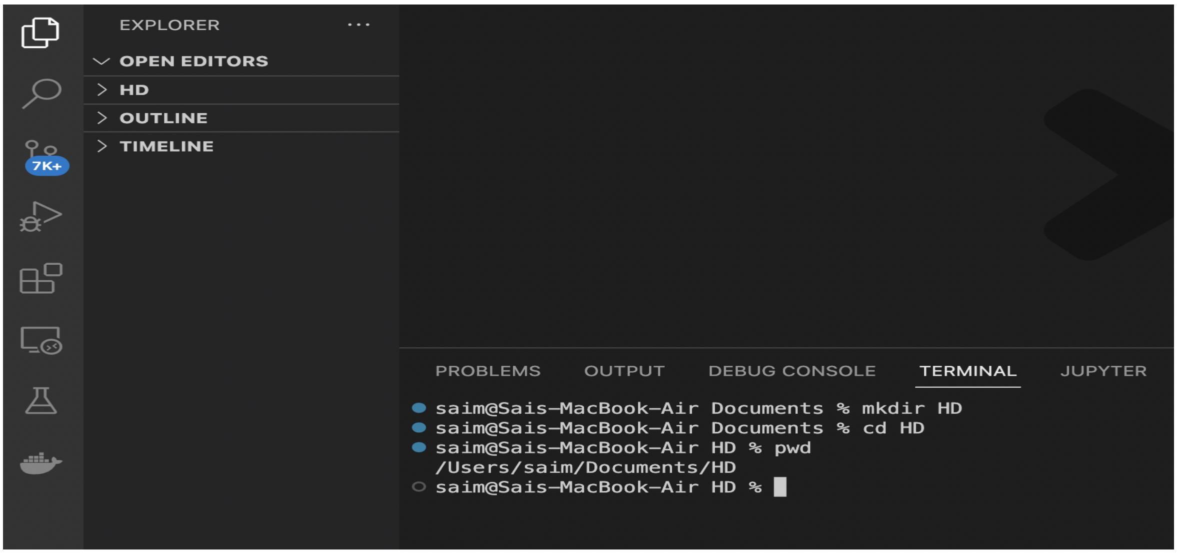
Instructions Document for Google Trends Solution using Airflow

1. **Create a directory in your local system and get inside the newly created directory using following commands**

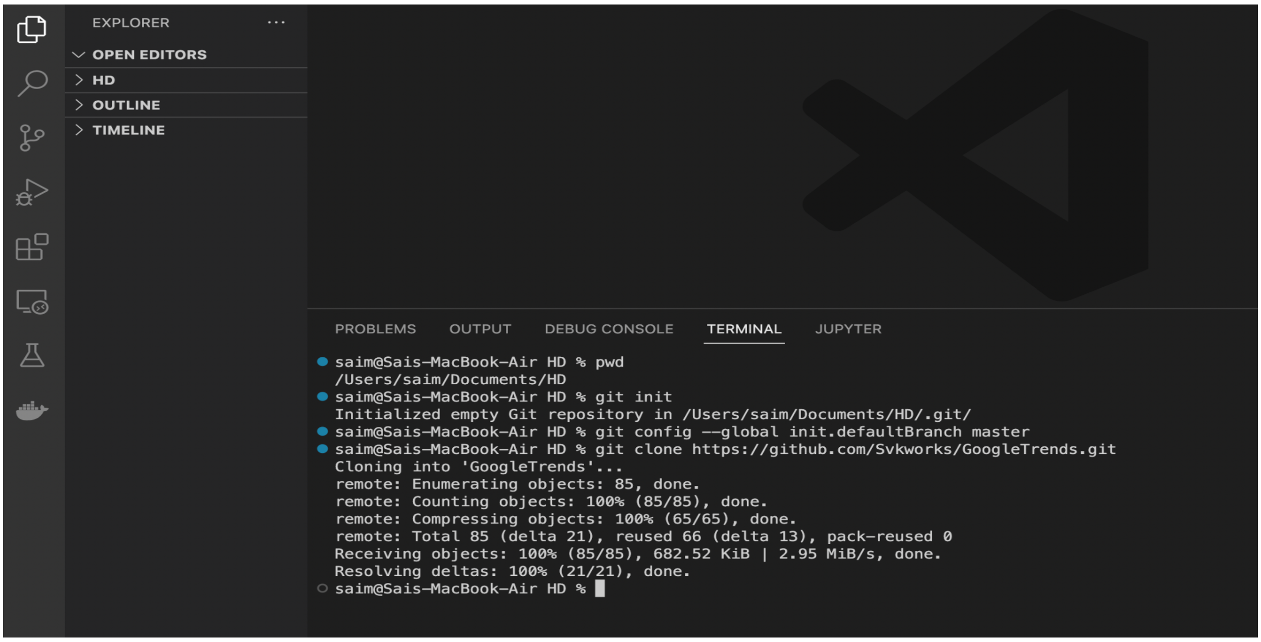
```
mkdir HD
cd HD
```



```
saim@Sais-MacBook-Air Documents % mkdir HD
saim@Sais-MacBook-Air Documents % cd HD
saim@Sais-MacBook-Air HD % pwd
/Users/saim/Documents/HD
saim@Sais-MacBook-Air HD %
```

