

ETL Job Execution (Post Terraform Deployment) – Screenshots

- Mira Radhakrishnan

Provisioned my own aws account with the terraform script and documented the outputs.

1. S3 buckets created – source, target and code bucket to have python scripts

General purpose buckets (3) [Info](#) [All AWS Regions](#)

[Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

Buckets are containers for data stored in S3.

	Name	AWS Region	IAM Access Analyzer	Creation date
<input type="radio"/>	my-poc-glue-code-bucket-unique-123	US East (N. Virginia) us-east-1	View analyzer for us-east-1	June 20, 2025, 16:23:12 (UTC-05:00)
<input type="radio"/>	my-poc-source-bucket-unique-123	US East (N. Virginia) us-east-1	View analyzer for us-east-1	June 20, 2025, 16:23:12 (UTC-05:00)
<input type="radio"/>	my-poc-target-bucket-unique-123	US East (N. Virginia) us-east-1	View analyzer for us-east-1	June 20, 2025, 16:23:12 (UTC-05:00)

2. Glue crawlers created one for source and one for target

AWS Glue > Crawlers

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (2) [Info](#) [Last updated \(UTC\) June 20, 2025 at 21:36:00](#) [Action](#) [Run](#) [Create crawler](#)

View and manage all available crawlers.

	Name	State	Schedule	Last run	Last run...	Log	Table cha...
<input type="checkbox"/>	source-user...	Ready		June 20, 20...	View log	1 created	
<input type="checkbox"/>	target-resu...	Ready		June 20, 20...	View log	1 created	

3. Glue IAM role Created – glue_service_role

Identity and Access Management (IAM)

[Search IAM](#)

glue_service_role [Info](#) [Delete](#) [Edit](#)

Summary

Creation date
June 20, 2025, 16:23 (UTC-05:00)

ARN
[arn:aws:iam::474668381167:role/glue_service_role](#)

Last activity
[12 minutes ago](#)

Maximum session duration
1 hour

Permissions [Trust relationships](#) [Tags](#) [Last Accessed](#) [Revoke sessions](#)

Permissions policies (1) [Info](#) [Simulate](#) [Remove](#) [Add permissions](#)

You can attach up to 10 managed policies.

Filter by Type
All types

4. Glue workflow – to systematically run the crawler and etl job

The screenshot displays the AWS Glue console interface. The top navigation bar shows 'AWS Glue' and 'Workflows'. The left sidebar lists various AWS Glue services, with 'Workflows (orchestration)' selected. The main content area shows the 'daily-etl-workflow' workflow graph. The graph consists of four nodes: 'daily-source-crawler-trigger' (Start), 'source-user-crawler' (Crawler), 'source-crawler-success-to-event-transform-job' (Job), and 'event-transform-job' (Job). The workflow is shown in a 'Graph' view, with a 'Legend' indicating the status of each node (Start, Trigger, Job, Crawler, Incomplete, Error, Deleting). Below the graph, the 'Workflow details' tab is selected, showing the workflow's name, description, last run status (Completed), and other properties.

Workflow details:

Name	Description	Max concurrency	Last run status
daily-etl-workflow	-	-	Completed

Last run	Last modified	Blueprint name	Blueprint run id
June 20, 2025 at 21:31:22	June 20, 2025 at 21:23:11	-	-

5. Created Triggers to execute the flow in order

The screenshot displays the AWS Glue console interface, specifically the 'Triggers' page. The left sidebar shows 'AWS Glue' and 'Triggers'. The main content area shows the 'Triggers' page, which lists the triggers for the 'daily-etl-workflow'. The triggers are: 'daily-source-crawler-trigger' (Scheduled), 'etl-job-success-to-target' (Conditional), and 'source-crawler-success-to-event-transform-job' (Conditional). The triggers are shown in a table with columns for Name, Status, Type, Parameters, and Targets.

