

Hands-on Lab: Getting Started with Apache Airflow



Estimated time needed: **20** minutes

Introduction

In this lab, you will have the opportunity to work with Apache Airflow. You will work with both the Web user interface (UI) and command line interface (CLI) to explore the various ways you can use Airflow for workflow management.

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

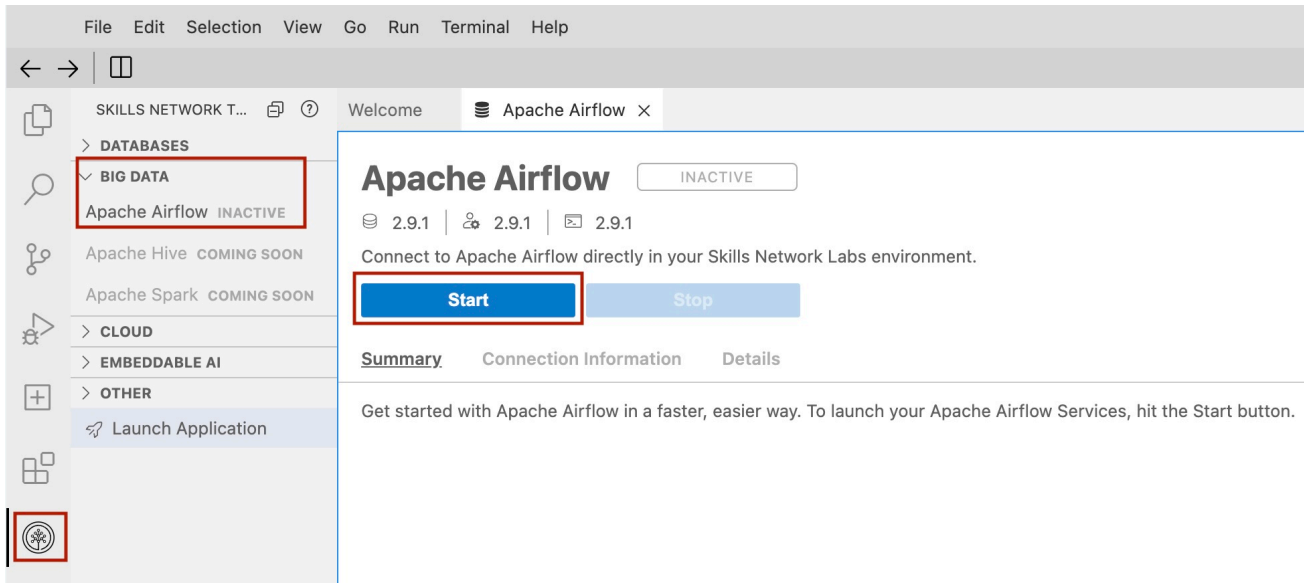
Skills Network Cloud IDE (based on Theia and Docker) provides an environment for hands on labs for course and project related labs. Theia is an open source IDE (Integrated Development Environment), that can be run on desktop or on the cloud. To complete this lab, we will be using the Cloud IDE based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