2. **To Clone the GoogleTrends, we need to initialise the git inside the "HD" directory**

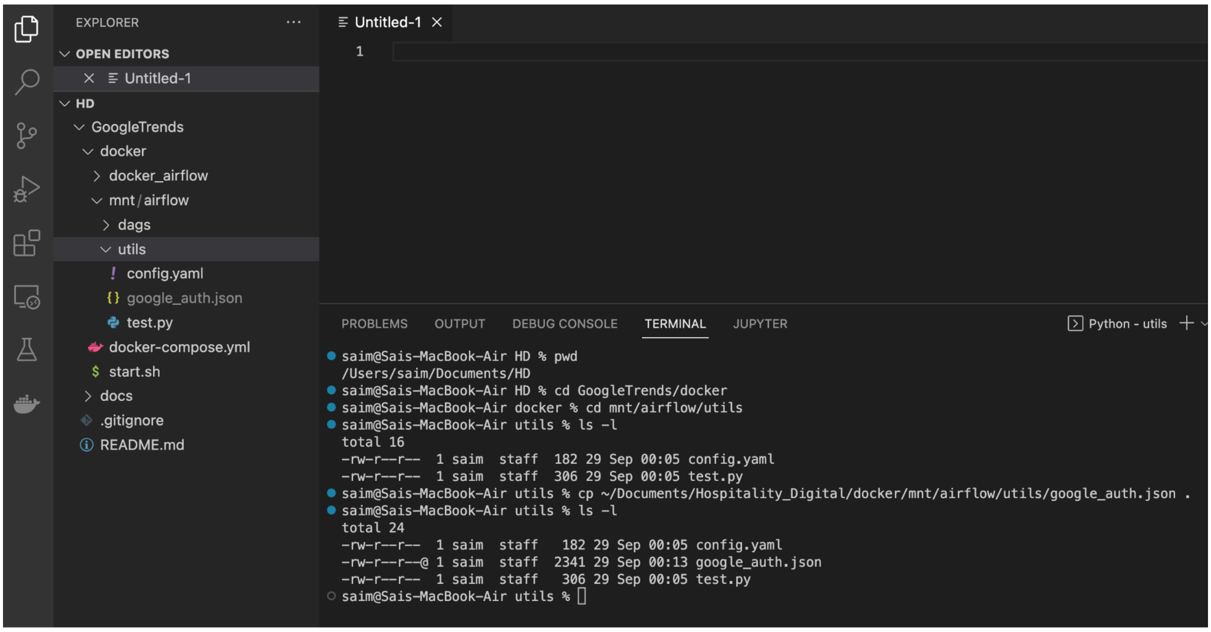
```
git init

git config --global init.defaultBranch master

git clone https://github.com/Svkworks/GoogleTrends.git
```



```
saim@Sais-MacBook-Air HD % pwd
/Users/saim/Documents/HD
saim@Sais-MacBook-Air HD % git init
Initialized empty Git repository in /Users/saim/Documents/HD/.git/
saim@Sais-MacBook-Air HD % git config --global init.defaultBranch master
saim@Sais-MacBook-Air HD % git clone https://github.com/Svkworks/GoogleTrends.git
Cloning into 'GoogleTrends'...
remote: Enumerating objects: 85, done.
remote: Counting objects: 100% (85/85), done.
remote: Compressing objects: 100% (65/65), done.
remote: Total 85 (delta 21), reused 66 (delta 13), pack-reused 0
Receiving objects: 100% (85/85), 682.52 KiB | 2.95 MiB/s, done.
Resolving deltas: 100% (21/21), done.
saim@Sais-MacBook-Air HD %
```

