

Installation of Airflow on the local machine

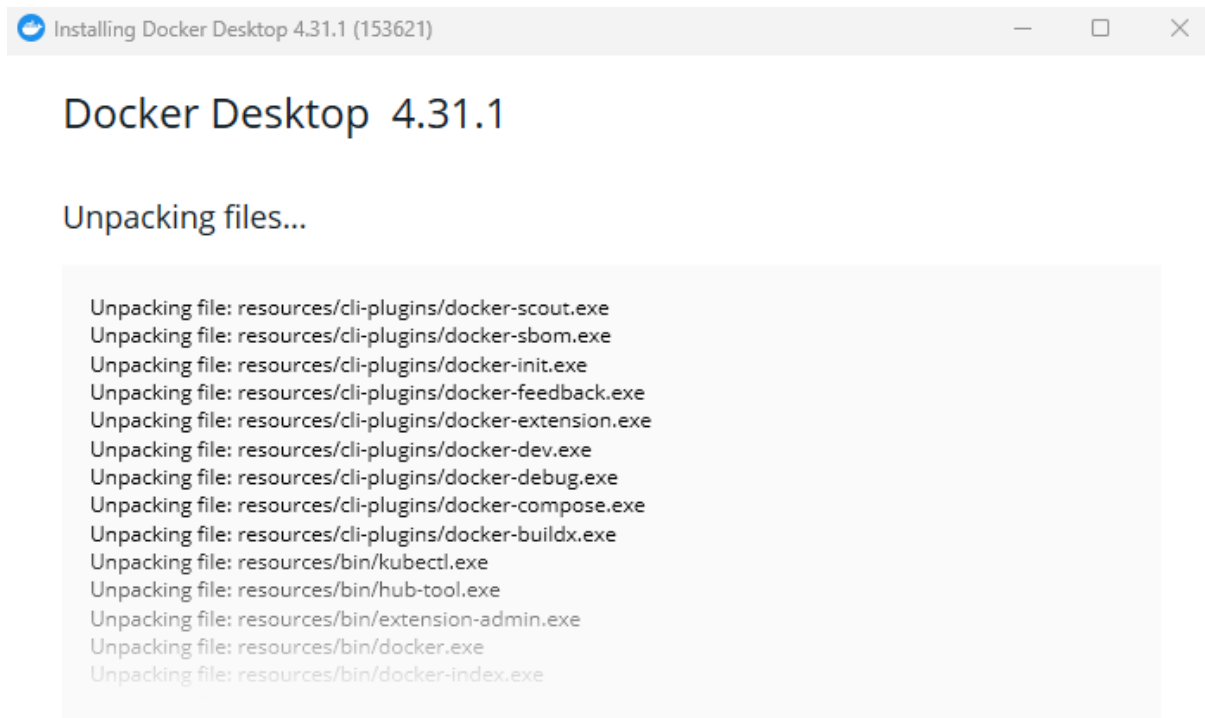
Authored by: Tushar Sharma (Tushar.techlab@gmail.com)

The document outlines the process to install Airflow on the local machine.

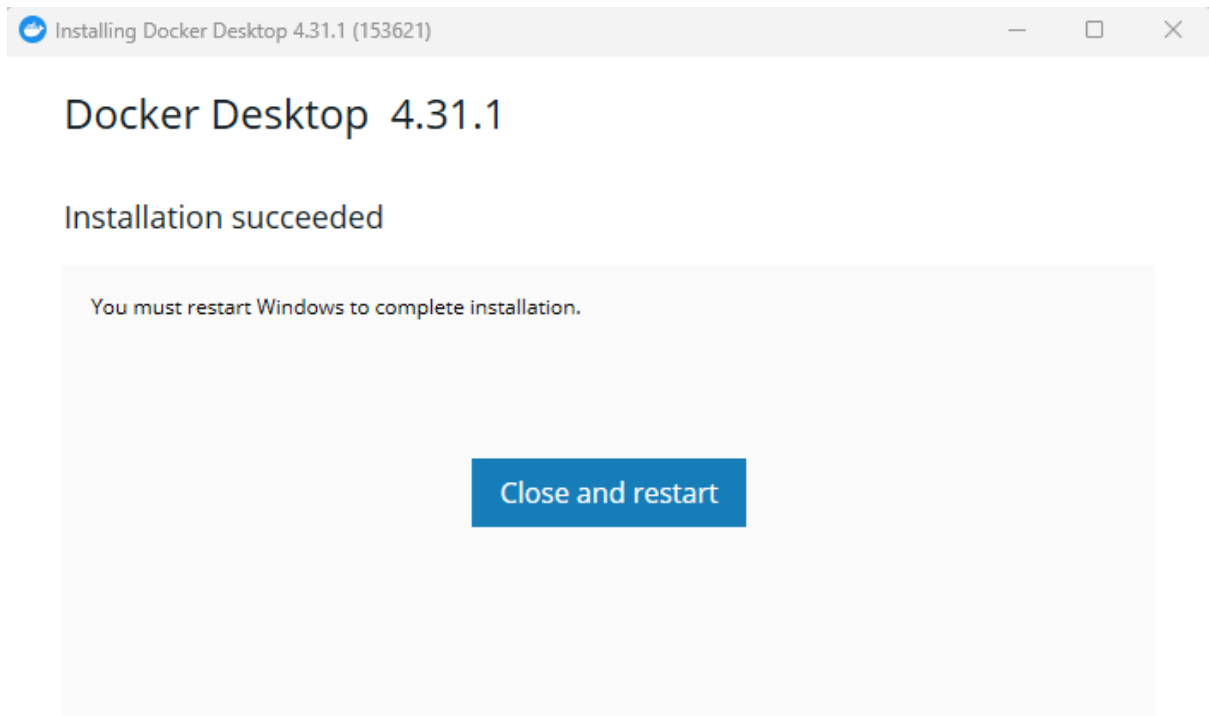
1. To install airflow on windows local machine we will need Docker first. Go the following URL and download docker installer.

