

# O'Reilly – Databricks Fundamentals Bootcamp

by Mohit Batra

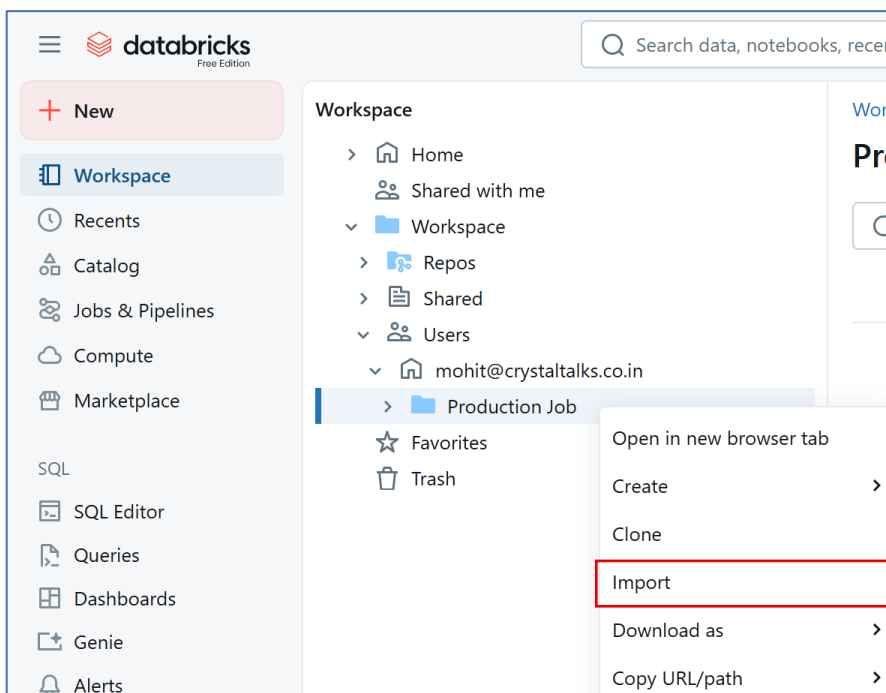
## Lab 5: Setup Databricks Job

### Contents

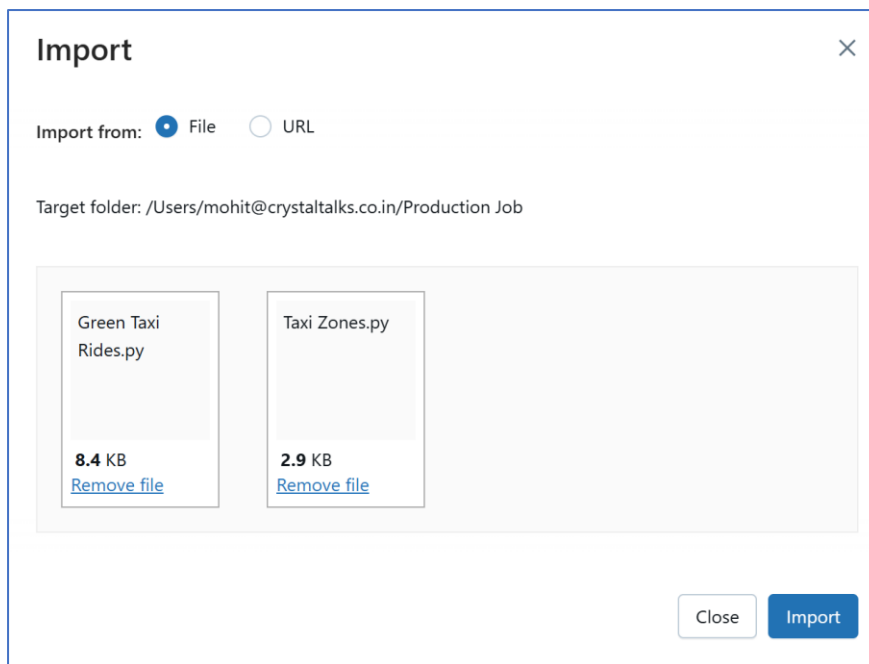
<b>Exercise 1 – Change Notebooks' Configurations</b> .....	2
<b>Exercise 2 – Create Databricks Job</b> .....	3

