

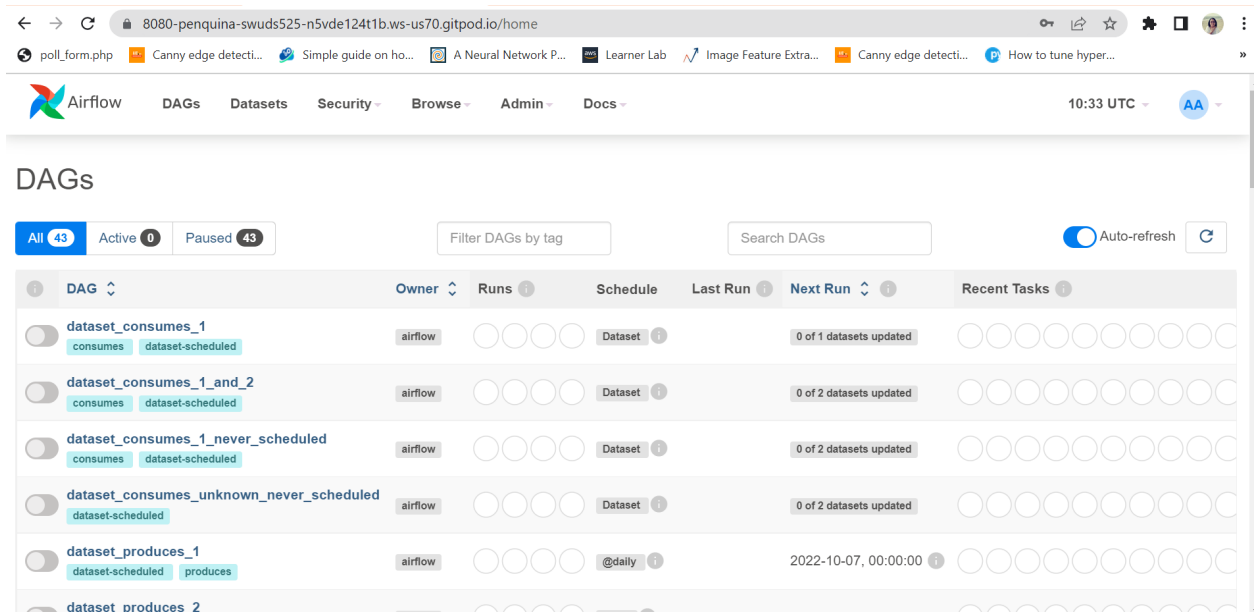
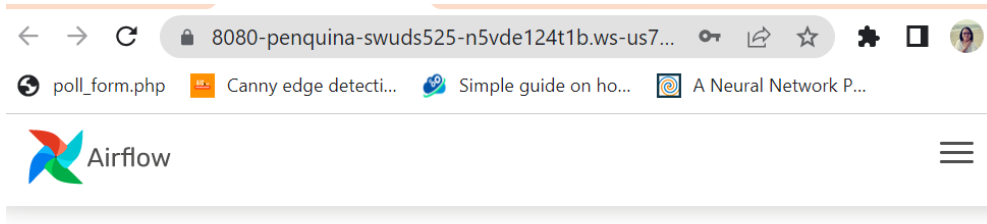
Project: Creating and Automating a Set of Data Pipelines with Airflow


สามารถเริ่มต้นจากโค้ดที่ <https://github.com/zkan/swu-ds525> ได้

สิ่งที่คาดหวังในโปรเจกต์นี้

1. นำโค้ดจาก [Project: Building a Data Modeling with Postgres \(SQL\)](#) มาสร้าง data pipeline โดยใช้ Airflow
2. มีการเขียน documentation อธิบายสิ่งที่ตัวเองทำลงไป รวมไปถึงการออกแบบ data model

3. มี instruction ในการรันโค้ดของตัวเอง



 Airflow

☐

example_branch_dop_operator_v3

airflow

*1 * * * *

☐

example_branch_labels

airflow

@daily

☐

example_branch_operator

airflow


example

example2

@daily

← → ↺ 8080-penquina-swuds525-n5vde124t1b.ws-us70.gitpod.io/dags/tutorial/code

poll_form.php Canny edge detecti... Simple guide on ho... A Neural Network P... Learner Lab Image Feature Extra... Canny edge detecti...

