

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

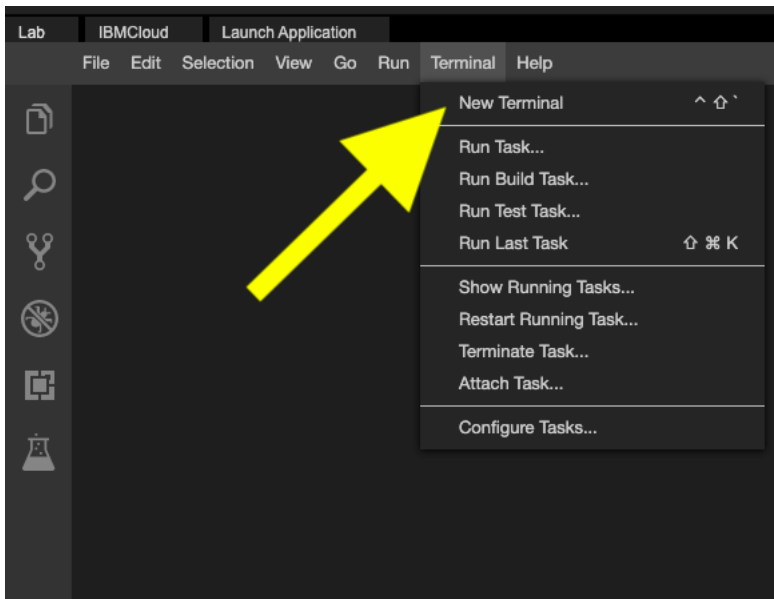
Skills Network Cloud IDE (based on Theia and Docker) provides an environment for hands on labs for course and project related labs. Theia is an open source IDE (Integrated Development Environment), that can be run on desktop or on the cloud. To complete this lab, we will be using the Cloud IDE based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

Important Notice about this lab environment

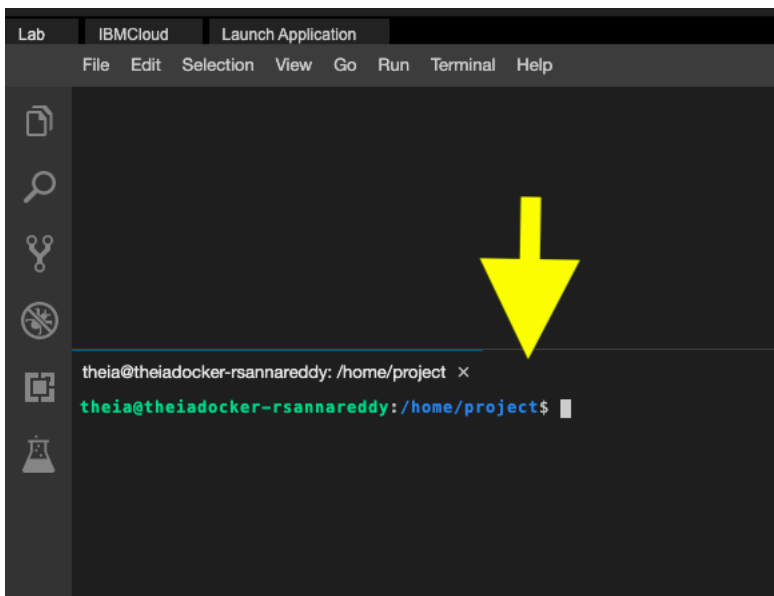
Please be aware that sessions for this lab environment are not persistent. A new environment is created for you every time you connect to this lab. Any data you may have saved in an earlier session will get lost. To avoid losing your data, please plan to complete these labs in a single session.

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.

```
theia@theiadocker-rsannareddy:/home/project$ start_airflow
Starting your airflow services....
This process can take a few minutes.

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

Your airflow server is now ready to use and available with username: airflow password: MTM4ODUtcnNhbm5h


You can access your Airflow Webserver at: https://rsannareddy-8080.theiadocker-5-labs-prod-th-eiak8s-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 2015-06-01
theia@theiadocker-rsannareddy:/home/project$
```

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:

 Airflow

Security

Browse

Admin

Docs


Skills Network Airflow

All 32Active 0Paused 32

Filter DAGs by tag

DAG	Owner	Runs	Schedule	Last Run	Recent Tasks
<div><div><input type="checkbox"/></div><div>example_bash_operator</div><div>exampleexample2</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>0 0 ***</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_branch_datetime_operator_2</div><div>example</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>@daily</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_branch_dop_operator_v3</div><div>example</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>*/1 * * * *</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_branch_labels</div><div></div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>@daily</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_branch_operator</div><div>exampleexample2</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>@daily</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_complex</div><div>exampleexample2example3</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>None</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_dag_decorator</div><div>example</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>None</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					
<div><div><input type="checkbox"/></div><div>example_external_task_marker_child</div><div>example2</div></div> <div>airflow</div> <div><div></div><div></div><div></div></div> <div>None</div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>					

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

 Airflow

Security

Browse

Admin

Docs

Skills Network Airflow


All 32

Active 0

Paused 32

DAG	Owner
<div><div><div>Pause/Unpause DAG</div><div><div><div></div></div></div><div><div>example_bash_operator</div><div>exampleexample2</div></div></div></div>	airflow
<div><div><div></div></div><div><div>example_branch_datetime_operator_2</div><div>example</div></div></div>	airflow

An unpaused DAG looks like this:

 Airflow

Security

Browse

Admin

Docs

Skills Network Airflow


All 32

Active 0

Paused 32

DAG	Owner
<div><div><div></div></div><div><div>example_bash_operator</div><div>exampleexample2</div></div></div>	airflow
<div><div><div></div></div><div><div>example_branch_datetime_operator_2</div><div>example</div></div></div>	airflow

Click on a DAG to explore more about the DAG.

 Airflow

Security

Browse

Admin

Docs

example_xcom

example

example_xcom_args

example

example_xcom_args_with_operators

example

latest_only

example2example3

latest_only_with_trigger

example3

test_utils

example

tutorial

example

tutorial_etl_dag

ETL DAG tutorial

airflowairflowairflowairflowairflowairflowairflow

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

The screenshot shows the Apache Airflow web interface. At the top, there's a navigation bar with links for Security, Browse, Admin, and Docs. Below this, the DAG 'dummy_dag' is displayed with the subtitle 'My first DAG'. The interface has tabs for Grid, Graph, Calendar, Task Duration, Task Tries, and Landing Times. A red arrow points to the 'Graph' button. Below the tabs, there's a filter bar with a date/time selector (04/01/2023, 11:40:04 am), a dropdown for '25', and dropdowns for 'All Run Types' and 'All Run States'. A 'Clear Filters' button is also present.

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

The screenshot shows the Apache Airflow web interface in the Graph view. At the top, there's a status bar with various task states: queued, running (highlighted in green), success, failed, up_for_retry, up_for_reschedule, upstream_failed, skipped, scheduled, and no_status. Below this, the DAG is visualized as a linear sequence of three tasks: task1, task2, and task3, connected by arrows. An 'Auto-refresh' toggle is visible in the top right corner.

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
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_branch_operator`.

1. 1
1. `airflow tasks list example_branch_operator`

Copied!

Here `example_branch_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
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.