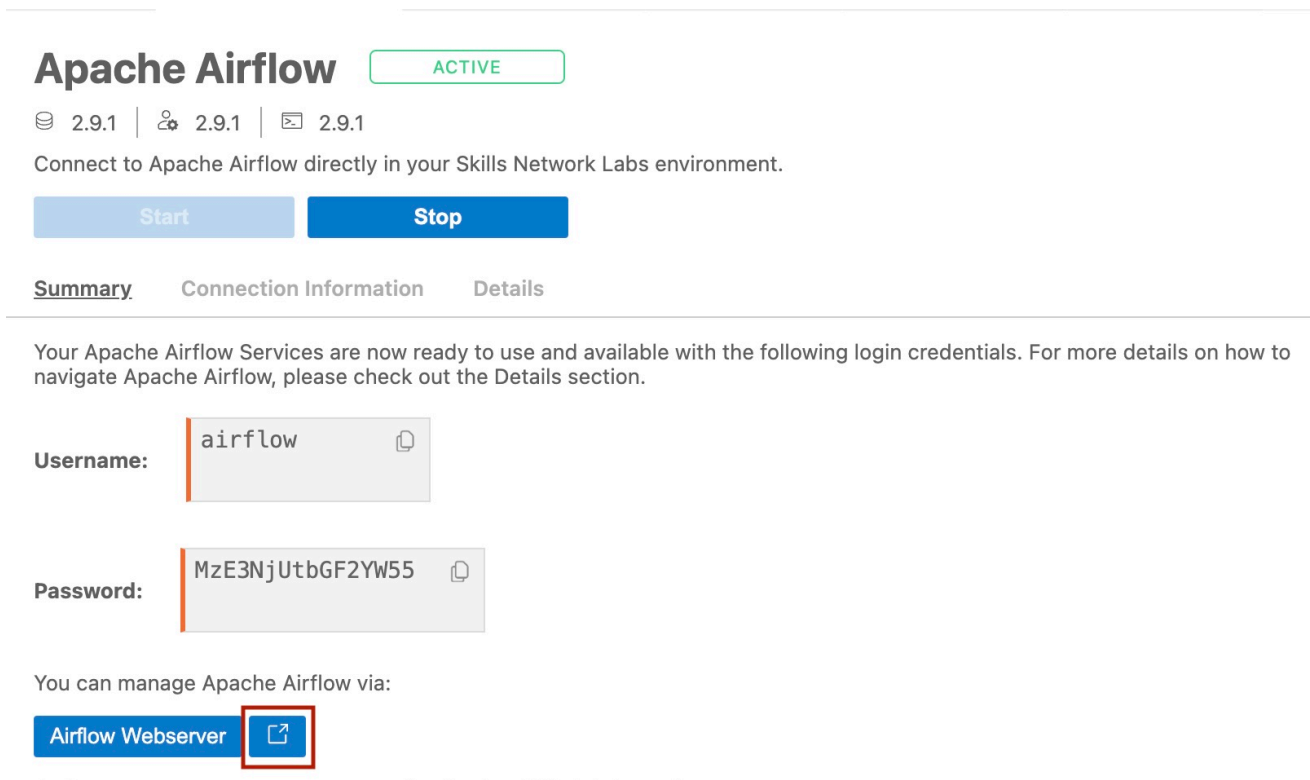
1. Click on **Skills Network Toolbox**.
2. From the **BIG DATA** section, click **Apache Airflow**.
3. Click **Start** to start the Apache Airflow.



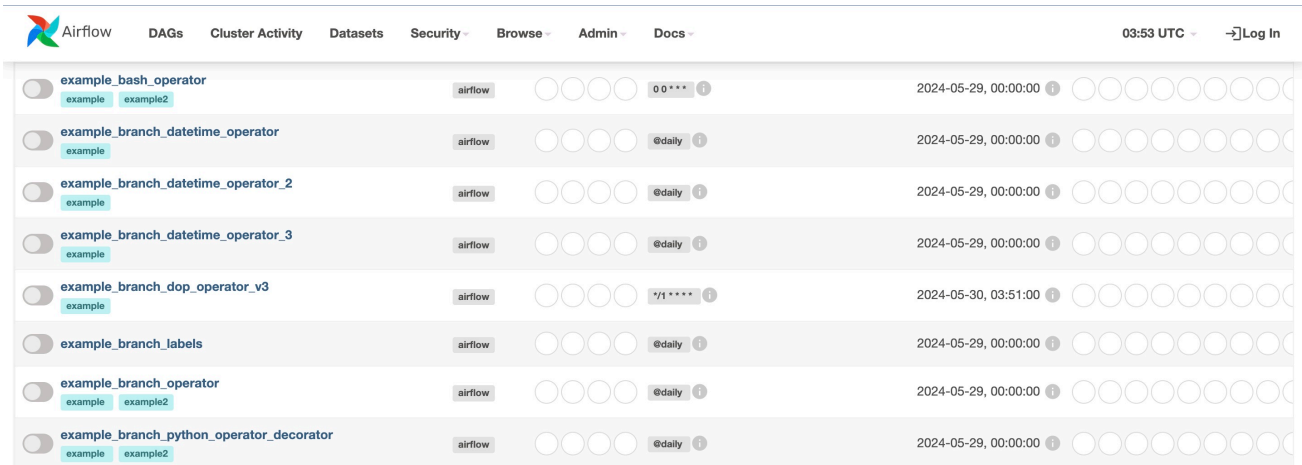
Note: Please be patient, it will take a few minutes for Airflow to start.

Exercise 2: Open the Airflow Web UI

1. When Airflow starts successfully, you should see an output similar to the one below. Once **Apache Airflow** has started, click on the highlighted icon to open **Apache Airflow Web UI** in the new window.