## Exercise 1 – Change Notebooks’ Configurations

1. Open Databricks Workspace created in previous labs.
2. Download notebooks under **Jobs** folder – **Taxi Zones.py** and **Green Taxi Rides.py** from GitHub repository.
3. In Databricks workspace, from the left menu, go to **Workspace** tab.
4. Click on Workspace folder → Users folder → Your user account.
5. Create a new folder → **Production Job**.
6. Right-click on the **Production Job** folder and click **Import**.



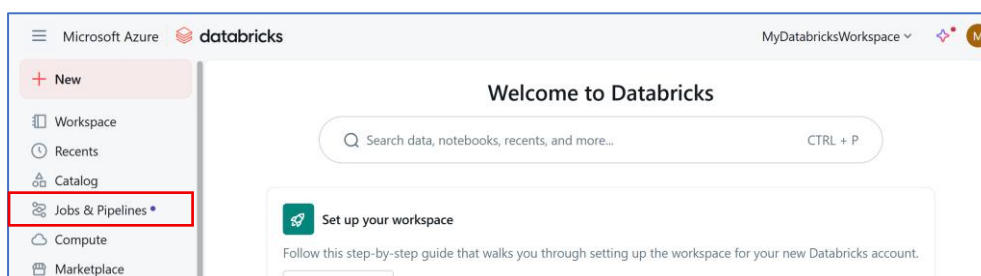
7. Upload the notebooks - **Taxi Zones.py** and **Green Taxi Rides.py** and click **Import**.



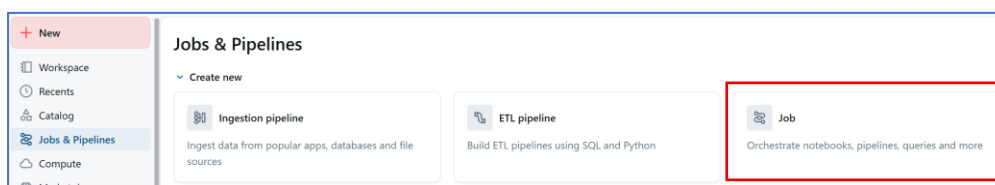
- Open the notebooks one by one and make changes to first 4 cells in each notebook by setting the right file paths and the catalog.

## Exercise 2 – Create Databricks Job

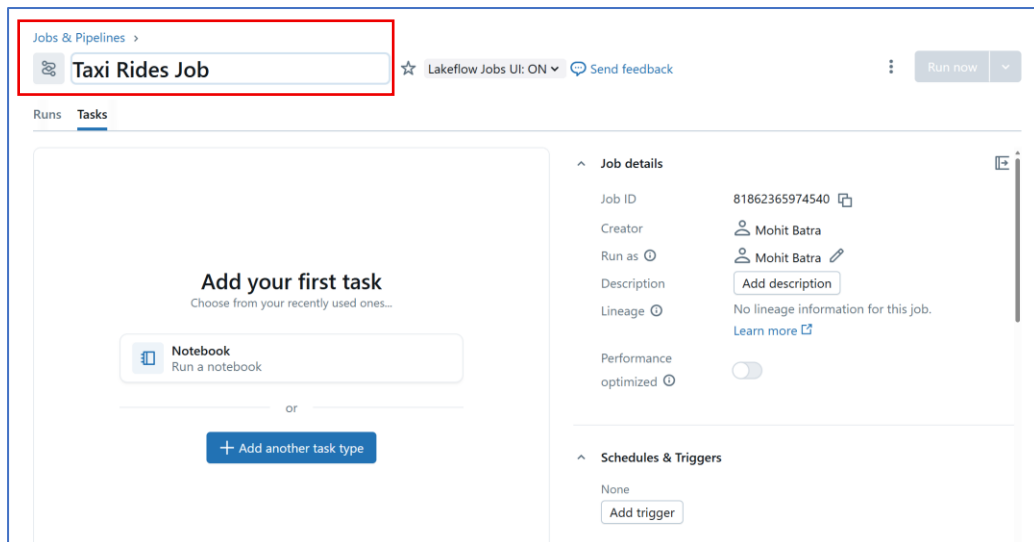
- In Databricks workspace, from the left menu, go to **Jobs & Pipelines** tab.