<https://docs.docker.com/desktop/install/windows-install/>

2. During the installation when prompted “Use WSL instead of Hyper-V”, please uncheck that checkbox and proceed for installation. The reason for this is in my previous installation experience I had faced a lot of issue in running airflow init command via docker compose. When I researched the reason was this. To get this setting reverse I used some powershell scripts that time but since this is a fresh installation and I am aware of that issue, it is better to uncheck this setting before moving forward.
3. You should see the following window during the installation process.



4. Something like this should be seen at the end of successful installation.



5. Verify the installation by running the following commands to check docker and docker compose are with proper versions on your system

```
C:\WINDOWS\system32\cmd. X + v
Microsoft Windows [Version 10.0.22631.3810]
(c) Microsoft Corporation. All rights reserved.

C:\Users\TUSHAR-PC>docker --version
Docker version 26.1.4, build 5650f9b

C:\Users\TUSHAR-PC>docker-compose --version
Docker Compose version v2.27.1-desktop.1
```

6. Create an airflow directory on your local machine for its installation in the way mentioned in the screenshot below:

```
C:\Users\TUSHAR-PC>mkdir airflow-docker
C:\Users\TUSHAR-PC>cd airflow-docker
```

7. Go the link mentioned below to download docker-compose.yaml file which will be used for setting up airflow in your environment. This is a preconfigured file which will be used as a base.

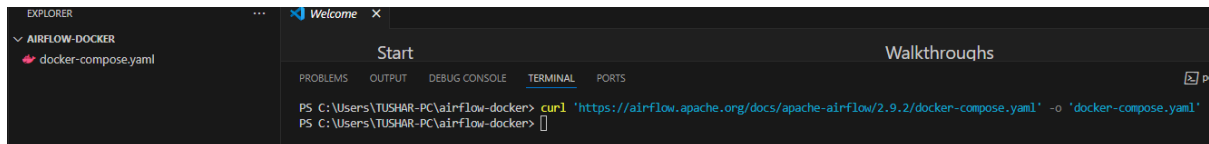
<https://airflow.apache.org/docs/apache-airflow/stable/howto/docker-compose/index.html#fetching-docker-compose-yaml>

8. In the screen you should see this command. Copy the link in the curl command which should be the following:

curl 'https://airflow.apache.org/docs/apache-airflow/2.9.2/docker-compose.yaml' -o 'docker-compose.yaml'