 DAGs Datasets Security Browse Admin Docs 10:41 UTC AA

☐ DAG: tutorial A simple tutorial DAG

Schedule: 1 day, 0:00:00 Next Run: 2022-10-07, 10:40:21

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details

<> Code Audit Log

1 #
2 # Licensed to the Apache Software Foundation (ASF) under one
3 # or more contributor license agreements. See the NOTICE file

Toggle Wrap

← → ↺ penquina-swuds525-n5vde124t1b.ws-us70.gitp...

poll_form.php Canny edge detecti... Simple guide on ho... A Neural Network P...

started README.md U docker-compose.yaml U my_dag.py U

05-creating-and-scheduling-data-pipelines > dags > my_dag.py

1 from airflow import DAG
2 from airflow.utils import timezone


```
#step 3
t1 >> t2
```

DAG: mydag

success Schedule: None Next Run: None

Grid Graph Calendar Task Duration Task Times Landing Times Gantt Details <> Code Audit Log

2022-10-08T11:04:14Z Runs 25 Run manual__2022-10-08T11:04:13.596122+00:00 Layout Left > Right Update Find Task...

EmptyOperator

deferred failed queued running scheduled skipped success up_for_reschedule up_for_retry upstream_failed no_status

Auto-refresh

```

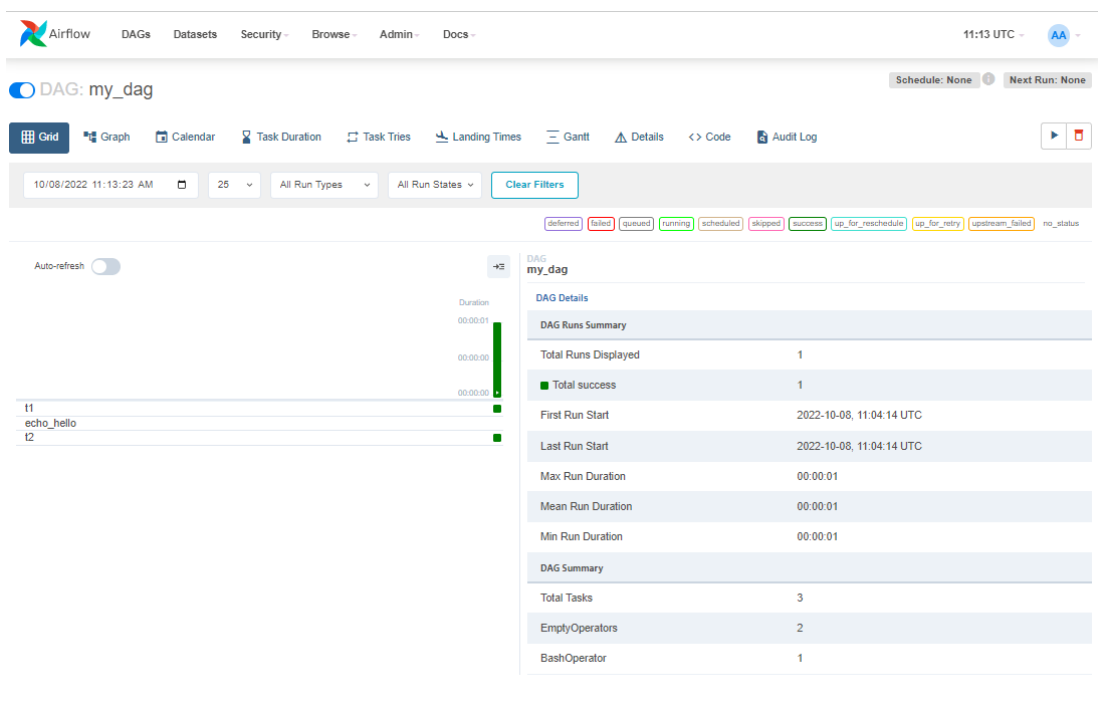
graph LR
    t1[t1] --> t2[t2]
  
```

```

# Task
#1st task ยังไม่ทำอะไร
# step 1
t1 = EmptyOperator( task_id = "t1")

# 2nd task
t2 = EmptyOperator( task_id = "t2")
#step 3
t1 >> t2

```



Airflow DAGs Datasets Security - Browse - Admin - Docs - 11:19 UTC - AA -

Triggered my_dag, it should start any moment now.

