

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.

The screenshot shows the Amazon S3 console. At the top, a green success message box displays: "Successfully deleted bucket 'aws-glue-assets-263108256482-us-east-1'". Below this, there are two tabs: "General purpose buckets" (which is selected) and "Directory buckets". Under the "General purpose buckets" tab, there is a sub-header "General purpose buckets (1)" with an "Info" link. Below this are several buttons: "Copy ARN", "Empty", "Delete", and "Create bucket". A note states: "Buckets are containers for data stored in S3." There is a search bar with the placeholder "Find buckets by name" and a pagination indicator showing "1". A table lists one bucket: "my-store-sales-pipeline-bucket" (Name), "US East (N. Virginia) us-east-1" (AWS Region), and "September 14, 2025, 21:00:55 (UTC+05:30)" (Creation date).

NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT

my-store-sales-pipeline-bucket [Info](#)

< [Objects](#) Metadata Properties Permissions Metrics Management >

Objects (3)

[C](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#)

[Actions ▾](#) [Create folder](#) [Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix < 1 > [⚙️](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	athena_results/	Folder	-	-	-
<input type="checkbox"/>	Processed/	Folder	-	-	-
<input type="checkbox"/>	RAW/	Folder	-	-	-

NAME:- Tejas Singh
Emp ID: AS1554

AIRFLOW ASSIGNMENT

RAW/

[Copy S3 URI](#)

[Objects](#) [Properties](#)

Objects (1)

[C](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#)

[Actions ▾](#) [Create folder](#) [Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 > | [⚙️](#)

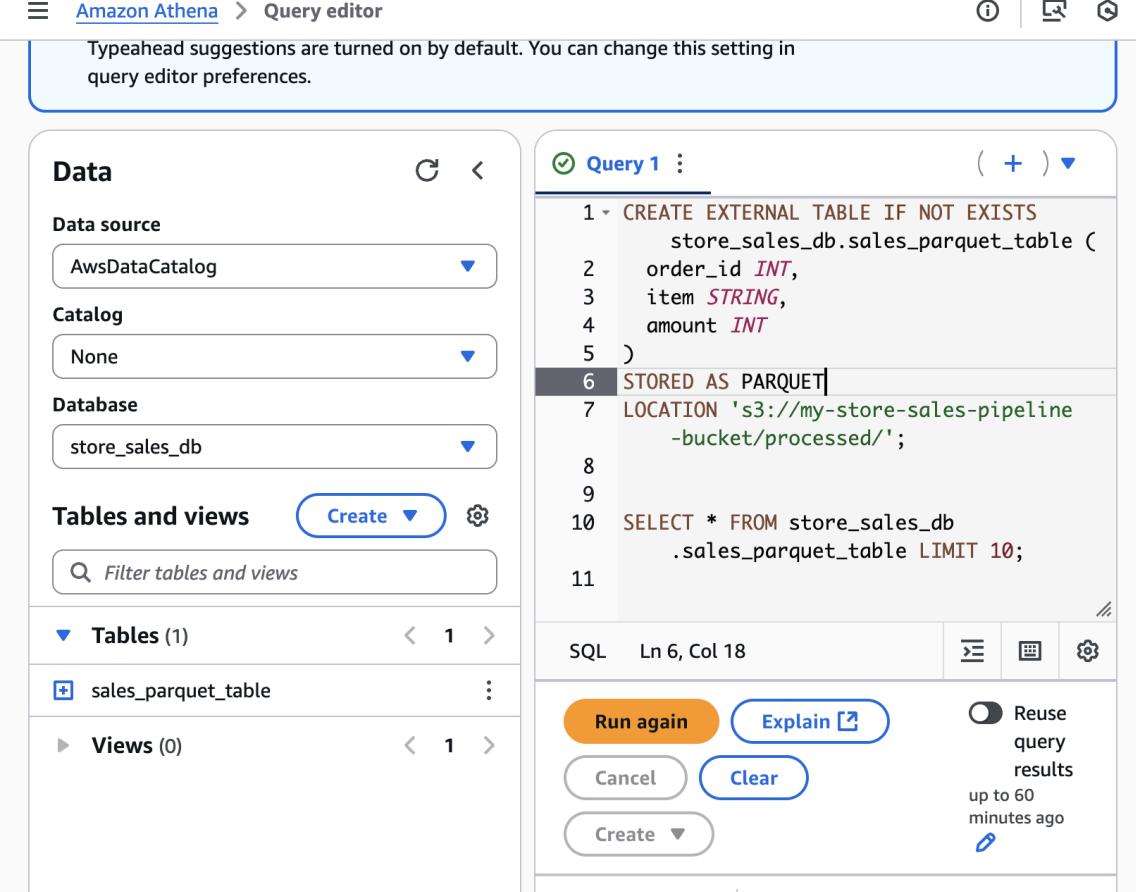
<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	productsdata.json	json	September 14, 2025, 21:04:52 (UTC+05:30)	282.0 B	Standard

