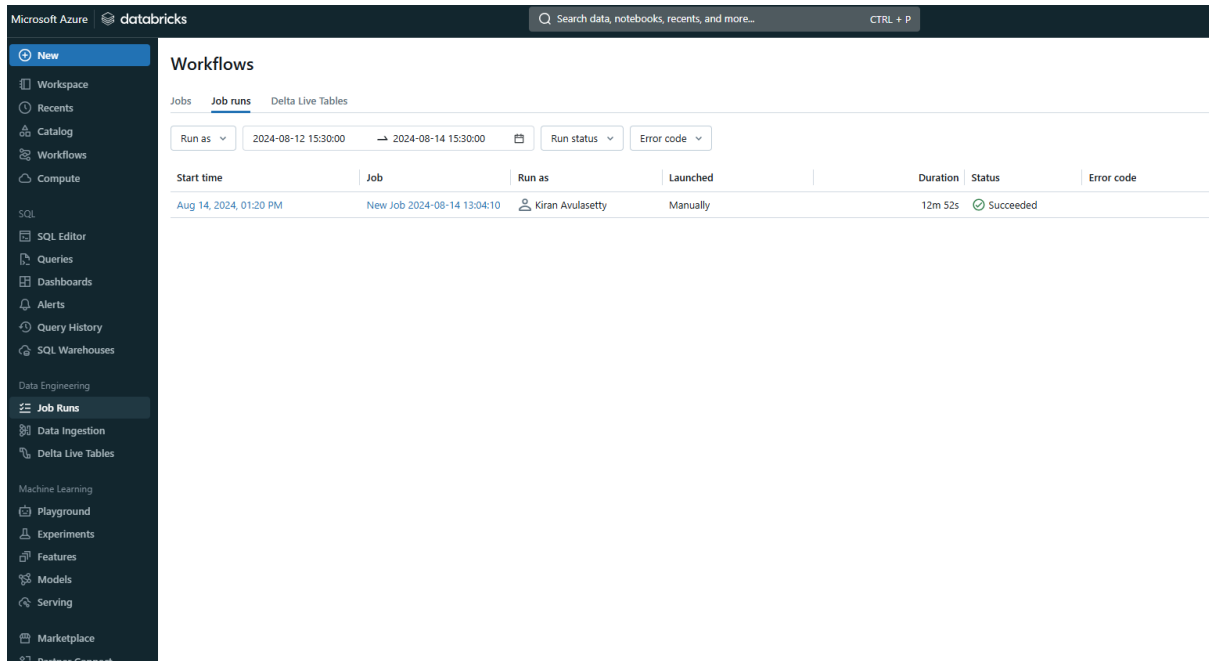


# Jobs in Databricks

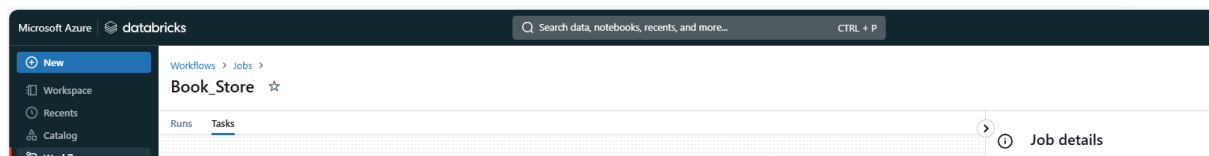
1. In this demo we are going to orchestrate our book\_store pipeline, with Databricks jobs.
2. For this we will create a multi-task job by clicking jobs runs under data engineering section in the sidebar.



3. Once you are on the job runs page, on the top right click on create job. This will open job creation page.



4. On the top left set the name as bookstore and start configuring the first job.



1. In the configuration page, for task 1, enter the following details:

- Task name: Landing\_New\_data
- Path: Click on the drop down to select Land\_New\_Data Notebook, then click on confirm.

Workspace
Recents

2 - Data Modeling
3 - Data Processing
4 - Improving Performance
5 - Databricks Tooling
6 - Security and Governance
7 - Testing and Deployment
Includes

5.2 - Multi-Task Jobs
5.4 - CLI

Task 1 - Land New Data
Task 2 - Bronze
Task 3 - Silver: Customers
Task 4 - Silver: Orders
Task 5 - Silver: Books
Task 6 - Silver: Customers Orders
Task 7 - Silver: Books Sales
Task 8 - Gold: Countries Statistics
Task 9 - Gold: Authors Statistics

Cancel
Confirm

- Parameters: Click add, in key type : number\_of\_files and in value type : 1.
- Cluster: create a new single node cluster to save the cost.

