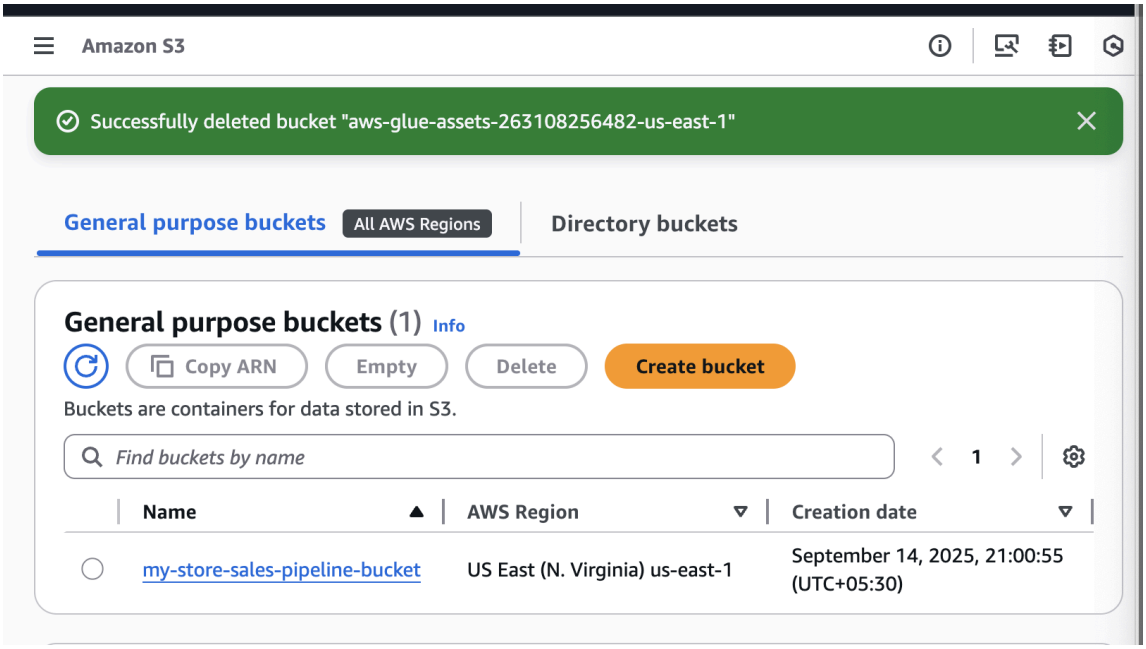
NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT

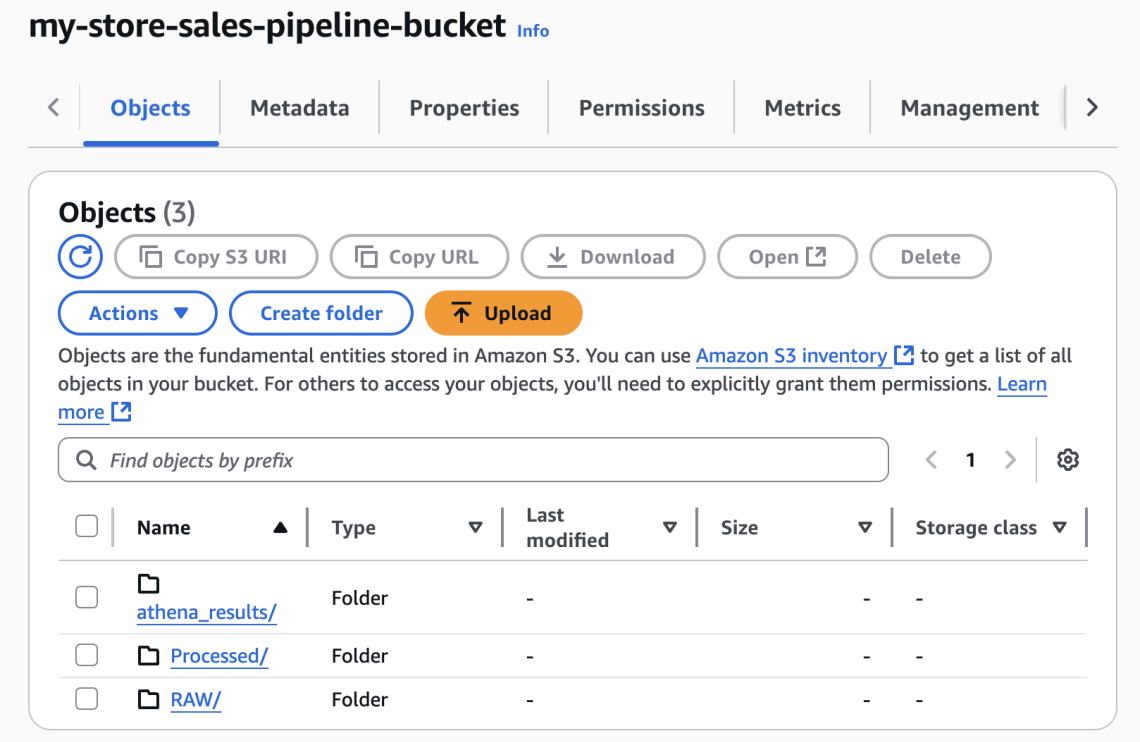
1. ***Why I pasted it:*** **To show the raw and processed storage locations I set up for my pipeline.**

