

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

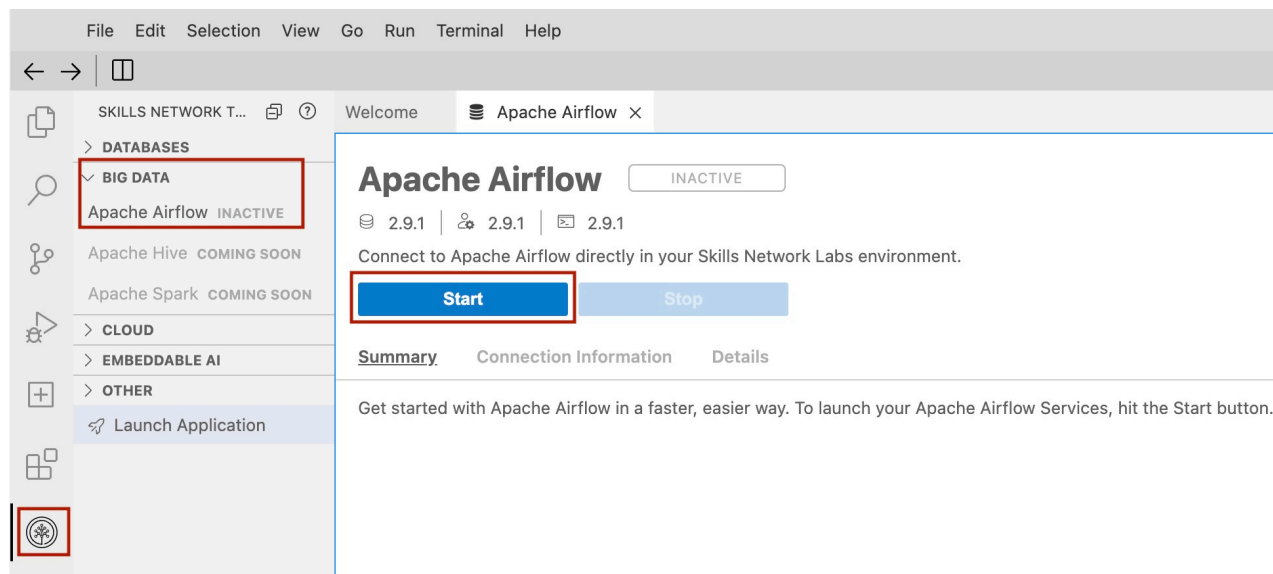
Skills Network Cloud IDE (based on Theia and Docker) provides an environment for hands on labs for course and project related labs. Theia is an open source IDE (Integrated Development Environment), that can be run on desktop or on the cloud. To complete this lab, we will be using the Cloud IDE based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

Apache Airflow

ACTIVE

2.9.1 | 2.9.1 | 2.9.1

Connect to Apache Airflow directly in your Skills Network Labs environment.

Start

Stop

Summary

Connection Information

Details

Your Apache Airflow Services are now ready to use and available with the following login credentials. For more details on how to navigate Apache Airflow, please check out the Details section.

Username:

airflow

Password:

MzE3NjUtbGF2YW55

You can manage Apache Airflow via:

Airflow Webserver



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	airflow	00***	1	2024-05-29, 00:00:00	1			
<input type="checkbox"/>	example_branch_datetime_operator	airflow	@daily	1	2024-05-29, 00:00:00	1			
<input type="checkbox"/>	example_branch_datetime_operator_2	airflow	@daily	1	2024-05-29, 00:00:00	1			
<input type="checkbox"/>	example_branch_datetime_operator_3	airflow	@daily	1	2024-05-29, 00:00:00	1			
<input type="checkbox"/>	example_branch_dop_operator_v3	airflow	*/1***	1	2024-05-30, 03:51:00	1			
<input type="checkbox"/>	example_branch_labels	airflow	@daily	1	2024-05-29, 00:00:00	1			
<input type="checkbox"/>	example_branch_operator	airflow	@daily	1	2024-05-29, 00:00:00	1			
<input type="checkbox"/>	example_branch_python_operator_decorator	airflow	@daily	1	2024-05-29, 00:00:00	1			

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

The button is grey when the DAG has been paused.

