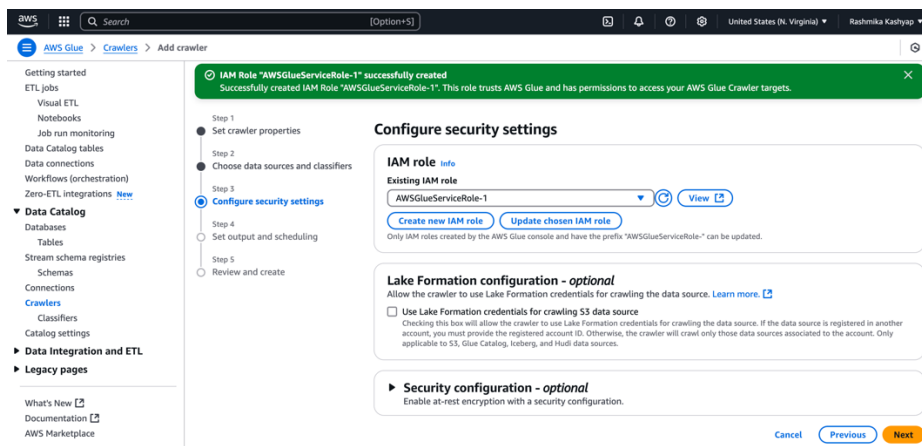
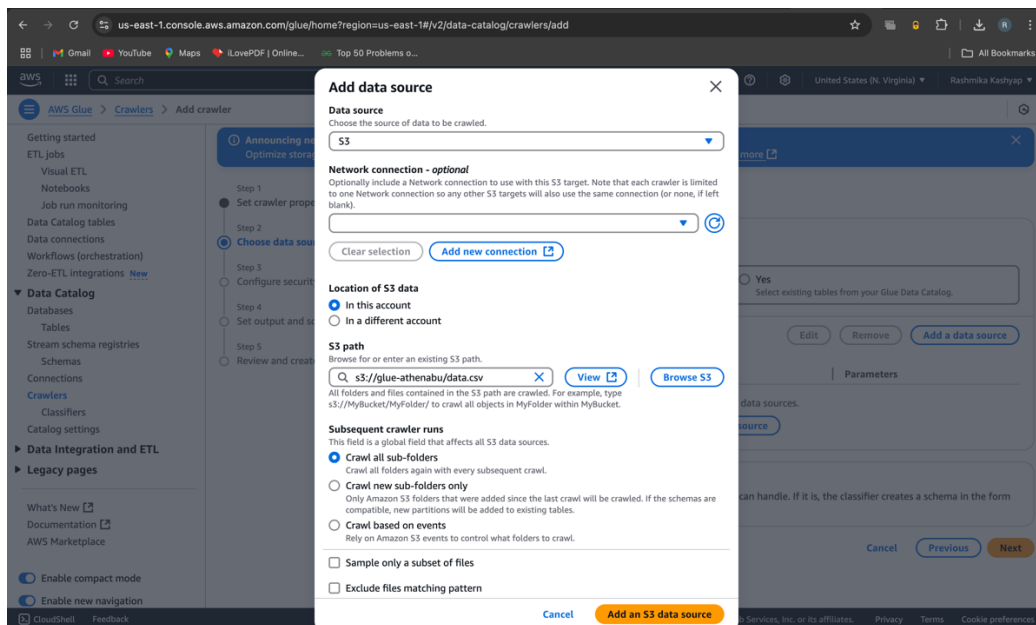
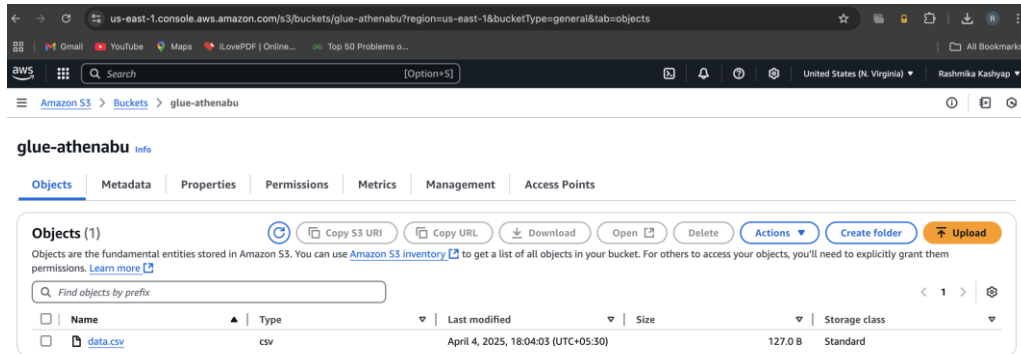


CLOUD LAB ASSIGNMENT 7

E22CSEU0550

RASHMIKA KASHYAP



Get started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations

Zero-ETL integrations

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

CloudShell

Feedback

Step 1

Set crawler properties

Step 2

Choose data sources and classifiers

Step 3

Configure security settings

Step 4

Set output and scheduling

Step 5

Review and create

Set output and scheduling

Output configuration

Target database

glue_db

Clear selection

Add database

Table name prefix - optional

Type a prefix added to table names

Maximum table threshold - optional

This field sets the maximum number of tables the crawler is allowed to generate. In the event that this number is surpassed, the crawl will fail with an error. If not set, the crawler will automatically generate the number of tables depending on the data schema.

Type a number greater than 0

Advanced options

Crawler schedule

You can define a time-based schedule for your crawlers and jobs in AWS Glue. The definition of these schedules uses the Unix-like cron syntax. Learn more

Frequency

On demand

Cancel

Previous

Next

Get started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations

Zero-ETL integrations

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

CloudShell

Feedback

One crawler successfully created

This following crawler is now created: "crawler"

Last updated (UTC)

April 4, 2025 at 12:47:38

Run crawler

Edit

Delete

crawler

Crawler properties

Name

crawler

IAM role

AWSGlueServiceRole-1

Database

glue_db

State

READY

Description

-

Security configuration

-

Lake Formation configuration

-

Table prefix

-

Maximum table threshold

-

Advanced settings

Crawler runs

Schedule

Data sources

Classifiers

Tags

Crawler runs (0)

The list of crawler runs for this crawler.

Filter data

Filter by a date and time range

Start time (UTC)

End time (UTC)

Current/last duration

Status

DPU hours

Table changes

You don't have any crawler runs.

Run crawler

Stop run

View CloudWatch logs

View run details

Get started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations

Zero-ETL integrations

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

CloudShell

Feedback

Standard AWS Glue table (default)

Create a standard AWS Glue table.

Apache Iceberg table

Create a table in Apache Iceberg table format.

Data store

Select the type of source

S3

Choose S3 path

S3 buckets

glue-athenabu

Objects (1/1)

Find object by prefix

Key

Last modified

Size

data.csv

April 4, 2025 at 12:34:03

127 Bytes

Cancel

Choose

Data format

Classification

Choose the format of the data in your table.

Avro

CSV

JSON

XML

Parquet

ORC

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations New

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

☒ Enable compact mode

☒ Enable new navigation

CloudShell Feedback

Set table properties

Step 2

Choose or define schema

Step 3

Review and create

Table details

Name

gluetable

If you plan to access the table from Amazon Athena, then the name should be under 256 characters and contain only lowercase letters (a-z), numbers (0-9), and underscore (_). For more information, see [Athena names](#).

Database

gluedb

Create database

Description - optional

Enter a description

Descriptions can be up to 2048 characters long.

Table format

Data Catalog managed tables support data optimization for Apache Iceberg table type. [Learn more](#)

☒ Standard AWS Glue table (default)

Create a standard AWS Glue table.

☐ Apache Iceberg table

Create a table in Apache Iceberg table format.

Data store

Select the type of source

☒ S3

☐ Kinesis

☐ Kafka

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations New

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

☒ Enable compact mode

☒ Enable new navigation

CloudShell Feedback

Standard AWS Glue table (default)

Create a standard AWS Glue table.

Apache Iceberg table

Create a table in Apache Iceberg table format.

Data store

Select the type of source

☒ S3

☐ Kinesis

☐ Kafka

Data location is specified in

☒ my account

☐ another account

Include path

s3://glue-athenabu/data.csv/

View

Browse S3

Path must be in the form s3://bucket/prefix/. It must end with a slash (/) and not include any files.

Data format

Classification

Choose the format of the data in your table.

☐ Avro

☒ CSV

☐ JSON

☐ XML

☐ Parquet

☐ ORC

Delimiter

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations New

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

☒ Enable compact mode

☒ Enable new navigation

CloudShell Feedback

Step 1

Set table properties

Step 2

Choose or define schema

Step 3

Review and create

Choose or define schema

Schema

☒ Define or upload schema

Manually define schema

☐ Choose from Glue Schema Registry

Select existing schema from your Glue Schema Registry.

Schema (1/3)

View and manage the table schema.

Filter schemas

#	Column name	Data type	Partition key	Comment
1	Name	string	-	-
2	salary	string	-	-
3	DepNo	string	-	-

Partition indexes - optional (0)

AWS Glue partition indexes are an important configuration to reduce overall data transfers and processing, and reduce query processing time. A maximum of 3 partition indexes can be created on a given table.

Index name	Index keys	Index status
No indexes		
Provide indexes		
Add indexes 		

Cancel Previous Next

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

QWS

Search

[Option+S]

United States (N. Virginia)

Rashmika Kashyap

Amazon Athena

Query editor

Manage settings

Manage settings

Query result location and encryption

Location of query result - optional

Enter an S3 prefix in the current region where the query result will be saved as an object.

s3://glue-athena-outputbu

View

Browse S3

You can create and manage lifecycle rules for this bucket

Use Amazon S3 lifecycle rules to store your query results and metadata cost effectively or to delete them after a period of time. Learn more

Lifecycle configuration

Expected bucket owner - optional

Specify the AWS account ID that you expect to be the owner of your query results output location bucket.

Enter AWS account ID

Assign bucket owner full control over query results

Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

Encrypt query results

Cancel

Save

QWS

Search

[Option+S]

United States (N. Virginia)

Rashmika Kashyap

Amazon Athena

Query editor

Editor

Recent queries

Saved queries

Settings

Workgroup

primary

Data

Data source

AwsDataCatalog

Catalog

None

Database

gluedb

Tables and views

Create

Filter tables and views

Tables (1)

glue-table

name

string

salary

string

deptno

string

Views (0)

Query 1

SELECT * FROM glue-table;

SQL

Ln 1, Col 31

Run again

Explain

Cancel

Clear

Create

Reuse query results

up to 60 minutes ago

Query results

Query stats

Completed

Time in queue: 123 ms

Run time: 331 ms

Data scanned: -

Results (0)

Copy

Download results CSV

aws

Search

[Option+S]

United States (N. Virginia)

Rashmika Kashyap

Amazon S3

Buckets

glue-athena-outputbu

glue-athena-outputbu

Info

Objects

Metadata

Properties

Permissions

Metrics

Management

Access Points

Objects (1)

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

Name

Type

Last modified

Size

Storage class

Unsaved/

Folder