3. Once you clone the GoogleTrends project, then move to "GoogleTrends/docker/mnt/airflow/utils", and Please "add your google_auth.json" file in following location

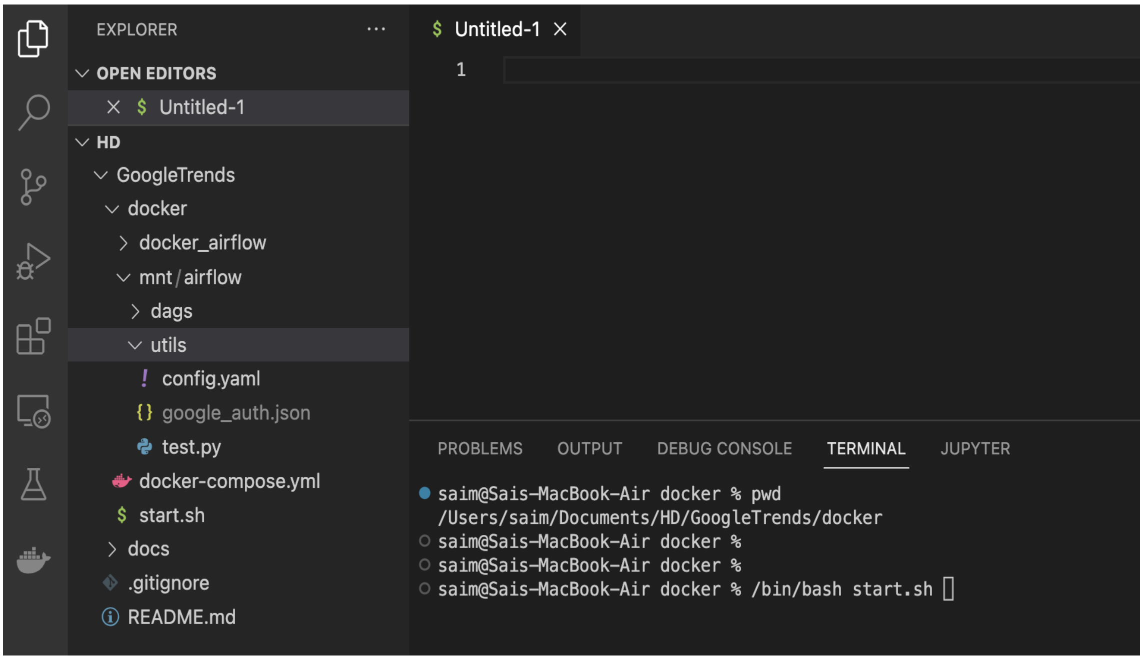
```
cd GoogleTrends/docker
cd mnt/airflow/utils
cp ~/Documents/Hospitality_Digital/docker/mnt/airflow/utils/google_auth.json .
```



```
saim@Sais-MacBook-Air HD % pwd
/Users/saim/Documents/HD
saim@Sais-MacBook-Air HD % cd GoogleTrends/docker
saim@Sais-MacBook-Air docker % cd mnt/airflow/utils
saim@Sais-MacBook-Air utils % ls -l
total 16
-rw-r--r--  1 saim  staff  182 29 Sep 00:05 config.yaml
-rw-r--r--  1 saim  staff  306 29 Sep 00:05 test.py
saim@Sais-MacBook-Air utils % cp ~/Documents/Hospitality_Digital/docker/mnt/airflow/utils/google_auth.json .
saim@Sais-MacBook-Air utils % ls -l
total 24
-rw-r--r--  1 saim  staff   182 29 Sep 00:05 config.yaml
-rw-r--r--@ 1 saim  staff  2341 29 Sep 00:13 google_auth.json
-rw-r--r--  1 saim  staff   306 29 Sep 00:05 test.py
saim@Sais-MacBook-Air utils % 
```