Cluster name

MyjobCluster

Policy ⓘ

Unrestricted

☐ Multi node
☒ Single node

Access mode ⓘ

Single user

Performance
Databricks runtime version ⓘ

Runtime: 14.3 LTS (Scala 2.12, Spark 3.5.0)

☒ Use Photon Acceleration ⓘ

Node type ⓘ

Standard\_D4ds\_v5 16 GB Memory, 4 Cores ⓘ

Tags ⓘ
Add tags

Key

Value

Add

> Automatically added tags

▶ Advanced options

Summary
1 Driver 16 GB Memory, 4 Cores
Runtime 14.3.x-scala2.12
Unity Catalog Photon Standard\_D4ds\_v5
2.5 DBU/h

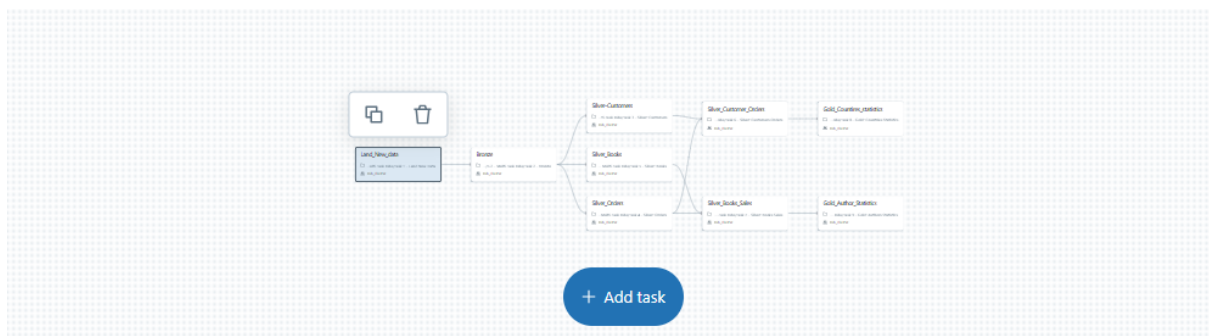
Cancel
Confirm

- Retries: Leave it default.
- Then Click create.

Task name* ⓘ	Land_New_data		
Type*	Notebook ▾		
Source* ⓘ	Workspace ▾		
Path* ⓘ	...ssional (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 1 - Land New Data <a href="#">🔗</a> <a href="#">🔗</a> ▾		
Cluster* ⓘ	Job_cluster	16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12 <a href="#">✎</a>	▾
Depends on	Select task dependencies... ▾		
Dependent libraries ⓘ	<a href="#">+ Add</a>		
Parameters ⓘ	<div> <div>UI JSON</div> <div> <div>number_of_files</div> <div>1</div> <div>{ }</div> <div>×</div> </div> </div>		

*In the job details, on the right, if you scroll down you can see the maximum concurrent runs that allow you to perform multiple runs of the same job concurrently.*

2. Now, let's add our second task, for that click +Add task and in the task configuration page enter the following details.



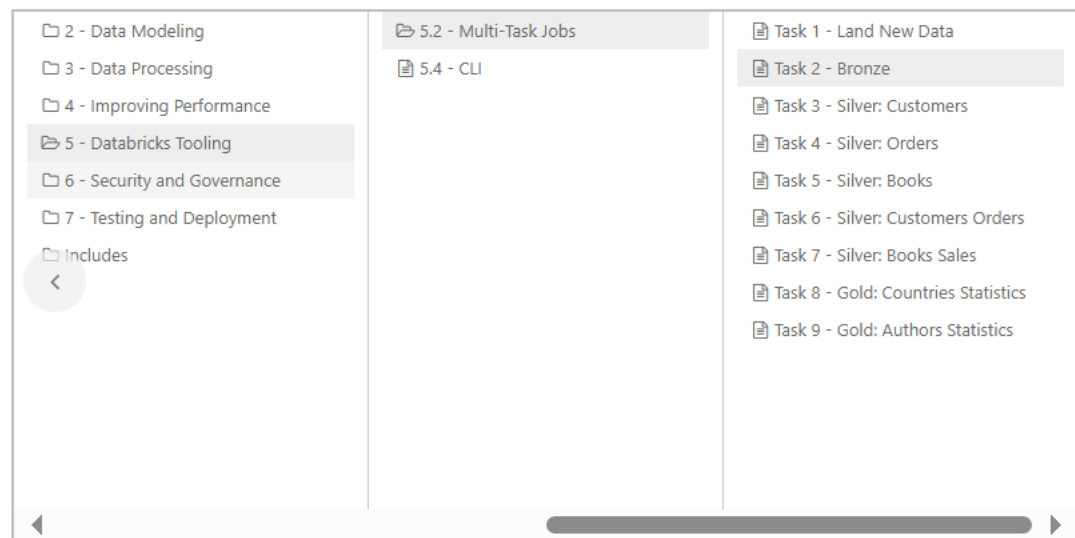
- Task name: Bronze
- Path: Click on the drop down to select Bronze Notebook, then click on confirm.

