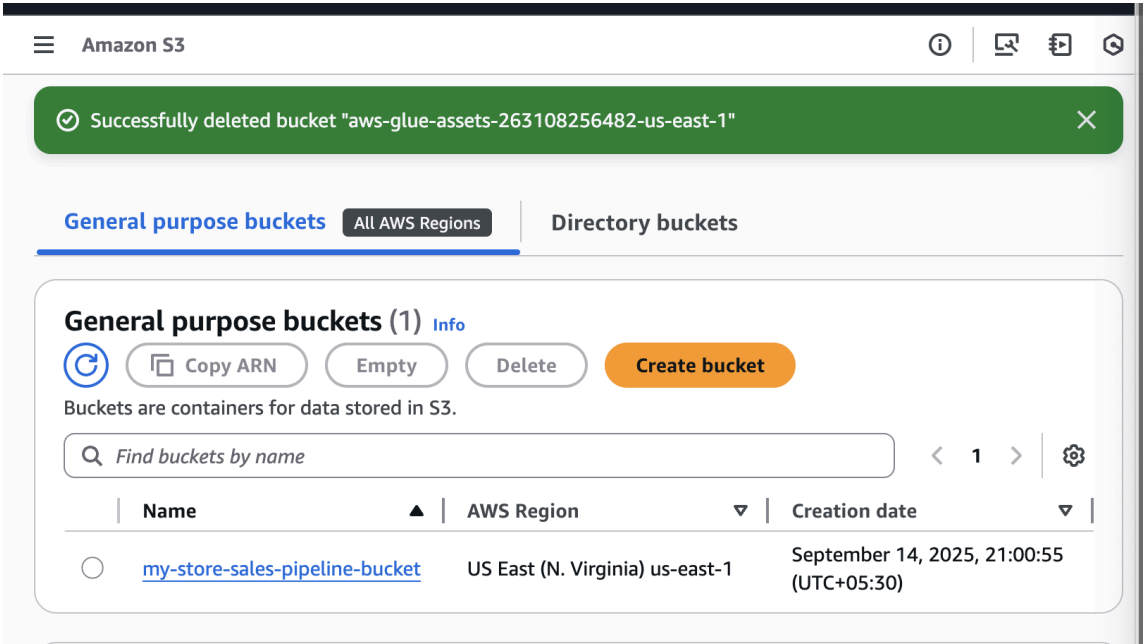


NAME:- Tejas Singh
Emp ID: AS1554

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

my-store-sales-pipeline-bucket

Info

<

Objects

Metadata

Properties

Permissions

Metrics

Management

>

Objects (3)

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

ObjectsProperties

Objects (1)

Refresh

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 >

Settings

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 productsdata.js n	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 displays the Amazon Athena Query Editor interface. At the top, a notification states: "Typeahead suggestions are turned on by default. You can change this setting in query editor preferences." The left sidebar contains a "Data" section with the following settings: "Data source" set to "AwsDataCatalog", "Catalog" set to "None", and "Database" set to "store_sales_db". Below these settings is a "Tables and views" section with a "Create" button and a search bar labeled "Filter tables and views". Under "Tables (1)", the table "sales_parquet_table" is listed. Under "Views (0)", no views are shown. The main editor area shows "Query 1" with the following SQL code:

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

Below the query editor, there are buttons for "Run again", "Explain", "Cancel", "Clear", and "Create". A "Reuse query results" toggle is also present, set to "off", with a note "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

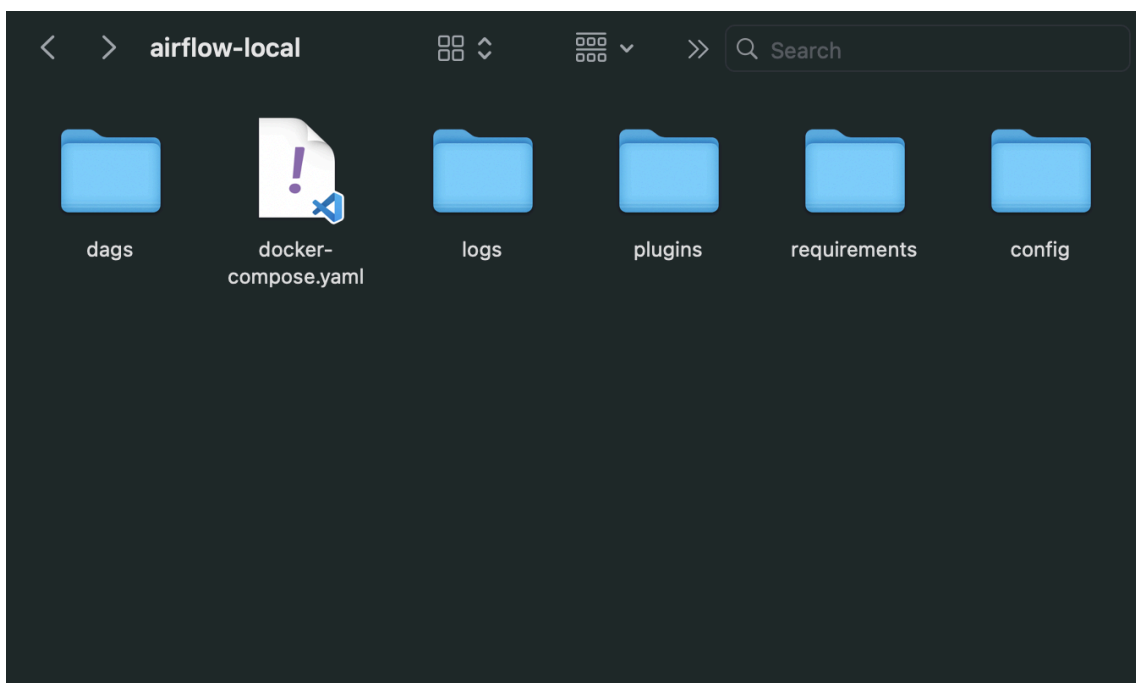
```
airflow-local — zsh — 80x24
Last login: Thu Sep 11 11:20:20 on ttys000
tejassingh@Tejass-MacBook-Air ~ % docker --version
docker compose version

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

% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
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,
```

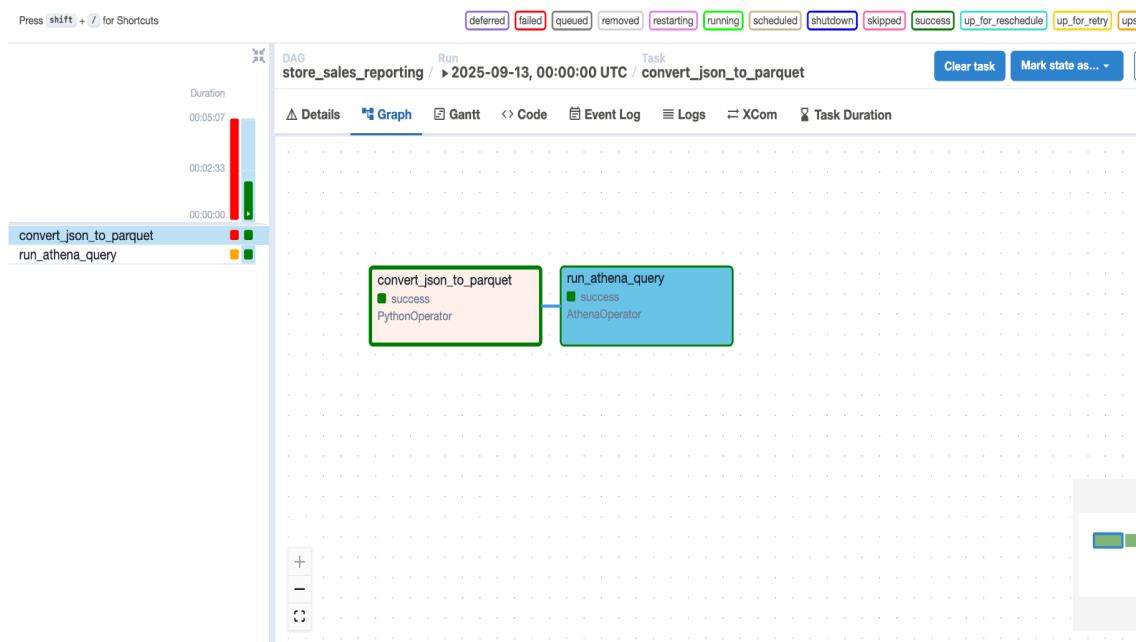


NAME:- Tejas Singh
Emp ID: AS1554

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

aws

United States (N. Virginia)

Account ID: 2631-0825-6482

Tejas_Singh123

Amazon S3

Buckets

my-store-sales-pipeline-bucket

Processed/

Amazon S3

General purpose buckets

Directory buckets

Table buckets

Vector buckets

Access Grants

Access Points (General Purpose Buckets, FSx file systems)

Access Points (Directory Buckets)

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Objects

Properties

Objects (1)

Copy S3 URIDivCopy URIDivDownloadDivOpenDivDeleteDivActionsDivCreate folderDivUpload

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>

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