What I did there: **Created an S3 bucket named `my-store-sales-pipeline` with folders `RAW/` for input JSON files, `Processed/` for Parquet outputs, and `athena_results/` for query results.**



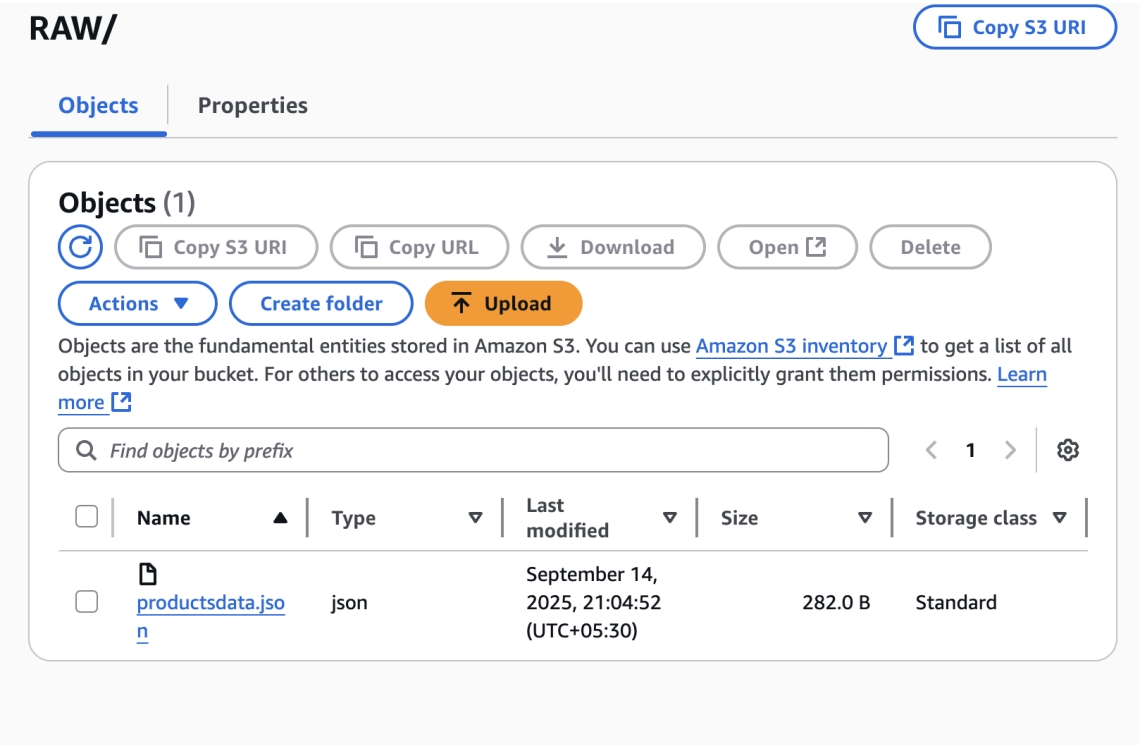
NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT



NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT

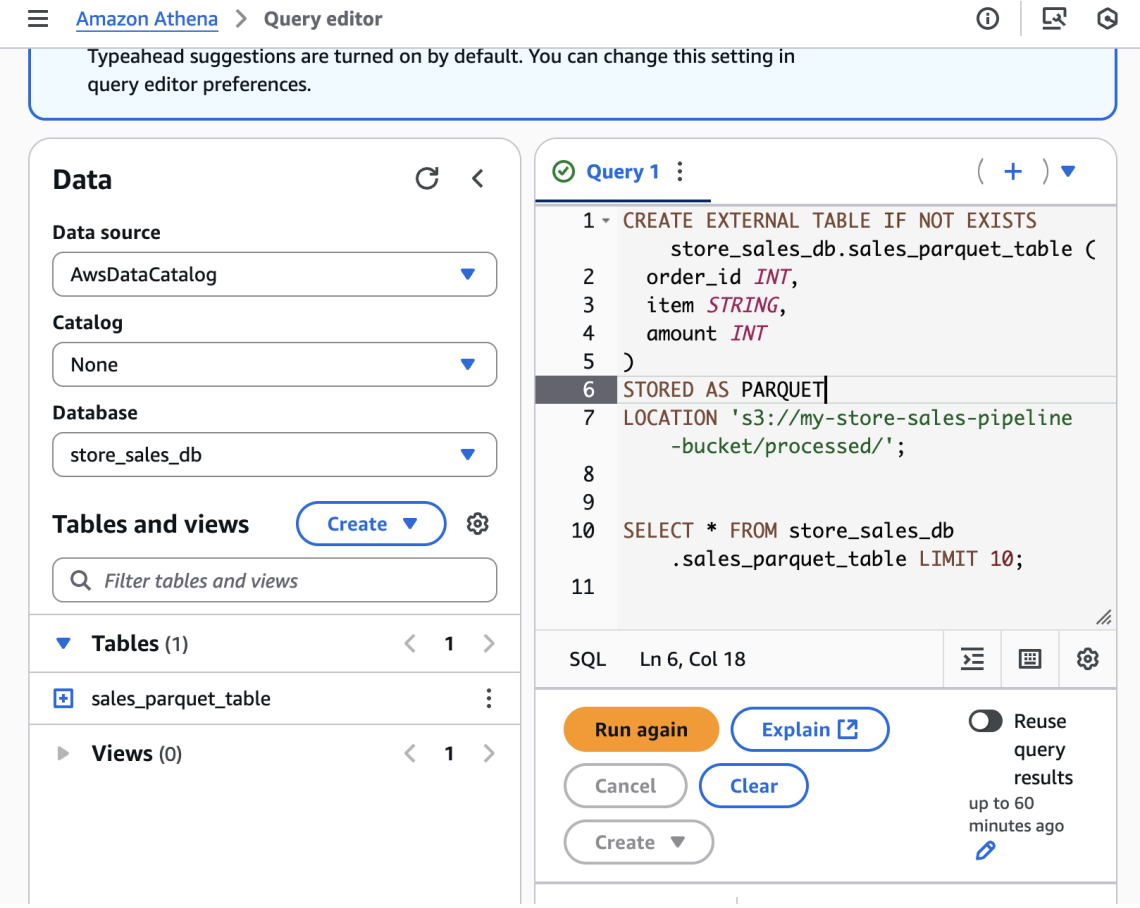


2. Why I pasted it: **To demonstrate that I configured Athena to use the Glue Data Catalog database for querying processed data.**

What I did there: **Opened Athena Query Editor and confirmed that the database `store_sales_db` is visible and available for running queries.**

NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT

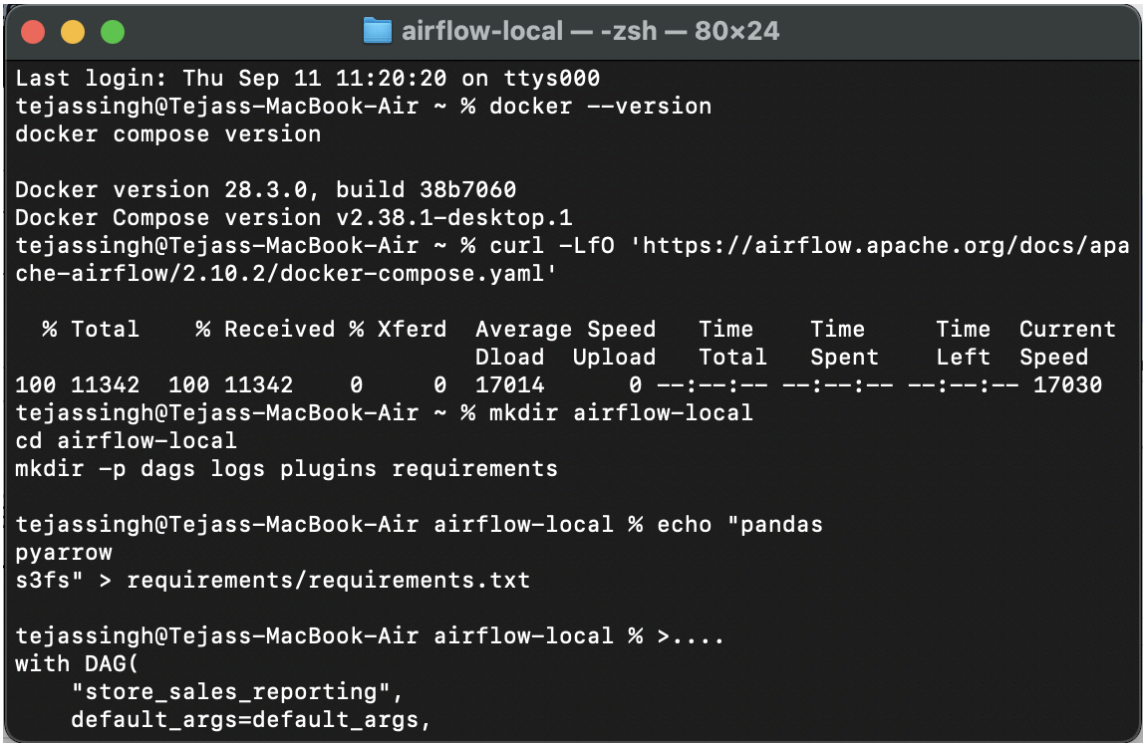


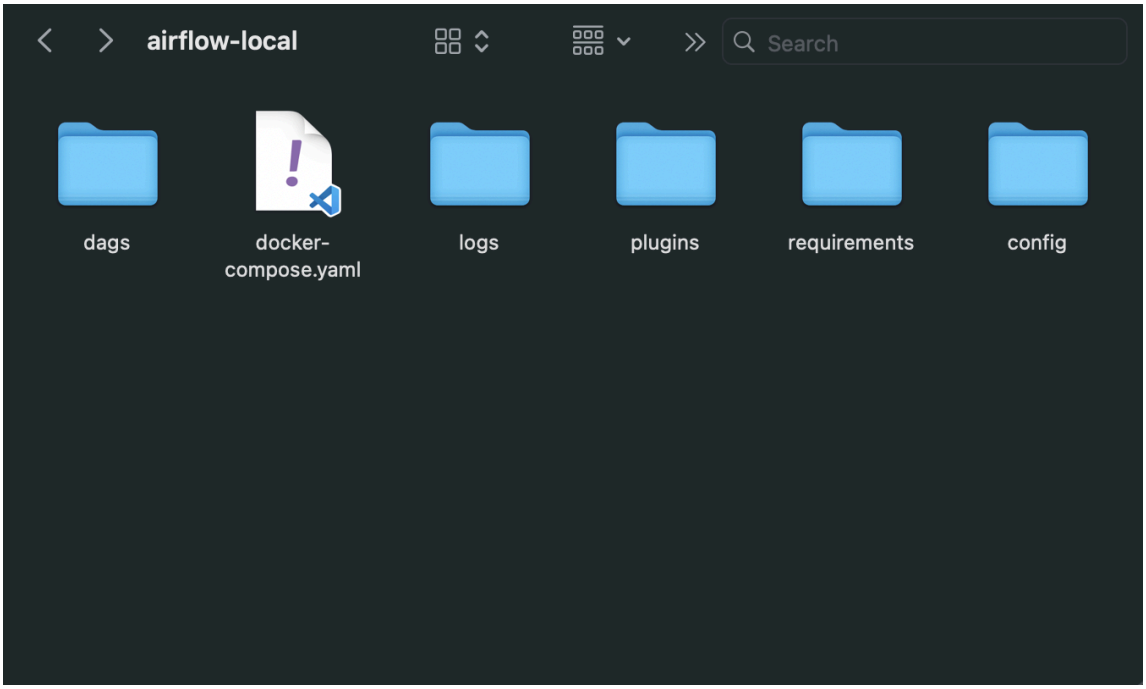
3. Why I pasted it: **To show the local environment setup needed to run Airflow without using paid MWAA.**

