INDIAN DATA CLUB

databricks

CODE BASICS

14 DAYS AI CHALLENGE

DAY 07

Workflows & Job Orchestration

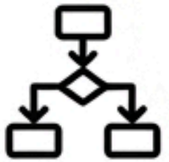
The Mission

Automating the Pipeline

Our goal is to transform a static notebook into a fully orchestrated automated factory. We must solve four specific challenges to pass Day 7.

Dynamic Inputs: Add parameter widgets to notebooks.

Orchestration: Create a multi-task job (`Bronze` → `Silver` → `Gold`).



Logic: Set up execution dependencies.

Automation: Schedule the execution.

Step 1: Injecting Dynamic Parameters

Making notebooks reusable with dbutils.widgets

```
1
2
3   # Read parameter values from widgets
4   layer_type = dbutils.widgets.get('layer_type')
5   source_path = dbutils.widgets.get('source_path')
6   process_date = dbutils.widgets.get('process_date')
7
```

Dynamic Variables

Instead of hard-coding 'Bronze', we pull the value from the job configuration at runtime.

Technical Editorial

Verifying Input Logic

Debugging print statements confirm the correct parameters are passed.

Code Input

```
print(f"Layer Type: {layer_type}")
print(f"Source Path: {source_path}")
print(f"Process Date: {process_date}")
```

Console Output

```
Layer Type: Bronze
Source Path: /mnt/data/bronze/
Process Date: 2026-01-15
```

Success

Step 2: Constructing the Multi-Task Job

A ‘Job’ acts as a container for multiple tasks. We configure three distinct tasks that reuse the same notebook but with different parameters.

Databricks Job: Day 7 Workflow

Task 1

Name: Bronze

Parameter:
`layer_type = 'Bronze'`

Task 2

Name: Silver

Parameter:
`layer_type = 'Silver'`

Task 3

Name: Gold

Parameter:
`layer_type = 'Gold'`

Technical Editorial

Step 3: Defining Dependencies

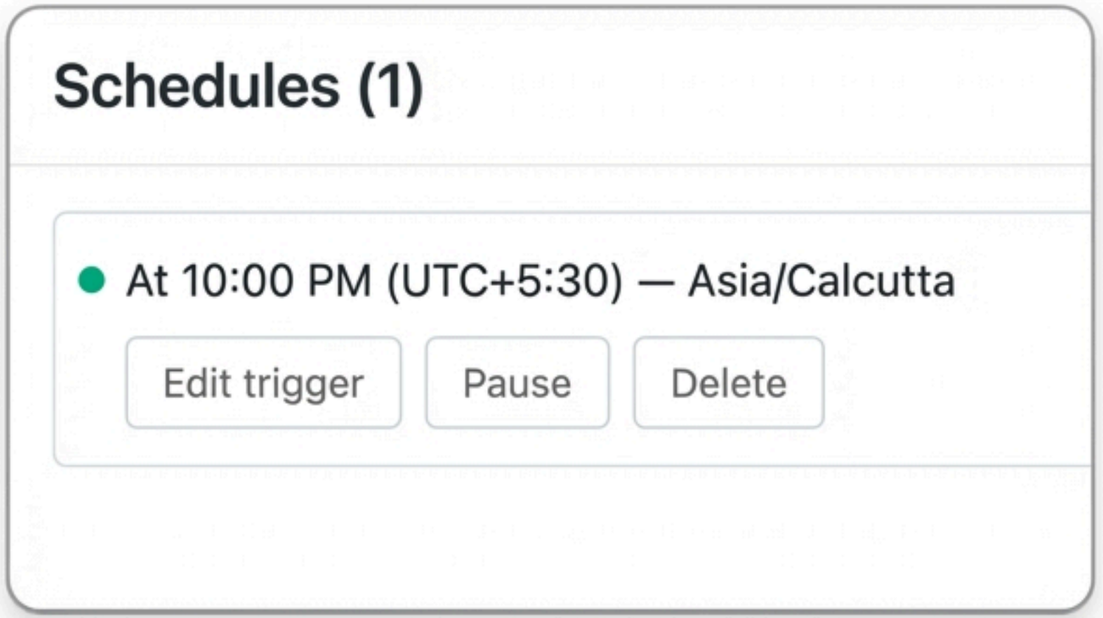
Enforcing the Directed Acyclic Graph (DAG)

Linear Progression.

The Gold layer is protected. It will never attempt to process if the Silver transformation fails, ensuring data quality downstream.

Step 4: Automating Execution

Cron Scheduling. By setting the timezone and trigger time, we convert the job from an ad-hoc experiment to a production-grade nightly pipeline.



Execution Results: The Run History

Runs

Start time	Run ID	Launched	Duration	Status
Jan 15, 2026, 11:02 PM	2176083517790...	Manually	35s	Succeeded

Optimized Performance. The entire multi-task workflow completed in just over half a minute.

Visualizing Success: The Completed Graph

Day 7 Challenge: Complete

The pipeline is now robust, automated, and ready for Day 8.

INDIAN DATA CLUB

- **1. Widgets:** Implemented dynamic variable passing for reusable code.
- **2. Orchestration:** Built a multi-task ETL job using the Jobs API.
- **3. Dependencies:** Enforced logical execution order (DAG) to protect data quality.
- **4. Scheduling:** Automated daily runs at 10:00 PM IST.