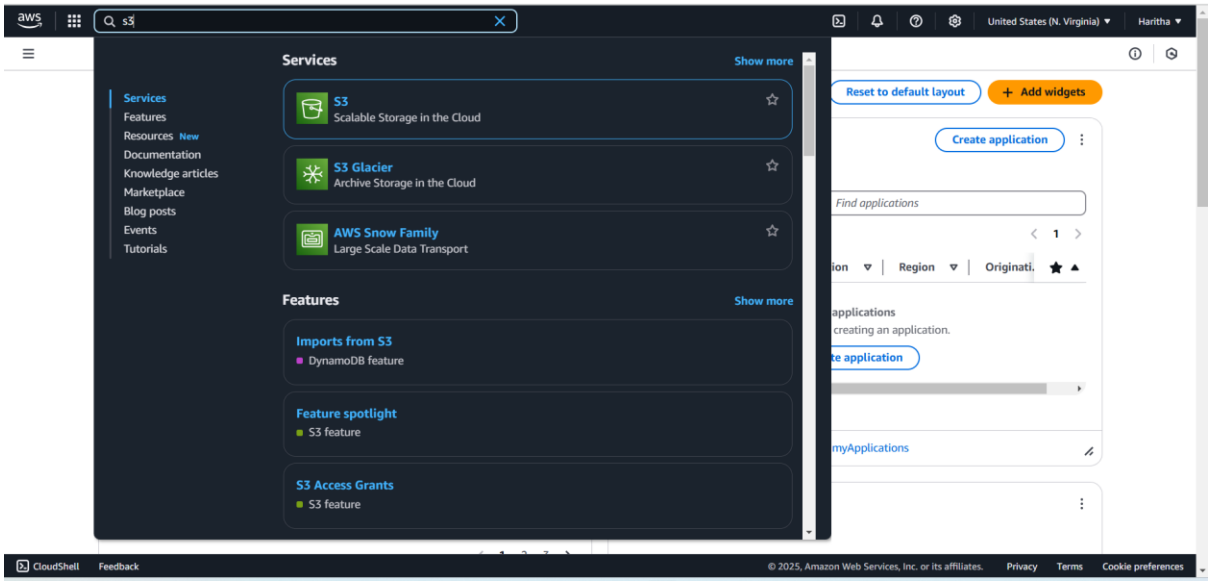
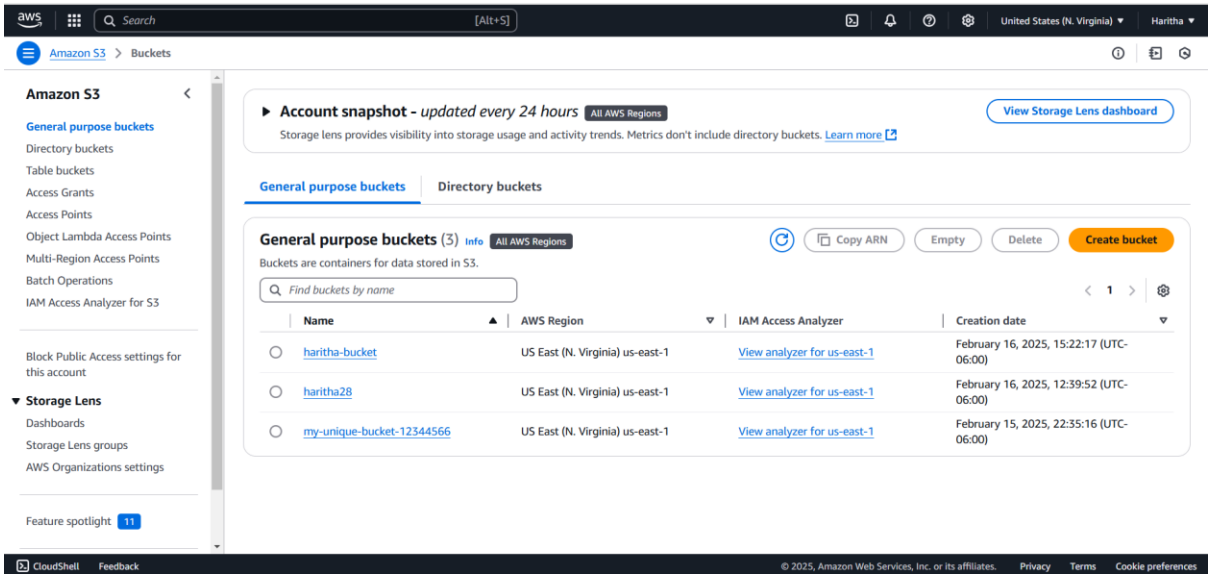
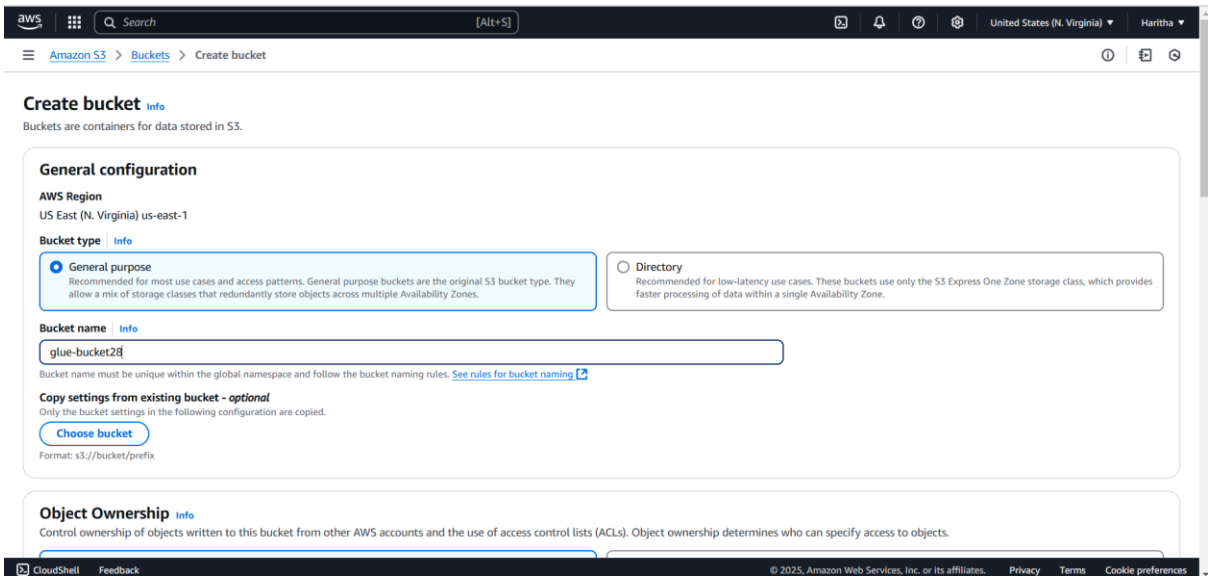
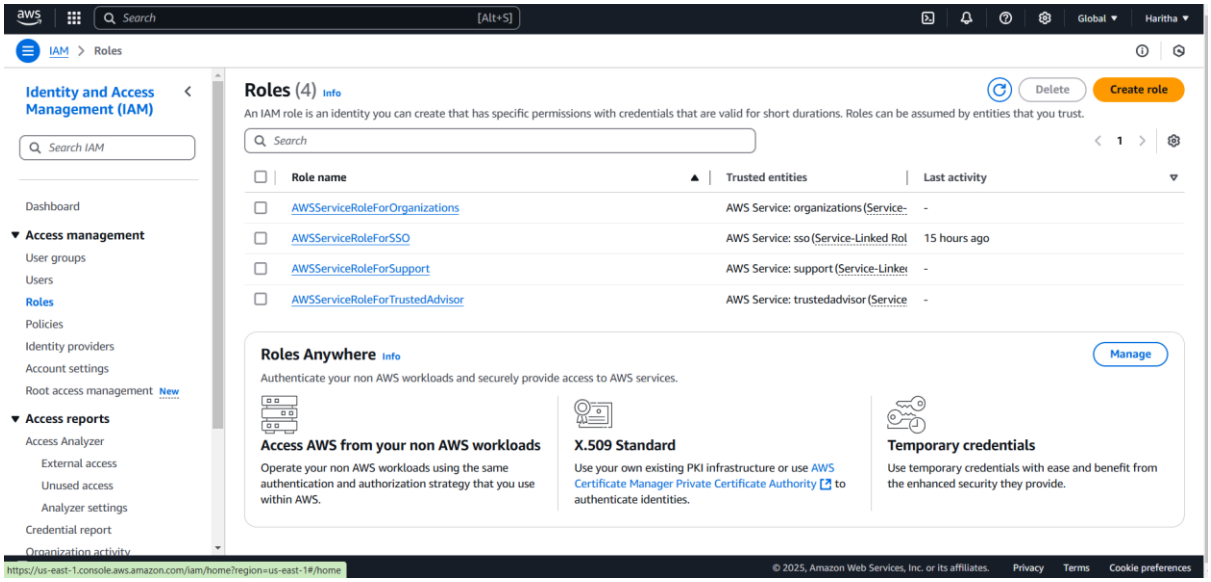
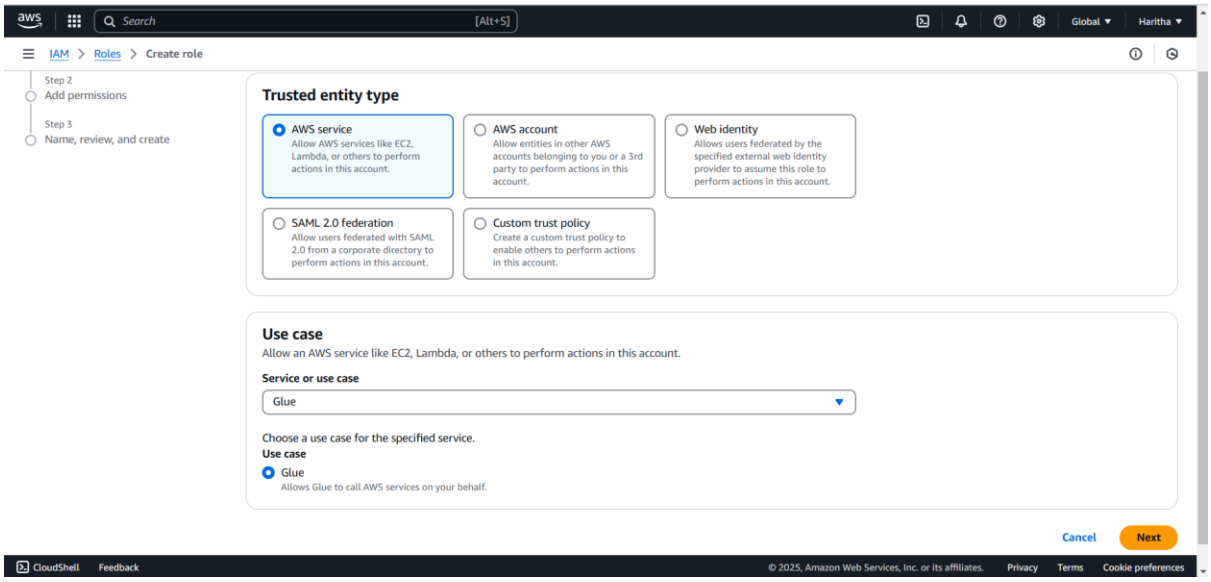
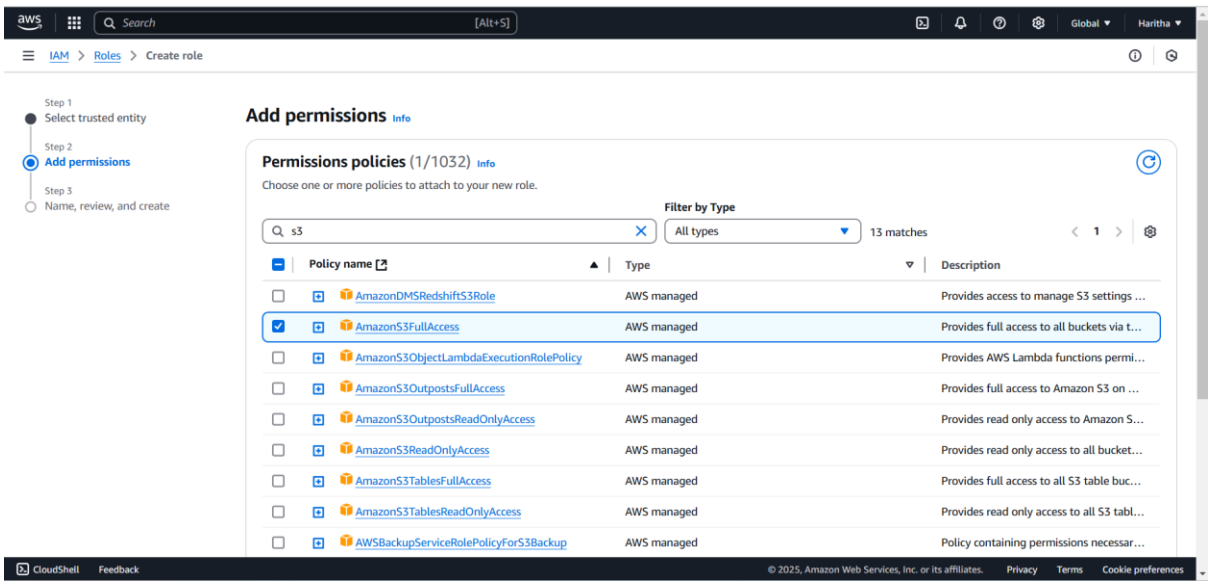

s3
United States (N. Virginia)
Haritha
Services
Show more
Services
Features
Resources New
Documentation
Knowledge articles
Marketplace
Blog posts
Events
Tutorials
S3
Scalable Storage in the Cloud
S3 Glacier
Archive Storage in the Cloud
AWS Snow Family
Large Scale Data Transport
Features
Show more
Imports from S3
DynamoDB feature
Feature spotlight
S3 feature
S3 Access Grants
S3 feature
Reset to default layout
Add widgets
Create application
Find applications
Region
creating an application.
myApplications
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

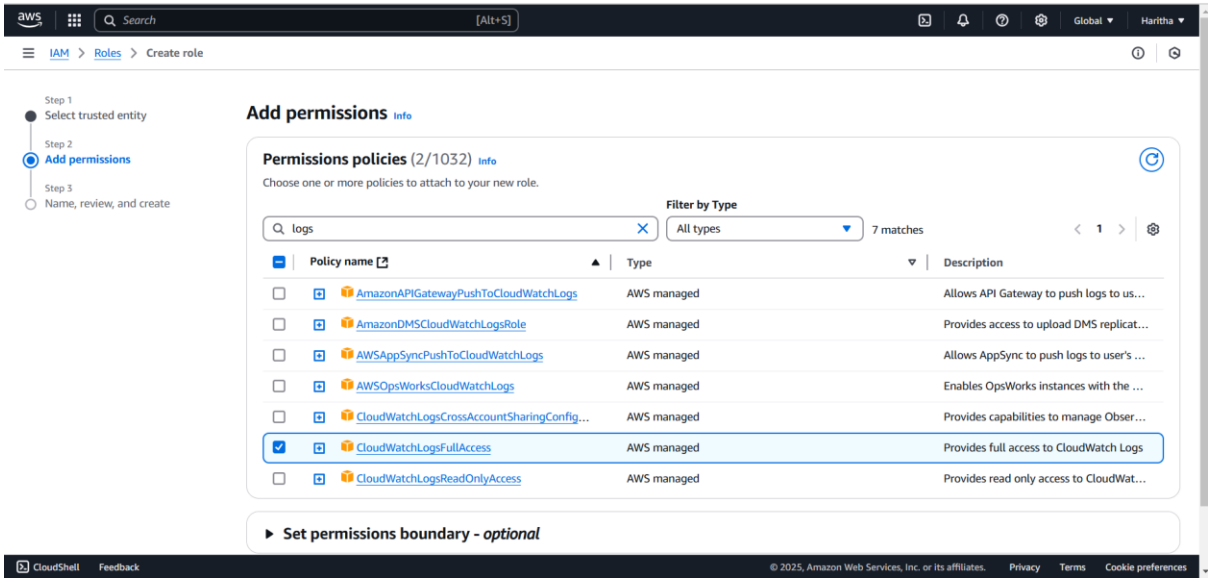

Search
[Alt+S]
United States (N. Virginia)
Haritha
Amazon S3
Buckets
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 11
Account snapshot - updated every 24 hours
All AWS Regions
Storage lens provides visibility into storage usage and activity trends. Metrics don't include directory buckets. Learn more
View Storage Lens dashboard
General purpose buckets
Directory buckets
General purpose buckets (3) Info
All AWS Regions
Buckets are containers for data stored in S3.
Copy ARN
Empty
Delete
Create bucket
Find buckets by name
Name
AWS Region
IAM Access Analyzer
Creation date
haritha-bucket
US East (N. Virginia) us-east-1
View analyzer for us-east-1
February 16, 2025, 15:22:17 (UTC-06:00)
haritha28
US East (N. Virginia) us-east-1
View analyzer for us-east-1
February 16, 2025, 12:39:52 (UTC-06:00)
my-unique-bucket-12344566
US East (N. Virginia) us-east-1
View analyzer for us-east-1
February 15, 2025, 22:35:16 (UTC-06:00)
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