9. Run this command in the VS Code terminal to get the yaml file

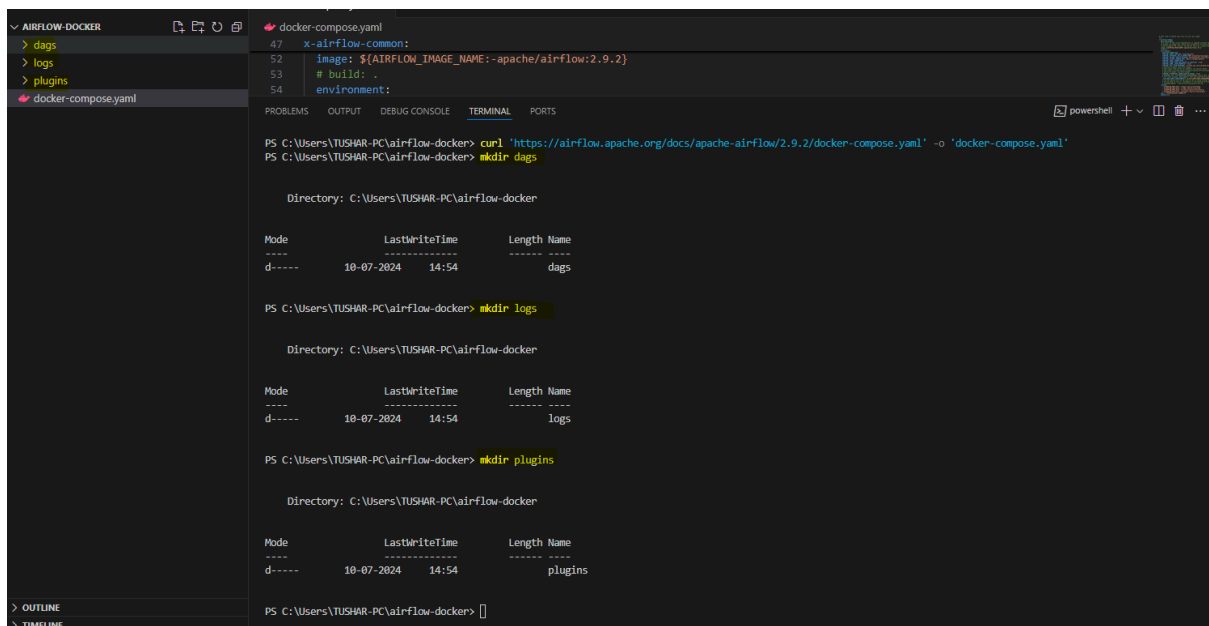
'https://airflow.apache.org/docs/apache-airflow/2.9.2/docker-compose.yaml'



10. You should get the docker-compose.yaml file which will have the something like this in the volumes section

```
74 | # AIRFLOW_CONFIG: /opt/airflow/config/airflow.cfg
75 | volumes:
76 |   - ${AIRFLOW_PROJ_DIR:-.}/dags:/opt/airflow/dags
77 |   - ${AIRFLOW_PROJ_DIR:-.}/logs:/opt/airflow/logs
78 |   - ${AIRFLOW_PROJ_DIR:-.}/config:/opt/airflow/config
79 |   - ${AIRFLOW_PROJ_DIR:-.}/plugins:/opt/airflow/plugins
80 | user: "${AIRFLOW_UID:-50000}:0"
```

11. Create the volumes mentioned in the yaml file in the docker-airflow directory as shown in the screenshot below:



12. Run the command “docker compose up airflow-init” command to initialize the airflow setup

```
[+] Running 23/23
  ✓ airflow-init Pulled 82.6s
  ✓ 09f376ebb190 Already exists 0.0s
  ✓ 578d3dae00a Already exists 0.0s
  ✓ 2614b8b77544 Already exists 0.0s
  ✓ 0d69f3ec269e Already exists 0.0s
  ✓ c9e6b31e1b2e Already exists 0.0s
  ✓ 94f546401a69 Already exists 0.0s
  ✓ 7f018bad6259 Already exists 0.0s
  ✓ cbee7f833443 Already exists 0.0s
  ✓ e670339f2103 Already exists 0.0s
  ✓ 84f0eab6a160 Already exists 0.0s
  ✓ 8020ee21fc8f Pull complete 78.4s
  ✓ a2efd3481353 Pull complete 78.5s
  ✓ 46a5cf400d64 Pull complete 78.6s
  ✓ c822cc61d3e7 Pull complete 78.7s
  ✓ 163c2ff48915 Pull complete 78.9s
  ✓ 8df33612dd90 Pull complete 79.0s
  ✓ 863682d23eb8 Pull complete 79.1s
  ✓ f624e456d5ad Pull complete 79.5s
  ✓ 1b652b7b83dc Pull complete 79.6s
  ✓ 4f40bee0102d Pull complete 79.6s
  ✓ 0b271170d28c Pull complete 79.8s
  ✓ 4f4fb700ef54 Pull complete 79.9s

[+] Running 5/5
  ✓ Network airflow-docker_default Created 0.1s
  ✓ Volume "airflow-docker_postgres-db-volume" Created 0.0s
  ✓ Container airflow-docker-postgres-1 Created 3.2s
  ✓ Container airflow-docker-redis-1 Created 3.2s
  ✓ Container airflow-docker-airflow-init-1 Created 5.8s

Attaching to airflow-init-1
```

13. You should see the following at the end of the command completion

```
airflow-init-1 | User "airflow" created with role "Admin"
airflow-init-1 | 2.9.2
airflow-init-1 exited with code 0
PS C:\Users\TUSHAR-PC\airflow-docker>
```