Airflow

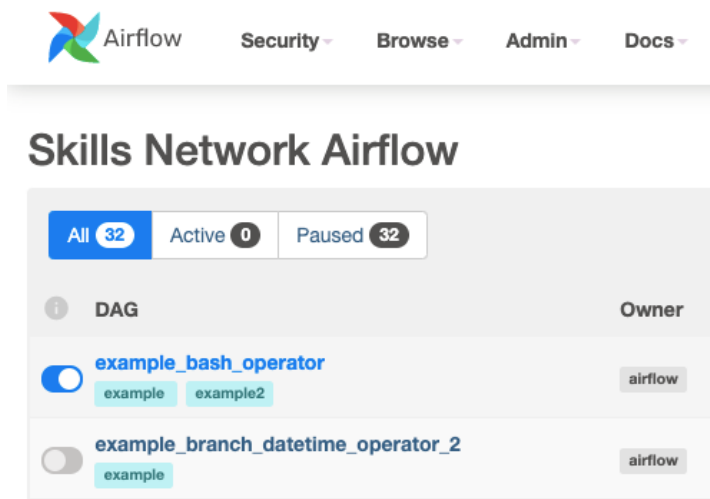
Security Browse Admin Docs

Skills Network Airflow

All 32 Active 0 Paused 32

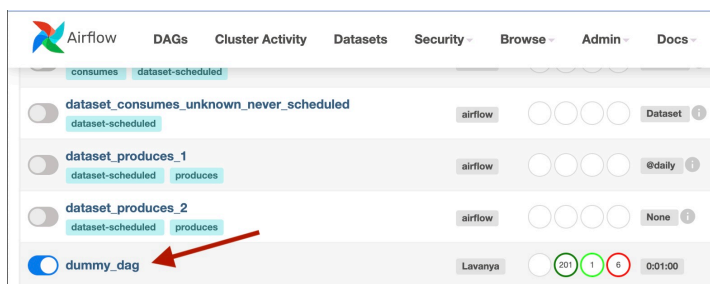
DAG	Owner
<input type="checkbox"/> example_bash_operator	airflow
<input type="checkbox"/> example_branch_datetime_operator_2	airflow

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



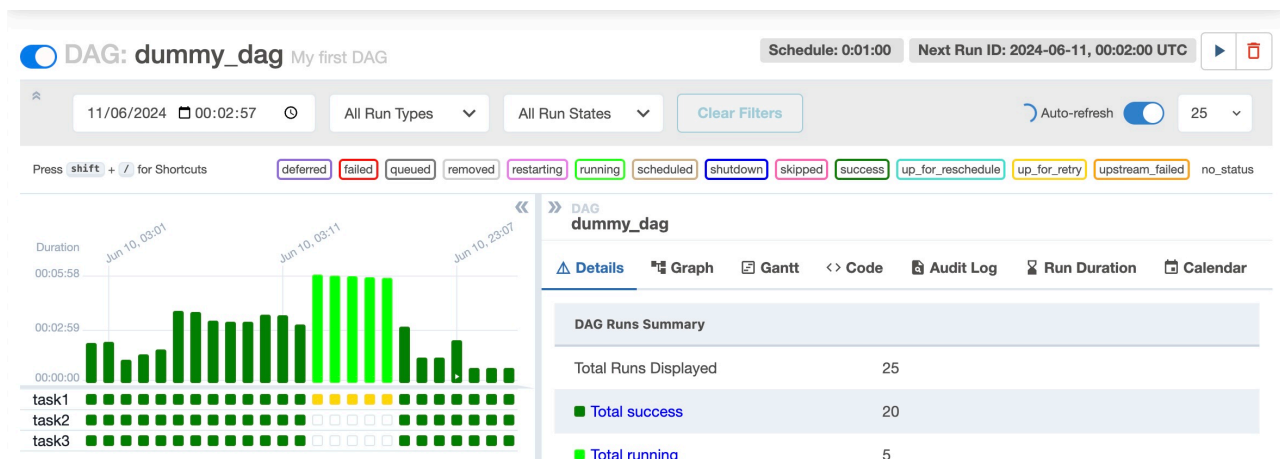
The screenshot shows the Airflow Skills Network interface. At the top, there's a navigation bar with the Airflow logo and links for Security, Browse, Admin, and Docs. Below this, the title "Skills Network Airflow" is displayed. A summary bar shows "All 32", "Active 0", and "Paused 32". The main section is a table with columns for "DAG" and "Owner". Two DAGs are listed: "example_bash_operator" and "example_branch_datetime_operator_2", both owned by "airflow". Each DAG has a toggle switch and a list of tasks (example, example2, example).

4. Click on a DAG to explore more.



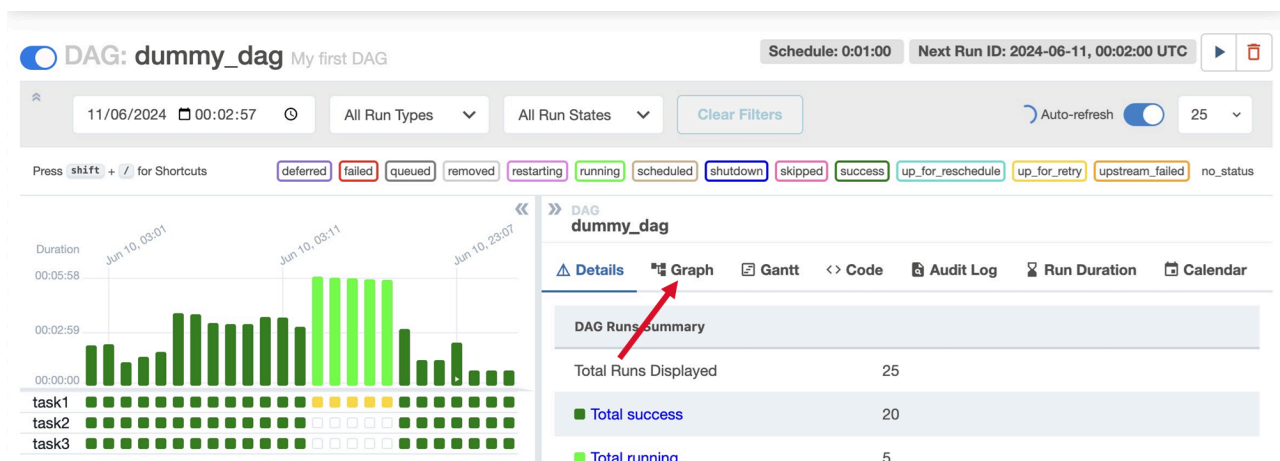
The screenshot shows the Airflow DAG list interface. The top navigation bar includes links for Airflow, DAGs, Cluster Activity, Datasets, Security, Browse, Admin, and Docs. Below this, there's a list of DAGs. The DAG "dummy_dag" is highlighted with a red arrow. It is owned by "Lavanya" and has a status of "running" (indicated by a green circle). The "Next Run ID" is "2024-06-11, 00:02:00 UTC".