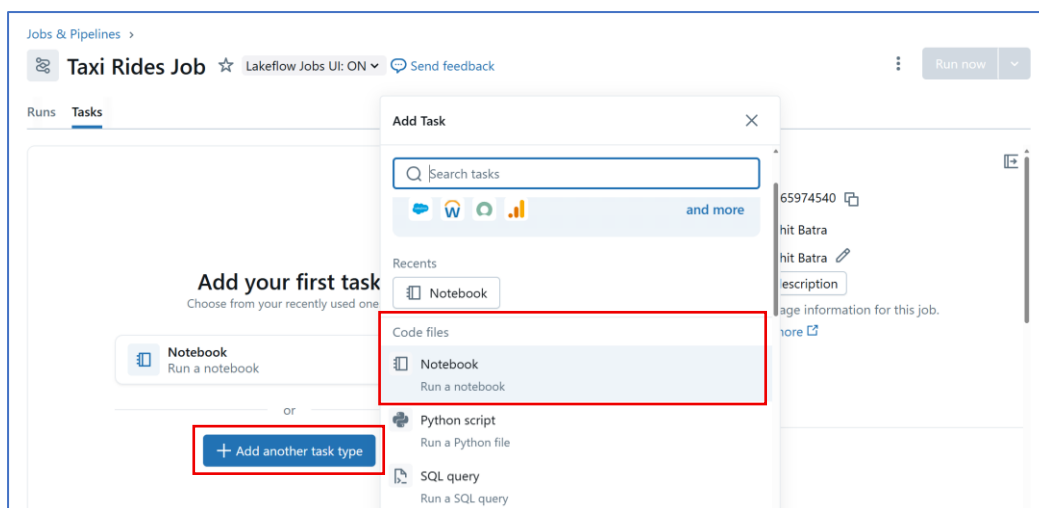
- Under **Create new** section, click on Job.



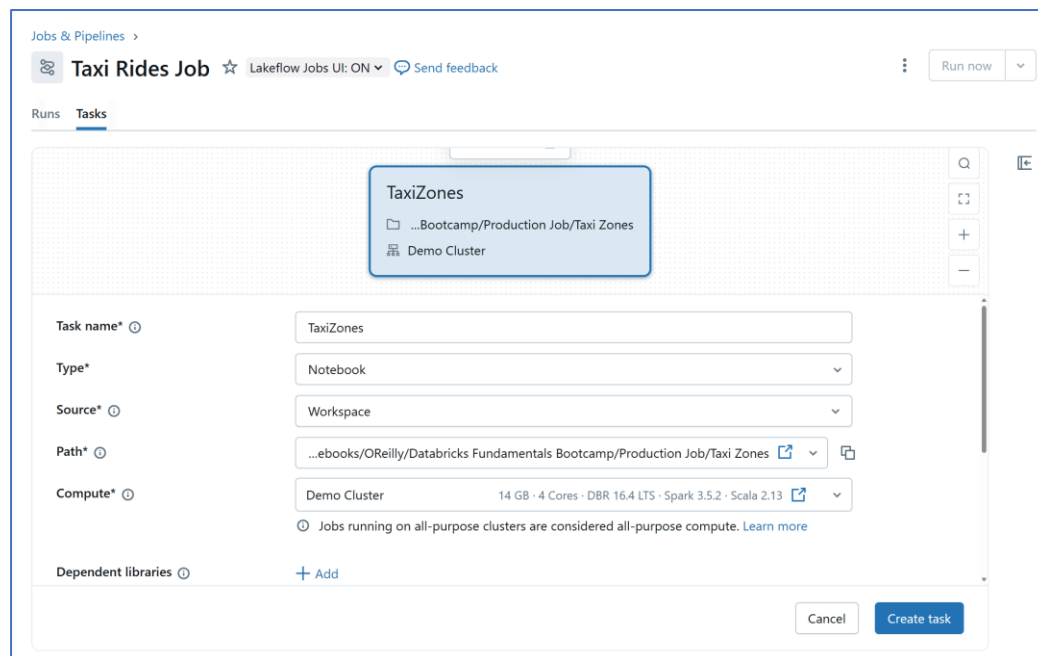
- Give a name to the job, **Taxi Rides Job**.