Triggers (3)

Name	Status	Type	Parameters	Targets
daily-source-crawler-trigger	Activated	Scheduled	At 12:00 AM	1 crawler: source-user-crawler
etl-job-success-to-target	Activated	Conditional	1 condition	1 crawler: target-result-crawler
source-crawler-success-to-event-transform-job	Activated	Conditional	1 condition	1 job: event-transform-job

6. Glue job created and executed

The screenshot displays the AWS Glue console interface, specifically the 'event-transform-job' job details page. The left sidebar shows 'AWS Glue' and 'Jobs'. The main content area shows the 'event-transform-job' job details, including the job's name, ID, start time, end time, Glue version, worker type, and status. The job is shown in a 'Runs' view, with a table of job runs.

Job runs (1/1)

Run details	Input arguments (11)	Logs	Run insights	Metrics	Troubleshooting and analysis
Job name: event-transform-job Id: jr_0a904dbd15203ec527aa9c2e d9621e74746b1e9a3b29a8ac69 cdc4b0ee2e2053	Start time (Local): 06/20/2025 16:29:44 End time (Local): 06/20/2025 16:31:22	Glue version: 4.0 Worker type: G.1X	Last modified on (Local): 06/20/2025 16:31:22 Log group name: /aws-glue/jobs	Run status: Succeeded Start-up time: 14 seconds Max capacity: 2 DPUs	Number of workers: 2

7. Created Catalog tables in catalog database– Source and Target by executing the flow

The screenshot shows the AWS Glue console interface. The top navigation bar includes the AWS logo, a search bar, and user information. The left sidebar shows the navigation menu with categories like 'Getting started', 'ETL jobs', 'Data Catalog tables', and 'Data Catalog'. The main content area is divided into two sections: 'Databases (1)' and 'Tables (2)'. The 'Databases (1)' section shows a table with one database named 'user' with a location URI of 's3://my-poc-source-bucket-uni'. The 'Tables (2)' section shows a table with two tables: 'my_poc_target_' and 'source', both with a location URI of 's3://my-poc-sol' and a classification of 'CSV'. The 'Tables (2)' section also includes a 'View data' link for each table.

Databases (1)

Name	Description	Location URI	Created on (UTC)
user	-	s3://my-poc-source-bucket-uni	June 20, 2025 at 21:23:14

Tables (2)

Name	Database	Location	Classific...	Depreca...	View data	Data quality
my_poc_target_	user	s3://my-poc-tar	CSV	-	Table data	View data qualit
source	user	s3://my-poc-sol	CSV	-	Table data	View data qualit

8. Output loaded to target S3 bucket

The screenshot shows the Amazon S3 console interface. The top navigation bar includes the Amazon S3 logo, a search bar, and user information. The left sidebar shows the navigation menu with categories like 'Directory buckets', 'Table buckets', 'Access Grants', and 'Storage Lens'. The main content area is divided into two sections: 'Objects (1)' and 'Properties'. The 'Objects (1)' section shows a table with one object named 'part-00000-6c551577-f978-4a01-8c60-e2fb6bb67783-' with a size of 371.0 B and a storage class of 'Standard'. The 'Properties' section shows the object's metadata, including its last modified date and time.

Objects (1)

Name	Type	Last modified	Size	Storage class
part-00000-6c551577-f978-4a01-8c60-e2fb6bb67783-	csv	June 20, 2025, 16:31:11 (UTC-05:00)	371.0 B	Standard

9. Source folder after successful execution of the job where archive folder is created and athena result is also captured.

- Directory buckets
- Table buckets
- Access Grants
- Access Points for general purpose buckets
- Access Points for 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

my-poc-source-bucket-unique-123 [Info](#)

[Objects](#) [Metadata](#) [Properties](#) [Permissions](#) [Metrics](#) [Management](#) [Access Points](#)

Objects (2)

🔄 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"/>	archive/	Folder	-	-	-
<input type="checkbox"/>	athena-results/	Folder	-	-	-