

Hands-on Lab: Getting started with Apache Airflow

Estimated time needed: **20** minutes

Objectives

After completing this lab you will be able to:

- Start Apache Airflow.
- Open the Airflow UI in a browser.
- List all the DAGs.
- List the tasks in a DAG.
- Explore a DAG in the UI.

About Skills Network Cloud IDE

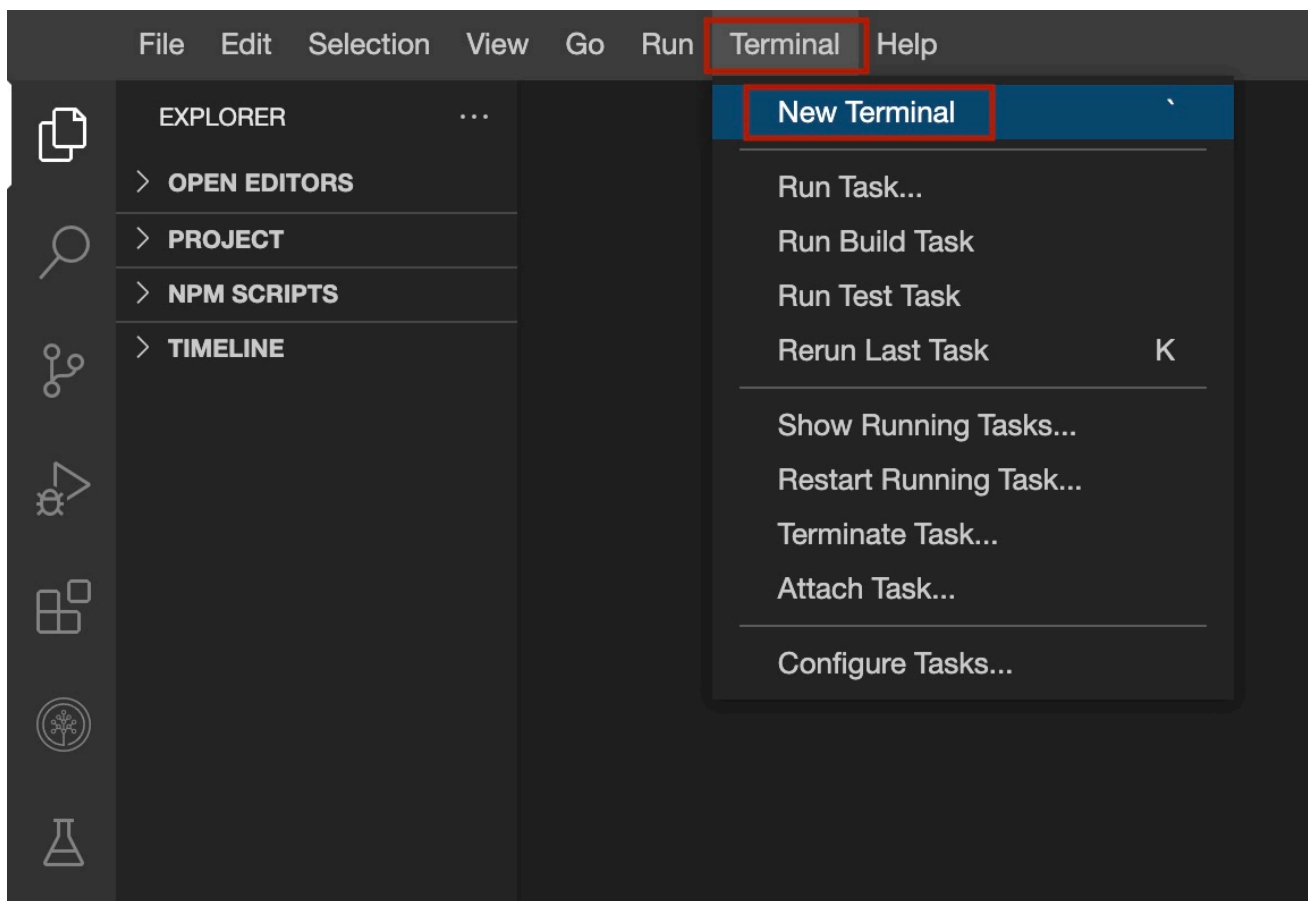
Skills Network Cloud IDE (based on Theia and Docker) provides an environment for hands on labs for course and project related labs. Theia is an open source IDE (Integrated Development Environment), that can be run on desktop or on the cloud. To complete this lab, we will be using the Cloud IDE based on Theia running in a Docker container.