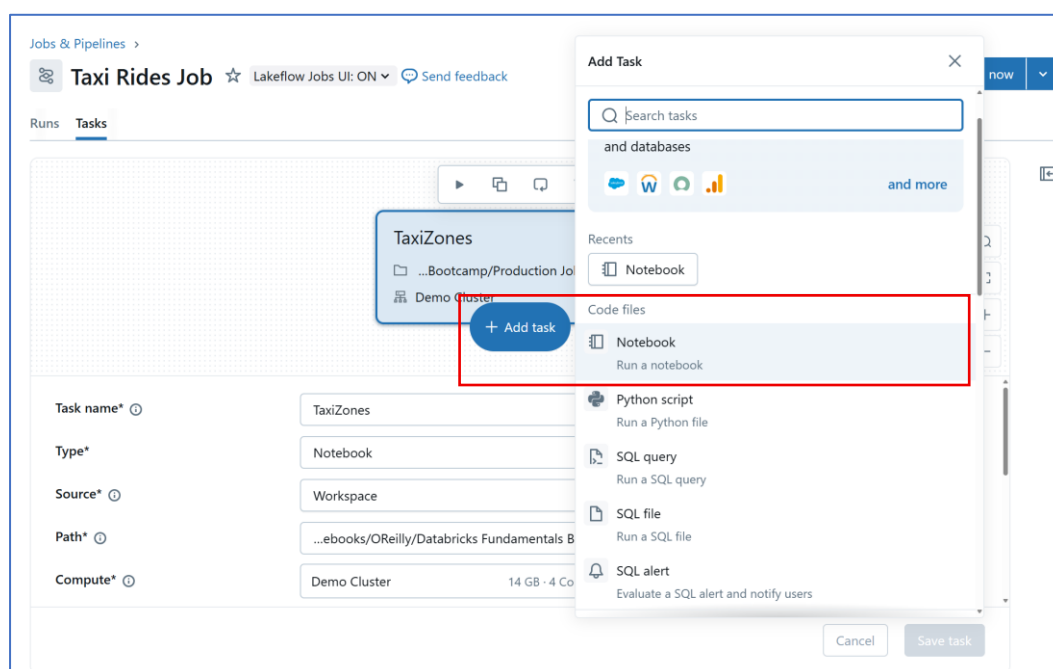
4. Click on **Add another task type**, and select **Notebook**.



