**demo\_etl\_pipeline**

Create an etl script (getData.py)

Create an ec2 instance and run airflow in it

* ssh into ec2 instance
* install your necessary dependencies
* test if airflow is installed and available, command: airflow
* run airflow using command: airflow standalone
  + copy instance public ip
  + paste public ip on the browser
  + log in to airflow

Create a dag file (getData\_dag.py)

Create the etl DAG

* have the etl script ready as well as the dag file ready
* import the files to airflow ->
  + check folders/files in your directory, command: ls
  + cd into airflow folder, command: cd airflow
  + edit the airflow.cfg file, command: sudo nano airflow.cfg
    - change folder name from dags to data\_dag
  + create a folder to store your files, command: mkdir files
  + copy files from local computer to files folder in airflow server
    - command: sudo nano getData.py (paste code and save)
    - command: sudo nano getData\_dag.py (paste code and save)