

Hands-on Lab: Getting Started with Apache Airflow



Skills
Network

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

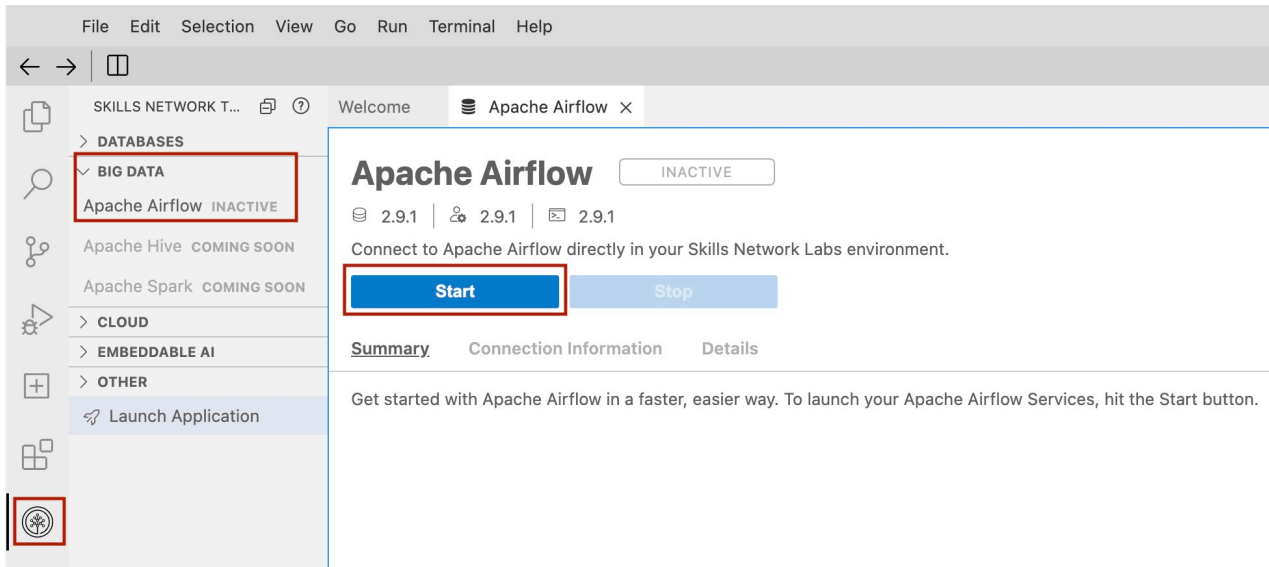
Skills Network Cloud IDE (based on Theia and Docker) provides an environment for hands on labs for course and project related labs. Theia is an open source IDE (Integrated Development Environment), that can be run on desktop or on the cloud. To complete this lab, we will be using the Cloud IDE based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