5. You will see the Grid view by default.



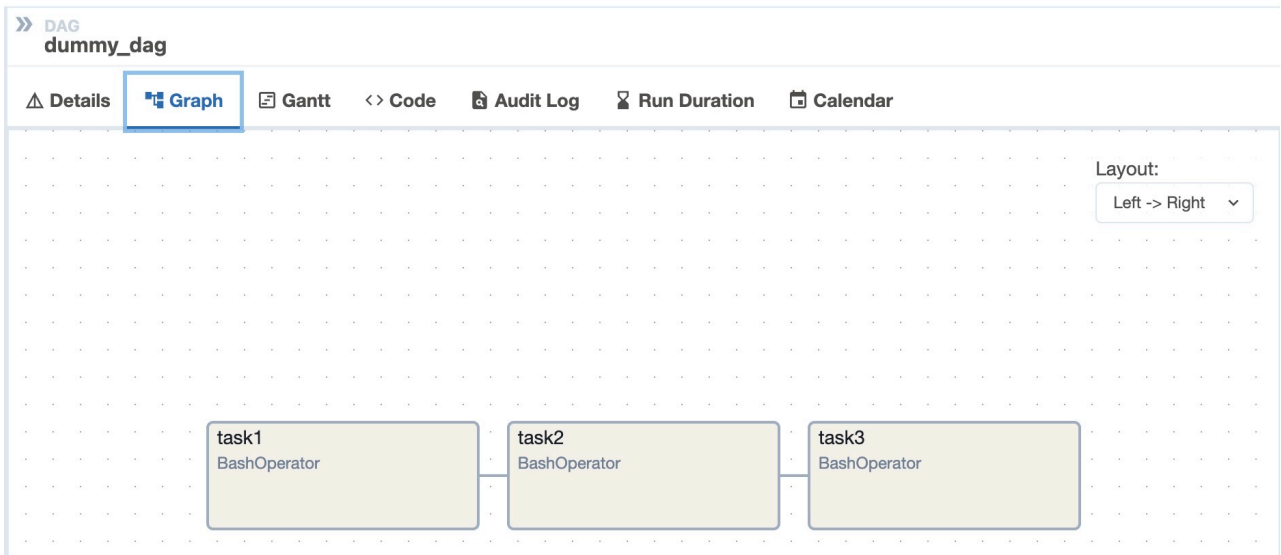
The screenshot shows the Airflow DAG details interface for "dummy_dag". The top bar displays the DAG name, "My first DAG", and the "Schedule: 0:01:00". The "Next Run ID" is "2024-06-11, 00:02:00 UTC". Below this, there's a filter bar with "All Run Types" and "All Run States" dropdowns, a "Clear Filters" button, and an "Auto-refresh" toggle. The main content area shows a "Grid" view of the DAG runs. The "Grid" view displays a bar chart of task durations and a table of task statuses. The "task1" row shows a series of green bars, indicating successful runs. The "task2" and "task3" rows show a mix of green and yellow bars, indicating some failed runs. The "task1" row has a "Total success" of 20 and a "Total running" of 5. The "task2" and "task3" rows have a "Total success" of 20 and a "Total running" of 5.

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



The screenshot shows the Airflow DAG details interface for "dummy_dag" in "Graph" view. The top bar displays the DAG name, "My first DAG", and the "Schedule: 0:01:00". The "Next Run ID" is "2024-06-11, 00:02:00 UTC". Below this, there's a filter bar with "All Run Types" and "All Run States" dropdowns, a "Clear Filters" button, and an "Auto-refresh" toggle. The main content area shows a "Graph" view of the DAG runs. The "Graph" view displays a bar chart of task durations and a table of task statuses. The "task1" row shows a series of green bars, indicating successful runs. The "task2" and "task3" rows show a mix of green and yellow bars, indicating some failed runs. The "task1" row has a "Total success" of 20 and a "Total running" of 5. The "task2" and "task3" rows have a "Total success" of 20 and a "Total running" of 5. A red arrow points to the "Graph" button in the top right corner of the DAG details section.

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.](#)
Use list option.
▼ [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.](#)

```
airflow dags unpause example_branch_labels
```
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.