

**Name: - Muhammad Huzaifa Waseem (2303-KHI-DEG-021)**

**Pair Partner 1: - Muhammad Faizan Rafique (2303.005.KHI.DEG)**

**Pair Partner 2: - Syed Muhammad Hammad Irshad(2303.KHI.DEG.032)**

## Assignment 5.2:

Output in S3 bucket where new data is after process from the job.

Amazon S3 > Buckets > faizanrafq-glue-data > output\_data/

output\_data/ [Copy S3 URI](#)

Objects Properties

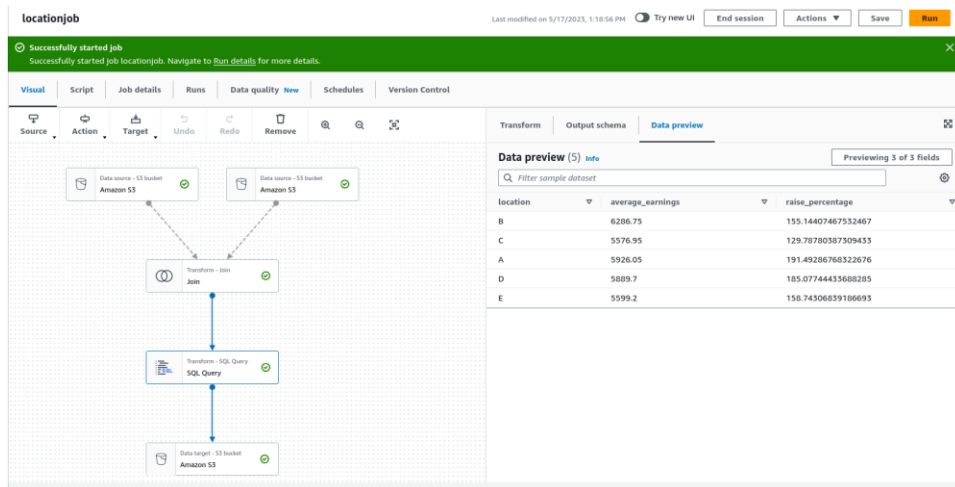
Objects (5)

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](#)

[Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	run-1684311635851-part-block-0-r-0002-snappy.parquet	parquet	May 17, 2023, 15:20:46 (UTC+05:00)	599.0 B	Standard
<input type="checkbox"/>	run-1684311635851-part-block-0-r-00014-snappy.parquet	parquet	May 17, 2023, 15:20:44 (UTC+05:00)	599.0 B	Standard
<input type="checkbox"/>	run-1684311635851-part-block-0-r-00021-snappy.parquet	parquet	May 17, 2023, 15:20:44 (UTC+05:00)	599.0 B	Standard
<input type="checkbox"/>	run-1684311635851-part-block-0-r-00025-snappy.parquet	parquet	May 17, 2023, 15:20:44 (UTC+05:00)	599.0 B	Standard
<input type="checkbox"/>	run-1684311635851-part-block-0-r-00031-snappy.parquet	parquet	May 17, 2023, 15:20:44 (UTC+05:00)	599.0 B	Standard

In this job I created two s3 source having location and earnings then join them with emp id then run sql query for final output.



Crawler for location.

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 (3) [info](#)

View and manage all available crawlers.

Last updated (UTC) May 17, 2023 at 08:25:36 [Action](#) [Run](#) [Create crawler](#)

<input type="checkbox"/>	Name	State	Schedule	Last run	Last run timestamp	Log	Table changes from last r...
<input type="checkbox"/>	faizanrafq-locations-crawler	Ready		Succeeded	May 17, 2023 at 07:40:35	<a href="#">View log</a>	1 created
<input type="checkbox"/>	faizanrafq_rds_employees...	Ready		Failed	May 17, 2023 at 05:19:58	<a href="#">View log</a>	-
<input type="checkbox"/>	faizanrafq_s3_earnings_cr...	Ready		Succeeded	May 17, 2023 at 05:32:58	<a href="#">View log</a>	1 created