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



The screenshot shows the Amazon Athena Query editor interface. On the left, the 'Data' sidebar is open, displaying the 'Data source' set to 'AwsDataCatalog', 'Catalog' set to 'None', and 'Database' set to 'store_sales_db'. Below this, the 'Tables and views' section shows 'Tables (1)' containing 'sales_parquet_table' and 'Views (0)'. On the right, the main area is titled 'Query 1' and contains the following SQL code:

```
1 CREATE EXTERNAL TABLE IF NOT EXISTS
  store_sales_db.sales_parquet_table (
2   order_id INT,
3   item STRING,
4   amount INT
5 )
6 STORED AS PARQUET
7 LOCATION 's3://my-store-sales-pipeline
-bucket/processed/';
8
9
10 SELECT * FROM store_sales_db
  .sales_parquet_table LIMIT 10;
11
```

The status bar at the bottom indicates 'SQL Ln 6, Col 18'. Below the status bar are buttons for 'Run again' (orange), 'Explain' (blue), 'Cancel' (gray), 'Clear' (blue), and 'Create' (gray). To the right of these buttons is a 'Reuse query results' toggle switch and a note stating 'up to 60 minutes ago'.

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

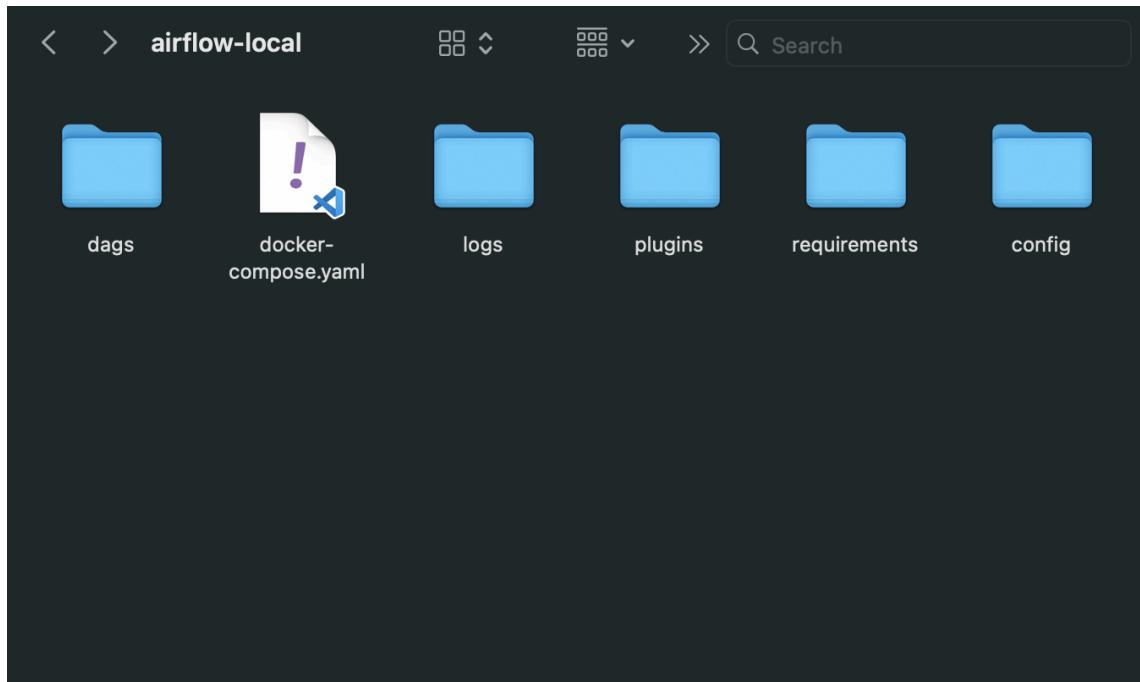
```
airflow-local — -zsh — 80x24
Last login: Thu Sep 11 11:20:20 on ttys000
tejassingh@Tejass-MacBook-Air ~ % docker --version
docker compose version

Docker version 28.3.0, build 38b7060
Docker Compose version v2.38.1-desktop.1
tejassingh@Tejass-MacBook-Air ~ % curl -Lf0 'https://airflow.apache.org/docs/apache-airflow/2.10.2/docker-compose.yaml'

% Total    % Received % Xferd  Average Speed   Time     Time      Time  Current
               Dload  Upload   Total   Spent   Left  Speed
100 11342  100 11342     0      0  17014       0 ---:---:--- ---:---:--- ---:---:--- 17030
tejassingh@Tejass-MacBook-Air ~ % mkdir airflow-local
cd airflow-local
mkdir -p dags logs plugins requirements

tejassingh@Tejass-MacBook-Air airflow-local % echo "pandas
pyarrow
s3fs" > requirements/requirements.txt

tejassingh@Tejass-MacBook-Air airflow-local % >....
with DAG(
    "store_sales_reporting",
    default_args=default_args,
```