## Select Notebook



Workspace

Recents



Cancel

Confirm

- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that Landing\_New\_data is selected
- click on create task.

Task name*	Bronze
Type*	Notebook
Source*	Workspace
Path*	...er-Professional (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 2 - Bronze
Cluster*	Job_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12
Depends on	Land_New_data
Run if dependencies	All succeeded
Dependent libraries	+ Add
Parameters	UI JSON

Cancel Save task

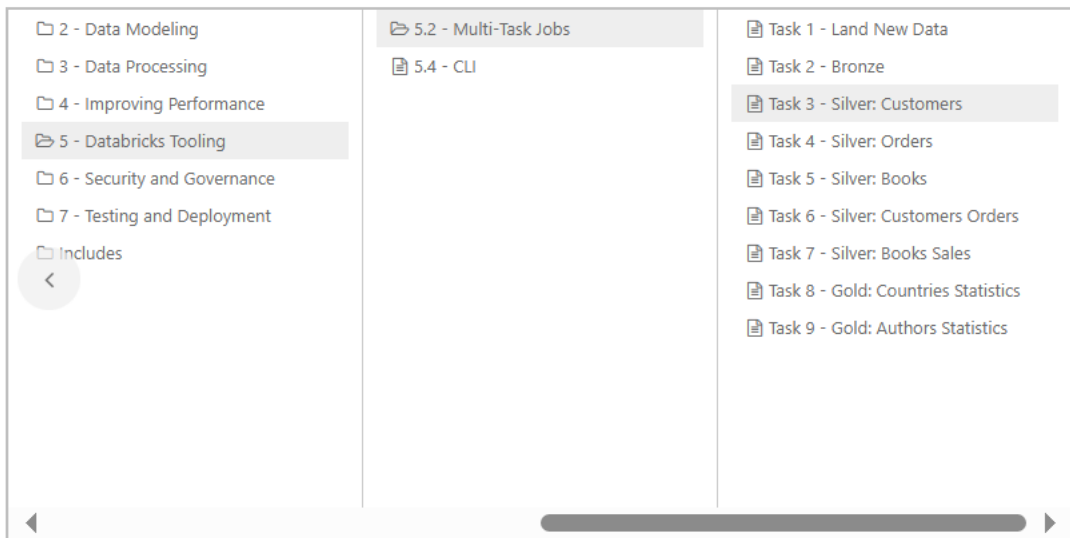
3. Now, lets add our third task, for that click +Add task and in the task configuration page enter the following details.

- Task name: Silver\_customers
- Path: Click on the drop down to select Silver\_customers Notebook, then click on confirm.

## Select Notebook



Workspace Recents



Cancel Confirm

- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that only Bronze task is selected
- click on create task.

Task name*	Silver-Customers
Type*	Notebook
Source*	Workspace
Path*	...sional (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 3 - Silver: Customers
Cluster*	Job_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12
Depends on	Bronze X
Run if dependencies	All succeeded
Dependent libraries	+ Add
Parameters	UI JSON

Cancel Save task

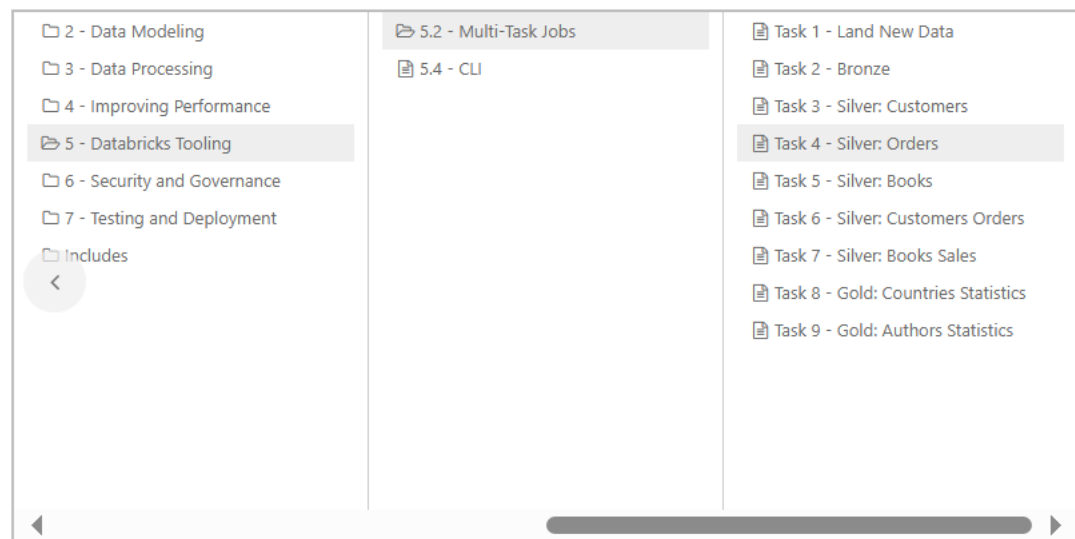
4. Now, lets add our fourth task, for that click +Add task and in the task configuration page enter the following details.

- Task name: Silver\_orders
- Path: Click on the drop down to select Silver\_orders Notebook, then click on confirm.

## Select Notebook



Workspace Recents



Cancel

Confirm

- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that only Bronze task is selected
- click on create task.

Task name*	<input type="text" value="Silver_Orders"/>
Type*	<input type="text" value="Notebook"/>
Source*	<input type="text" value="Workspace"/>
Path*	<input type="text" value="...fessional (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 4 - Silver: Orders"/>
Cluster*	<input type="text" value="Job_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12"/>
Depends on	<input type="text" value="Bronze"/>
Run if dependencies	<input type="text" value="All succeeded"/>
Dependent libraries	<a href="#">+ Add</a>
Parameters	<input type="text" value="UI"/> <input type="text" value="JSON"/>

Cancel Save task

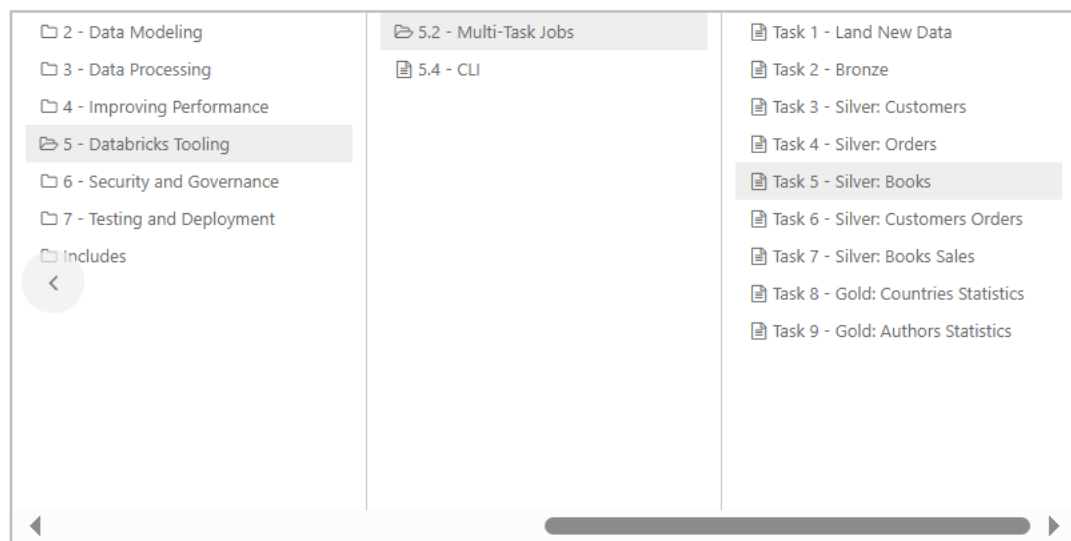
5. Now, lets add our fifth task, for that click +Add task and in the task configuration page enter the following details.

- Task name: Silver\_books
- Path: Click on the drop down to select Silver\_books Notebook, then click on confirm.

## Select Notebook



Workspace Recents



- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that only Bronze task is selected
- click on create task.

Task name\* ⓘ Silver\_Books

Type\* Notebook

Source\* ⓘ Workspace

Path\* ⓘ ...ofessional (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 5 - Silver: Books ⓘ ⓘ

Cluster\* ⓘ Job\_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12 ⓘ

Depends on Bronze X

Run if dependencies ⓘ All succeeded

Dependent libraries ⓘ + Add

Parameters ⓘ UI JSON

Cancel Save task

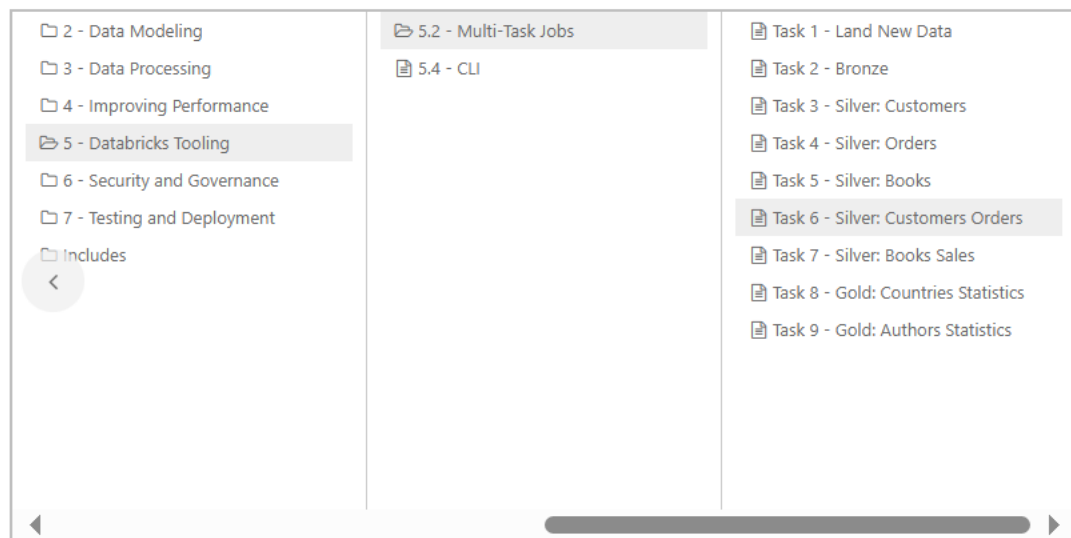
6. Now, lets add our next task, for that click +Add task and in the task configuration page enter the following details.

- Task name: Silver\_customers\_orders
- Path: Click on the drop down to select Silver\_customers\_orders Notebook, then click on confirm.

## Select Notebook



Workspace Recents



- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that both Silver\_customers and Silver\_orders tasks are selected
- click on create task.

Task name\* ⓘ Silver\_Customer\_Orders

Type\* Notebook

Source\* ⓘ Workspace

Path\* ⓘ .../5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 6 - Silver: Customers Orders ⓘ ⓘ

Cluster\* ⓘ Job\_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12 ⓘ

Depends on Silver-Customers X Silver\_Orders X

Run if dependencies ⓘ All succeeded

Dependent libraries ⓘ + Add

Parameters ⓘ UI JSON

Cancel Save task

7. Now, lets add our next task, for that click +Add task and in the task configuration page enter the following details.

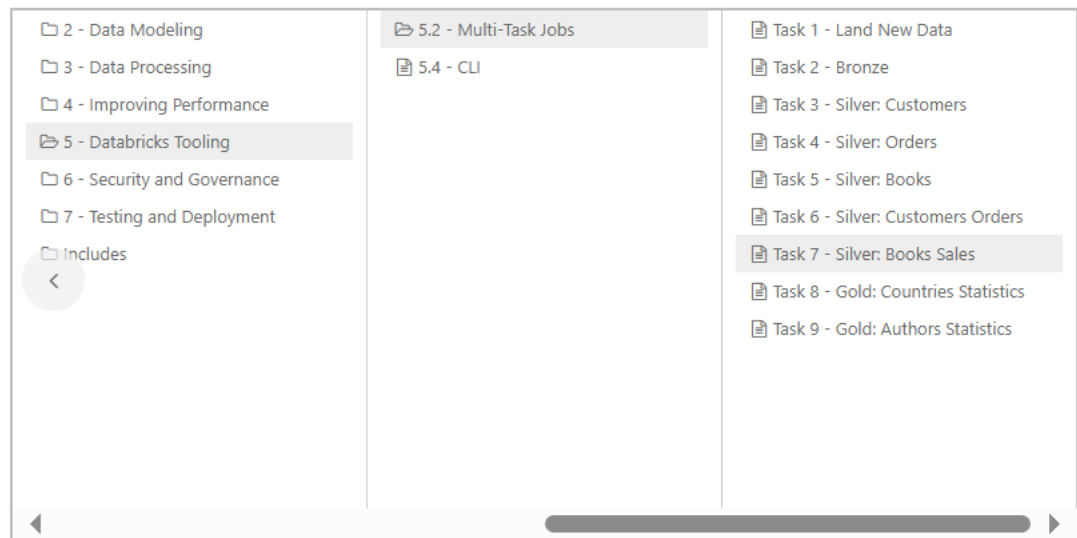
- Task name: Silver\_Book\_sales
- Path: Click on the drop down to select Silver\_Book\_sales Notebook, then click on confirm.



## Select Notebook



Workspace Recents



Cancel

Confirm

- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that both Silver\_Orders and silver\_Books tasks are selected
- click on create task.

Task name*	<input type="text" value="Silver_Books_Sales"/>
Type*	<input type="text" value="Notebook"/>
Source*	<input type="text" value="Workspace"/>
Path*	<input type="text" value="...ional (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 7 - Silver: Books Sales"/>
Cluster*	<input type="text" value="Job_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12"/>
Depends on	<input type="text" value="Silver_Orders X Silver_Books X"/>
Run if dependencies	<input type="text" value="All succeeded"/>
Dependent libraries	<a href="#">+ Add</a>
Parameters	<input type="text" value="UI"/> <input type="text" value="JSON"/>

Cancel Save task

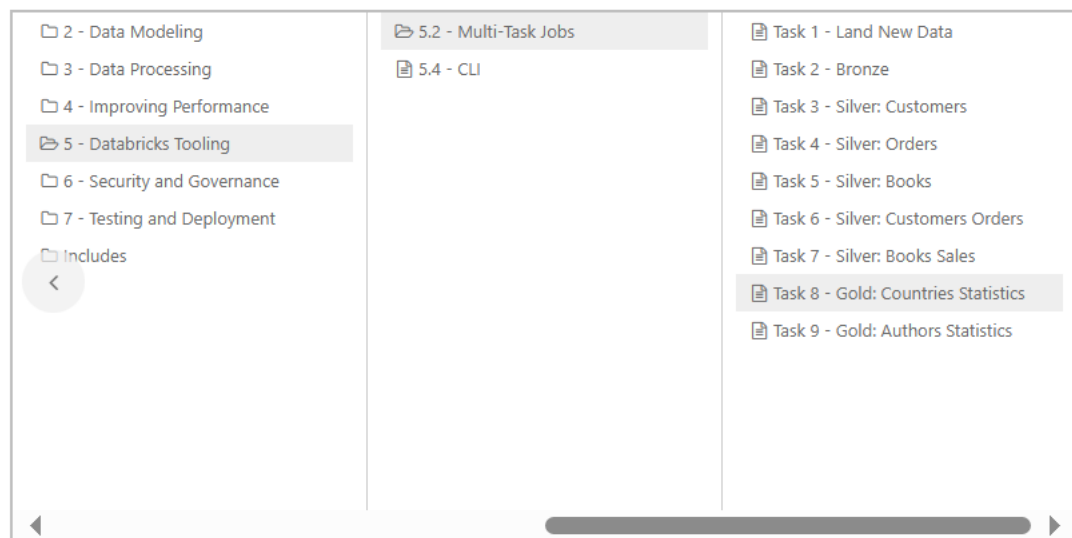
8. Now, lets add our next task, for that click **+Add task** and in the task configuration page enter the following details.

- Task name: Gold\_countries\_Statistics
- Path: Click on the drop down to select Gold \_countries\_Statistics Notebook, then click on confirm.

## Select Notebook



Workspace Recents



- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that both Silver\_Orders and silver\_Customers tasks are selected
- click on create task.

Task name*	Gold_Countires_statistics
Type*	Notebook
Source*	Workspace
Path*	.../5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 8 - Gold: Countries Statistics
Cluster*	Job_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12
Depends on	Silver_Customer_Orders X
Run if dependencies	All succeeded
Dependent libraries	+ Add
Parameters	UI JSON

Cancel Save task

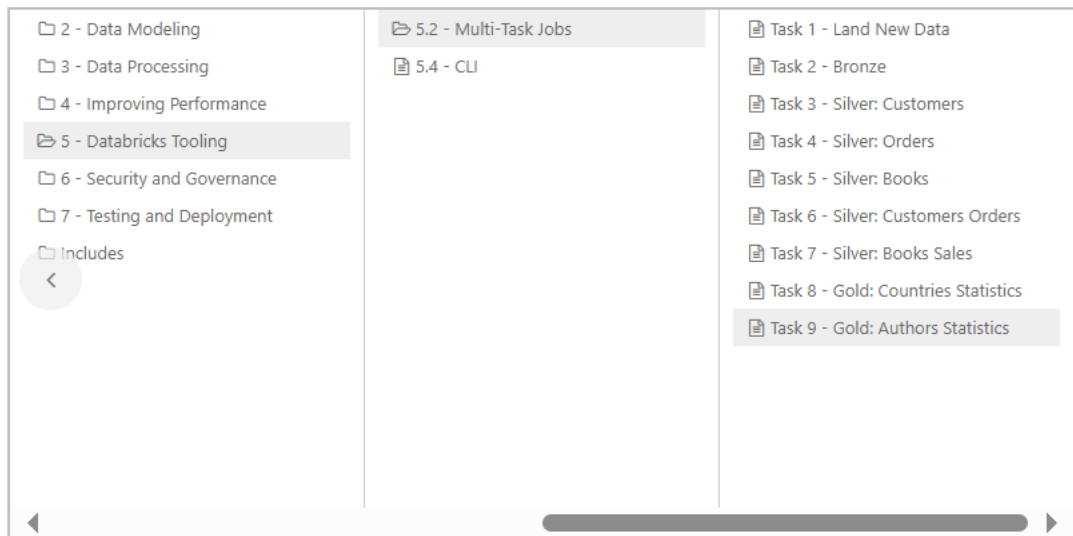
9. Finally, lets add our last task, for that click **+Add task** and in the task configuration page enter the following details.

- Task name: Gold\_Authors\_Statistics
- Path: Click on the drop down to select Gold \_Authors\_Statistics Notebook, then click on confirm.

## Select Notebook



Workspace Recents



Cancel

Confirm

- 
- Cluster: select the same cluster which we created in the previous task creation.
- Depends: Ensure that only Silver\_Book\_sales task is selected
- click on create task.

Task name*	Gold_Author_Statistics
Type*	Notebook
Source*	Workspace
Path*	... (2)/5 - Databricks Tooling/5.2 - Multi-Task Jobs/Task 9 - Gold: Authors Statistics
Cluster*	Job_cluster 16 GB · 4 Cores · DBR 14.3 LTS · Photon · Spark 3.5.0 · Scala 2.12
Depends on	Silver_Books_Sales X
Run if dependencies	All succeeded
Dependent libraries	+ Add
Parameters	UI JSON