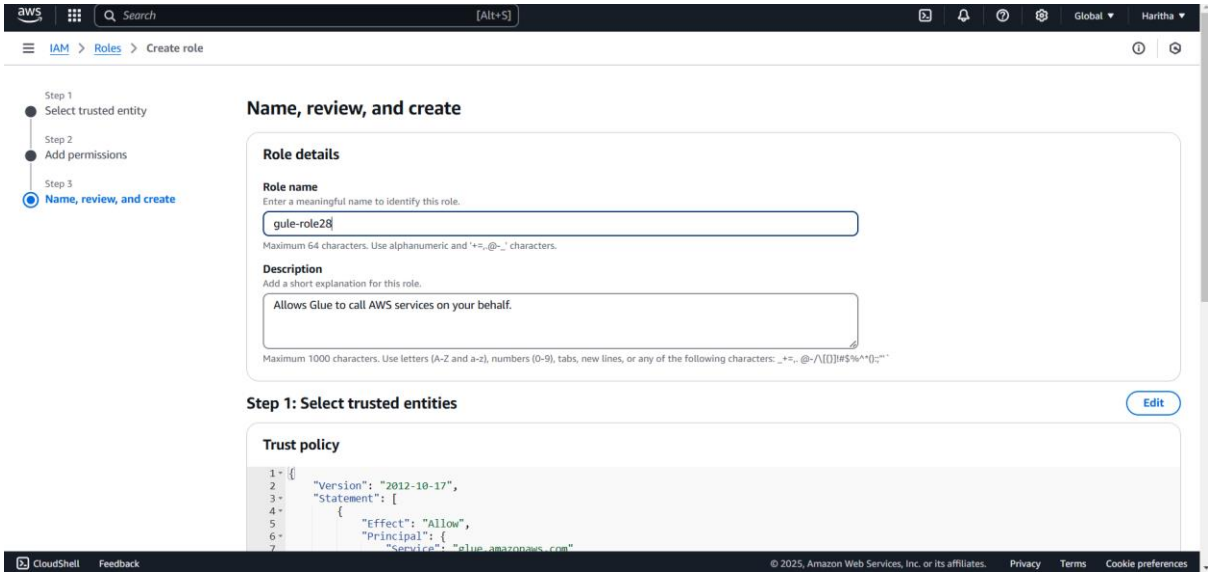

Search
[Alt+S]
United States (N. Virginia)
Haritha
Amazon S3
Buckets
Create bucket
Create bucket Info
Buckets are containers for data stored in S3.
General configuration
AWS Region
US East (N. Virginia) us-east-1
Bucket type Info
General purpose
Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.
Directory
Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.
Bucket name Info
glue-bucket28
Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming
Copy settings from existing bucket - optional
Only the bucket settings in the following configuration are copied.
Choose bucket
Format: s3://bucket/prefix
Object Ownership Info
Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

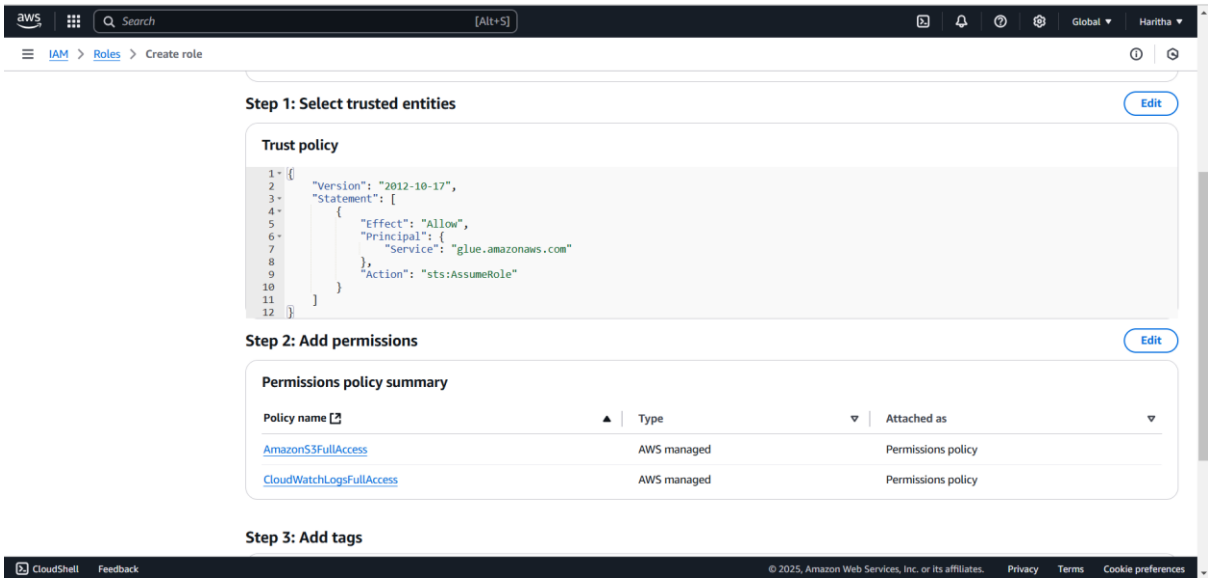

Search [Alt+S]
Global
Haritha
IAM > Roles
Identity and Access Management (IAM)
Search IAM
Dashboard
Access management
User groups
Users
Roles
Policies
Identity providers
Account settings
Root access management New
Access reports
Access Analyzer
External access
Unused access
Analyzer settings
Credential report
Organization activity
Roles (4) Info
An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.
Delete
Create role
Search
Role name
Trusted entities
Last activity
AWSServiceRoleForOrganizations
AWS Service: organizations (Service-
-
AWSServiceRoleForSSO
AWS Service: sso (Service-Linked Rol
15 hours ago
AWSServiceRoleForSupport
AWS Service: support (Service-Linke
-
AWSServiceRoleForTrustedAdvisor
AWS Service: trustedadvisor (Service
-
Roles Anywhere Info
Manage
Authenticate your non AWS workloads and securely provide access to AWS services.
Access AWS from your non AWS workloads
Operate your non AWS workloads using the same authentication and authorization strategy that you use within AWS.
X.509 Standard
Use your own existing PKI infrastructure or use AWS Certificate Manager Private Certificate Authority to authenticate identities.
Temporary credentials
Use temporary credentials with ease and benefit from the enhanced security they provide.
https://us-east-1.console.aws.amazon.com/iam/home?region=us-east-1#/home
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

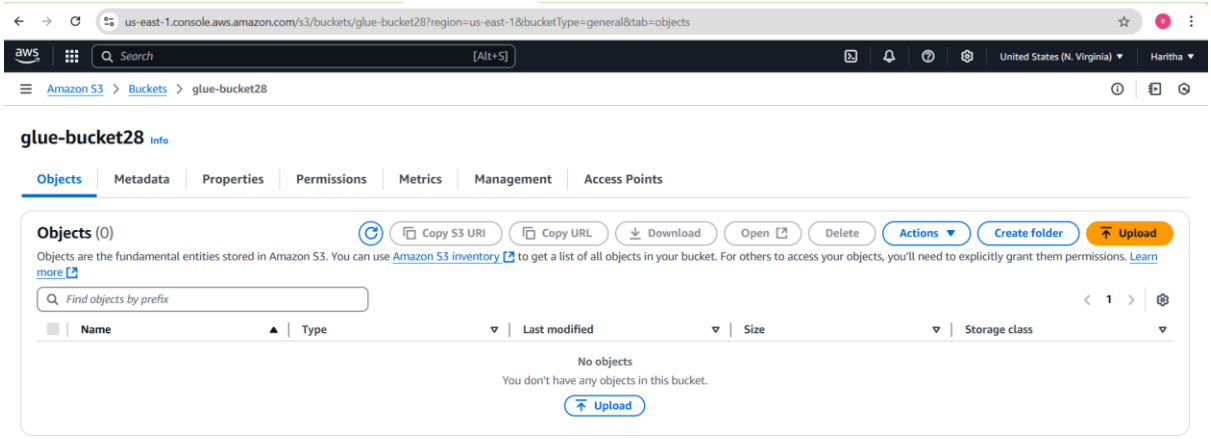

Search [Alt+S]
Global
Haritha
IAM > Roles > Create role
Step 2
Add permissions
Step 3
Name, review, and create
Trusted entity type
AWS service
Allow AWS services like EC2, Lambda, or others to perform actions in this account.
AWS account
Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.
Web identity
Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.
SAML 2.0 federation
Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.
Custom trust policy
Create a custom trust policy to enable others to perform actions in this account.
Use case
Allow an AWS service like EC2, Lambda, or others to perform actions in this account.
Service or use case
Glue
Choose a use case for the specified service.
Use case
Glue
Allows Glue to call AWS services on your behalf.
Cancel
Next
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

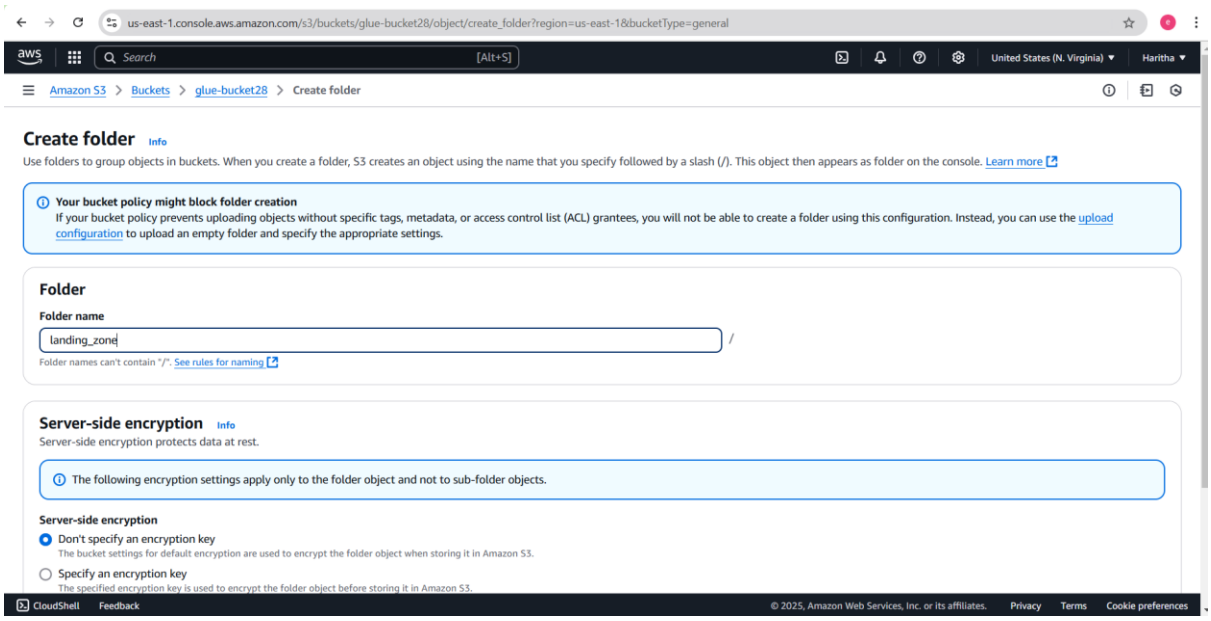

Search [Alt+S]
Global
Haritha
IAM > Roles > Create role
Step 1
Select trusted entity
Step 2
Add permissions
Step 3
Name, review, and create
Add permissions Info
Permissions policies (1/1032) Info
Choose one or more policies to attach to your new role.
Filter by Type
s3
All types
13 matches
Policy name
Type
Description
AmazonDMSRedshiftS3Role
AWS managed
Provides access to manage S3 settings ...
AmazonS3FullAccess
AWS managed
Provides full access to all buckets via t...
AmazonS3ObjectLambdaExecutionRolePolicy
AWS managed
Provides AWS Lambda functions permi...
AmazonS3OutpostsFullAccess
AWS managed
Provides full access to Amazon S3 on ...
AmazonS3OutpostsReadOnlyAccess
AWS managed
Provides read only access to Amazon S...
AmazonS3ReadOnlyAccess
AWS managed
Provides read only access to all bucket...
AmazonS3TablesFullAccess
AWS managed
Provides full access to all S3 table buc...
AmazonS3TablesReadOnlyAccess
AWS managed
Provides read only access to all S3 tabl...
AWSBackupServiceRolePolicyForS3Backup
AWS managed
Policy containing permissions necessar...
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

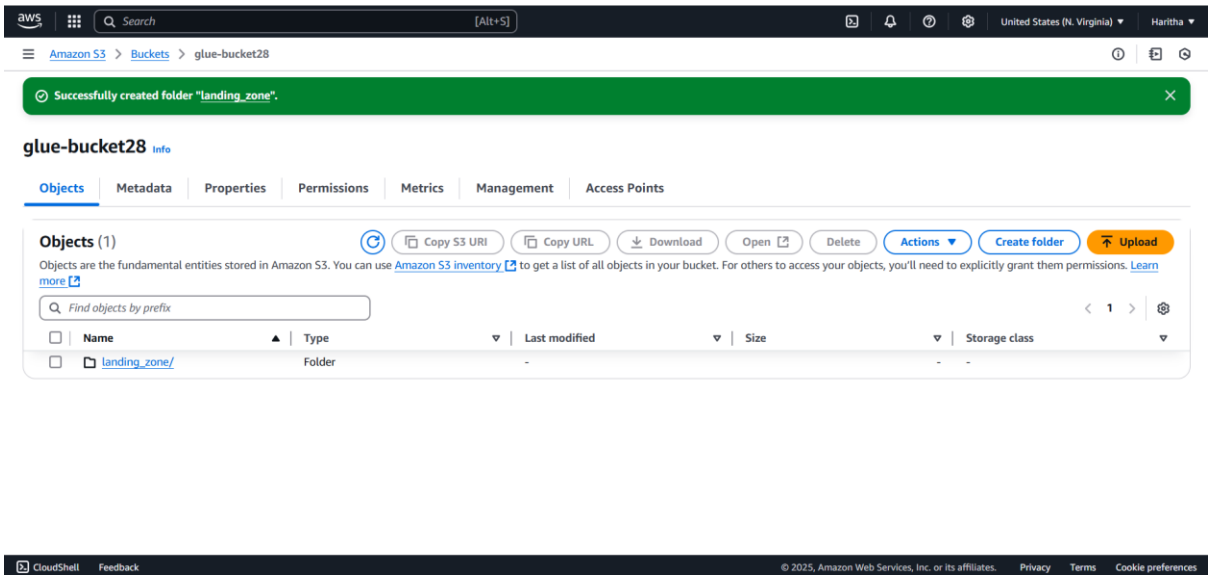

Search [Alt+S]
Global
Haritha
IAM > Roles > Create role
Step 1
Select trusted entity
Step 2
Add permissions
Step 3
Name, review, and create
Add permissions Info
Permissions policies (2/1032) Info
Choose one or more policies to attach to your new role.
Filter by Type
logs
All types
7 matches
Policy name
Type
Description
AmazonAPIGatewayPushToCloudWatchLogs
AWS managed
Allows API Gateway to push logs to us...
AmazonDMSCloudWatchLogsRole
AWS managed
Provides access to upload DMS replicat...
AWSAppSyncPushToCloudWatchLogs
AWS managed
Allows AppSync to push logs to user's ...
AWSOpsWorksCloudWatchLogs
AWS managed
Enables OpsWorks instances with the ...
CloudWatchLogsCrossAccountSharingConfig...
AWS managed
Provides capabilities to manage Obser...
CloudWatchLogsFullAccess
AWS managed
Provides full access to CloudWatch Logs
CloudWatchLogsReadOnlyAccess
AWS managed
Provides read only access to CloudWat...
Set permissions boundary - optional
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

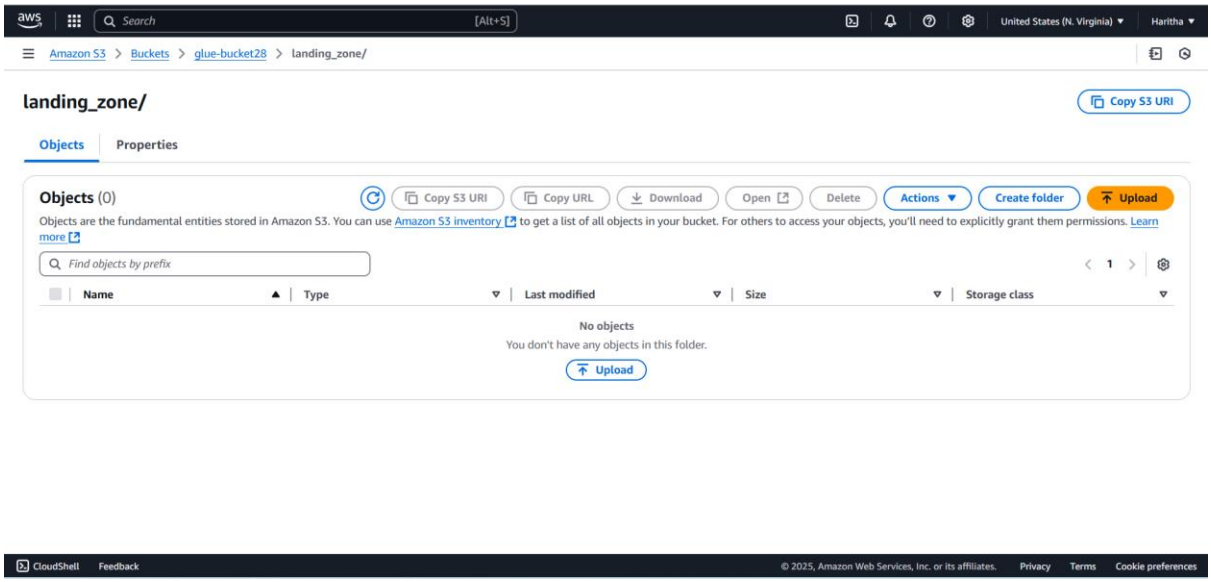

Search [Alt+S]
Global
Haritha
IAM > Roles > Create role
Step 1
Select trusted entity
Step 2
Add permissions
Step 3
Name, review, and create
Name, review, and create
Role details
Role name
Enter a meaningful name to identify this role.
gule-role28
Maximum 64 characters. Use alphanumeric and '+=,.@-_' characters.
Description
Add a short explanation for this role.
Allows Glue to call AWS services on your behalf.
Maximum 1000 characters. Use letters (A-Z and a-z), numbers (0-9), tabs, new lines, or any of the following characters: _+=,. @-/\[{}]!#$%^*();:"'`
Step 1: Select trusted entities
Edit
Trust policy
{
"Version": "2012-10-17",
"Statement": [
{
"Effect": "Allow",
"Principal": {
"Service": "glue.amazonaws.com"
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