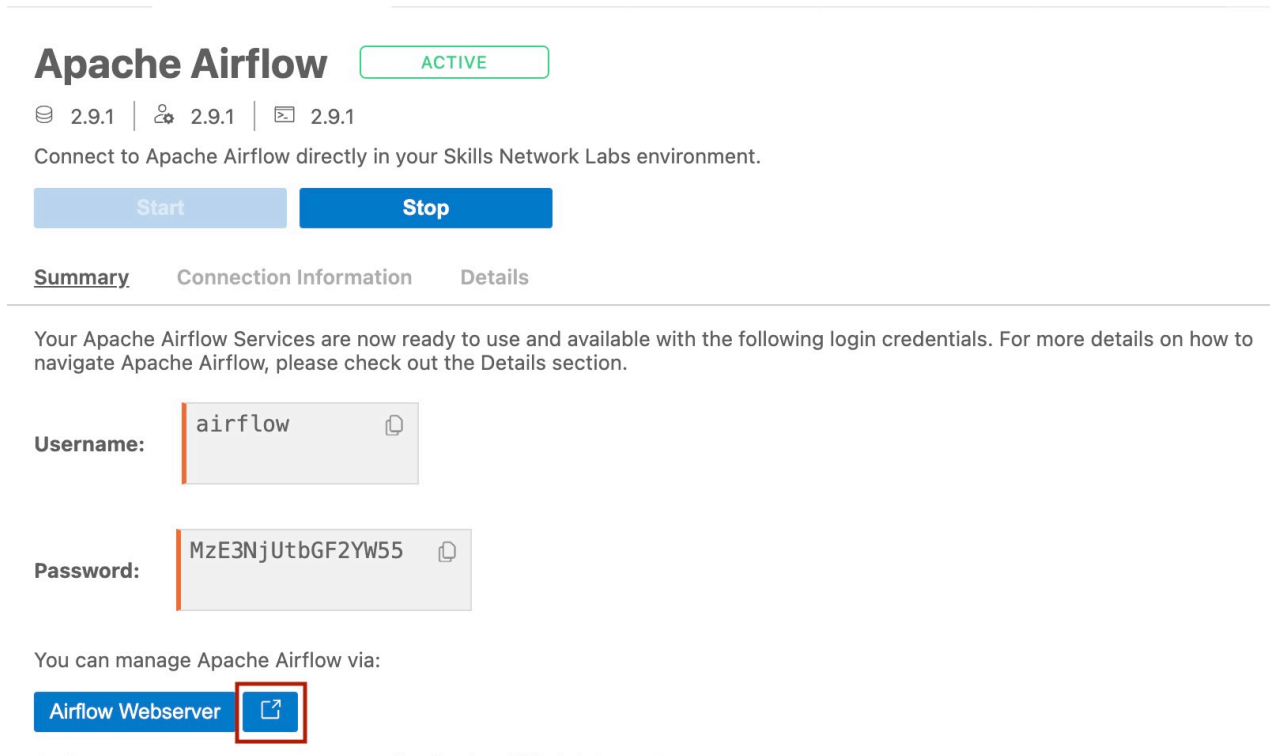
1. Click on **Skills Network Toolbox**.
2. From the **BIG DATA** section, click **Apache Airflow**.
3. Click **Start** to start the Apache Airflow.



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.

Airflow		DAGs	Cluster Activity	Datasets	Security	Browse	Admin	Docs	03:53 UTC		Log In
<input type="checkbox"/>	example_bash_operator	example example2	airflow					0 0 * * *	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_datetime_operator	example	airflow					@daily	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_datetime_operator_2	example	airflow					@daily	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_datetime_operator_3	example	airflow					@daily	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_dop_operator_v3	example	airflow					* / 1 * * * *	<input type="checkbox"/>	2024-05-30, 03:51:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_labels		airflow					@daily	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_operator	example example2	airflow					@daily	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>
<input type="checkbox"/>	example_branch_python_operator_decorator	example example2	airflow					@daily	<input type="checkbox"/>	2024-05-29, 00:00:00	<input type="checkbox"/>

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


The button is grey when the DAG has been paused.

Skills Network Airflow

All 32 Active 0 Paused 32

	DAG	Owner
<input type="checkbox"/>	example_bash_operator example example2	airflow
<input type="checkbox"/>	example_branch_datetime_operator_2 example	airflow

