vcs_ignore=False, global=False)
seeder FromAppData(download=False, pip=bundle, setuptools=bundle, wheel=bundle, via=copy, app_data_dir=C:\Users\raufa\AppData\Local\pypa\virtualenv)
added seed packages: pip==23.2.1, setuptools==68.2.0, wheel==0.41.2
activators BashActivator,BatchActivator,FishActivator,NushellActivator,PowerShellActivator,PythonActivator
```

## Activamos el entorno virtual

```
C:\Users\raufa>airflow_env\Scripts\activate  
  
(airflow_env) C:\Users\raufa>
```

## Instalo el airflow

```
Building wheel for python-nvd3 (setup.py) ... done  
Created wheel for python-nvd3: filename=python_nvd3-0.15.0-py3-none-any.whl size=38186 sha256=8ae64c837aa5aa446ed86c0e974d0bc21a615518e0992778463b2c5e1064a995  
Stored in directory: c:\users\raufa\appdata\local\pip\cache\wheels\1e\8a\ef\61e11108e04dac50d0ac846cea099675548974495b40870edb  
Building wheel for unicodcsv (setup.py) ... done  
Created wheel for unicodcsv: filename=unicodcsv-0.14.1-py3-none-any.whl size=10759 sha256=54ef6354ca3fb48d5d40b55d319cc2822b95722a6da0e56cc6559a2e94842c9a  
Stored in directory: c:\users\raufa\appdata\local\pip\cache\wheels\ec\03\6f\d2e0162d94c0d451556fa43dd4d5531457245c34a36b41ef4a  
Successfully built cron-descriptor python-nvd3 unicodcsv  
Installing collected packages: unicodcsv, text-unidecode, pytz, lockfile, cron-descriptor, zipp, wrapt, urllib3, uc-micro-py, typing-extensions, termcolor, tenacity, tabulate, sqlparse, sniffio, six, setproctitle, rpsds-py, PyYAML, pytzdata, python-slugify, pyjwt, pygments, pycparser, psutil, protobuf, pluggy, pathspec, packaging, ordered-set, opentelemetry-semantic-conventions, multidict, mdurl, markupsafe, markdown, lazy-object-proxy, itsdangerous, inflection, importlib-resources, idna, h11, grpcio, greenlet, graphviz, google-re2, frozenlist, docutils, dnspython, dill, configupdater, colorama, charset-normalizer, certifi, cachelib, blinker, backoff, Babel, attrs, async-timeout, asgiref, argcomplete, annotated-types, yarl, WTForms, werkzeug, sqlalchemy, rfc3339-validator, requests, referencing, python-dateutil, python-daemon, pydantic-core, prison, opentelemetry-proto, marshmallow, markdown-it-py, Mako, linkify-it-py, Jinja2, importlib-metadata, gunicorn, googleapis-common-protos, email-validator, deprecated, colorlog, click, cffi, cattrs, apispec, anyio, aiosignal, sqlalchemy-utils, sqlalchemy-jsonfield, rich, requests-toolbelt, python-nvd3, pydantic, pendulum, opentelemetry-exporter-otlp-proto-common, opentelemetry-api, mdit-py-plugins, marshmallow-oneofschema, limits, jsonschema-specifications, httpcore, flask, cryptography, croniter, clickclick, alembic, aiohttp, rich-argparse, opentelemetry-sdk, jsonschema, httpx, flask-wtf, Flask-SQLAlchemy, flask-session, flask-login, Flask-Limiter, Flask-JWT-Extended, flask-caching, Flask-Babel, opentelemetry-exporter-otlp-proto-http, opentelemetry-exporter-otlp-proto-grpc, flask-appbuilder, connexion, opentelemetry-exporter-otlp, apache-airflow-providers-common-sql, apache-airflow-providers-sqlite, apache-airflow-providers-imap, apache-airflow-providers-http, apache-airflow-providers-ftp, apache-airflow
```

## Inicializamos la base de datos

```
Successfully installed pip-23.3  
  
(airflow_env) C:\Users\raufa>airflow db init  
WARNING:root:OSError while attempting to symlink the latest log directory  
C:\Users\raufa\airflow_env\Lib\site-packages\airflow\cli\commands\db_command.py:43 DeprecationWarning: `db init` is deprecated. Use `db migrate` instead to migrate the db and/or airflow connections create-default-connections to create the default connections  
DB: sqlite:///C:\Users\raufa\airflow\airflow.db  
[2023-10-17T00:21:05.582-0500] {migration.py:213} INFO - Context impl SQLiteImpl.  
[2023-10-17T00:21:05.607-0500] {migration.py:216} INFO - Will assume non-transactional DDL.  
INFO [alembic.runtime.migration] Context impl SQLiteImpl.  
INFO [alembic.runtime.migration] Will assume non-transactional DDL.  
INFO [alembic.runtime.migration] Running stamp_revision -> 405de8318b3a  
WARNI [airflow.models.crypto] empty cryptography key - values will not be stored encrypted.  
Initialization done
```

## Creamos al usuario y verificamos que este correctamente

```
(airflow_env) C:\Users\raufa>airflow users create --username admin --password admin --firstname admin --lastname admin --role Admin --email raufalfonso@gmail.com
WARNING:root:OSError while attempting to symlink the latest log directory
C:\Users\raufa\airflow_env\Lib\site-packages\flask_limiter\extension.py:336 UserWarning: Using the in-memory storage for tracking rate limits as no storage was explicitly specified. This is not recommended for production use. See: https://flask-limiter.readthedocs.io#configuring-a-storage-backend for documentation about configuring the storage backend.
[2023-10-17T00:26:26.072-0500] {manager.py:670} WARNING - No user yet created, use flask fab command to do it.
[2023-10-17T00:26:26.852-0500] {manager.py:211} INFO - Added user %s
User "admin" created with role "Admin"

(airflow_env) C:\Users\raufa>
(airflow_env) C:\Users\raufa>airflow users list
WARNING:root:OSError while attempting to symlink the latest log directory
id | username | email | first_name | last_name | roles
=====+=====+=====+=====+=====+=====
1 | admin | raufalfonso@gmail.com | admin | admin | Admin
```

## Conclusión

Airflow es una plataforma poderosa para automatizar flujos de trabajo, pero su configuración y puesta en marcha pueden ser desafiantes debido a sus numerosos requisitos y configuraciones. A pesar de los obstáculos, el proceso de trabajar con Airflow puede ser una experiencia educativa, ya que te brinda un alto grado de control y personalización. Airflow es una plataforma el cual me costo mucho trabajo ponerla andar debido a que tenia muchos requisitos y al final no llegue a los resultados esperados, sin embargo se disfruto y se llego hasta el final, sin embargo, hay muchas mas plataforma para poder automatizar los trabajos y mucho mas fáciles de usar