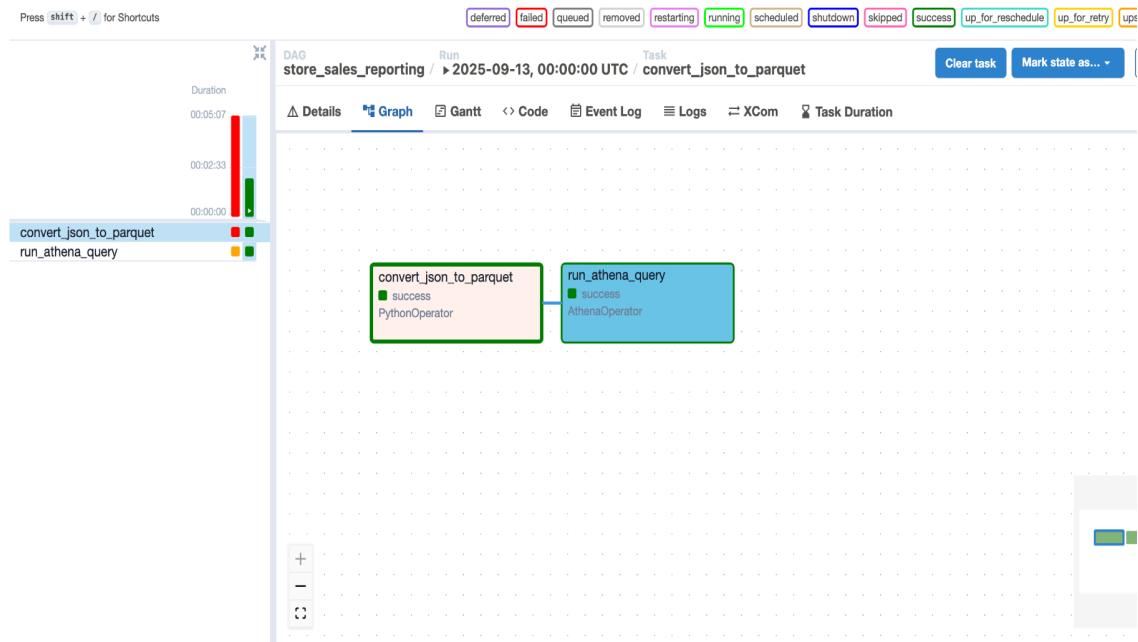


NAME:- Tejas Singh
Emp ID: AS1554

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

The screenshot shows the AWS S3 console interface. The top navigation bar includes the AWS logo, search, notifications, and account information (Account ID: 2631-0825-6482, Tejas_Singh123). The left sidebar under 'Amazon S3' lists various bucket types and specific buckets like 'my-store-sales-pipeline-bucket'. A collapsed section for 'Storage Lens' is also visible. The main content area is titled 'Objects' and displays one object: 'productsdata.parquet'. The object details show it is a parquet file, last modified on September 14, 2025, at 22:37:41 (UTC+05:30). The 'Actions' menu includes options for Copy S3 URI, Copy URL, Download, Open, Delete, Create folder, and Upload (which is highlighted).

Name	Type	Last modified	Size
productsdata.parquet	parquet	September 14, 2025, 22:37:41 (UTC+05:30)	2