Cancel Save task

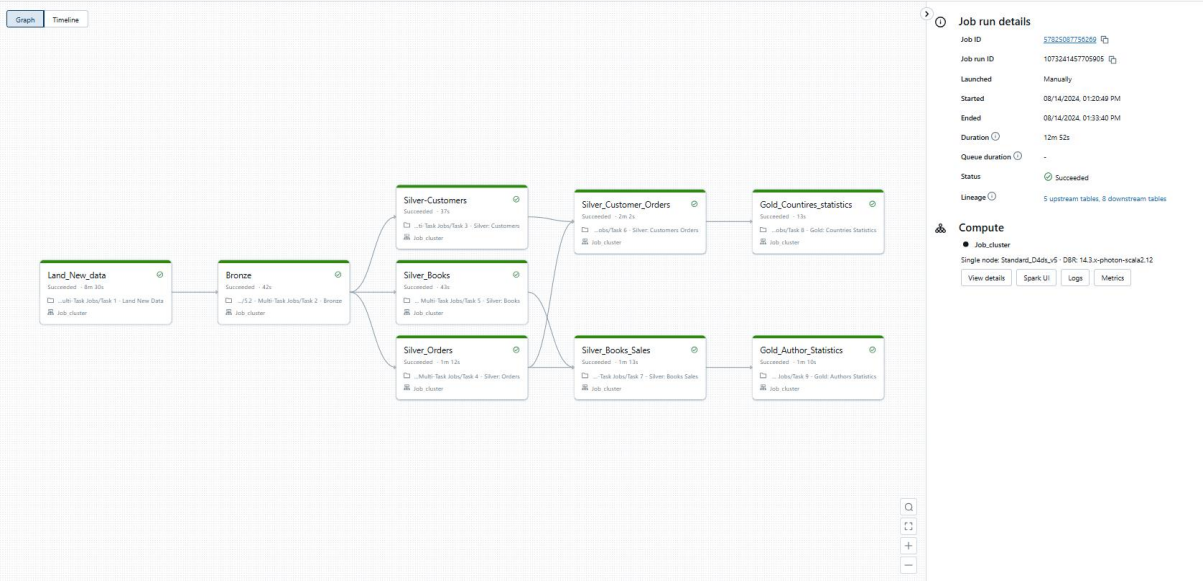
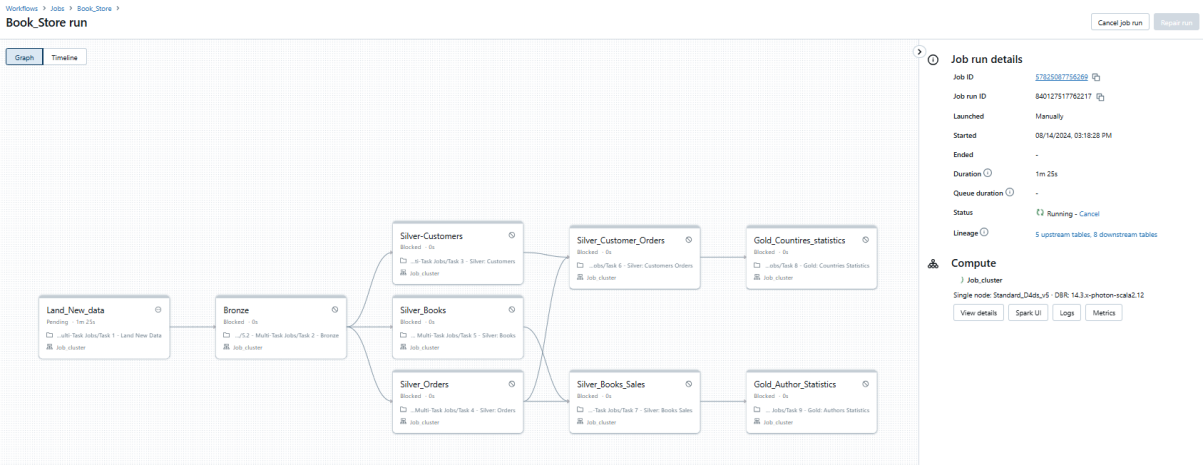
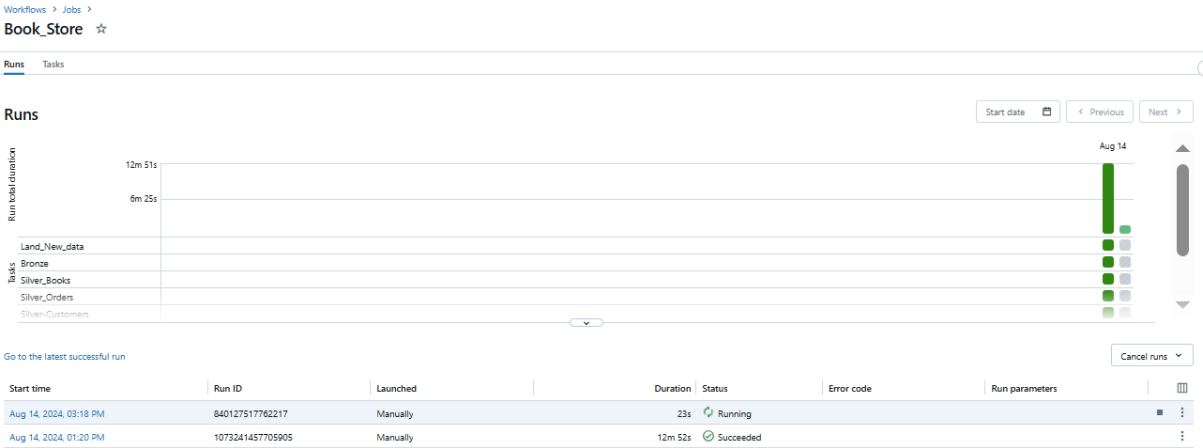
Now, all the tasks and dependencies between them are configured now. Let's go ahead and run this job.

Click **Run Now**.



Navigate to the **Runs** tab and observe that first our compute is getting created, then the

job the will get executed. Once the job is executed you will be able to see the following screen.



New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Playground

Experiments

Features

Models

Serving

Marketplace

Partner Connect

Collaborate (more)

Workflows > Jobs >

Book\_Store ☆

RunsTasks

Start datePreviousNext

Runs

Run total duration

Aug 14

12m 51s

6m 25s

Land\_New\_data

Bronze

Silver\_Books

Silver\_Orders

Silver-Customers

Go to the latest successful run

Cancel runs

Start time	Run ID	Launched	Duration	Status	Error code	Run parameters
Aug 14, 2024, 01:20 PM	1073241457705905	Manually	12m 52s	Succeeded		

Job details

Job ID5782508756269

Creator

Run as

TagsAdd tag

DescriptionAdd description

Lineage5 upstream tables, 8 downstream tables

Git

Not configured

Add Git settings

Schedules & Triggers

Every day, next run at Aug 15, 2024, 01:20 PM

Edit triggerPauseDelete

Compute

Job\_cluster

Single node: Standard\_D4ds\_v5 · DBR: 14.3.x-photon-scala2.12

ConfigureSwap

Job parameters

No job parameters are defined for this job