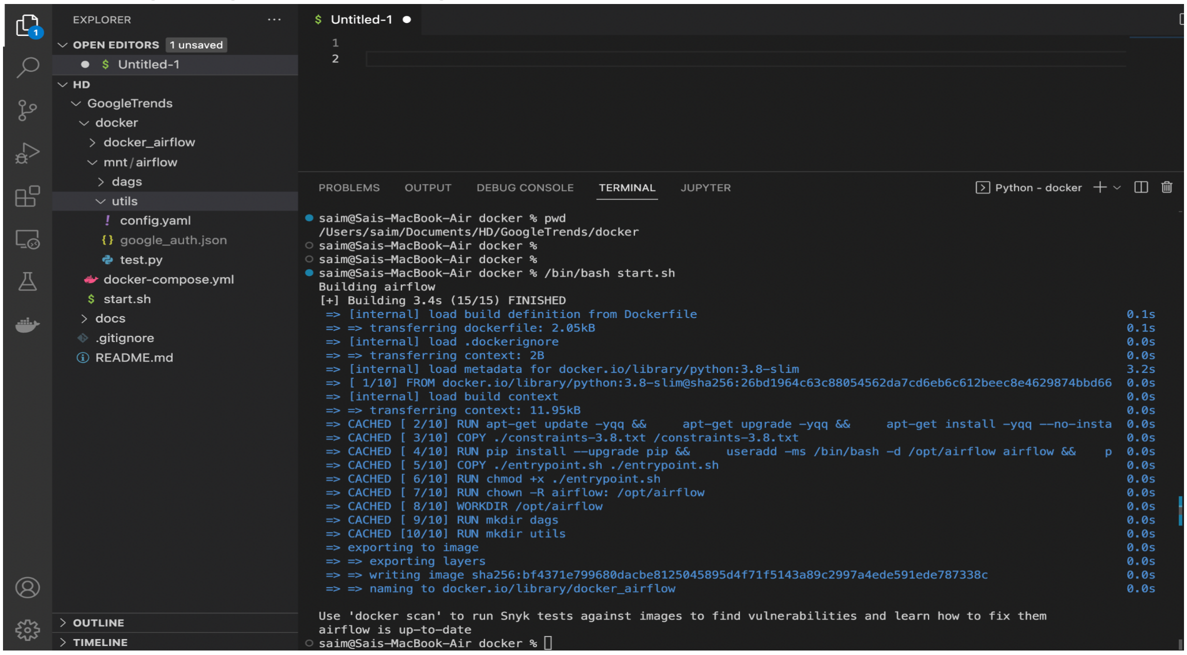
4. Once you add google_auth.json, Please run the following commands to go to "docker" directory. Following command will create airflow docker, (Note: Your docker folder should be your present working directory (PWD))

```
cd ../../../

/bin/bash start.sh
```



```
saim@Sais-MacBook-Air docker % pwd
/Users/saim/Documents/HD/GoogleTrends/docker
saim@Sais-MacBook-Air docker %
saim@Sais-MacBook-Air docker %
saim@Sais-MacBook-Air docker % /bin/bash start.sh 
```