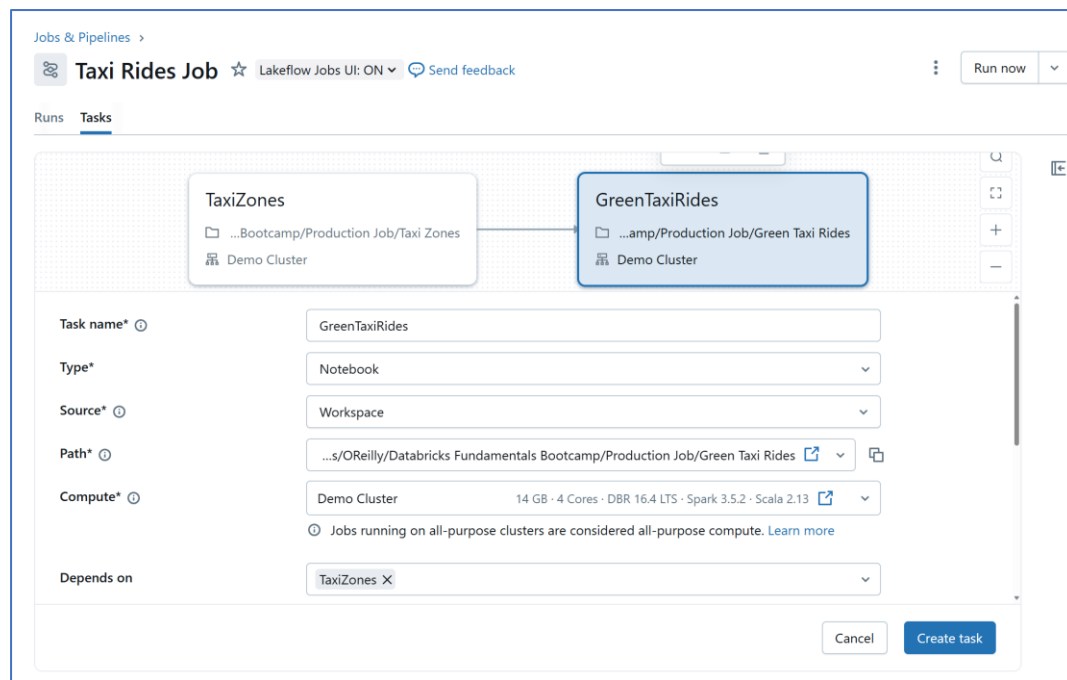
5. Fill up the task details for Taxi Zones, and create the task:
  - a. Task name: TaxiZones
  - b. Type: Notebook
  - c. Source: Workspace
  - d. Path: Select the path of Taxi Zones.py notebook.
  - e. Compute: Either select Serverless OR Demo Cluster (previously created).



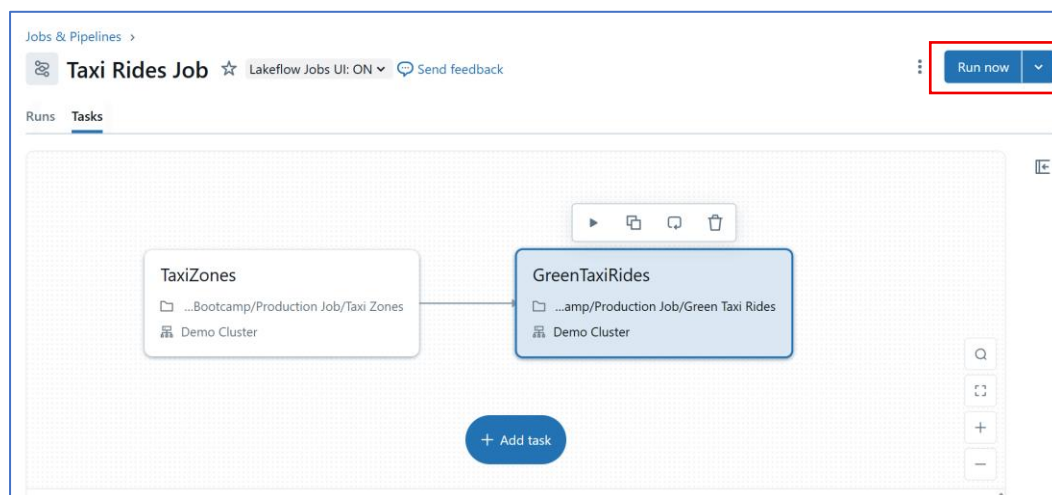
6. Click on **Add task** to add a new one, and select **Notebook**.