14. Run `docker compose up` to start services, where you should see something like this in the terminal.

```
[+] Running 7/7
✓ Container airflow-docker-postgres-1      Running
✓ Container airflow-docker-redis-1         Running
✓ Container airflow-docker-airflow-init-1   Created
✓ Container airflow-docker-airflow-scheduler-1 Created
✓ Container airflow-docker-airflow-worker-1 Created
✓ Container airflow-docker-airflow-triggerer-1 Created
✓ Container airflow-docker-airflow-webserver-1 Created
Attaching to airflow-init-1, airflow-scheduler-1, airflow-triggerer-1, airflow-webserver-1, airflow-worker-1, postgres-1, redis-1
```

[illegible]

15. After the completion of that command, run “docker ps” to check if all components of airflow are working fine. It should show healthy in status column.

```
PS C:\Users\TUSHAR-PC\airflow-docker> docker ps
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NA
21e7b6b60d36	apache/airflow:2.9.2	"/usr/bin/dumb-init ..."	7 minutes ago	Up 6 minutes (healthy)	8080/tcp	ai
rflow-docker-airflow-scheduler-1						
a9dac77771dd	apache/airflow:2.9.2	"/usr/bin/dumb-init ..."	7 minutes ago	Up 4 minutes (health: starting)	0.0.0.0:8080->8080/tcp	ai
rflow-docker-airflow-webserver-1						
c5a5b18129c	apache/airflow:2.9.2	"/usr/bin/dumb-init ..."	7 minutes ago	Up 6 minutes (healthy)	8080/tcp	ai
rflow-docker-airflow-triggerer-1						
d5dd9764c01a	apache/airflow:2.9.2	"/usr/bin/dumb-init ..."	7 minutes ago	Up 4 minutes (health: starting)	8080/tcp	ai
rflow-docker-airflow-worker-1						
5182966277fc	redis:7.2-bookworm	"docker-entrypoint.s..."	10 minutes ago	Up 10 minutes (healthy)	6379/tcp	ai
rflow-docker-redis-1						
450f514bd200	postgres:13	"docker-entrypoint.s..."	10 minutes ago	Up 10 minutes (healthy)	5432/tcp	ai
rflow-docker-postgres-1						

```
PS C:\Users\TUSHAR-PC\airflow-docker>
```

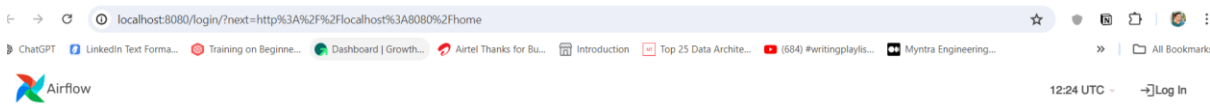
16. Docker & Airflow process may take long time or throw weird errors in between. Please give more resources to it from docker desktop for it to run fine on your machine. This can

be done in the settings panel of the docker desktop app, which runs in system tray. As shown below:



You can then restart docker and check then.

17. After all the installation is done, you should see this screen in your browser when you hit localhost:8080 in the browser. It can be an issue if other applications are using port 8080 already. You might want to give another port to the app in the docker-compose.yaml file or in airflow.cfg if you are not using docker.



The 'Sign In' form is a simple web interface with a title bar, a prompt to enter login and password, and two input fields for 'Username' and 'Password'. A 'Sign In' button is located at the bottom.

Sign In

Enter your login and password below:

Username:

Password:

Sign In

18. The login and password by default are "airflow" and "airflow" which can be changed later. After login if you clicks on "DAGS", you should see the system DAGs list below:

DAG ↕	Owner ↕	Runs ⓘ	Schedule	Last Run ↕	Next Run ↕ ⓘ	Recent Tasks ⓘ	Actions
<input type="checkbox"/> conditional_dataset_and_time_based_timetable <small>dataset-time-based-timetable</small>	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset or 0 1 * * * 3 ⓘ		2024-07-10, 01:00:00 ⓘ	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> consume_1_and_2_with_dataset_expressions	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		0 of 2 datasets updated	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> consume_1_or_2_with_dataset_expressions	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		0 of 2 datasets updated	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> consume_1_or_both_2_and_3_with_dataset_expressions	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		0 of 3 datasets updated	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> dataset_consumes_1 <small>consumes dataset-scheduled</small>	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		On s3:/dag1/output_1.txt	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> dataset_consumes_1_and_2 <small>consumes dataset-scheduled</small>	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		0 of 2 datasets updated	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> dataset_consumes_1_never_scheduled <small>consumes dataset-scheduled</small>	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		0 of 2 datasets updated	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> dataset_consumes_unknown_never_scheduled <small>dataset-scheduled</small>	airflow	<div><div></div><div></div><div></div><div></div></div>	Dataset ⓘ		0 of 2 datasets updated	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>
<input type="checkbox"/> dataset_produces_1 <small>dataset-scheduled produces</small>	airflow	<div><div></div><div></div><div></div><div></div></div>	@daily ⓘ		2024-07-09, 00:00:00 ⓘ	<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><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div></div>