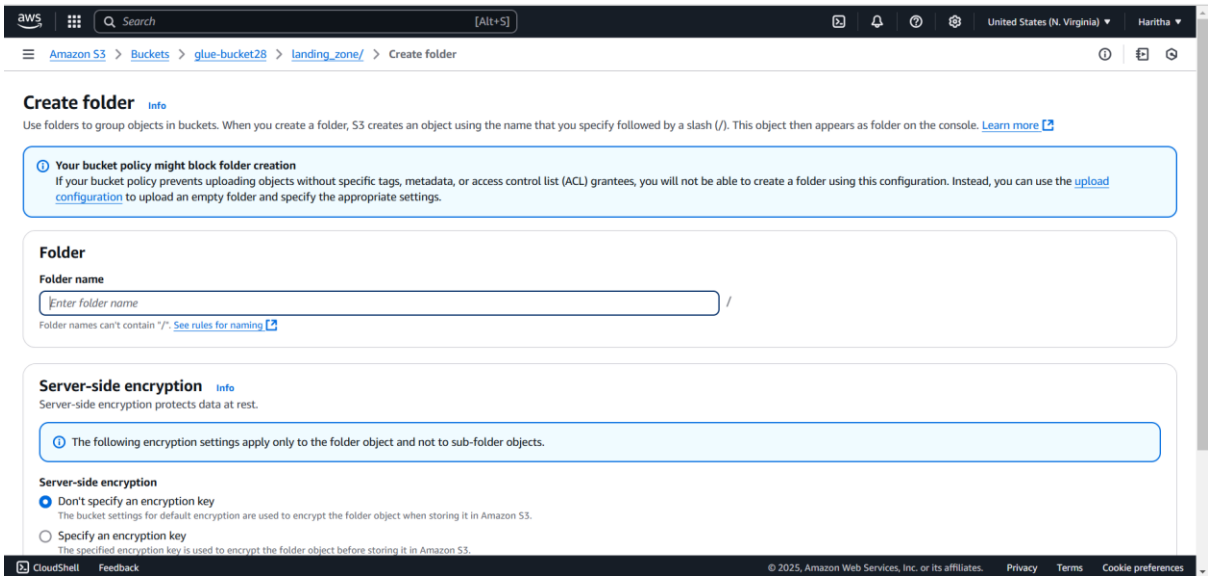

Search [Alt+S]
Global
Haritha
IAM > Roles > Create role
Step 1: Select trusted entities
Edit
Trust policy
{
"Version": "2012-10-17",
"Statement": [
{
"Effect": "Allow",
"Principal": {
"Service": "glue.amazonaws.com"
},
"Action": "sts:AssumeRole"
}
]
}
Step 2: Add permissions
Edit
Permissions policy summary
Policy name
Type
Attached as
AmazonS3FullAccess
AWS managed
Permissions policy
CloudWatchLogsFullAccess
AWS managed
Permissions policy
Step 3: Add tags
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

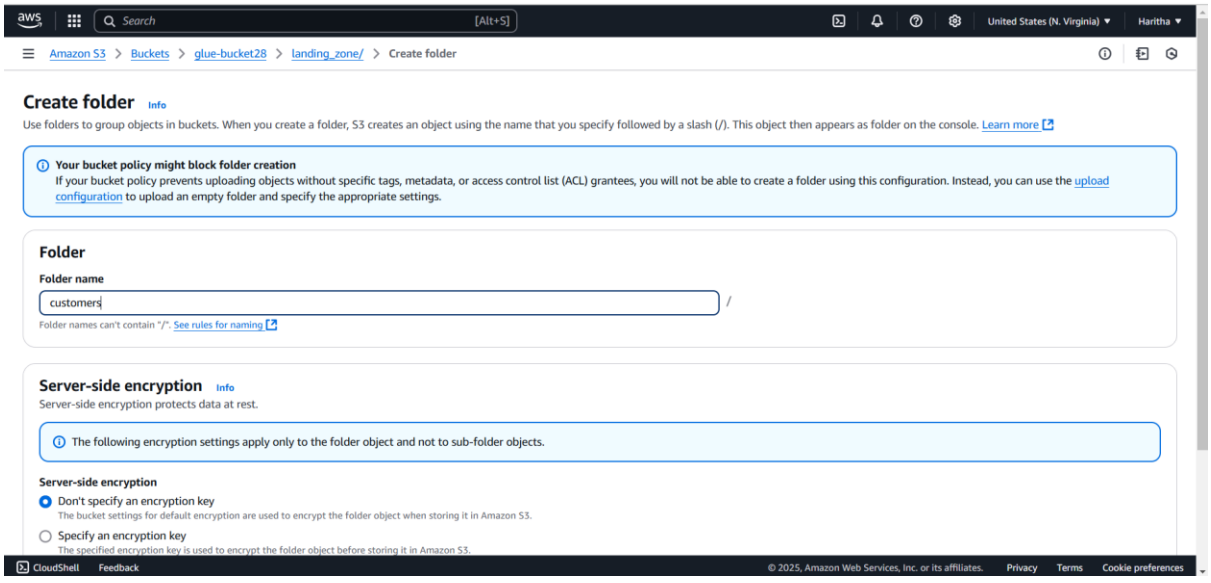

us-east-1.console.aws.amazon.com/s3/buckets/glue-bucket28?region=us-east-1&bucketType=general&tab=objects
Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28
glue-bucket28 Info
Objects Metadata Properties Permissions Metrics Management Access Points
Objects (0)
Copy 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
Name Type Last modified Size Storage class
No objects
You don't have any objects in this bucket.
Upload
CloudShell Feedback
© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

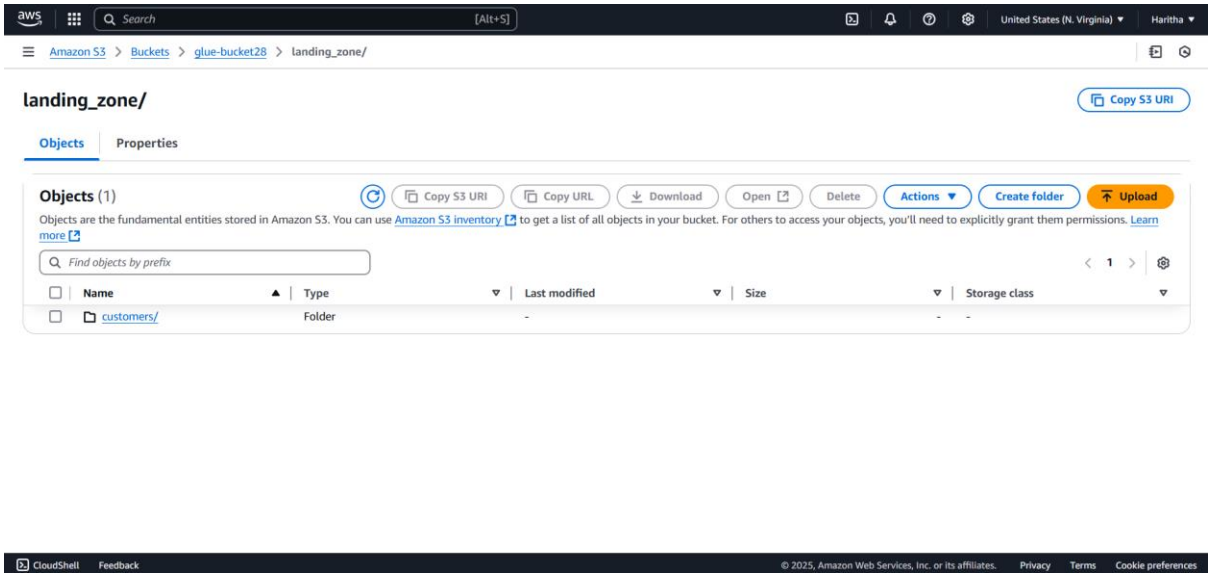

us-east-1.console.aws.amazon.com/s3/buckets/glue-bucket28/object/create_folder?region=us-east-1&bucketType=general
Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > Create folder
Create folder Info
Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more
Your bucket policy might block folder creation
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.
Folder
Folder name
landing_zone
/
Folder names can't contain "/". See rules for naming
Server-side encryption Info
Server-side encryption protects data at rest.
The following encryption settings apply only to the folder object and not to sub-folder objects.
Server-side encryption
Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.
Specify an encryption key
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.
CloudShell Feedback
© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

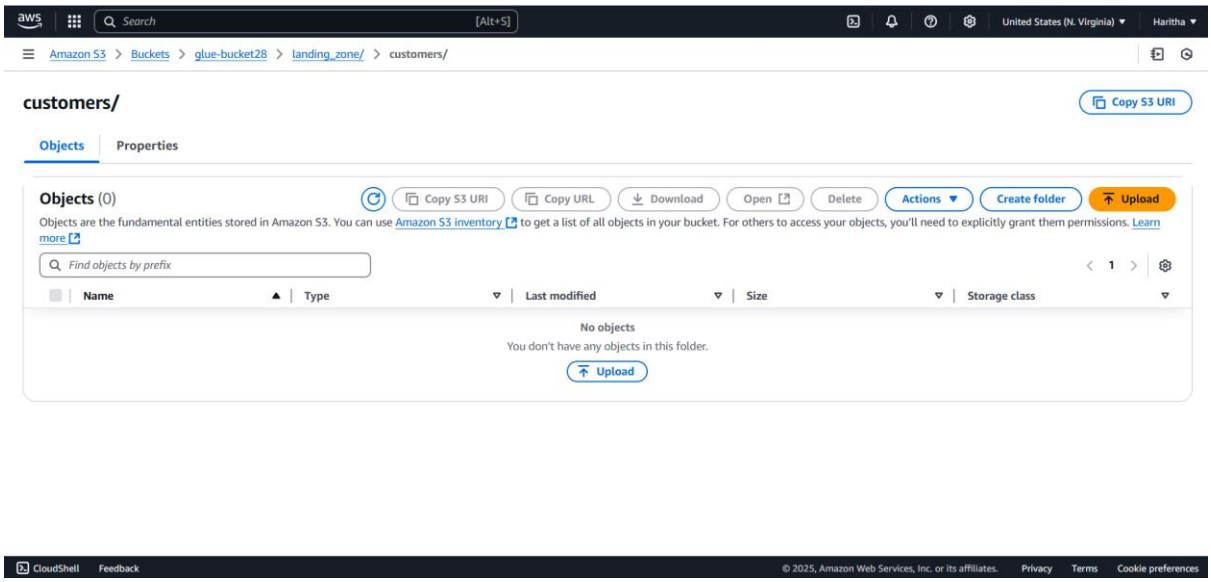

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28
Successfully created folder "landing_zone".
glue-bucket28 Info
Objects Metadata Properties Permissions Metrics Management Access Points
Objects (1)
Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload
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
Name Type Last modified Size Storage class
landing_zone/ Folder - - -
CloudShell Feedback
© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

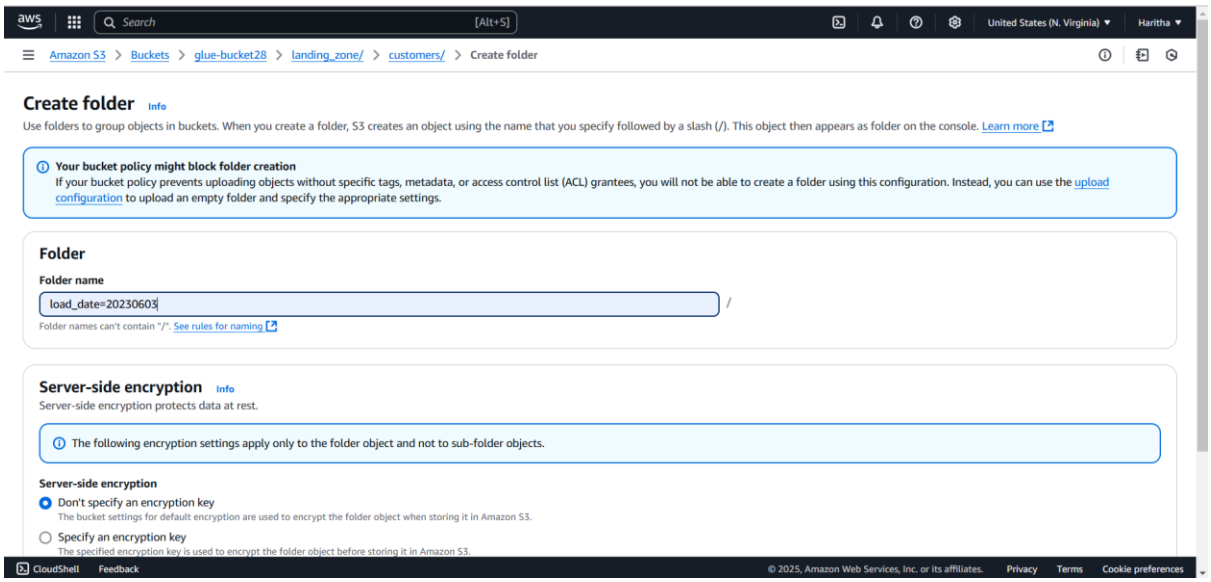

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/
landing_zone/
Copy S3 URI
Objects
Properties
Objects (0)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You can use 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
Name
Type
Last modified
Size
Storage class
No objects
You don't have any objects in this folder.
Upload
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

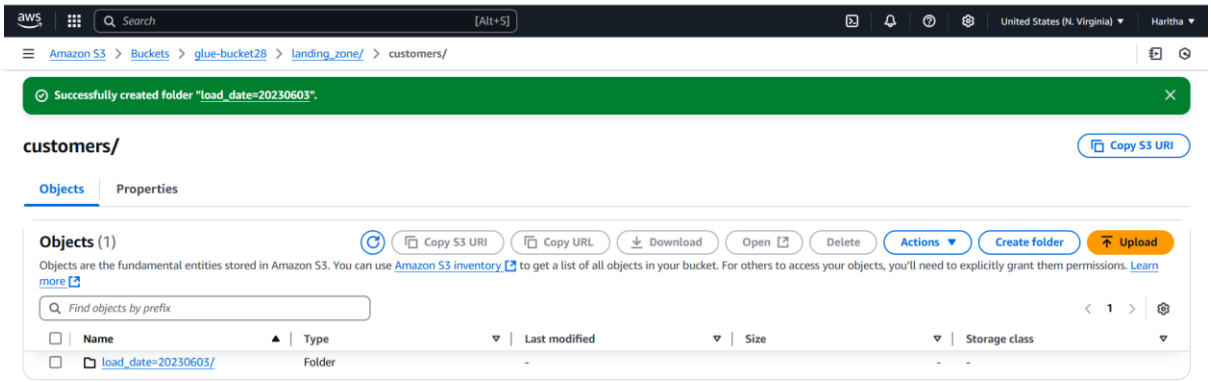

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > Create folder
Create folder
Info
Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more
Your bucket policy might block folder creation
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.
Folder
Folder name
Enter folder name
/
Folder names can't contain "/". See rules for naming
Server-side encryption
Info
Server-side encryption protects data at rest.
The following encryption settings apply only to the folder object and not to sub-folder objects.
Server-side encryption
Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.
Specify an encryption key
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

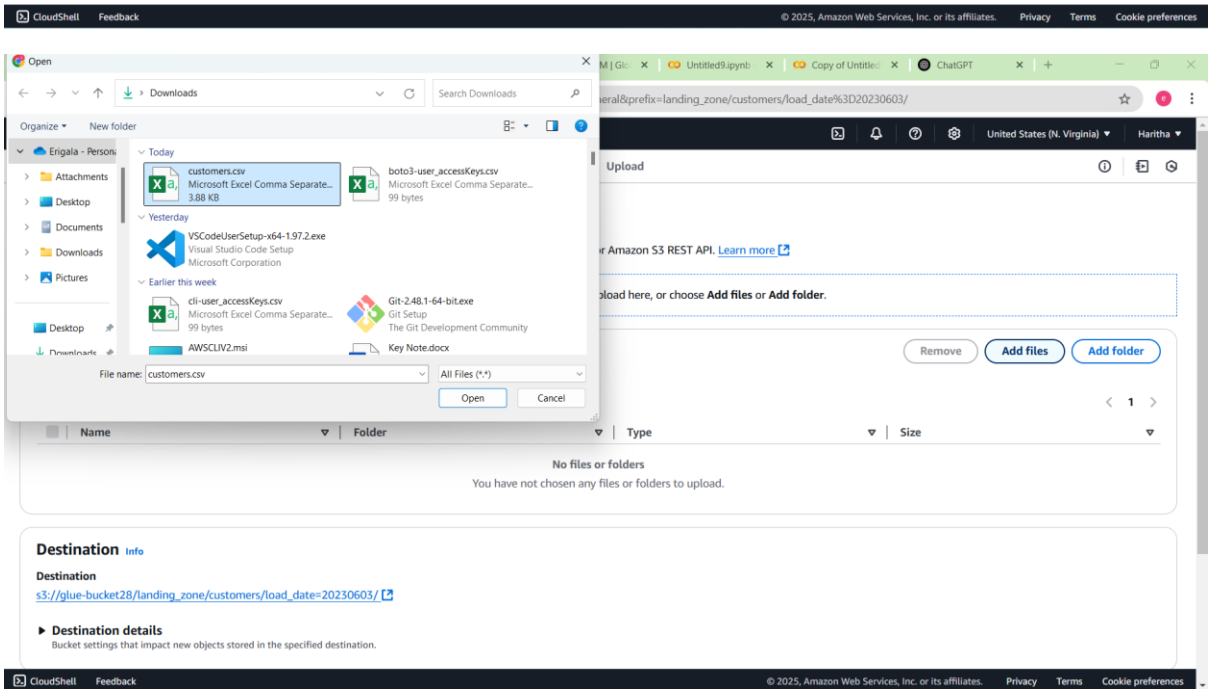

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > Create folder
Create folder
Info
Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more
Your bucket policy might block folder creation
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.
Folder
Folder name
customers
/
Folder names can't contain "/". See rules for naming
Server-side encryption
Info
Server-side encryption protects data at rest.
The following encryption settings apply only to the folder object and not to sub-folder objects.
Server-side encryption
Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.
Specify an encryption key
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

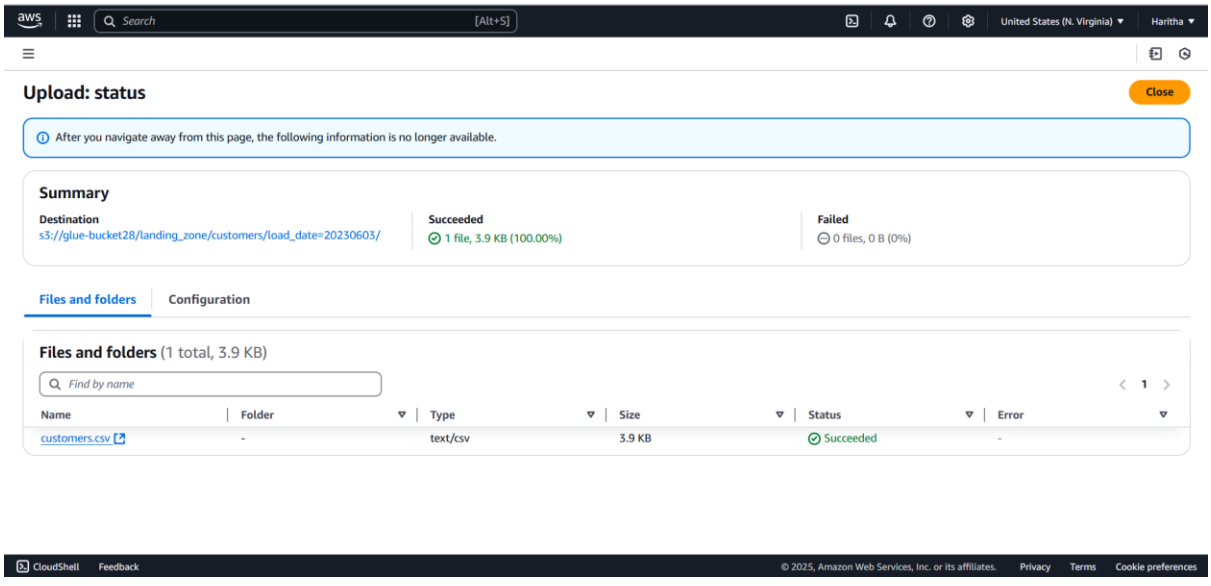

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/
landing_zone/
Copy S3 URI
Objects
Properties
Objects (1)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You can use 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
Name
Type
Last modified
Size
Storage class
customers/
Folder
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

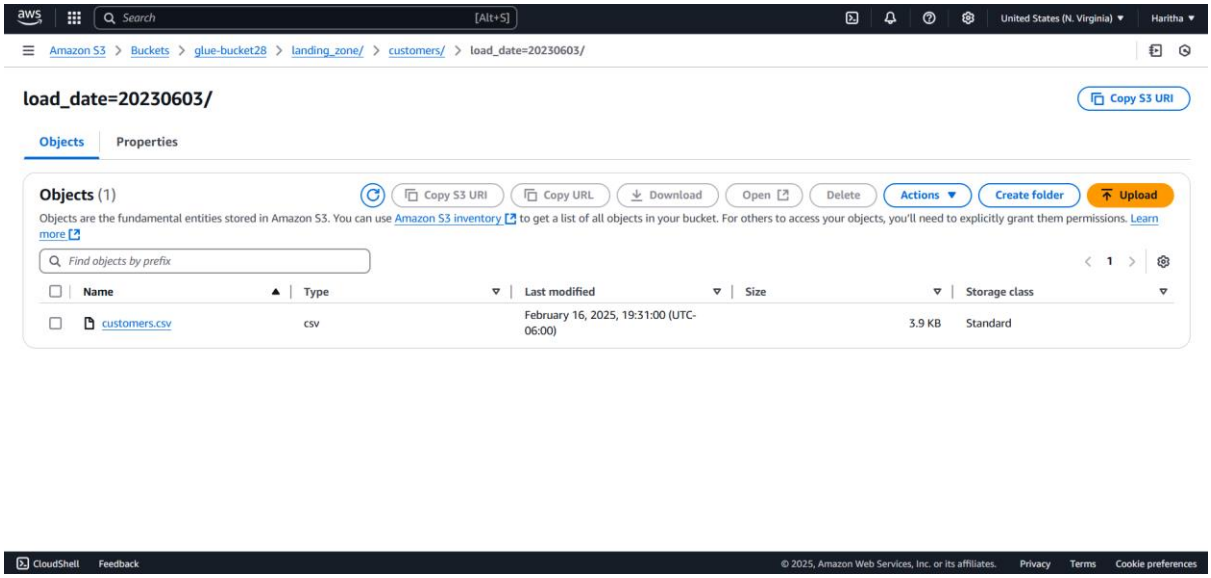

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > customers/
customers/
Copy S3 URI
Objects
Properties
Objects (0)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You 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
Name
Type
Last modified
Size
Storage class
No objects
You don't have any objects in this folder.
Upload
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > customers/ > Create folder
Create folder Info
Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more
Your bucket policy might block folder creation
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.
Folder
Folder name
load_date=20230603
/
Folder names can't contain "/". See rules for naming
Server-side encryption Info
Server-side encryption protects data at rest.
The following encryption settings apply only to the folder object and not to sub-folder objects.
Server-side encryption
Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.
Specify an encryption key
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > customers/
Successfully created folder "load_date=20230603".
customers/
Copy S3 URI
Objects
Properties
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
Name
Type
Last modified
Size
Storage class
load_date=20230603/
Folder


CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Open
Downloads
Search Downloads
Organize
New folder
Today
customers.csv
Microsoft Excel Comma Separate...
3.88 KB
boto3-user_accessKeys.csv
99 bytes
Yesterday
VSCodeUserSetup-x64-1.97.2.exe
Visual Studio Code Setup
Microsoft Corporation
Earlier this week
cli-user_accessKeys.csv
Git-2.48.1-64-bit.exe
Git Setup
The Git Development Community
AWSCLIV2.msi
Key Note.docx
File name: customers.csv
All Files (*.*)
Open
Cancel
Upload
Amazon S3 REST API. Learn more
Add files or Add folder.
Remove
Add files
Add folder
Name
Folder
Type
Size
No files or folders
You have not chosen any files or folders to upload.
Destination Info
Destination
s3://glue-bucket28/landing_zone/customers/load_date=20230603/
Destination details
Bucket settings that impact new objects stored in the specified destination.


Search [Alt+S]
United States (N. Virginia)
Haritha
Upload: status
Close
After you navigate away from this page, the following information is no longer available.
Summary
Destination
s3://glue-bucket28/landing_zone/customers/load_date=20230603/
Succeeded
1 file, 3.9 KB (100.00%)
Failed
0 files, 0 B (0%)
Files and folders
Configuration
Files and folders (1 total, 3.9 KB)
Find by name
Name
Folder
Type
Size
Status
Error
customers.csv
-
text/csv
3.9 KB
Succeeded
-
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences



AWS Glue

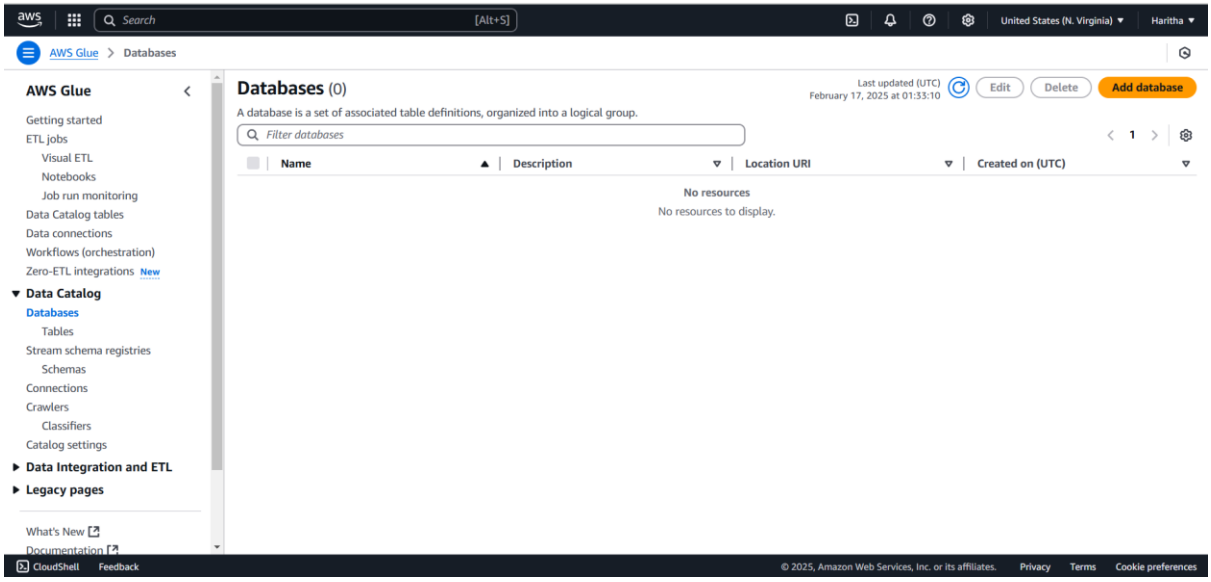
AWS Glue is a serverless data integration service that allows users to prepare, transform, and analyze data. It is commonly used for ETL (Extract, Transform, Load) workflows.

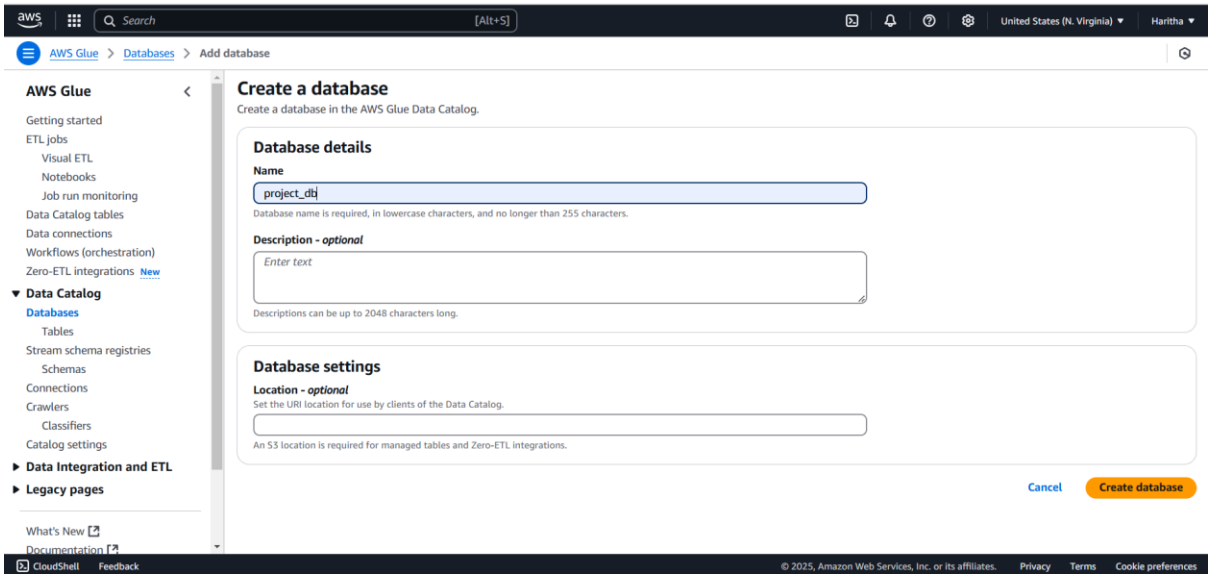
Key Components of AWS Glue:

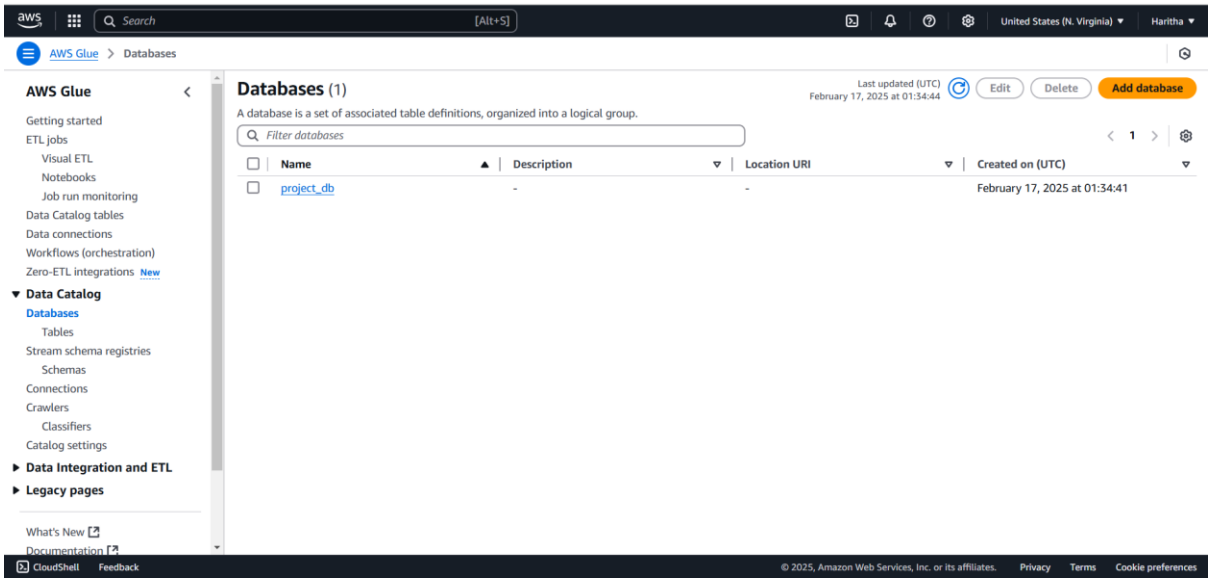
- **Glue Data Catalog** – A centralized metadata repository for managing datasets.
- **Glue Crawlers** – Automatically discover and catalog datasets.
- **Glue Jobs** – ETL scripts that process data.
- **Glue Triggers** – Automate job execution based on schedules or events.

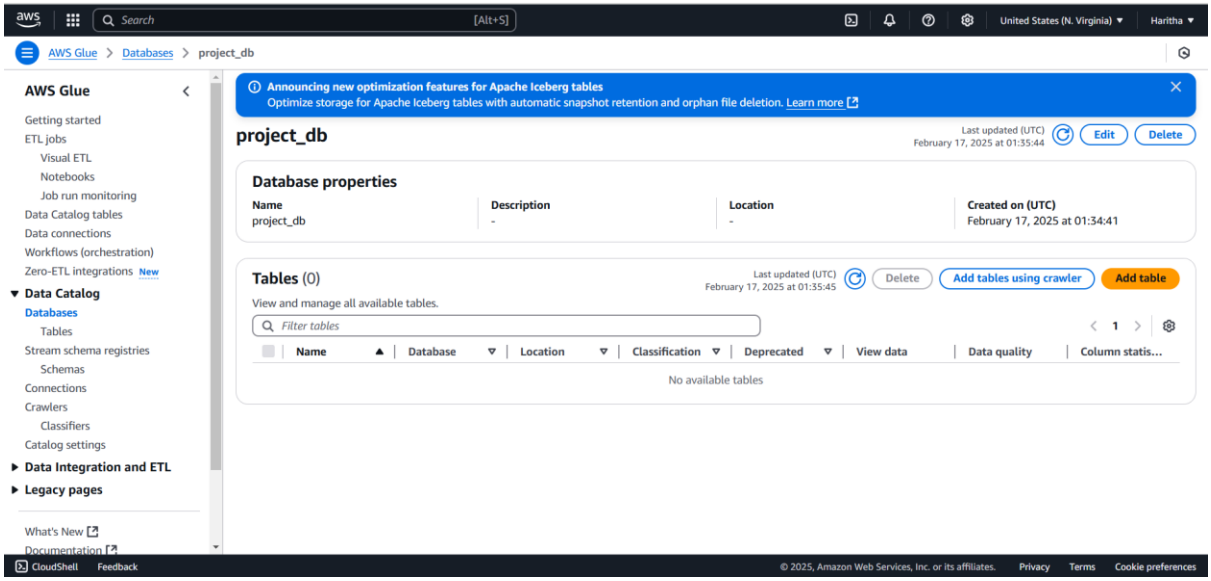
AWS Glue ETL (Extract, Transform, Load):

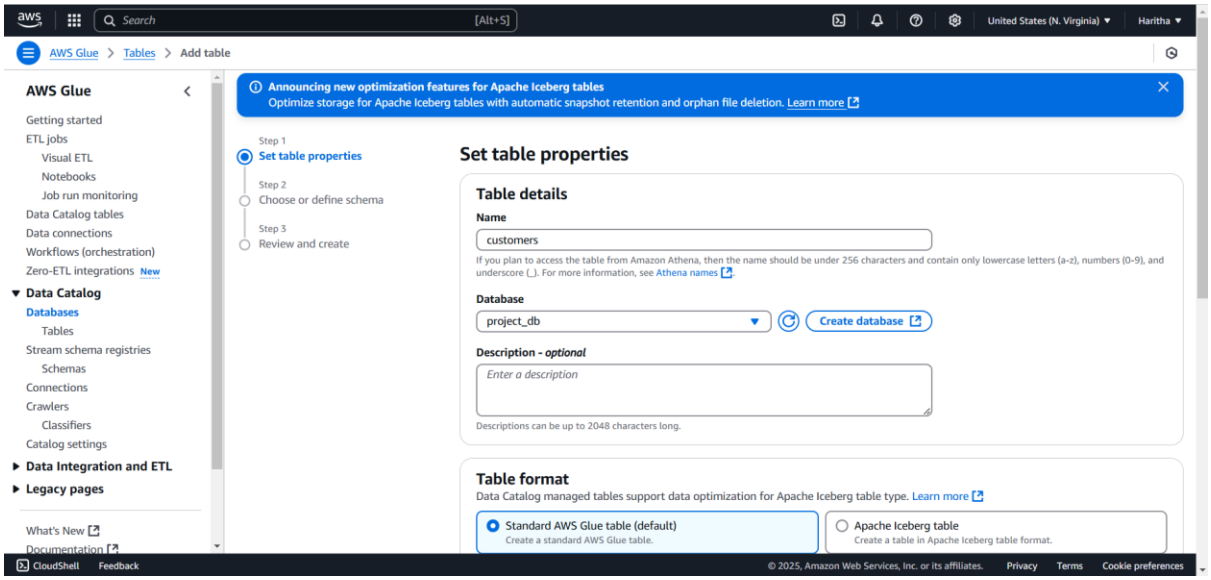
AWS Glue ETL is used to process and transform large-scale data from various sources. It supports Python and Spark-based transformations.

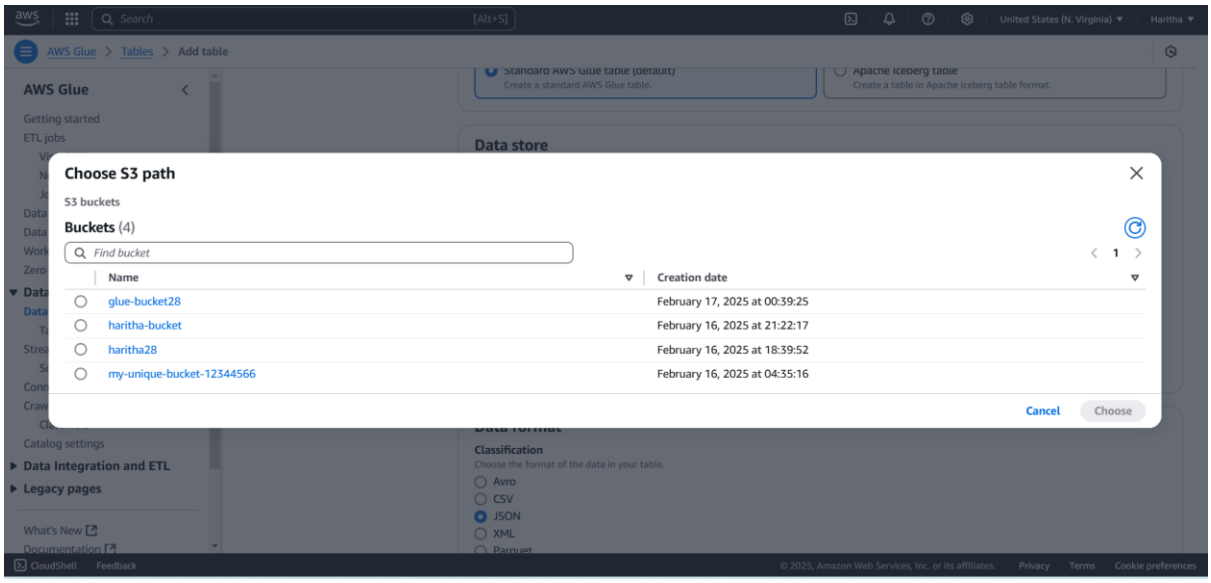


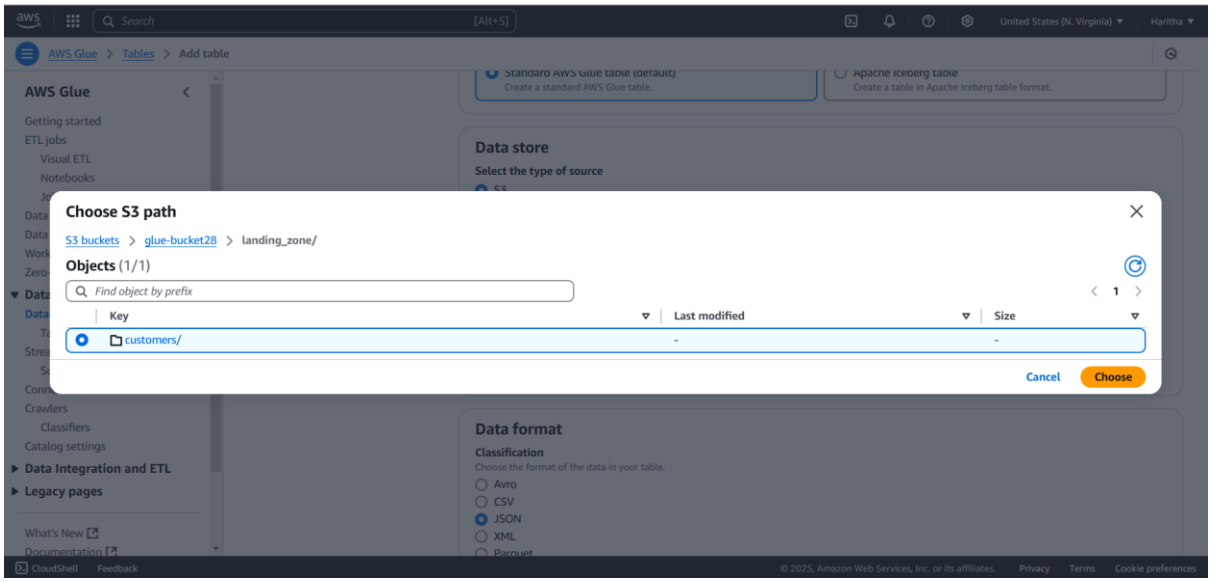

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Databases > Add database
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
Create a database
Create a database in the AWS Glue Data Catalog.
Database details
Name
project_db
Database name is required, in lowercase characters, and no longer than 255 characters.
Description - optional
Enter text
Descriptions can be up to 2048 characters long.
Database settings
Location - optional
Set the URI location for use by clients of the Data Catalog.
An S3 location is required for managed tables and Zero-ETL integrations.
Cancel
Create database
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

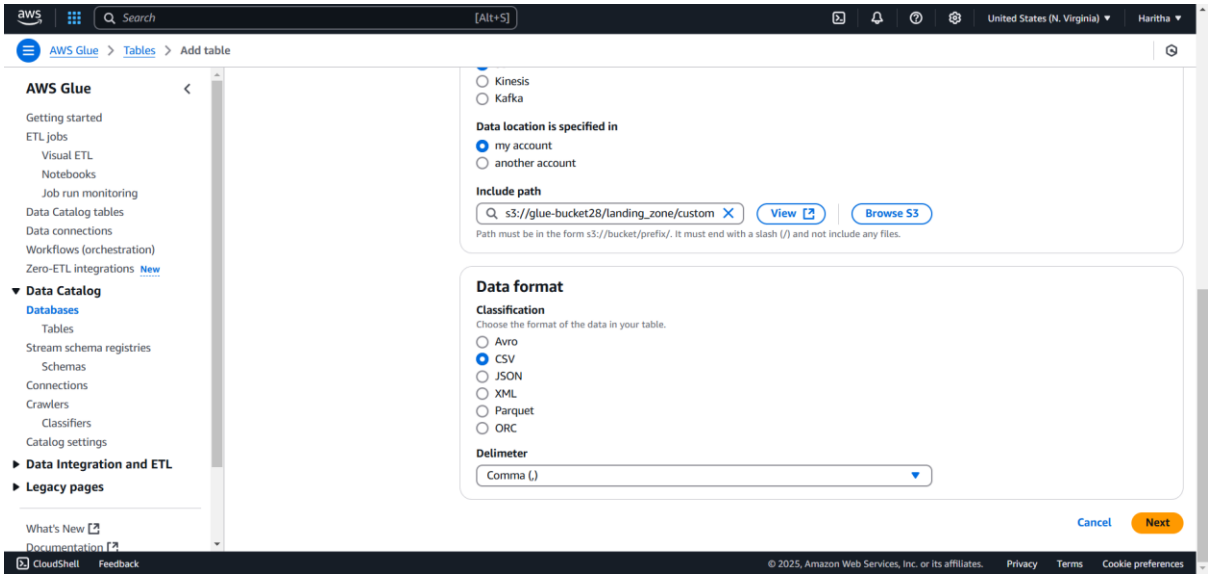

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Databases
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
Databases (1)
A database is a set of associated table definitions, organized into a logical group.
Last updated (UTC)
February 17, 2025 at 01:34:44
Edit
Delete
Add database
Filter databases
1
Name
Description
Location URI
Created on (UTC)
project_db
-
-
February 17, 2025 at 01:34:41
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

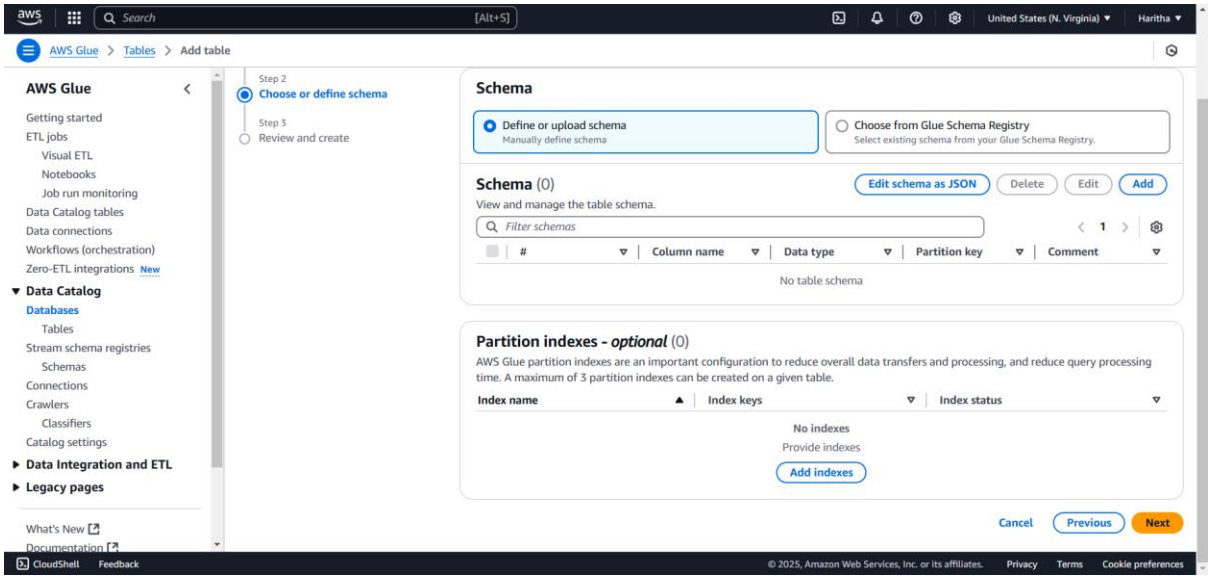

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Databases > project_db
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
Announcing new optimization features for Apache Iceberg tables
Optimize storage for Apache Iceberg tables with automatic snapshot retention and orphan file deletion. Learn more
project_db
Last updated (UTC)
February 17, 2025 at 01:35:44
Edit
Delete
Database properties
Name
project_db
Description
-
Location
-
Created on (UTC)
February 17, 2025 at 01:34:41
Tables (0)
View and manage all available tables.
Last updated (UTC)
February 17, 2025 at 01:35:45
Delete
Add tables using crawler
Add table
Filter tables
1
Name
Database
Location
Classification
Deprecated
View data
Data quality
Column statis...
No available tables
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

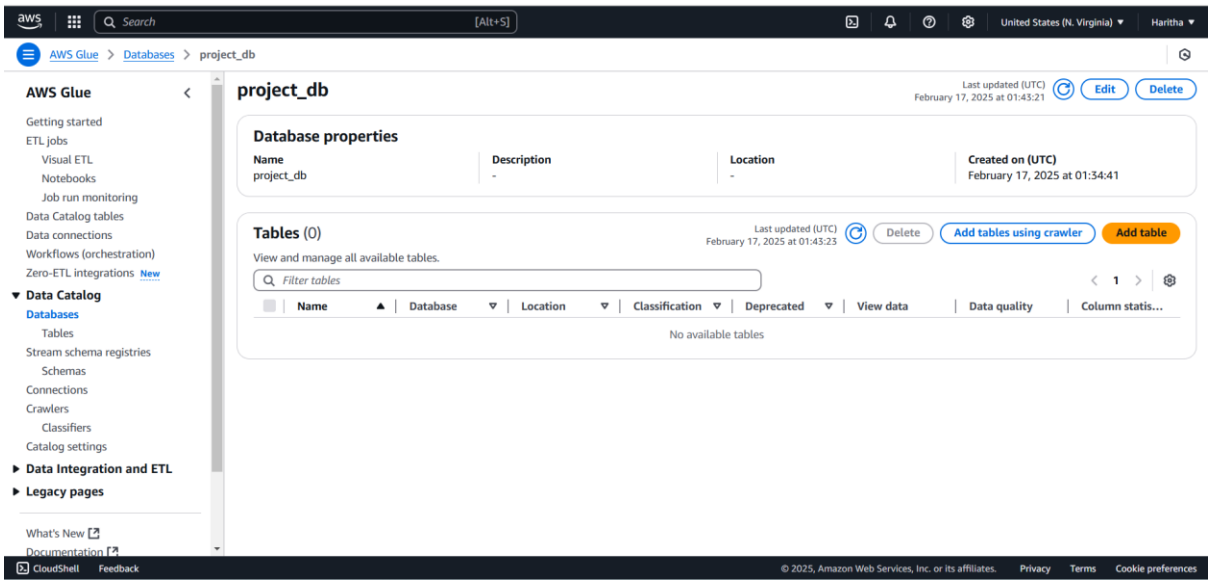

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Tables > Add table
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
Announcing new optimization features for Apache Iceberg tables
Optimize storage for Apache Iceberg tables with automatic snapshot retention and orphan file deletion. Learn more
Step 1
Set table properties
Step 2
Choose or define schema
Step 3
Review and create
Set table properties
Table details
Name
customers
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
project_db
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.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

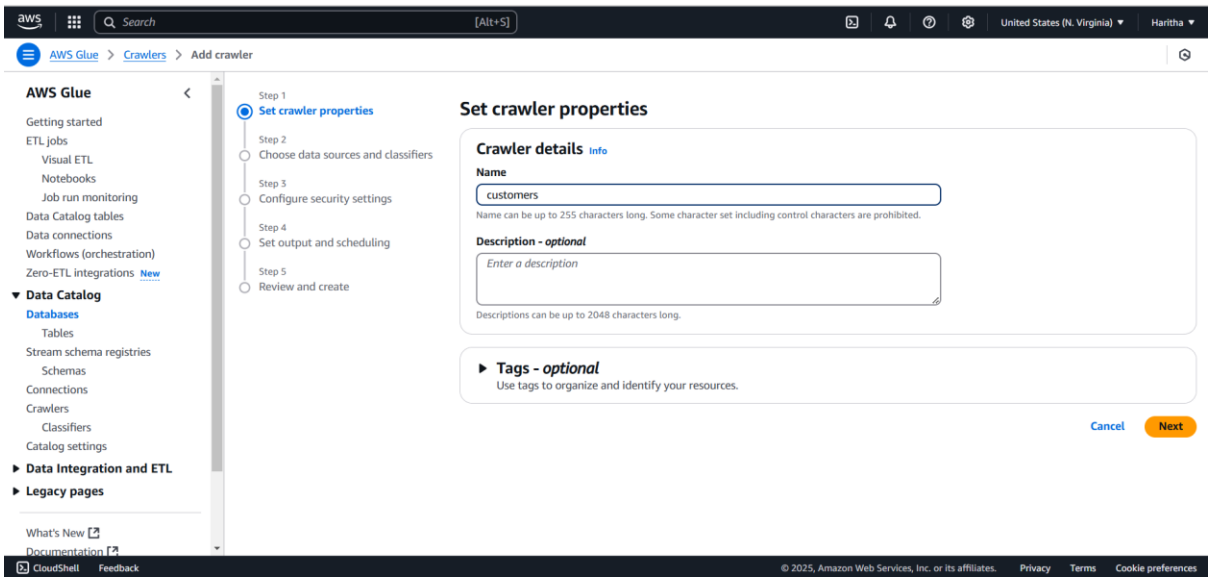

Choose S3 path
S3 buckets
Buckets (4)
Find bucket
Name
Creation date
glue-bucket28
February 17, 2025 at 00:39:25
haritha-bucket
February 16, 2025 at 21:22:17
haritha28
February 16, 2025 at 18:39:52
my-unique-bucket-12344566
February 16, 2025 at 04:35:16
Cancel
Choose
Data store
Classification
Avro
CSV
JSON
XML
Parquet

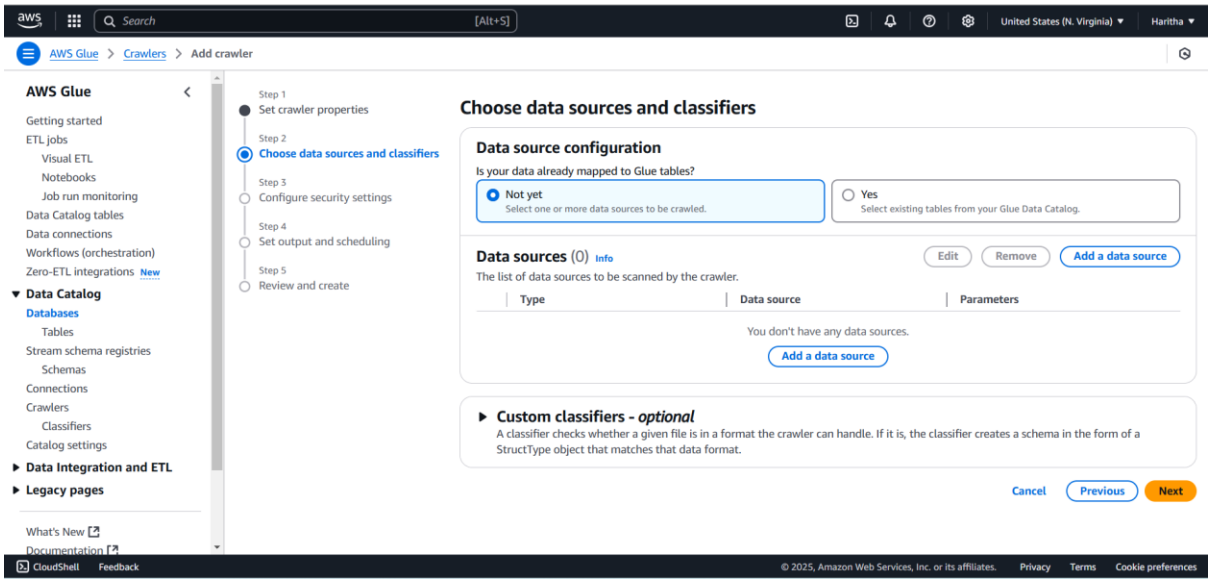

Choose S3 path
S3 buckets > glue-bucket28 > landing_zone/
Objects (1/1)
Find object by prefix
Key
Last modified
Size
customers/
Cancel
Choose
Data store
Select the type of source
Data format
Classification
Avro
CSV
JSON
XML
Parquet

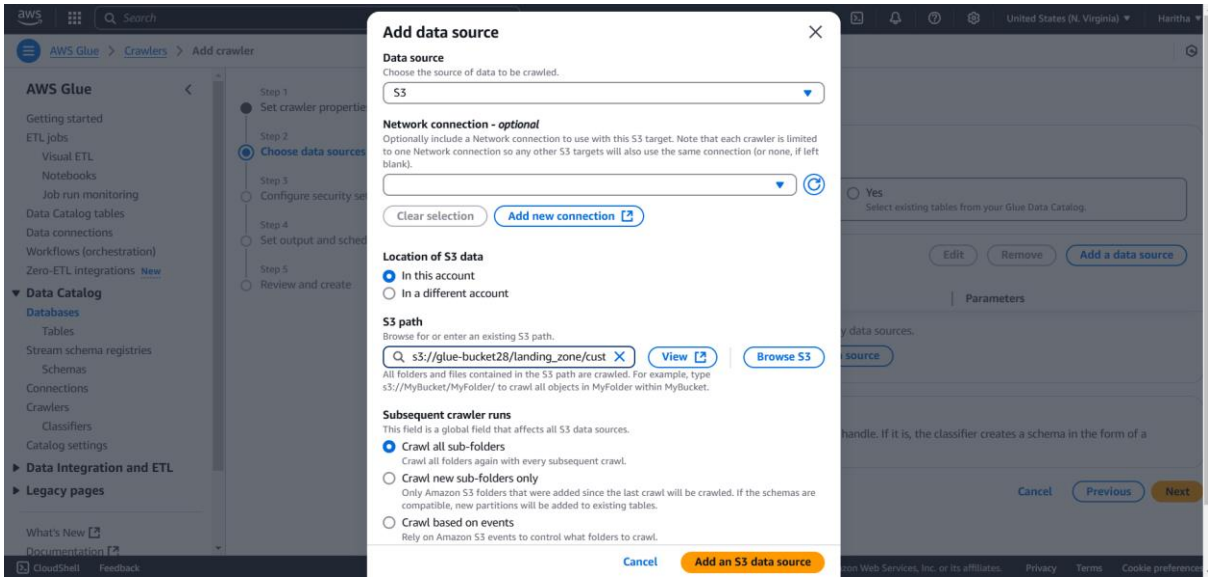

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Tables > Add table
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
Kinesis
Kafka
Data location is specified in
my account
another account
Include path
s3://glue-bucket28/landing_zone/custom
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
Delimeter
Comma (,)
Cancel
Next
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

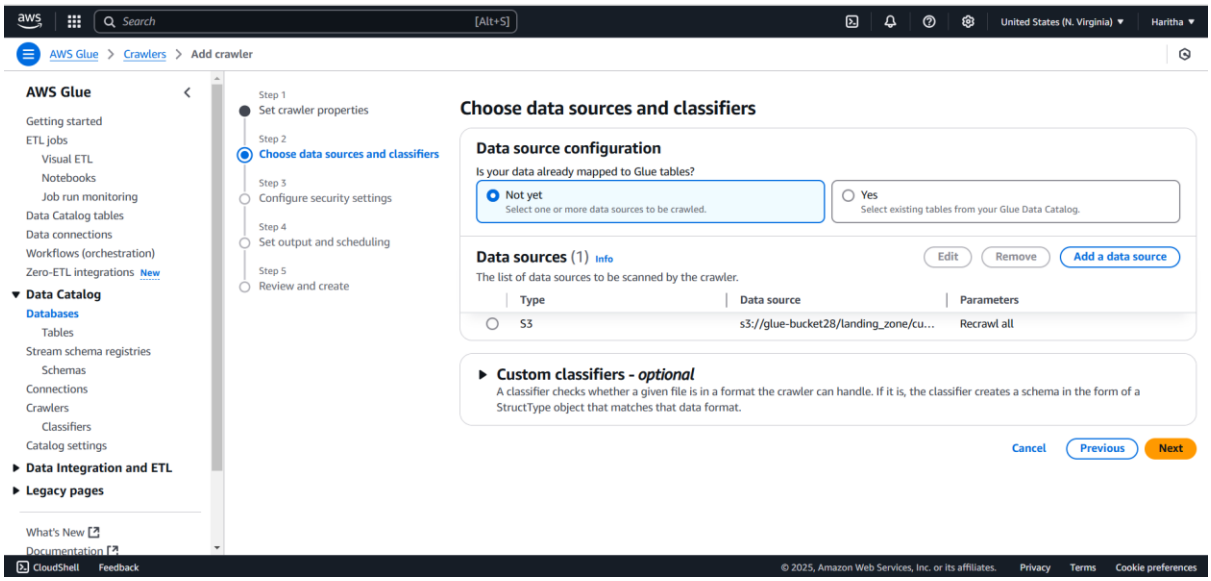

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Tables > Add table
AWS Glue
Step 2
Choose or define schema
Step 3
Review and create
Schema
Define or upload schema
Manually define schema
Choose from Glue Schema Registry
Select existing schema from your Glue Schema Registry.
Schema (0)
View and manage the table schema.
Edit schema as JSON
Delete
Edit
Add
Filter schemas
| Column name | Data type | Partition key | Comment
No table schema
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
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

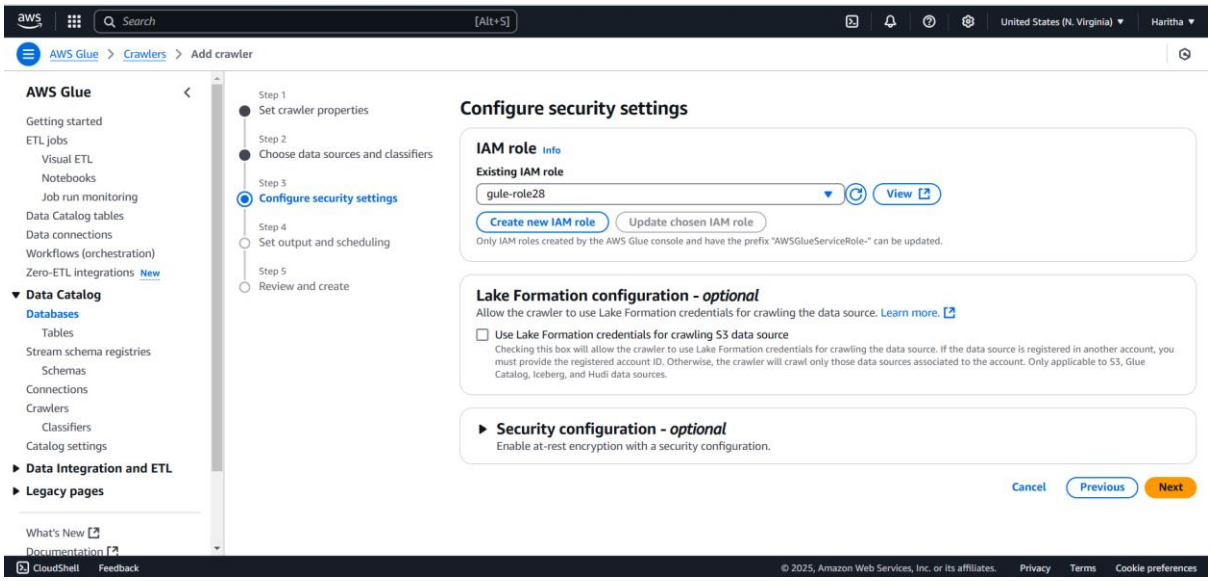

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Databases > project_db
AWS Glue
project_db
Last updated (UTC)
February 17, 2025 at 01:43:21
Edit
Delete
Database properties
Name
project_db
Description
-
Location
-
Created on (UTC)
February 17, 2025 at 01:34:41
Tables (0)
Last updated (UTC)
February 17, 2025 at 01:43:23
Delete
Add tables using crawler
Add table
View and manage all available tables.
Filter tables
Name | Database | Location | Classification | Deprecated | View data | Data quality | Column statis...
No available tables
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

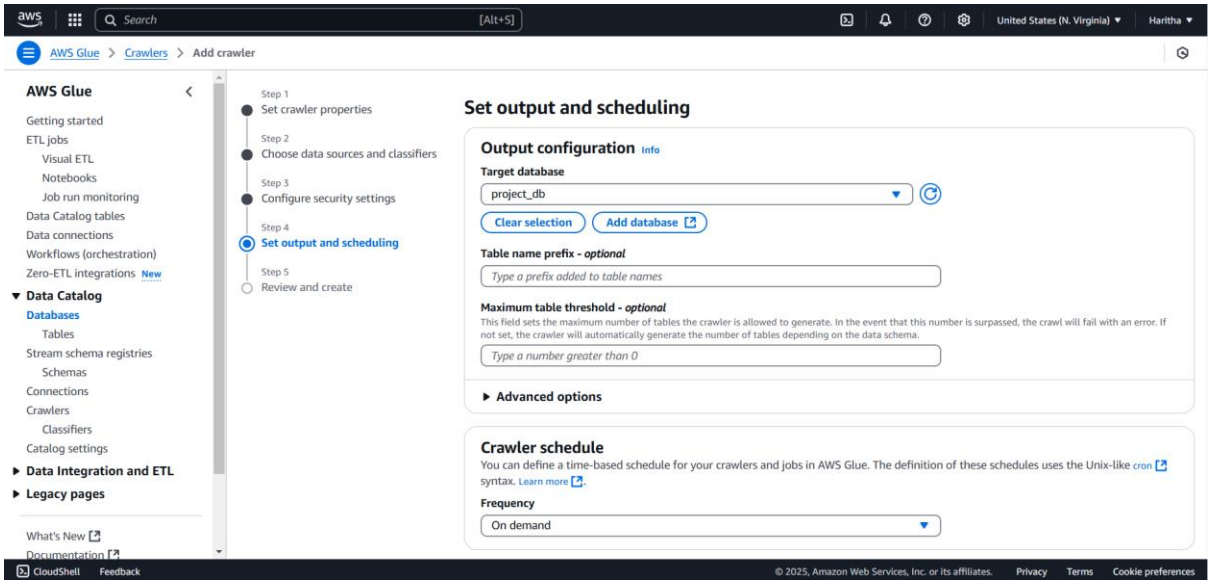

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > Add crawler
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
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 crawler properties
Crawler details Info
Name
customers
Name can be up to 255 characters long. Some character set including control characters are prohibited.
Description - optional
Enter a description
Descriptions can be up to 2048 characters long.
Tags - optional
Use tags to organize and identify your resources.
Cancel
Next
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

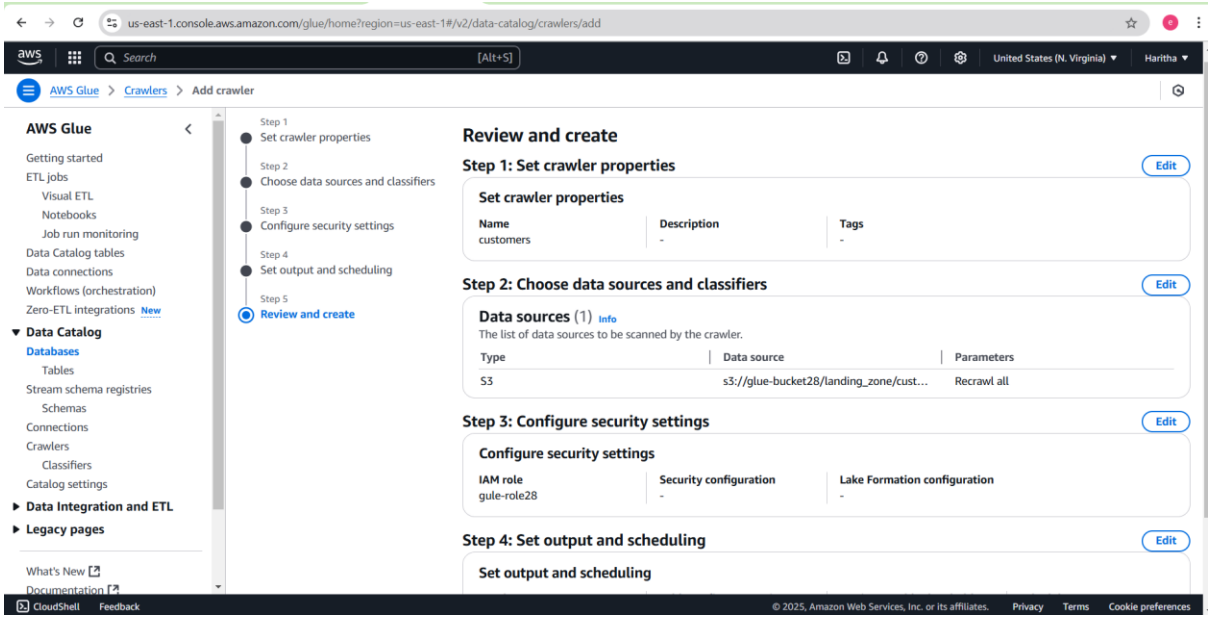

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > Add crawler
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
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
Choose data sources and classifiers
Data source configuration
Is your data already mapped to Glue tables?
Not yet
Select one or more data sources to be crawled.
Yes
Select existing tables from your Glue Data Catalog.
Data sources (0) Info
Edit
Remove
Add a data source
The list of data sources to be scanned by the crawler.
Type
Data source
Parameters
You don't have any data sources.
Add a data source
Custom classifiers - optional
A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.
Cancel
Previous
Next
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