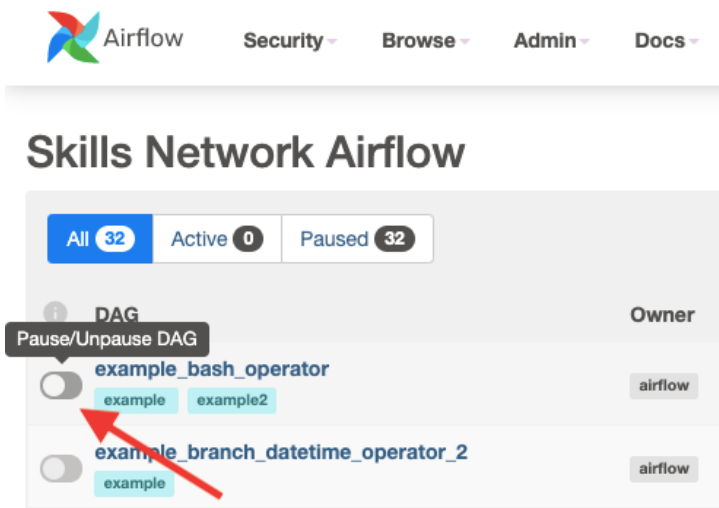
2. You will land on a page that displays various DAGs and associated options.



DAG	Owner	Frequency	Last Run Time
example_bash_operator	airflow	00:00:00	2024-05-29, 00:00:00
example_branch_datetime_operator	airflow	@daily	2024-05-29, 00:00:00
example_branch_datetime_operator_2	airflow	@daily	2024-05-29, 00:00:00
example_branch_datetime_operator_3	airflow	@daily	2024-05-29, 00:00:00
example_branch_dop_operator_v3	airflow	*1:00:00	2024-05-30, 03:51:00
example_branch_labels	airflow	@daily	2024-05-29, 00:00:00
example_branch_operator	airflow	@daily	2024-05-29, 00:00:00
example_branch_python_operator_decorator	airflow	@daily	2024-05-29, 00:00:00

3. You can unpause and pause a DAG using the **Unpause/Pause** toggle button.

The button is grey when the DAG has been paused.



The button is not greyed out when the DAG is running.



