

Name: Osama Abdul Razzak(2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: M Humza Moeen(2303.KHI.DEG.019)

Assignment 5.2

Using the earnings CSV as a base, prepare a new data file with employees' office locations. Make sure there are 5-6 distinct locations that are shared between employees.

Create a Glue job that aggregates the data based on the office location to calculate average salaries and raise percentages for these locations.

Pending reply from Instructor



Your answer



Modify your answer

First, we S3 bucket push the csv file init

The screenshot shows the AWS S3 console interface. The top navigation bar includes the AWS logo, Services menu, a search bar, and user information (Global, osama-xloop @ 9074-6904-0602). The left sidebar shows the 'Amazon S3' section with options like Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, IAM Access Analyzer for S3, Block Public Access settings, Storage Lens, Dashboards, AWS Organizations settings, and Feature spotlight.

The main content area displays the 'Account snapshot' section, which is last updated on May 17, 2023. It provides metrics: Total storage (1.6 KB), Object count (9), and Average object size (179.9 B). A link to 'View Storage Lens dashboard' is also present.

Below the snapshot is the 'Buckets (4)' section. It includes a search bar 'Find buckets by name' and buttons for 'Copy ARN', 'Empty', 'Delete', and 'Create bucket'. A table lists the buckets:

Name	AWS Region	Access	Creation date
assignment-glue-data	US East (N. Virginia) us-east-1	Bucket and objects not public	May 18, 2023, 09:08:43 (UTC+05:00)
osama-module4-day5	US East (N. Virginia) us-east-1	Bucket and objects not public	May 18, 2023, 12:40:49 (UTC+05:00)
osama-razzak-glue-data	US East (N. Virginia) us-east-1	Bucket and objects not public	May 17, 2023, 08:25:14 (UTC+05:00)

Name: Osama Abdul Razzak(2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: M Humza Moeen(2303.KHI.DEG.019)

follow instruction as per mention in readme file and and put the csv file into input directory

Objects (3)

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

< 1 >

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	glue_data/	Folder	-	-	-
<input type="checkbox"/>	input_data/	Folder	-	-	-
<input type="checkbox"/>	output_data/	Folder	-	-	-

data = 2023-05-17/

Copy S3 URI

Objects

Properties

Objects (3)

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

< 1 >

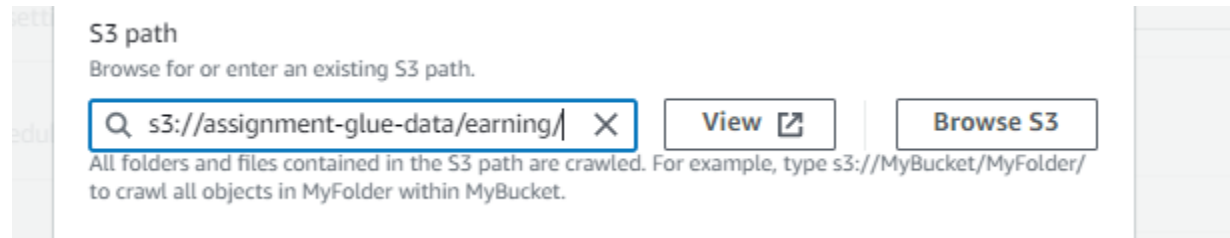
<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	earnings_1.csv	csv	May 18, 2023, 11:51:46 (UTC+05:00)	1.2 KB	Standard
<input type="checkbox"/>	earnings_2.csv	csv	May 18, 2023, 11:51:46 (UTC+05:00)	1.2 KB	Standard
<input type="checkbox"/>	locations.csv	csv	May 18, 2023, 11:51:47 (UTC+05:00)	916.0 B	Standard

Name: Osama Abdul Razzak(2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: M Humza Moeen(2303.KHI.DEG.019)

Now create AWS Glue crawler name 'my_s3_crawler'

Adding the data source

In data source, provide the path of s3 bucket

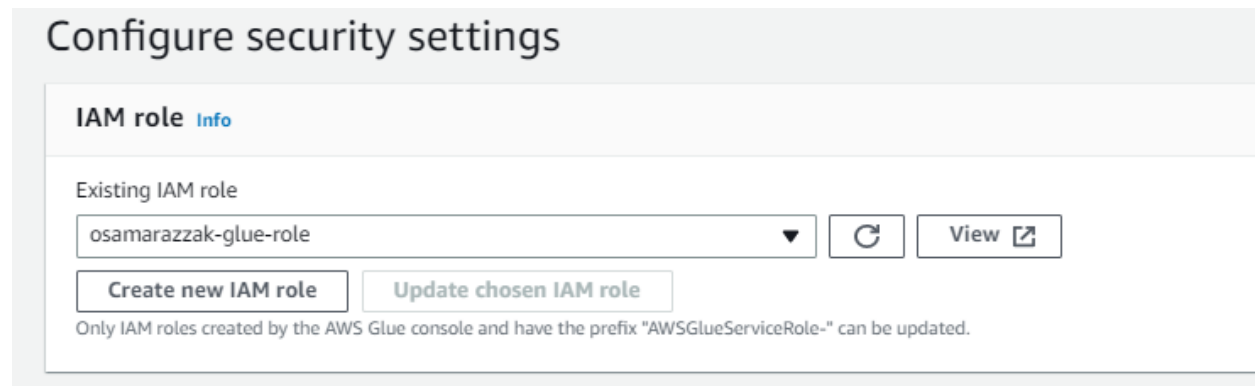


S3 path
Browse for or enter an existing S3 path.

All folders and files contained in the S3 path are crawled. For example, type s3://MyBucket/MyFolder/ to crawl all objects in MyFolder within MyBucket.

[View](#) [Browse S3](#)

Provide the iam role



Configure security settings

IAM role [Info](#)

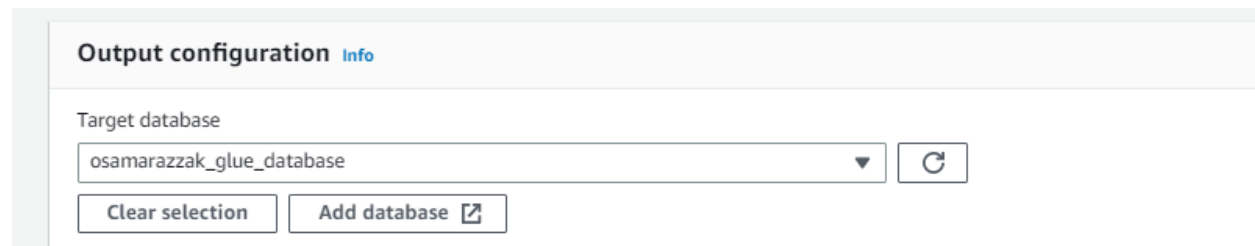
Existing IAM role

[View](#)

[Create new IAM role](#) [Update chosen IAM role](#)

Only IAM roles created by the AWS Glue console and have the prefix "AWSGlueServiceRole-" can be updated.

In output configuration add the target data base



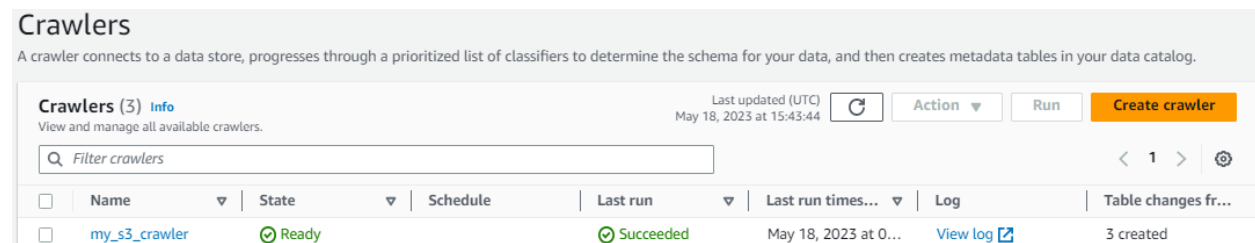
Output configuration

[Info](#)

Target database

[Clear selection](#) [Add database](#)

Successfully create and run the crawler



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 18, 2023 at 15:43:44 [Action](#) [Run](#) [Create crawler](#)

<input type="checkbox"/>	Name	State	Schedule	Last run	Last run times...	Log	Table changes fr...
<input type="checkbox"/>	my_s3_crawler	Ready			May 18, 2023 at 0...	View log	3 created

Name: Osama Abdul Razzak(2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui(2303.KHI.DEG.030)
Peer Name: M Humza Moeen(2303.KHI.DEG.019)

Now, navigate to AWS glue studio and create job 'visual with blank canvas'

AWS Glue > Jobs

AWS Glue Studio [Info](#)

Create job [Info](#)

☐ Visual with a source and target
Start with a source, ApplyMapping transform, and target.

☒ Visual with a blank canvas
Author using an interactive visual interface.

☐ Spark script editor
Write or upload your own Spark code.

☐ Python Shell script editor
Write or upload your own Python shell script.

☐ Jupyter Notebook
Write your own code in a Jupyter Notebook for interactive development.

☐ Ray script editor **New**
Write your own code to run on Ray.

[Create](#)

Your jobs (0) [Info](#)

Job name

Type

Last modified

AWS Glue version

No jobs
You have not created a job yet.
[Create job from a blank graph](#)

Add data source earning and location

Then join it

And run the following query

Untitled job [↗](#) ⚠ Job has not been saved 🔄 Try new UI End session Actions ▾ Save Run

+

Data source - S3 bucket
Amazon S3

✓

Data source - S3 bucket
Amazon S3

✓

Transform - Join
Join

✓

Transform - SQL Query
SQL Query

✓

Visual

Script

Job details

Runs

Data quality **New**

Schedules

Version Control

Transform

Output schema

Data preview

Data preview (5) [Info](#)

location	average_earnings	raise_percentage
B	6286.75	155.14407467532467
C	5576.95	129.78780387309433
A	5926.05	191.49286768322676
D	5889.7	185.07744433688285
E	5599.2	158.74306839186693