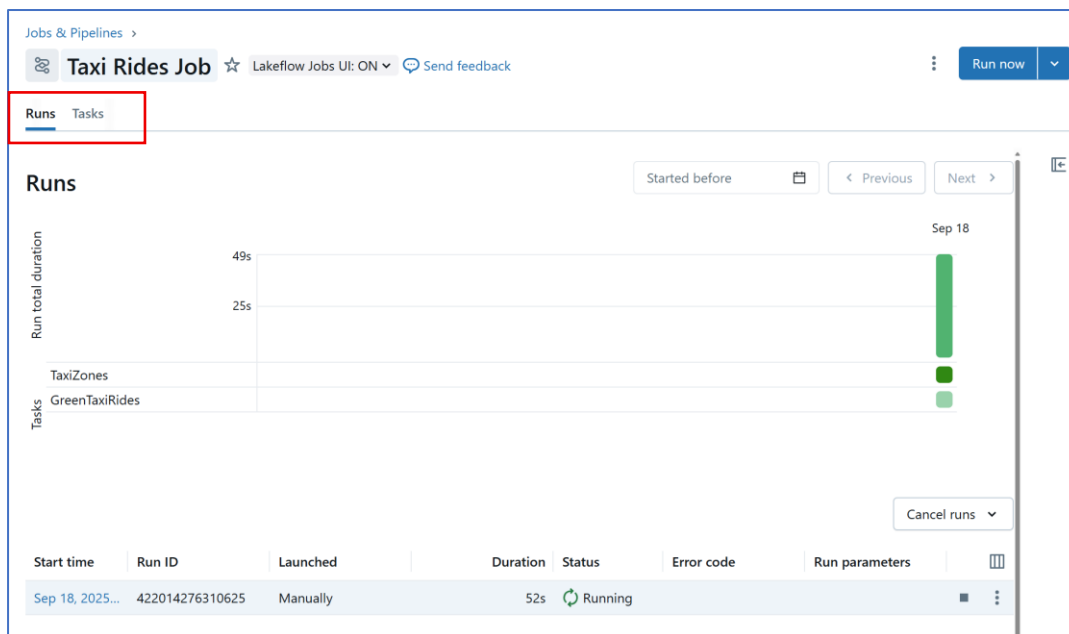
7. Fill up the task details for Green Taxi Rides, and create the task:
  - a. Task name: GreenTaxiRides
  - b. Type: Notebook
  - c. Source: Workspace
  - d. Path: Select the path of Green Taxi Rides.py notebook.
  - e. Compute: Either select Serverless OR Demo Cluster (previously created).
  - f. Depends on: Select TaxiZones



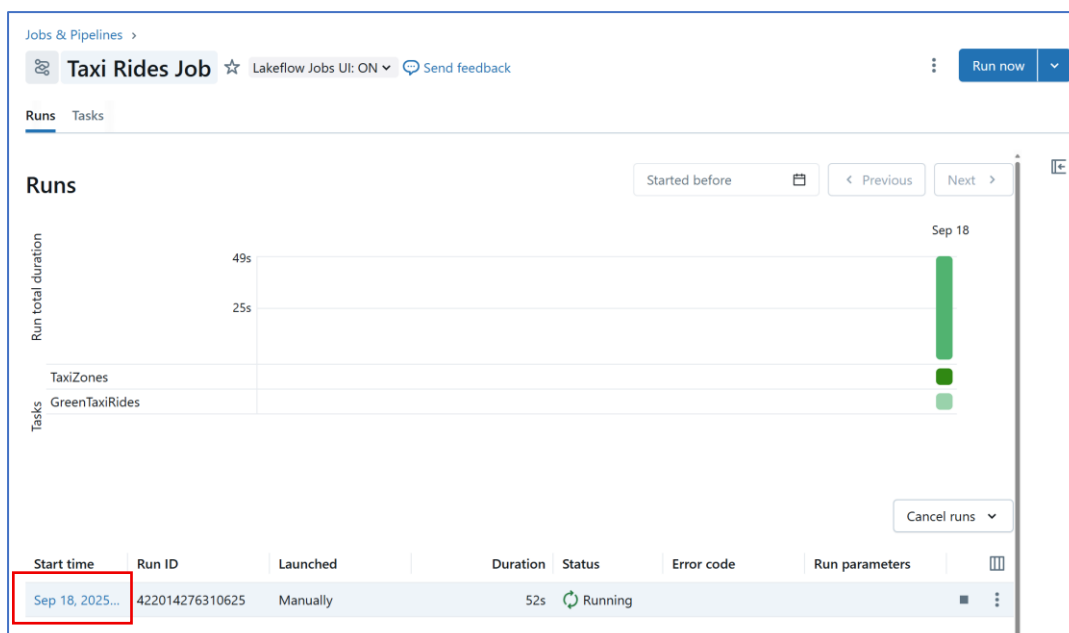
8. From top-right, click on **Run now** to run the job.



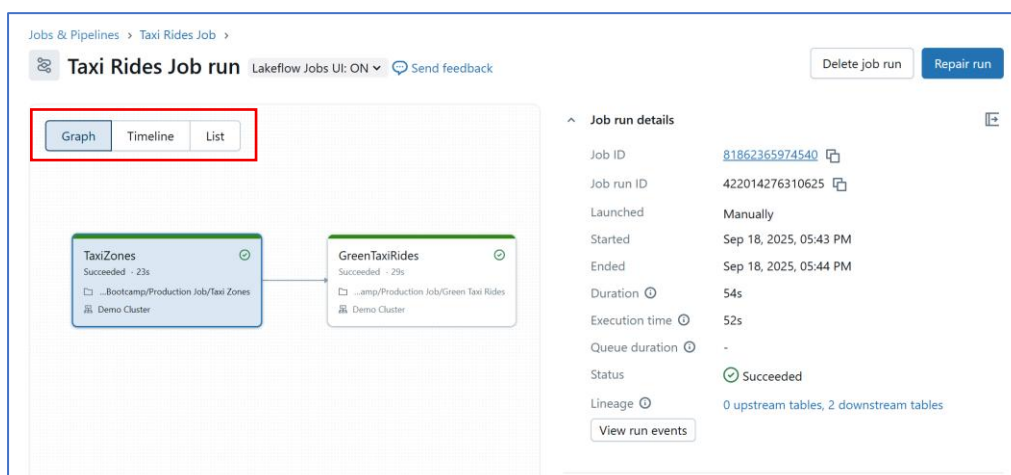
9. Go to **Runs** tab and check the execution of job.



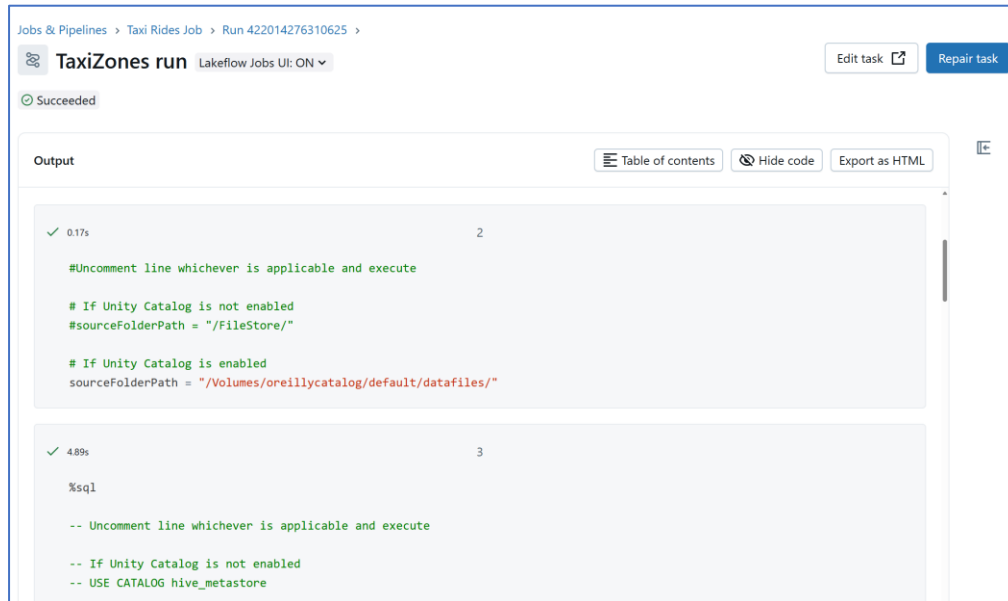
10. Click on the job to see execution details.