Important Notice about this lab environment

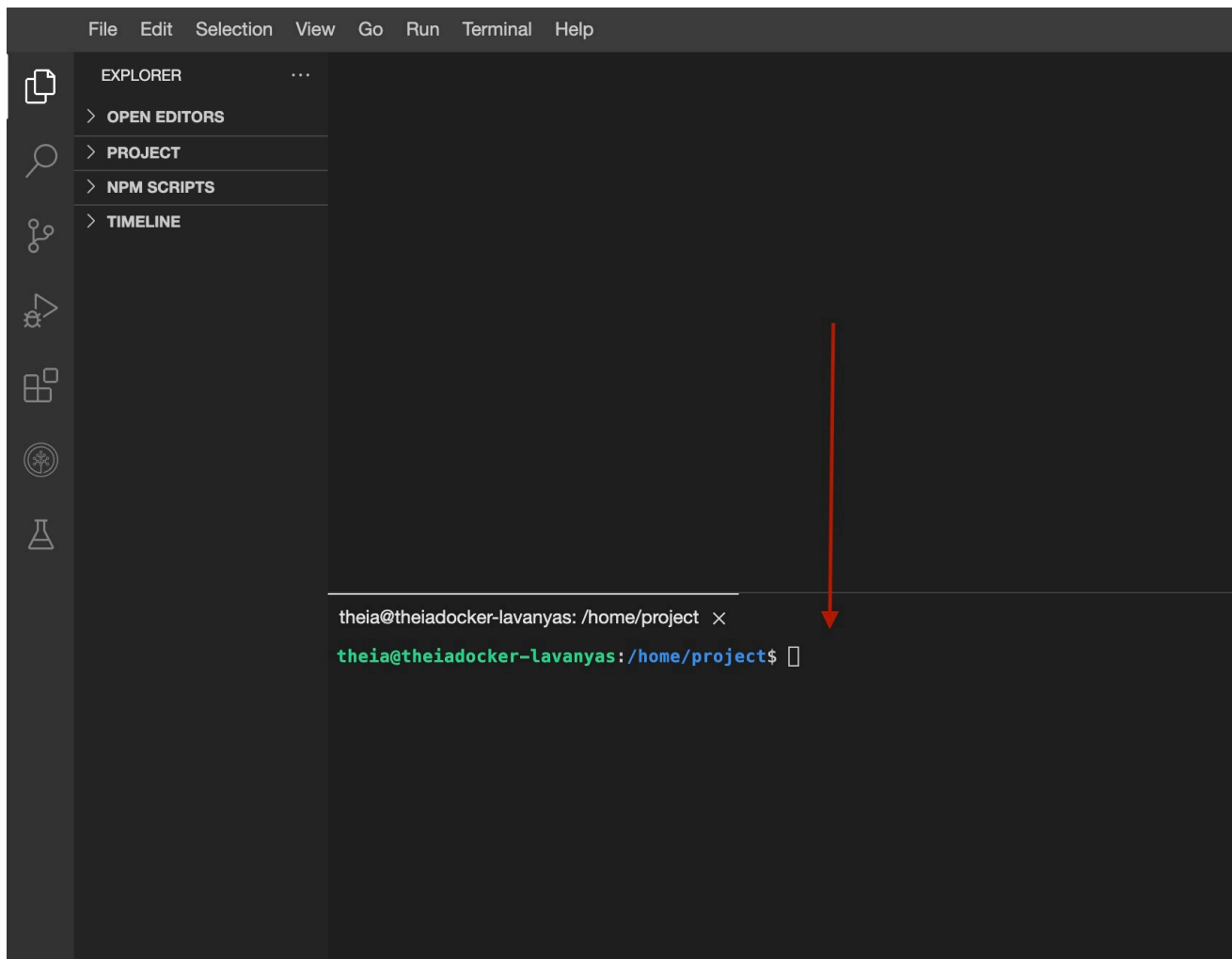
Please be aware that sessions for this lab environment are not persistent. A new environment is created for you every time you connect to this lab. Any data you may have saved in an earlier session will get lost. To avoid losing your data, please plan to complete these labs in a single session.

Exercise 1 - Start Apache Airflow

Open a new terminal by clicking on the menu bar and selecting **Terminal->New Terminal**, as shown in the image below.



This will open a new terminal at the bottom of the screen as in the image below.



Run the commands below on the newly opened terminal. (You can copy the code by clicking on the little copy button on the bottom right of the codeblock below and paste it wherever you wish.)

Run the command below in the terminal to start Apache Airflow.

1. 1

1. start_airflow

Copied!

Please be patient, it may take upto 5 minutes for airflow to get started.

When airflow starts successfully, you should see an output similar to the one below.

Airflow started, waiting for all services to be ready....

Your airflow server is now ready to use and available with username: `airflow` password: `MjQ1NzktbGF2YW55`

You can access your Airflow Webserver at: `https://lavanyas-8080.theiadocker-2-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai`

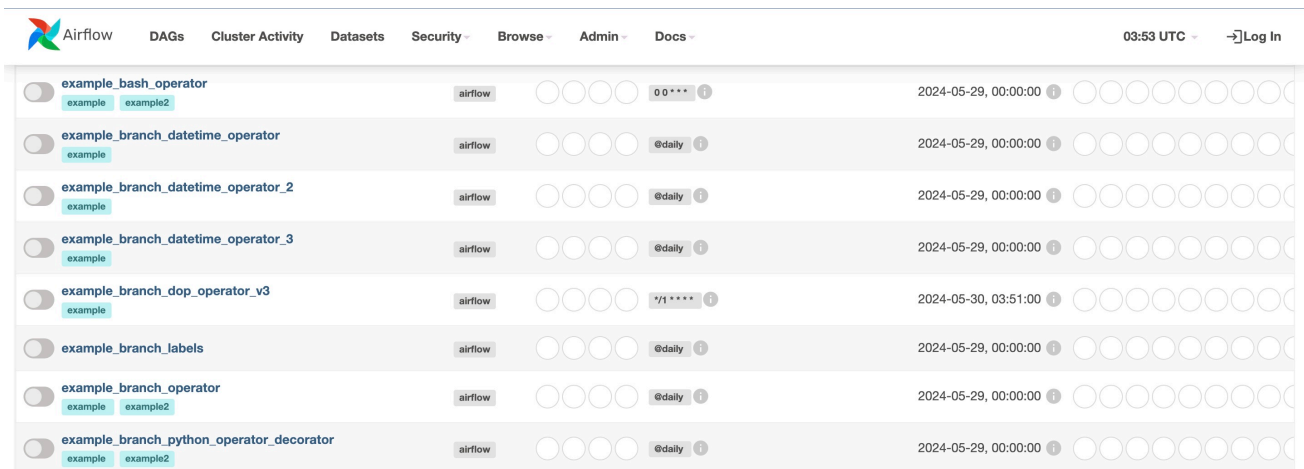
CommandLine:

- List DAGs: `airflow dags list`
- List Tasks: `airflow tasks list example_bash_operator`
- Run an example task: `airflow tasks test example_bash_operator runme_1 2024-05-01`

Exercise 2 - Open the Airflow Web UI

Copy the Web-UI URL and paste it on a new browser tab. You can also click on the URL by holding the control key (Command key in case of a Mac).

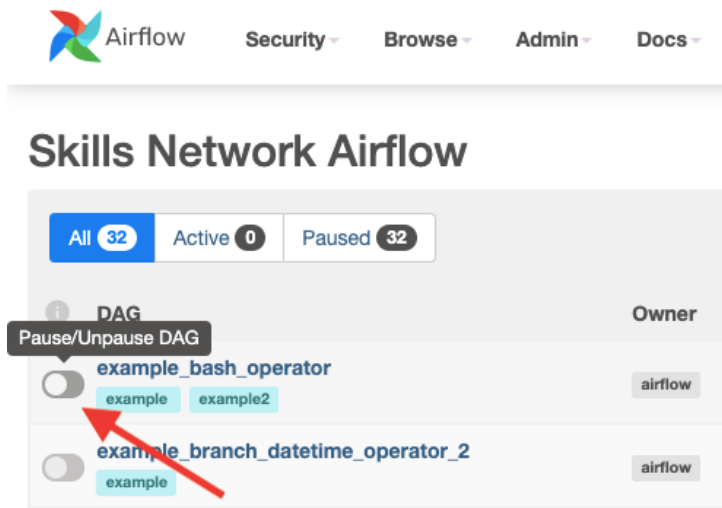
You should land at a page that looks like this:



The screenshot shows the Airflow Web UI interface. At the top, there is a navigation bar with the Airflow logo and several menu items: DAGs, Cluster Activity, Datasets, Security, Browse, Admin, and Docs. On the right side of the navigation bar, it displays the current time as 03:53 UTC and a Log In button. Below the navigation bar, there is a table listing various DAGs. Each row represents a DAG and includes a toggle switch to pause or unpause it, the DAG name, a list of example tasks, the DAG's owner (airflow), a status indicator (four circles), a frequency (e.g., @daily), a last run time, and a series of small circles representing the DAG's history.