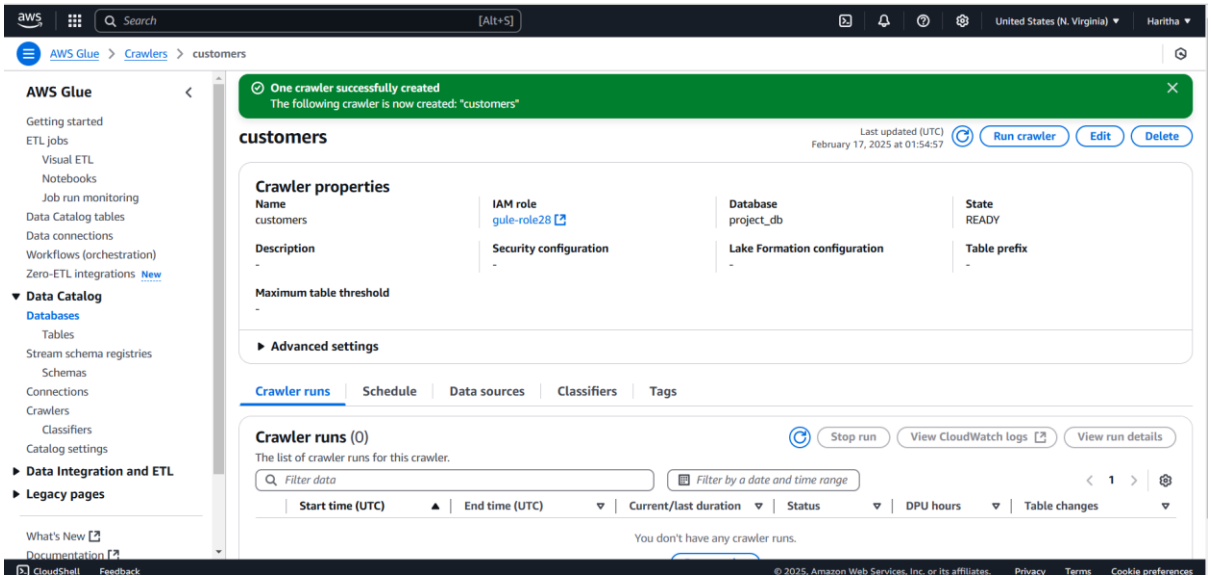

Add data source
Data source
Choose the source of data to be crawled.
S3
Network connection - optional
Optionally include a Network connection to use with this S3 target. Note that each crawler is limited to one Network connection so any other S3 targets will also use the same connection (or none, if left blank).
Clear selection
Add new connection
Location of S3 data
In this account
In a different account
S3 path
Browse for or enter an existing S3 path.
s3://glue-bucket28/landing_zone/cust
View
Browse S3
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.
Subsequent crawler runs
This field is a global field that affects all S3 data sources.
Crawl all sub-folders
Crawl all folders again with every subsequent crawl.
Crawl new sub-folders only
Only Amazon S3 folders that were added since the last crawl will be crawled. If the schemas are compatible, new partitions will be added to existing tables.
Crawl based on events
Rely on Amazon S3 events to control what folders to crawl.
Cancel
Add an S3 data source

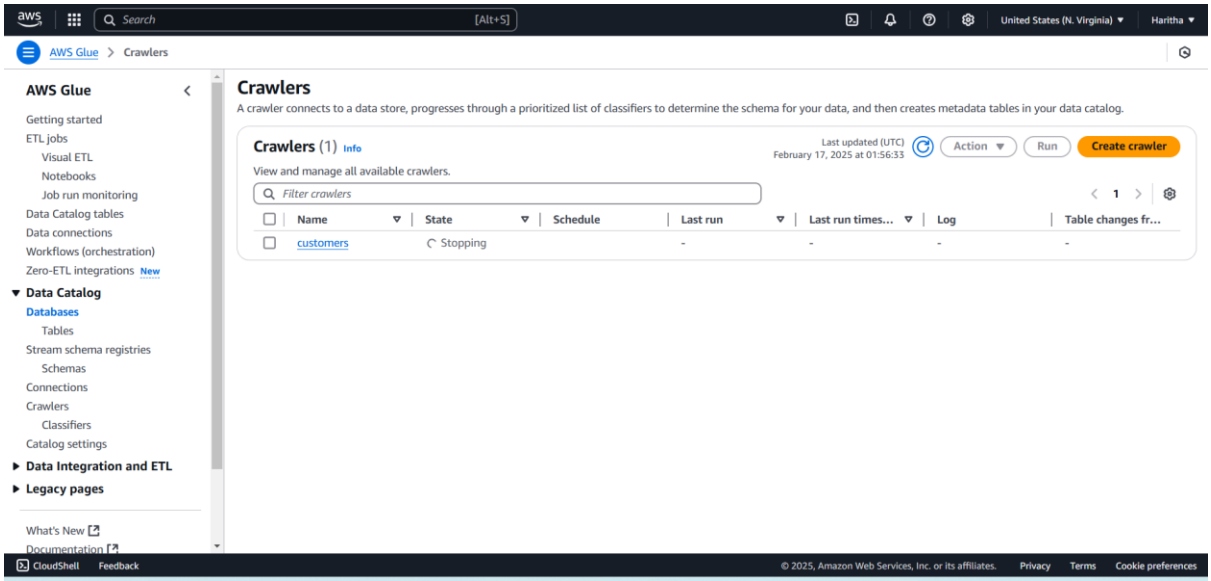

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > Add crawler
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
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
Choose data sources and classifiers
Data source configuration
Is your data already mapped to Glue tables?
Not yet
Select one or more data sources to be crawled.
Yes
Select existing tables from your Glue Data Catalog.
Data sources (1) Info
The list of data sources to be scanned by the crawler.
Edit
Remove
Add a data source
Type
Data source
Parameters
S3
s3://glue-bucket28/landing_zone/cu...
Recrawl all
Custom classifiers - optional
A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.
Cancel
Previous
Next
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

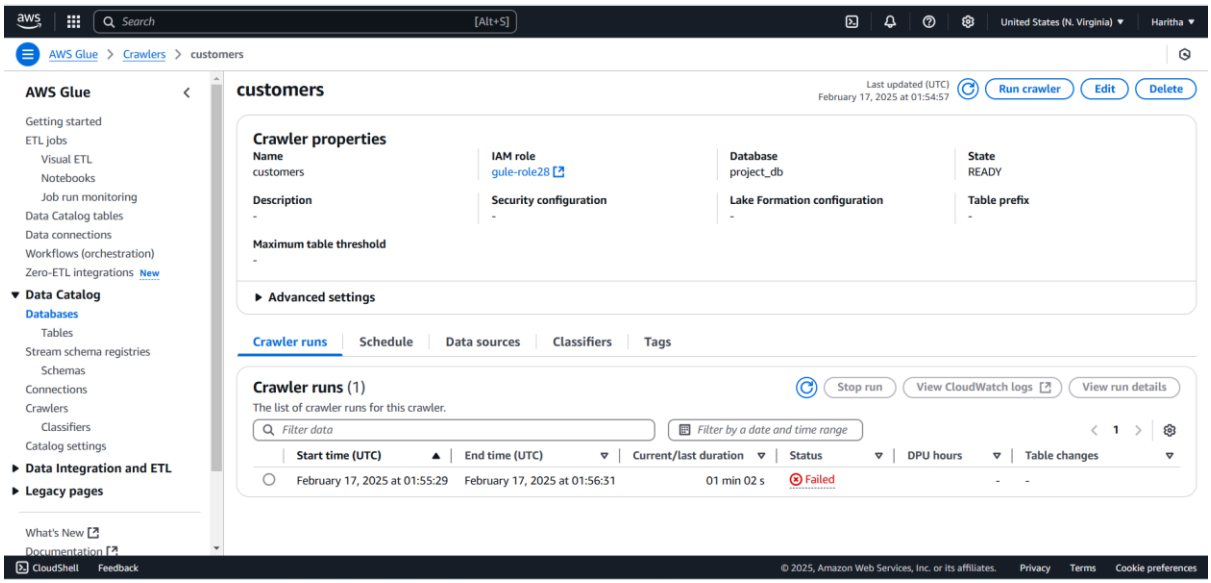

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > Add crawler
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
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
Configure security settings
IAM role Info
Existing IAM role
gule-role28
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.
Lake Formation configuration - optional
Allow the crawler to use Lake Formation credentials for crawling the data source. Learn more.
Use Lake Formation credentials for crawling S3 data source
Checking this box will allow the crawler to use Lake Formation credentials for crawling the data source. If the data source is registered in another account, you must provide the registered account ID. Otherwise, the crawler will crawl only those data sources associated to the account. Only applicable to S3, Glue Catalog, Iceberg, and Hudi data sources.
Security configuration - optional
Enable at-rest encryption with a security configuration.
Cancel
Previous
Next
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

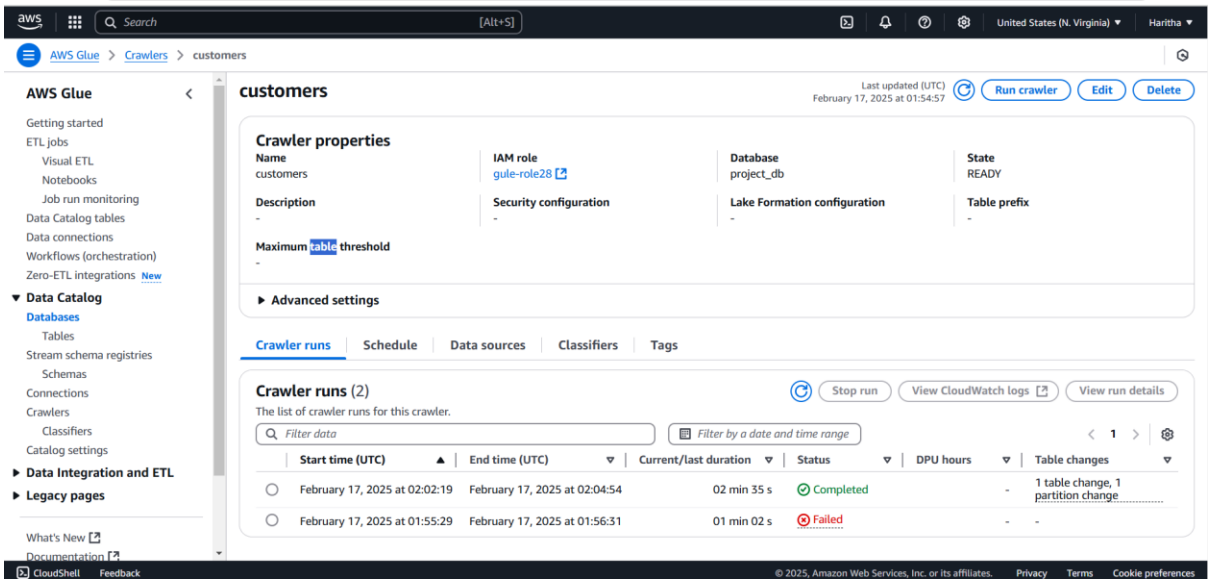

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > Add crawler
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
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 Info
Target database
project_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
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

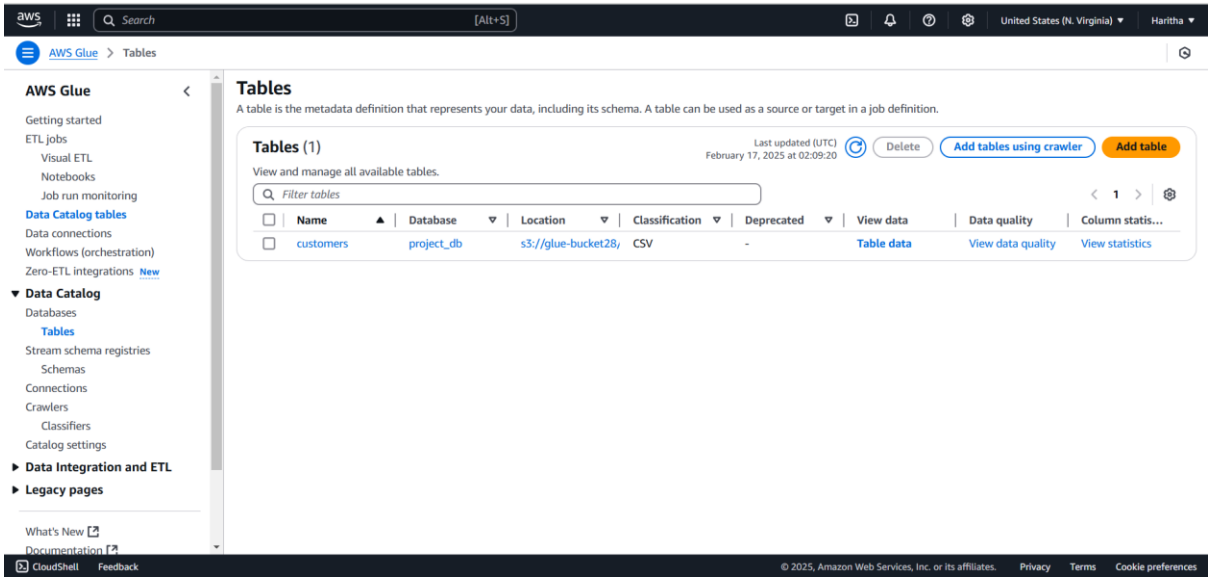

us-east-1.console.aws.amazon.com/glue/home?region=us-east-1#/v2/data-catalog/crawlers/add
AWS Glue > Crawlers > Add crawler
Step 1 Set crawler properties
Step 2 Choose data sources and classifiers
Step 3 Configure security settings
Step 4 Set output and scheduling
Step 5 Review and create
Review and create
Step 1: Set crawler properties
Edit
Name customers
Description -
Tags -
Step 2: Choose data sources and classifiers
Data sources (1) Info
The list of data sources to be scanned by the crawler.
Type S3
Data source s3://glue-bucket28/landing_zone/cust...
Parameters Recrawl all
Step 3: Configure security settings
IAM role gule-role28
Security configuration -
Lake Formation configuration -
Step 4: Set output and scheduling
Set output and scheduling

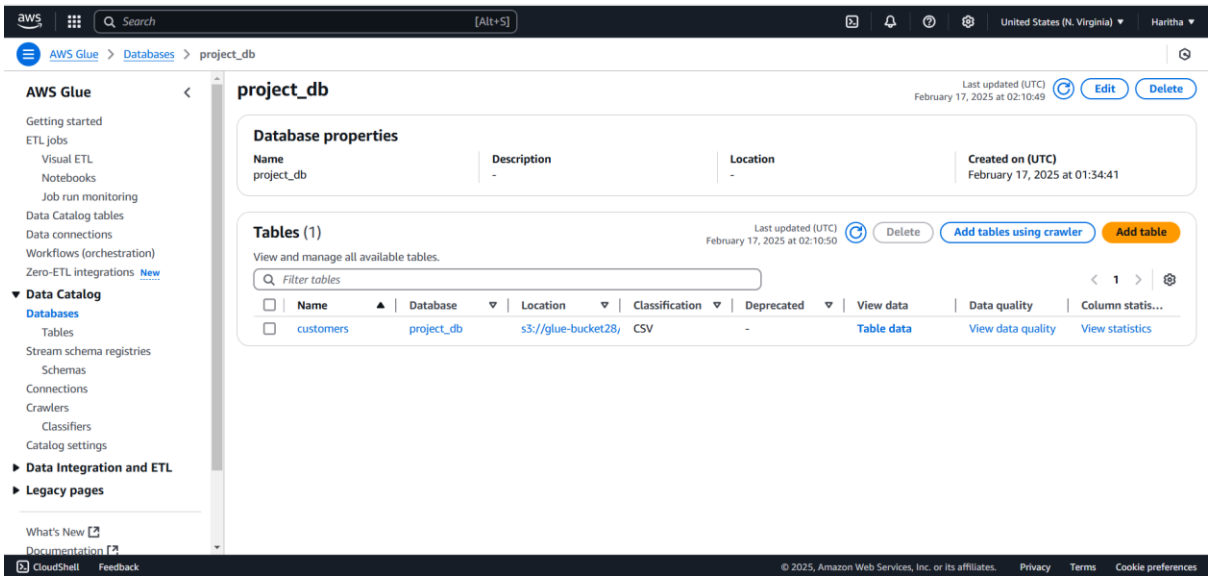

AWS Glue > Crawlers > customers
One crawler successfully created
The following crawler is now created: "customers"
customers
Last updated (UTC) February 17, 2025 at 01:54:57
Run crawler
Edit
Delete
Crawler properties
Name customers
IAM role gule-role28
Database project_db
State READY
Description -
Security configuration -
Lake Formation configuration -
Table prefix -
Maximum table threshold -
Advanced settings
Crawler runs
Schedule
Data sources
Classifiers
Tags
Crawler runs (0)
Stop run
View CloudWatch logs
View run details
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.