Airflow

Security

Browse

Admin

Docs

Skills Network Airflow

All 32

Active 0

Paused 32



DAG

Owner



example_bash_operator

example example2

airflow



example_branch_datetime_operator_2

example

airflow

4. Click on a DAG to explore more.



Airflow

DAGs

Cluster Activity

Datasets

Security

Browse

Admin

Docs

consumes dataset-scheduled



dataset_consumes_unknown_never_scheduled

airflow

Dataset



dataset_produces_1

dataset-scheduled produces

airflow

@daily



dataset_produces_2

dataset-scheduled produces

airflow

None



dummy_dag

Lavanya

20 1 5

0:01:00

5. You will see the Grid view by default.



DAG: dummy_dag My first DAG

Schedule: 0:01:00

Next Run ID: 2024-06-11, 00:02:00 UTC



11/06/2024 00:02:57



All Run Types

All Run States

Clear Filters



Auto-refresh

25



Press shift + / for Shortcuts

deferred

failed

queued

removed

restarting

running

scheduled

shutdown

skipped

success

up_for_reschedule

up_for_retry

upstream_failed

no_status

Duration

Jun 10, 03:01

Jun 10, 03:11

Jun 10, 23:07

00:05:58

00:02:59

00:00:00

task1

task2

task3



DAG

dummy_dag



Details



Graph



Gantt



Code



Audit Log



Run Duration



Calendar

DAG Runs Summary

Total Runs Displayed

25

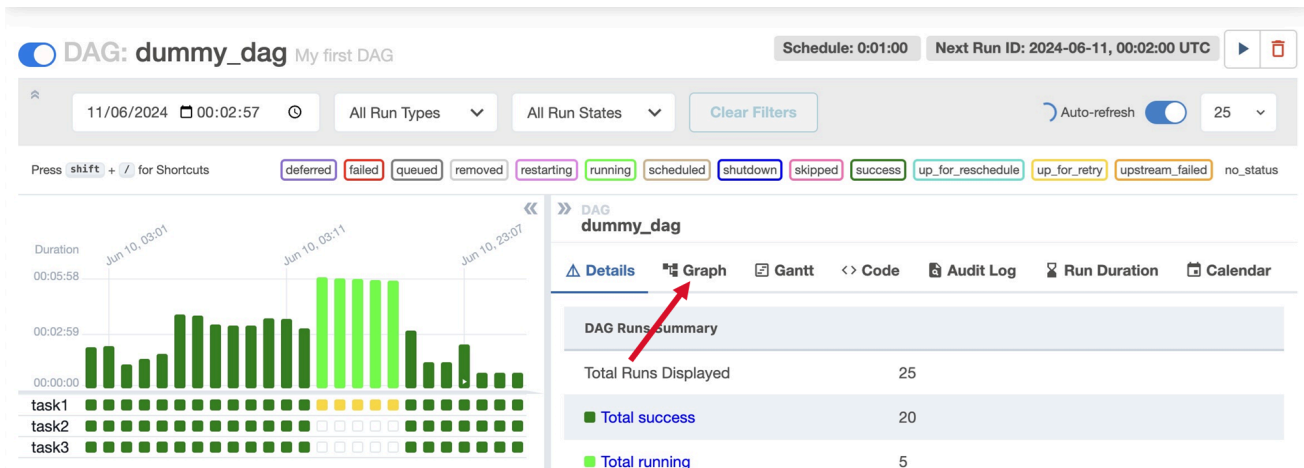
Total success

20

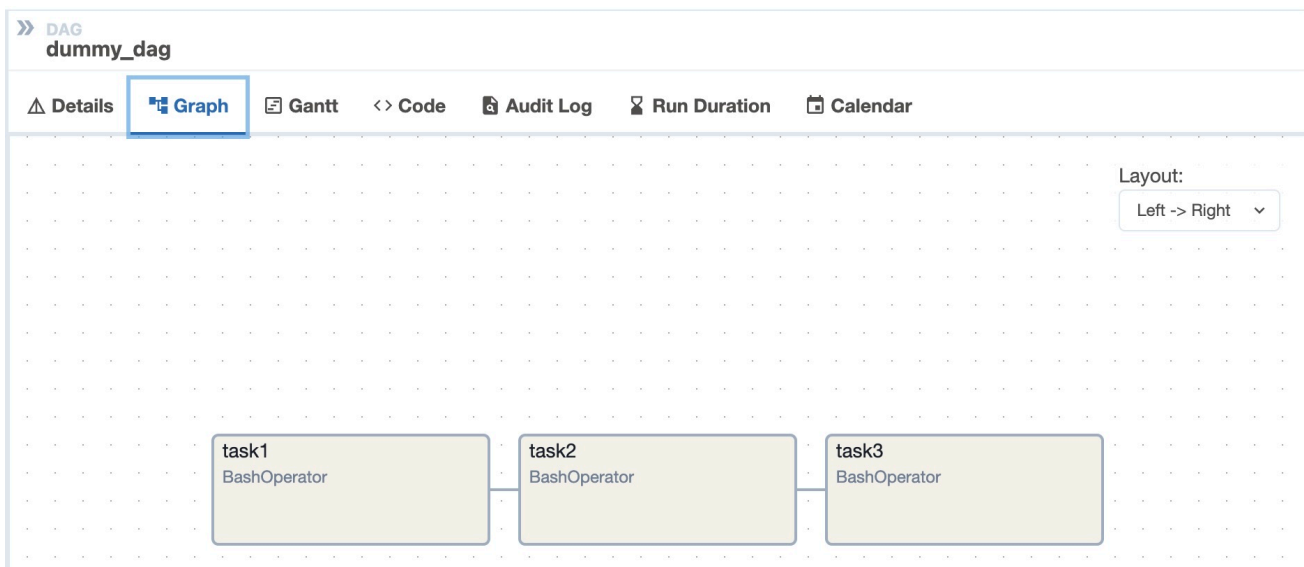
Total running

5

6. Next, click on the **Graph** button.



7. Notice the graph view of the DAG will display.



Exercise 3: Apache Airflow CLI

Apache Airflow provides some command line options.

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

```
1. 1
1. airflow dags list
```

Copied! Executed!

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

```
1. 1
1. airflow tasks list example_bash_operator
```

Copied! Executed!

Practice Exercise

1. Run a command to list all tasks for the DAG named tutorial.
 - ▶ [Click here for the solution.](#)

Exercise 4: Pause or Unpause a DAG

1. Run the command below in the terminal to unpause a DAG named tutorial.

```
1. 1
```

```
1. airflow dags unpause tutorial
```

Copied!

2. Run the command to pause the DAG.

```
1. 1
```

```
1. airflow dags pause tutorial
```

Copied!

Practice Exercise

1. Run a command to unpause the DAG named example_branch_operator.
 - ▶ [Click here for the solution.](#)

Practice exercises

1. List tasks for the DAG example_branch_labels.

- ▶ [Click here for hint.](#)
- ▶ [Click here for the solution.](#)

2. Unpause the DAG example_branch_labels.

- ▶ [Click here for hint.](#)
- ▶ [Click here for the solution.](#)

3. Pause the DAG example_branch_labels.

- ▶ [Click here for hint.](#)
- ▶ [Click here for the solution.](#)

Authors

Ramesh Sannareddy
[Lavanya T S](#)

Other Contributors

Rav Ahuja

© IBM Corporation. All rights reserved.