DAG Name	Example Tasks	Owner	Status	Frequency	Last Run Time	History
example_bash_operator	example, example2	airflow	Unpaused	@daily	2024-05-29, 00:00:00	
example_branch_datetime_operator	example	airflow	Unpaused	@daily	2024-05-29, 00:00:00	
example_branch_datetime_operator_2	example	airflow	Unpaused	@daily	2024-05-29, 00:00:00	
example_branch_datetime_operator_3	example	airflow	Unpaused	@daily	2024-05-29, 00:00:00	
example_branch_dop_operator_v3	example	airflow	Unpaused	@daily	2024-05-30, 03:51:00	
example_branch_labels		airflow	Unpaused	@daily	2024-05-29, 00:00:00	
example_branch_operator	example, example2	airflow	Unpaused	@daily	2024-05-29, 00:00:00	
example_branch_python_operator_decorator	example, example2	airflow	Unpaused	@daily	2024-05-29, 00:00:00	

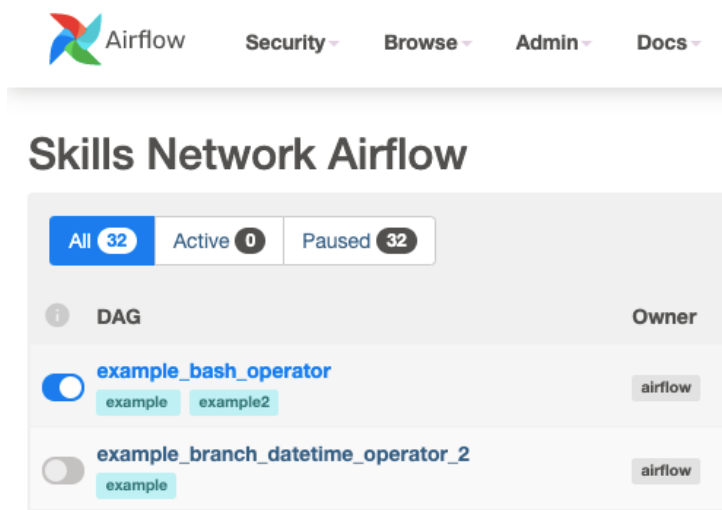
You can Unpause/Pause a DAG using the Unpause/Pause toggle button.



The screenshot shows the Airflow Skills Network interface. At the top, there's a navigation bar with the Airflow logo and links for Security, Browse, Admin, and Docs. Below this, the title "Skills Network Airflow" is displayed. A filter bar shows "All 32", "Active 0", and "Paused 32". A table lists DAGs with columns for "DAG" and "Owner". The first row is "example_bash_operator" with owner "airflow", showing two task instances: "example" and "example2". The second row is "example_branch_datetime_operator_2" with owner "airflow", showing one task instance: "example". A tooltip "Pause/Unpause DAG" is visible over the first row, and a red arrow points to the "example" task instance in the second row.

DAG	Owner
example_bash_operator	airflow
example_branch_datetime_operator_2	airflow

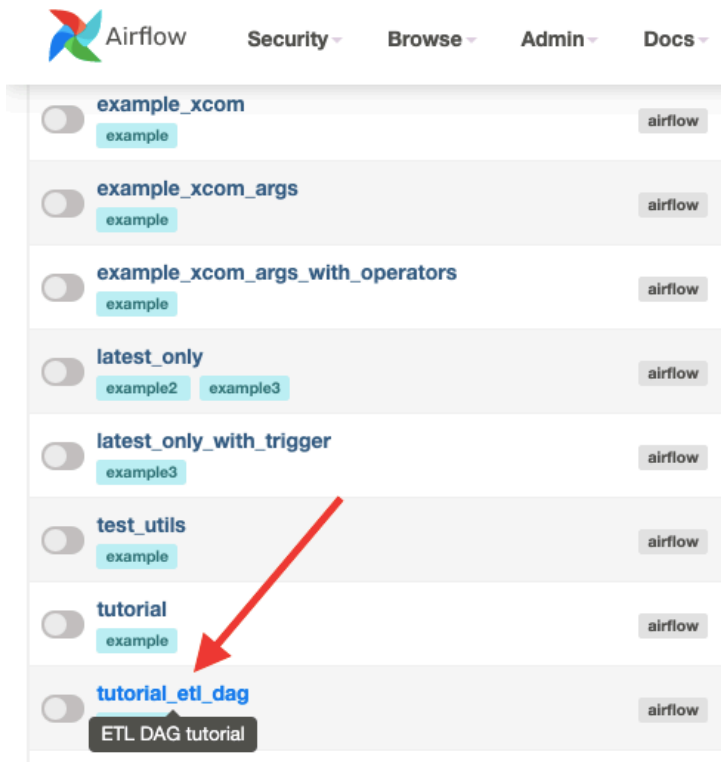
An unpaused DAG looks like this:



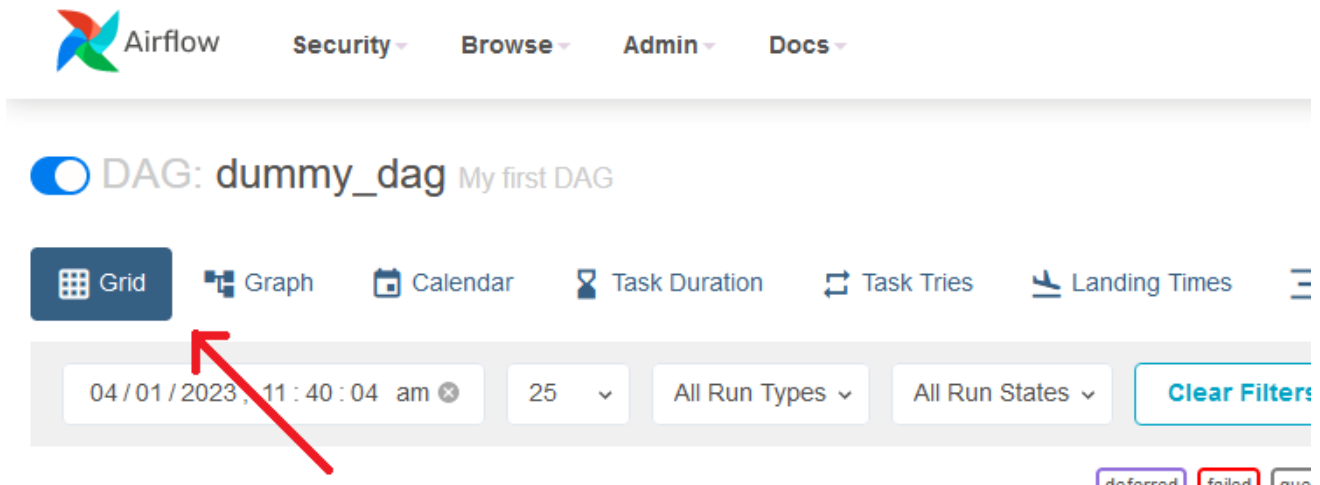
The screenshot shows the Airflow Skills Network interface with the same navigation bar and title. The filter bar shows "All 32", "Active 0", and "Paused 32". The table lists DAGs with columns for "DAG" and "Owner". The first row is "example_bash_operator" with owner "airflow", showing two task instances: "example" and "example2". The second row is "example_branch_datetime_operator_2" with owner "airflow", showing one task instance: "example". The "example_bash_operator" row is highlighted, indicating it is the selected DAG.

DAG	Owner
example_bash_operator	airflow
example_branch_datetime_operator_2	airflow

Click on a DAG to explore more about the DAG.



Click on the Grid button to see the grid view of the DAG.



Click on the Graph button to see the graph view of the DAG.



Exercise 3 - List all DAGs

Apache airflow gives us some handy command line options to work with.

Run the command below in the terminal to list out all the existing DAGs.

- 1
1. `airflow dags list`

Copied!

Exercise 4 - List tasks in a DAG

Run the command below in the terminal to list out all the tasks in the DAG named `example_bash_operator`.

- 1
1. `airflow tasks list example_bash_operator`

Copied!

Here `example_bash_operator` is the name of the DAG.

Try listing out the tasks for the DAG `tutorial`.

Exercise 5 - Pause/Unpause a DAG

Run the command below in the terminal to unpause a DAG.

- 1
1. `airflow dags unpause tutorial`

Copied!

Here `tutorial` is the name of the DAG.

Pause a DAG.

1. 1

1. airflow dags pause tutorial

Copied!

Try to unpause the DAG named `example_branch_operator`.

Practice exercises

1. Problem:

List tasks for the DAG `example_branch_Labels`.

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

2. Problem:

Unpause the DAG `example_branch_Labels`.

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

3. Problem:

Pause the DAG `example_branch_Labels`.

- ▶ [Click here for Hint](#)
- ▶ [Click here for Solution](#)

Authors

Ramesh Sannareddy

Other Contributors

Rav Ahuja

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2023-05-04	0.2	Benny Li	Minor formatting changes
2021-07-05	0.1	Ramesh Sannareddy	Created initial version of the lab

Copyright (c) 2023 IBM Corporation. All rights reserved.