What I did there: **Installed Docker Desktop, pulled the official Airflow Docker Compose setup, and prepared the `dags/`, `plugins/`, and `requirements/` folders for the project.**

NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT



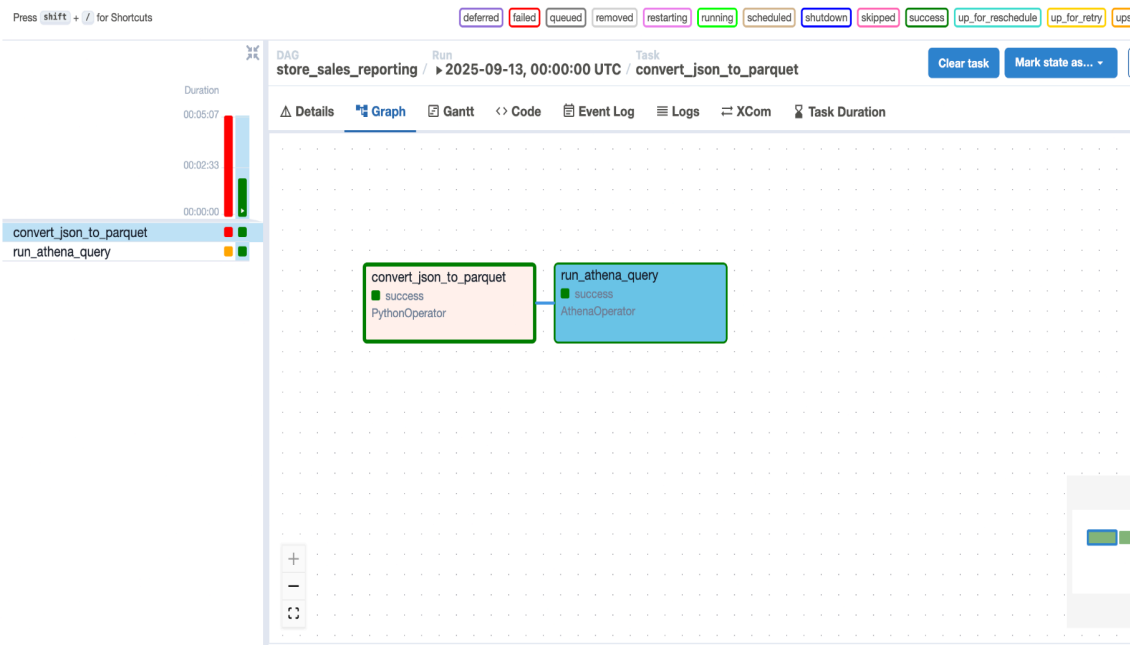


NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT

4. Why I pasted it: **To demonstrate that my DAG executed successfully, orchestrating the JSON-to-Parquet conversion and Athena table refresh.**

What I did there: **Triggered the `store_sales_reporting` DAG in the Airflow UI, and both tasks (`convert_json_to_parquet` and `run_athena_query`) completed successfully.**



5. Why I pasted it: **To validate that the DAG successfully transformed raw JSON files into Parquet format and stored them back in S3.**

What I did there: **Checked the `Processed/` folder inside my S3 bucket and confirmed that new `.parquet` files were created after the DAG run.**

NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT