Airflow Mini-Project System Set-up Steps

1. Install docker and Airflow (celery)

● Install docker on WSL Ubuntu

Link	https://www.digitalocean.com/community/tutorials/how-to-install-and-use-docker-on-ubuntu-20-04
Steps	# Update your existing list of packages: sudo apt update #Install a few prerequisite packages sudo apt install apt-transport-https ca-certificates curl software-properties-common #Add the GPG key for the official Docker repository to your system: curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key add - #Add the Docker repository to APT sources: sudo add-apt-repository "deb [arch=amd64] https://download.docker.com/linux/ubuntu focal stable" #Make sure you are about to install from the Docker repo instead of the default Ubuntu repo: apt-cache policy docker-ce #IInstall Docker sudo apt install docker-ce #Check status sudo service docker status #Add user 'chu' to group docker sudo usermod -aG docker chu

- Install docker-compose (v1.29.1 or newer)

Steps	[How To Install and Use Docker Compose on Ubuntu 20.04 | DigitalOcean](#) sudo curl -L "https://github.com/docker/compose/releases/download/1.29.2/docker-compose-$(uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose sudo chmod +x /usr/local/bin/docker-compose /usr/local/bin/docker-compo

- Install airflow using docker container

  This step is to set up airflow using container in folder ' ~/airflow_docker'. We will follow the steps provided by Airflow official website, which will set up Airflow using Celery mode.

Link	[Running Airflow in Docker — Airflow Documentation (apache.org)](#)
Set up	# create folder ~/airflow_docker su chu cd ~ mkdir airflow_docker cd airflow_docker # fetch docker-compose.yaml. `curl -LfO 'https://airflow.apache.org/docs/apache-airflow/2.3.2/docker-compose.yaml'`

Customize	1. Modify docker-compose.yaml file
	   a. Add one entry in 'volumes' section (for setting up '/tmp' folder in docker container)
	
	     volumes:
	      - ./dags:/opt/airflow/dags
	      - ./logs:/opt/airflow/logs
	      - ./plugins:/opt/airflow/plugins
	      - /data/tmp:/tmp

	b. Modify '_PIP_ADDITIONAL_REQUIREMENTS' section (to install python modules that will be used)
	
	…: ${_PIP_ADDITIONAL_REQUIREMENTS:- **pandas yfinance** }
	
	2. Create folders
	    a. In folder '~/chu/airflow_docker', create following 3 folders:
	    ./dags - you can put your DAG files here.
	    ./logs - contains logs from task execution and scheduler.
	
	    ./plugins - you can put your custom plugins here.
	
	    b. Create folder '/data/tmp' (to be mapped to '/tmp' in container) and sub-folders
	
	mkdir /data/tmp
	chmod +777 /data/tmp
	
	mkdir /data/tmp/data
	
	mkdir /data/tmp/query
	chmod +777 /data/tmp
	
	3. Set an env variable
	    .Type 'nano ~/.bashrc'
	    <u>.Add the following command at the</u> end:
	
	export AIRFLOW_UID=50000

Initialize the database	Execute following command to run database migrations and create the first user account ('airflow'): docker-compose up airflow-init
Run airflow	Start the service using following command docker-compose up

2. Deploy

.copy 'dag_minimain.py' to Airflow dags folder .

```
. Login unbun as user 'chu'
.Copy  'dag_minimain.py to ~/airflow_docker/dags folder
```

3. Run

1) Make sure Docker is running

```
#Check if Docker is running
sudo service docker status

#If Docker is  not running, type following command to start it
sudo service docker start
```

2) Start Airflow related docker contains to start the Airflow services

```
cd ~/airflow_docker

# The following command will allow start the dockers and attached to the
webserver container
docker-compose up
```

4. Monitor

```
 Launch a  web browser,  visit 'localhost:8080' . Enter username ('airflow') and
password ('airflow').
```