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

i

DAG

Owner

☒

example_bash_operator

exampleexample2

airflow

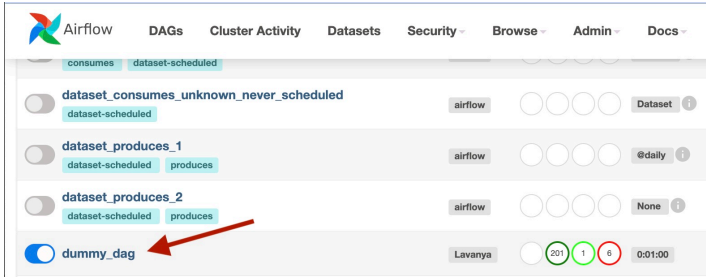
☐

example_branch_datetime_operator_2

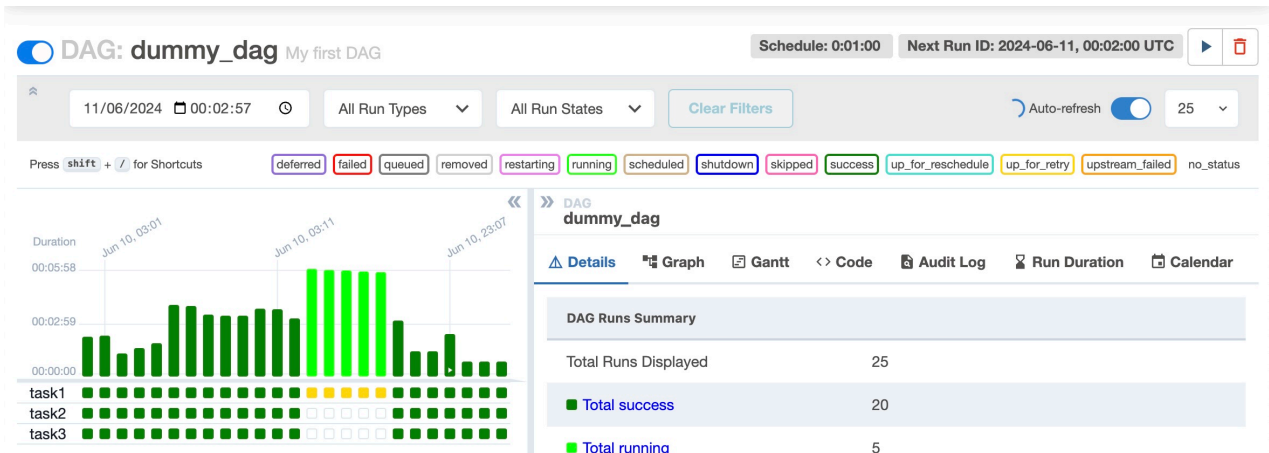
example

airflow

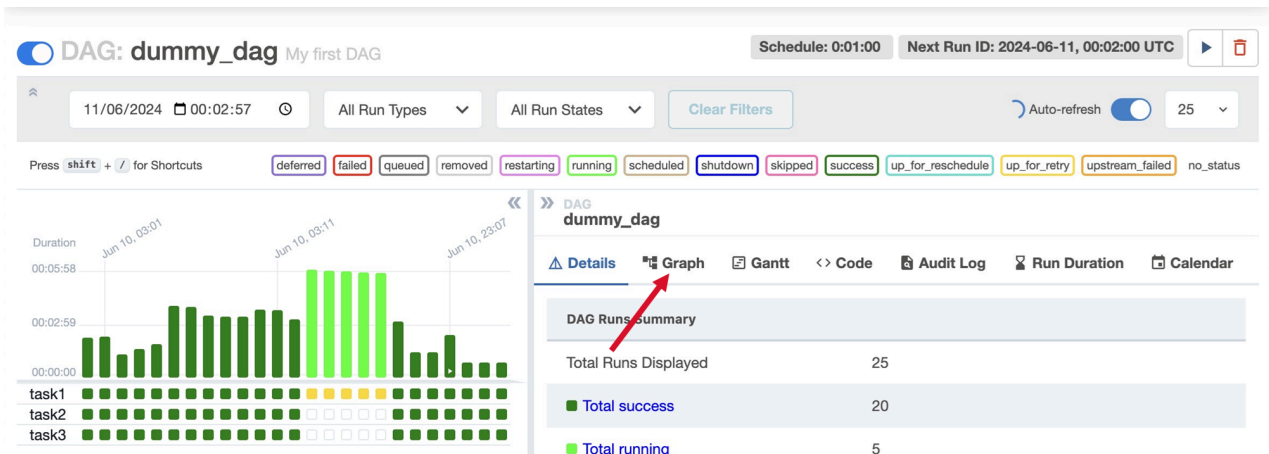
4. Click on a DAG to explore more.