DAG: my_dag Schedule: None Next Run: None

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details <> Code Audit Log

10/08/2022 11:18:11 AM 25 All Run Types All Run States Clear Filters

Auto-refresh

Duration

11 echo_hello 12

DAG my_dag Run Task echo_hello

Task Instance Details Rendered Template Log XCom List Instances, all runs Filter Upstream

Details Logs

(by attempts)

1

All Levels All File Sources Wrap Full Logs Download See More

```
[2022-10-08, 11:18:14 UTC] (taskinstance.py:285) WARNING - Dependency <task(_PythonDecoratedOperator): load>, transform a
[2022-10-08, 11:18:14 UTC] (taskinstance.py:285) WARNING - Dependency <task(_PythonDecoratedOperator): transform>, load a
[2022-10-08, 11:18:14 UTC] (taskinstance.py:285) WARNING - Dependency <task(_PythonDecoratedOperator): load>, transform a
[2022-10-08, 11:18:15 UTC] (taskinstance.py:184) INFO - Running <taskinstance: my_dag:echo_hello manual_2022-10-08T11
[2022-10-08, 11:18:15 UTC] (taskinstance.py:1592) INFO - Exporting the following env vars:
AIRFLOW_CTX_DAG_OWNER=***
AIRFLOW_CTX_DAG_ID=my_dag
AIRFLOW_CTX_TASK_ID=echo_hello
AIRFLOW_CTX_EXECUTION_DATE=2022-10-08T11:18:13.196026+00:00
AIRFLOW_CTX_TRY_NUMBER=1
AIRFLOW_CTX_DAG_RUN_ID=manual_2022-10-08T11:18:13.196026+00:00
[2022-10-08, 11:18:15 UTC] (subprocess.py:63) INFO - Temp dir root location:
/tmp
[2022-10-08, 11:18:15 UTC] (subprocess.py:75) INFO - Running command: ['bin/bash', '-c', 'echo 'hello''']
[2022-10-08, 11:18:15 UTC] (subprocess.py:86) INFO - Output:
[2022-10-08, 11:18:15 UTC] (subprocess.py:93) INFO - hello
[2022-10-08, 11:18:15 UTC] (subprocess.py:97) INFO - Command exited with return code 0
[2022-10-08, 11:18:15 UTC] (taskinstance.py:1486) INFO - Marking task as SUCCESS, dag_id=my_dag, task_id=echo_hello, o
[2022-10-08, 11:18:15 UTC] (local_task_job.py:164) INFO - Task exited with return code 0
[2022-10-08, 11:18:15 UTC] (local_task_job.py:275) INFO - 1 downstream tasks scheduled from follow-on schedule check
```

Airflow DAGs Datasets Security - Browse - Admin - Docs - 11:25 UTC - AA -

Triggered my_dag, it should start any moment now.

DAG: my_dag Schedule: None Next Run: None

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details <> Code Audit Log

10/08/2022 11:24:48 AM 25 All Run Types All Run States Clear Filters

Auto-refresh

Duration

11 echo_hello print_hey 12

DAG my_dag

DAG Details

DAG Runs Summary

Total Runs Displayed	5
Total success	3
Total failed	2
First Run Start	2022-10-08, 11:15:36 UTC
Last Run Start	2022-10-08, 11:24:47 UTC
Max Run Duration	00:00:03
Mean Run Duration	00:00:02
Min Run Duration	00:00:02

DAG Summary

Total Tasks	4
EmptyOperators	2
BashOperator	1
PythonOperator	1

← → ↻ crontab.guru/#30_8_**_*

poll_form.php Canny edge detecti... Simple guide on ho... A Neural Network P... Learner Lab

Cronitor Cron Job Monitoring

crontab guru

The quick and simple editor for cron schedule expressions by [Cronitor](#)

“At 08:30.”

[next](#) at 2022-10-09 08:30:00 [random](#)

30 8 * * *

minute	hour	day (month)	month	day (week)
*		any value		
		value list		
		separator		
		range of values		

```
with DAG(  
    "my_dag",  
    start_date = timezone.datetime(2022, 10, 8),  
    schedule = "* * * * *",  
    tags = ["workshop"],  
):
```