11. Check **Graph**, **Timeline** and **List** tabs.



12. Click on the tasks one by one to see execution of each cell in the notebook.



Jobs & Pipelines > Taxi Rides Job > Run 422014276310625 >

**TaxiZones run** Lakeflow Jobs UI: ON

✓ Succeeded

Output

Table of contents Hide code Export as HTML

```
✓ 0.17s 2
# Uncomment line whichever is applicable and execute

# If Unity Catalog is not enabled
#sourceFolderPath = "/FileStore/"

# If Unity Catalog is enabled
sourceFolderPath = "/Volumes/oreillicatalog/default/datafiles/"

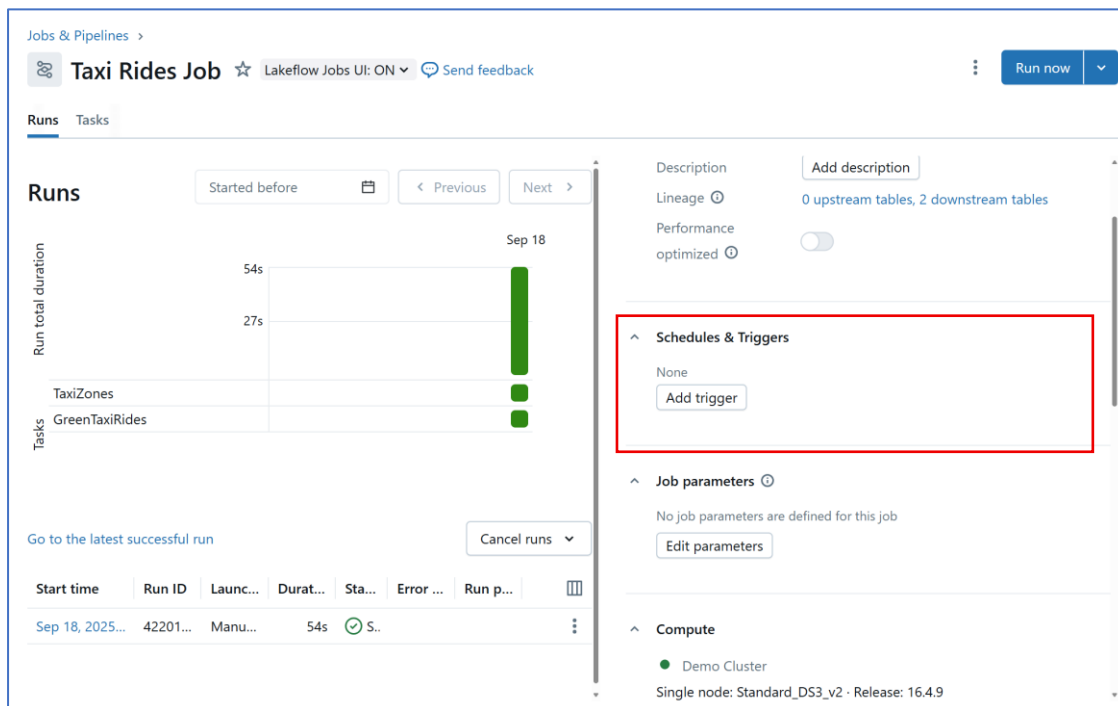
✓ 4.89s 3
%sql

-- Uncomment line whichever is applicable and execute

-- If Unity Catalog is not enabled
-- USE CATALOG hive_metastore
```

13. Go back to Taxi Rides Job.

14. From right pane in the job, notice the option of **Schedules and Triggers**. Set a schedule for the job. And notice other properties of the job.



Jobs & Pipelines >

**Taxi Rides Job** ☆ Lakeflow Jobs UI: ON Send feedback

Run now

Runs Tasks

Runs

Started before

< Previous Next >

Run total duration

54s

27s

Sep 18

Tasks

TaxiZones

GreenTaxiRides

Go to the latest successful run

Cancel runs

Start time Run ID Launc... Durat... Sta... Error ... Run p...

Sep 18, 2025... 42201... Manu... 54s S..

Description Add description

Lineage 0 upstream tables, 2 downstream tables

Performance optimized

**Schedules & Triggers**

None

Add trigger

**Job parameters**

No job parameters are defined for this job

Edit parameters

**Compute**

Demo Cluster

Single node: Standard\_DS3\_v2 · Release: 16.4.9