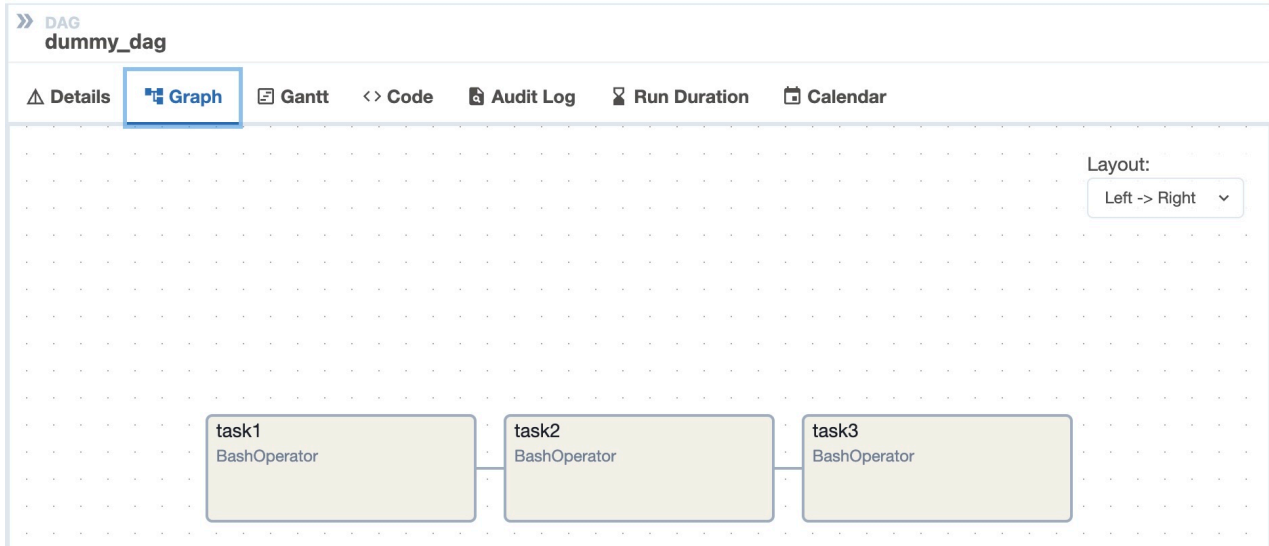
5. You will see the Grid view by default.



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.

```
airflow dags list
```

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

```
airflow tasks list example_bash_operator
```

Practice Exercise

1. Run a command to list all tasks for the DAG named `tutorial`.

▼ [Click here for the solution.](#)

```
airflow tasks list tutorial
```

Exercise 4: Pause or Unpause a DAG

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

```
airflow dags unpause tutorial
```

2. Run the command to pause the DAG.

```
airflow dags pause tutorial
```

Practice Exercise

1. Run a command to unpause the DAG named `example_branch_operator`.

▼ [Click here for the solution.](#)

```
airflow dags unpause example_branch_operator
```

Practice exercises

1. List tasks for the DAG `example_branch_labels`.

► [Click here for hint.](#)

▼ [Click here for the solution.](#)

```
airflow tasks list example_branch_labels
```

2. Unpause the DAG `example_branch_labels`.

▼ [Click here for hint.](#)

Use the unpause option.

► [Click here for the solution.](#)

3. Pause the DAG `example_branch_labels`.

▼ [Click here for hint.](#)

Use the pause option.

▼ [Click here for the solution.](#)

```
airflow dags pause example_branch_labels
```

Authors

Ramesh Sannareddy

[Lavanya T S](#)

Other Contributors

Rav Ahuja

© IBM Corporation. All rights reserved.