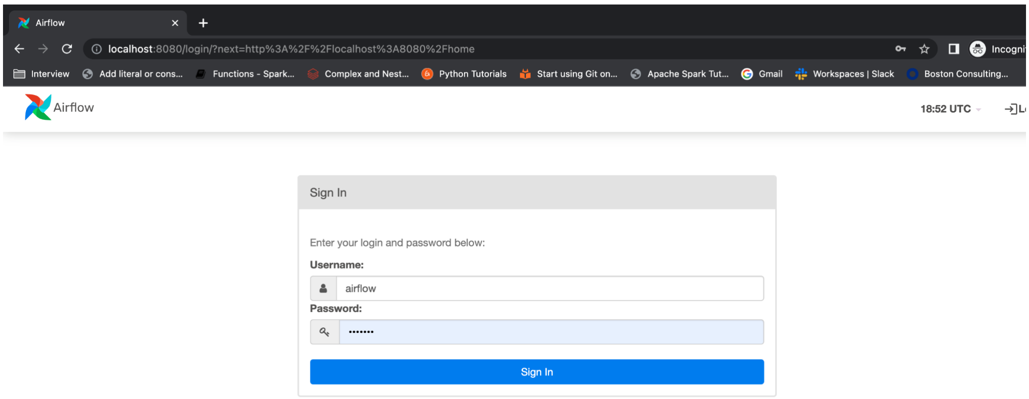
5. Once you run start.sh file in dockor directory, it will create airflow docker image, and we will be getting below message.



```
saim@Sais-MacBook-Air docker % pwd
/Users/saim/Documents/HD/GoogleTrends/docker
saim@Sais-MacBook-Air docker %
saim@Sais-MacBook-Air docker %
saim@Sais-MacBook-Air docker % /bin/bash start.sh
Building airflow
[+] Building 3.4s (15/15) FINISHED
 => [internal] load build definition from Dockerfile                                                         0.1s
 => => transferring dockerfile: 2.05kB                                                                       0.1s
 => [internal] load .dockerignore                                                                            0.0s
 => => transferring context: 2B                                                                              0.0s
 => [internal] load metadata for docker.io/library/python:3.8-slim                                           3.2s
 => [ 1/10] FROM docker.io/library/python:3.8-slim@sha256:26bd1964c63c88054562da7cd6eb6c612beec8e4629874bbd66 0.0s
 => [internal] load build context                                                                            0.0s
 => => transferring context: 11.95kB                                                                         0.0s
 => CACHED [ 2/10] RUN apt-get update -yqq &&     apt-get upgrade -yqq &&     apt-get install -yqq --no-insta 0.0s
 => CACHED [ 3/10] COPY ./constraints-3.8.txt /constraints-3.8.txt                                           0.0s
 => CACHED [ 4/10] RUN pip install --upgrade pip &&     useradd -ms /bin/bash -d /opt/airflow airflow &&    p 0.0s
 => CACHED [ 5/10] COPY ./entrypoint.sh ./entrypoint.sh                                                      0.0s
 => CACHED [ 6/10] RUN chmod +x ./entrypoint.sh                                                              0.0s
 => CACHED [ 7/10] RUN chown -R airflow: /opt/airflow                                                        0.0s
 => CACHED [ 8/10] WORKDIR /opt/airflow                                                                      0.0s
 => CACHED [ 9/10] RUN mkdir dags                                                                            0.0s
 => CACHED [10/10] RUN mkdir utils                                                                           0.0s
 => exporting to image                                                                                       0.0s
 => => exporting layers                                                                                      0.0s
 => => writing image sha256:bf4371e799680dacbe8125045895d4f71f5143a89c2997a4ede591ede787338c                 0.0s
 => => naming to docker.io/library/docker_airflow                                                            0.0s

Use 'docker scan' to run Snyk tests against images to find vulnerabilities and learn how to fix them
airflow is up-to-date
saim@Sais-MacBook-Air docker %
```

6. Once you run the above command, it will create a snowflake docker container, to verify the container status please go to "http://localhost:8080" You can see airflow login page, and please use the following credentials to access the airflow webUI

```
username : airflow

password : airflow
```



Sign In

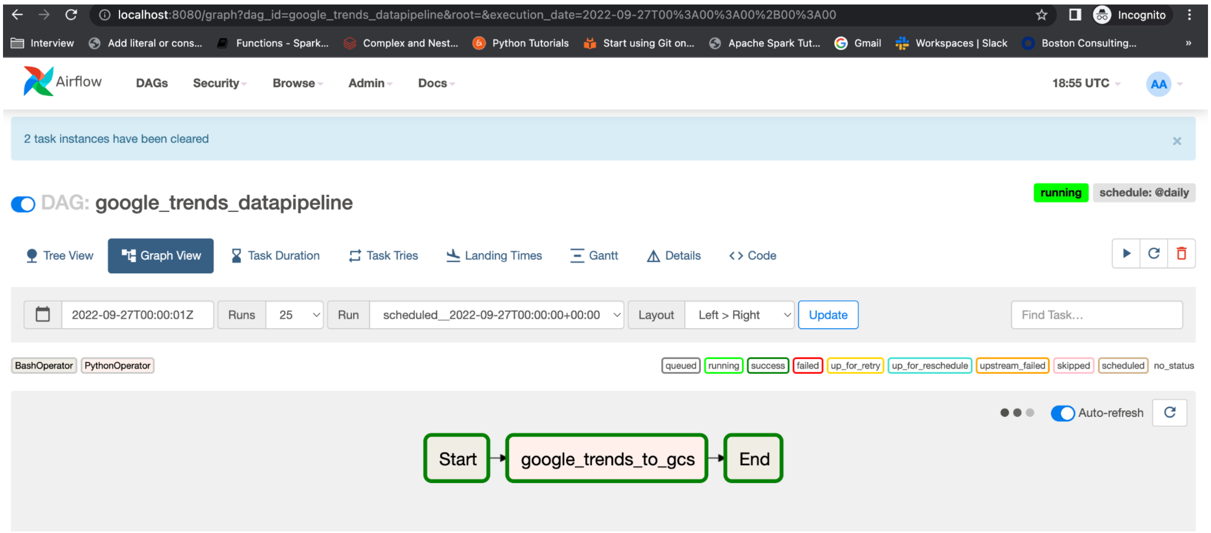
Enter your login and password below:

Username: airflow

Password: •••••••

Sign In

7. **Once you login in airflow webUI, we can see “google_trends_datapipeline”, clear the logs and trigger the dag, then the DAG will be executed successfully**



localhost:8080/graph?dag_id=google_trends_datapipeline&root=&execution_date=2022-09-27T00%3A00%3A00%2B00%3A00

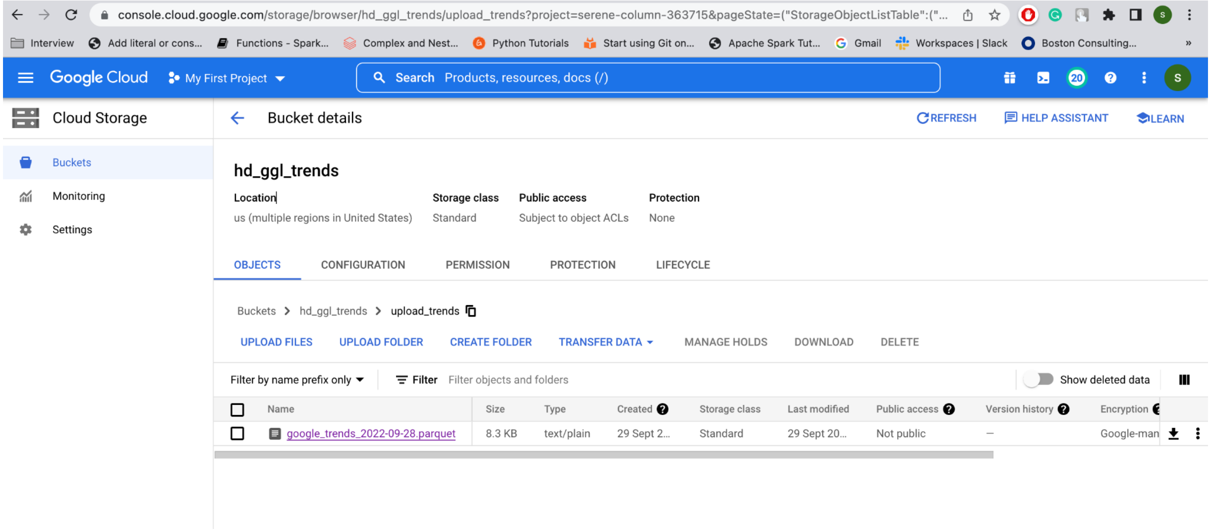
2 task instances have been cleared

DAG: google_trends_datapipeline

running schedule: @daily

2022-09-27T00:00:01Z Runs 25 Run scheduled__2022-09-27T00:00:00+00:00 Layout Left > Right

8. **Please go to your Google cloud Bucket and verify the file ingestion. There we can see the file as below.**



console.cloud.google.com/storage/browser/hd_ggl_trends/upload_trends?project=serene-column-363715&pageState=("StorageObjectListTable":("...

Bucket details

hd_ggl_trends

Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Subject to object ACLs	None

Buckets > hd_ggl_trends > upload_trends

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption
google_trends_2022-09-28.parquet	8.3 KB	text/plain	29 Sept 2...	Standard	29 Sept 20...	Not public	–	Google-man