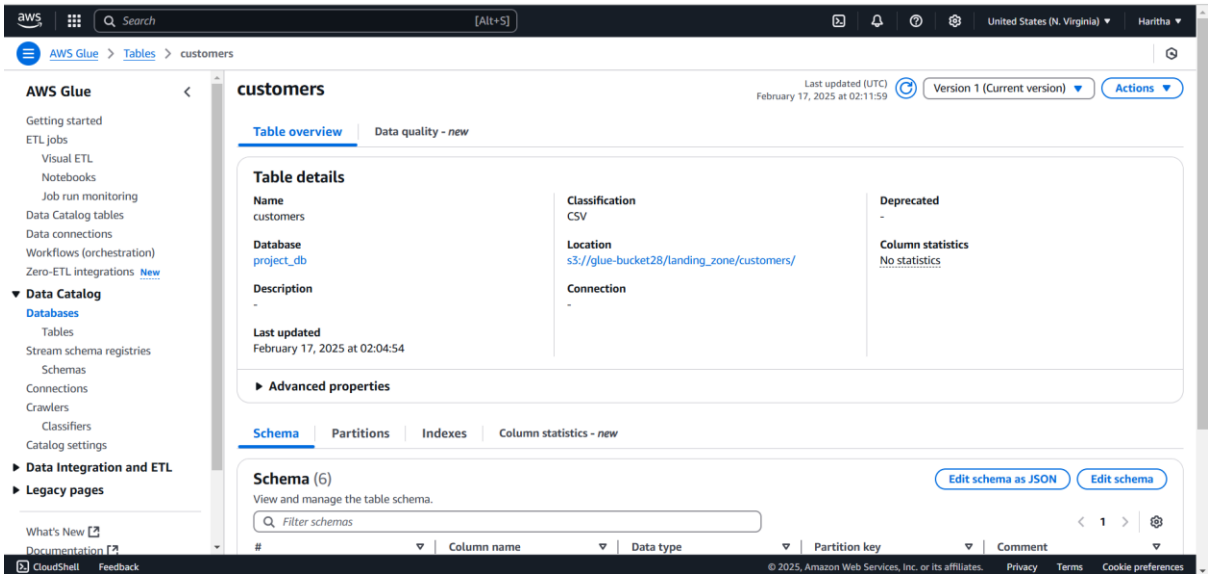

AWS Glue > Crawlers
Crawlers
A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.
Crawlers (1) Info
Last updated (UTC) February 17, 2025 at 01:56:33
Action
Run
Create crawler
View and manage all available crawlers.
Filter crawlers
Name
State
Schedule
Last run
Last run times...
Log
Table changes fr...
customers
Stopping

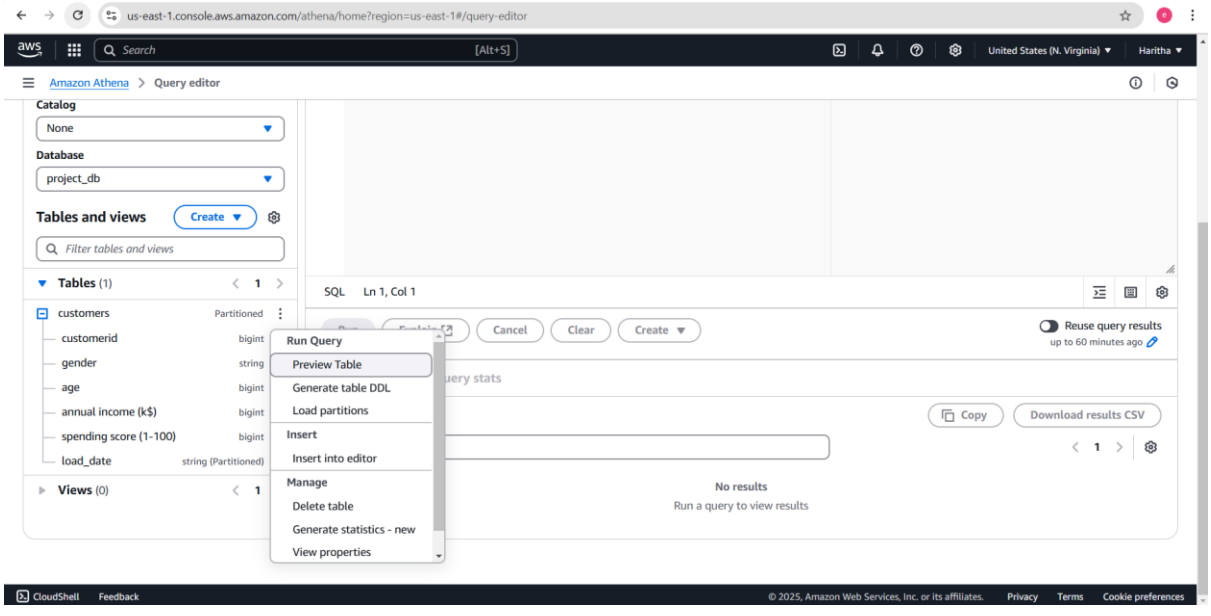

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > customers
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
customers
Last updated (UTC) February 17, 2025 at 01:54:57
Run crawler
Edit
Delete
Crawler properties
Name
customers
IAM role
gule-role28
Database
project_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 (1)
The list of crawler runs for this crawler.
Stop run
View CloudWatch logs
View run details
Filter data
Filter by a date and time range
Start time (UTC)
End time (UTC)
Current/last duration
Status
DPU hours
Table changes
February 17, 2025 at 01:55:29
February 17, 2025 at 01:56:31
01 min 02 s
Failed
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

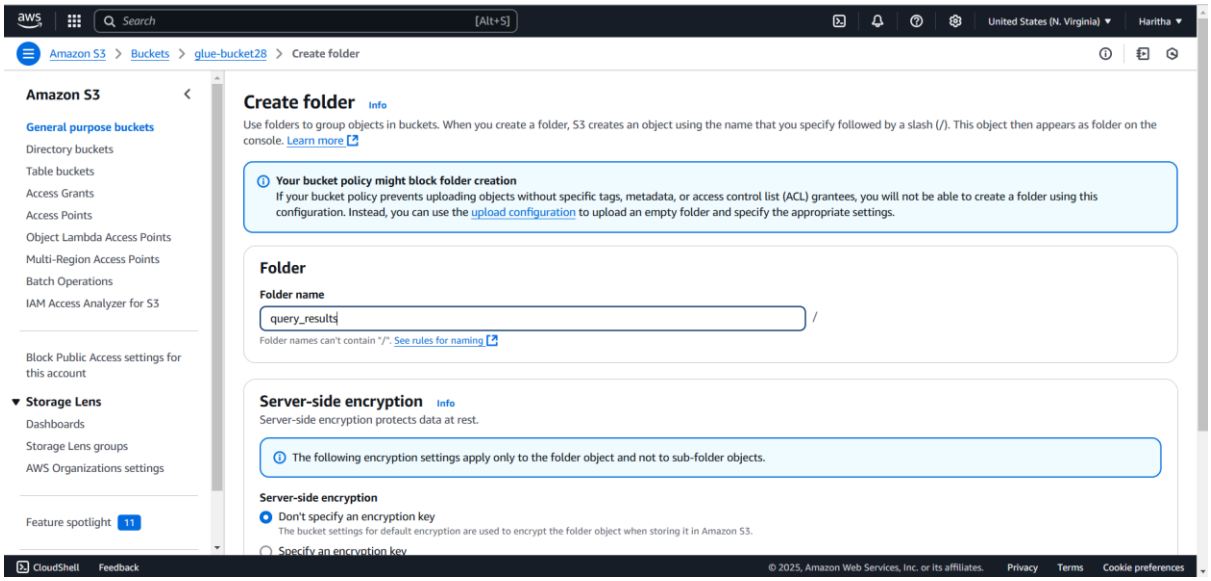

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Crawlers > customers
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
customers
Last updated (UTC) February 17, 2025 at 01:54:57
Run crawler
Edit
Delete
Crawler properties
Name
customers
IAM role
gule-role28
Database
project_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 (2)
The list of crawler runs for this crawler.
Stop run
View CloudWatch logs
View run details
Filter data
Filter by a date and time range
Start time (UTC)
End time (UTC)
Current/last duration
Status
DPU hours
Table changes
February 17, 2025 at 02:02:19
February 17, 2025 at 02:04:54
02 min 35 s
Completed
1 table change, 1 partition change
February 17, 2025 at 01:55:29
February 17, 2025 at 01:56:31
01 min 02 s
Failed
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

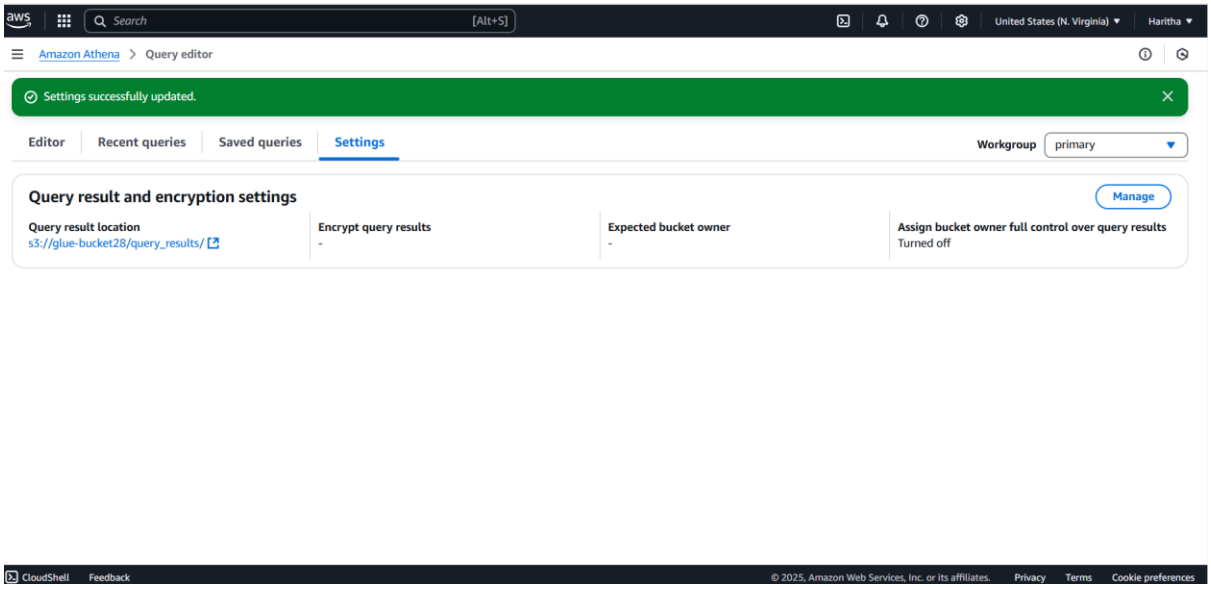

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Tables
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
Tables
A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.
Tables (1)
View and manage all available tables.
Last updated (UTC) February 17, 2025 at 02:09:20
Delete
Add tables using crawler
Add table
Filter tables
Name
Database
Location
Classification
Deprecated
View data
Data quality
Column statis...
customers
project_db
s3://glue-bucket28/
CSV
-
Table data
View data quality
View statistics
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

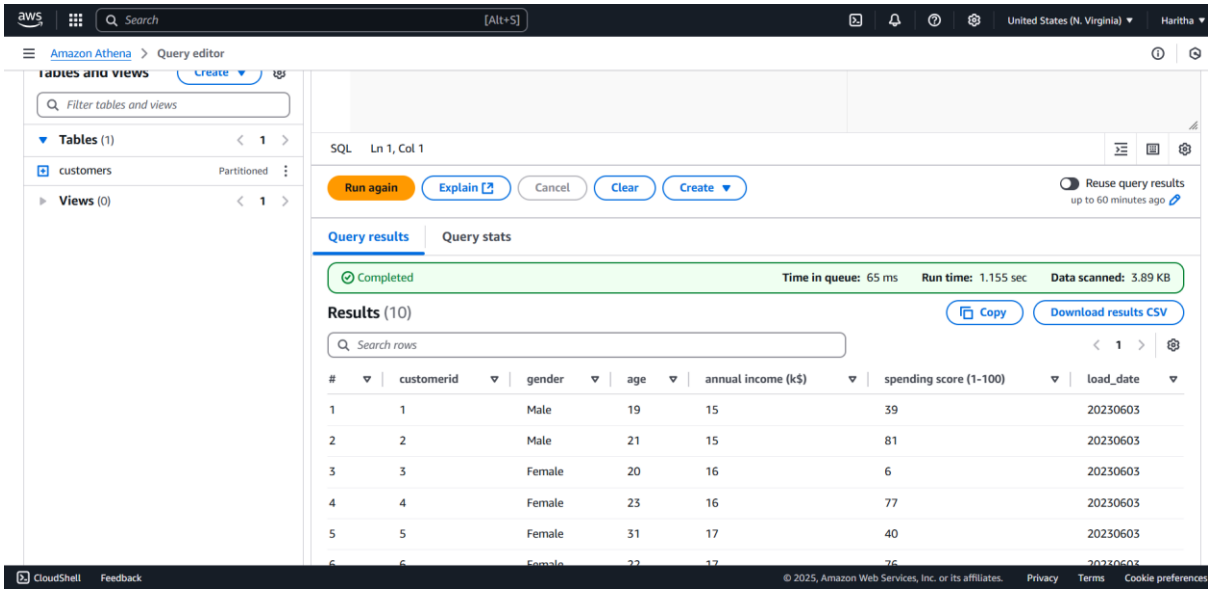


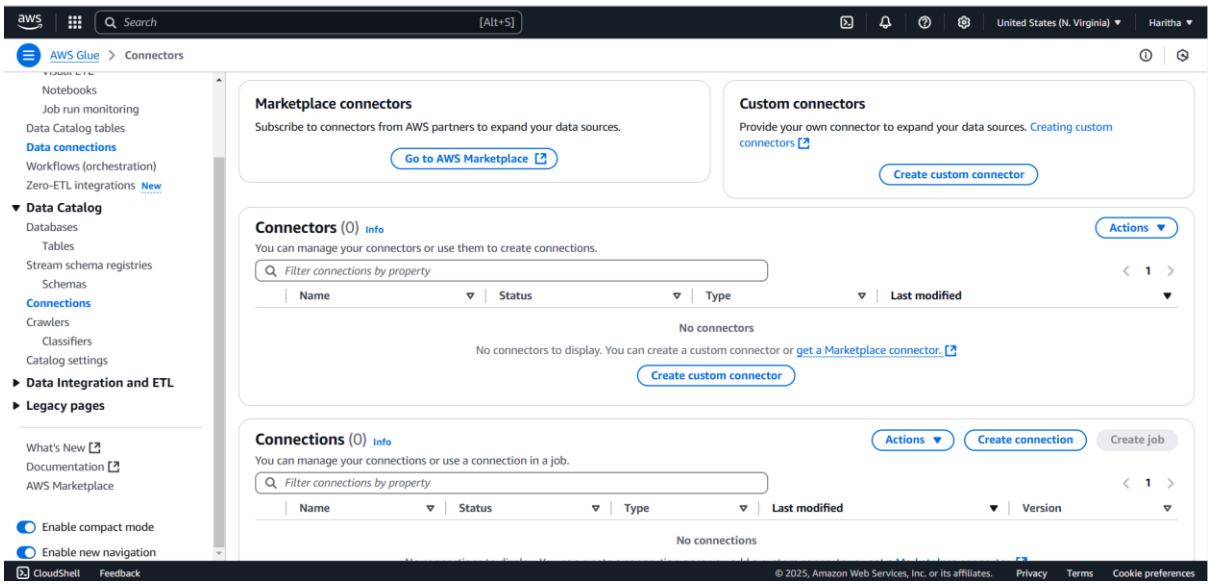


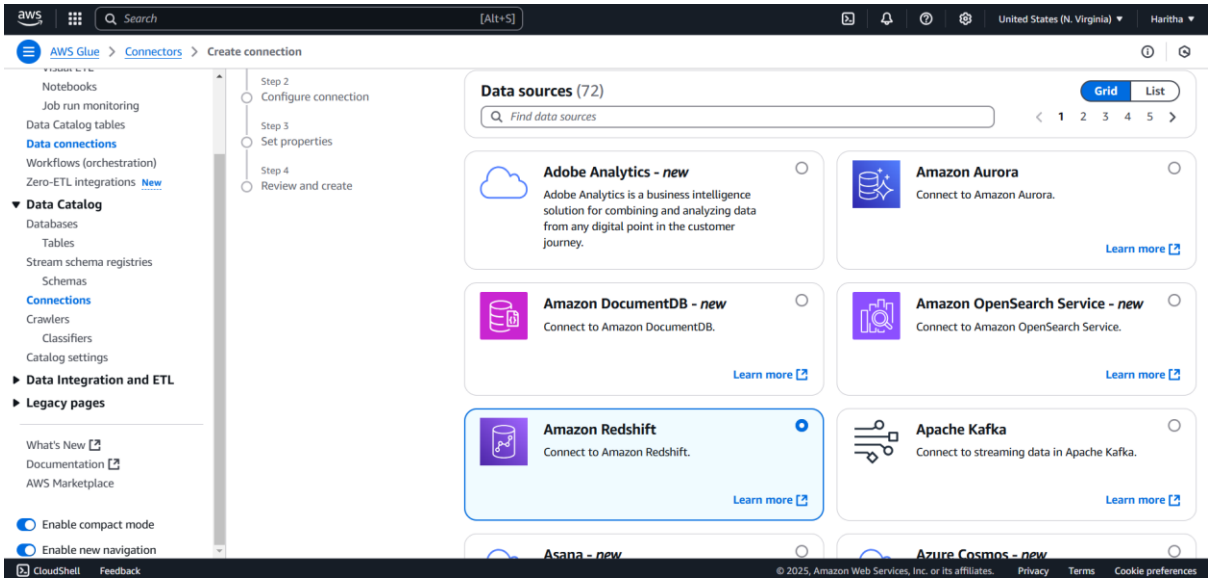


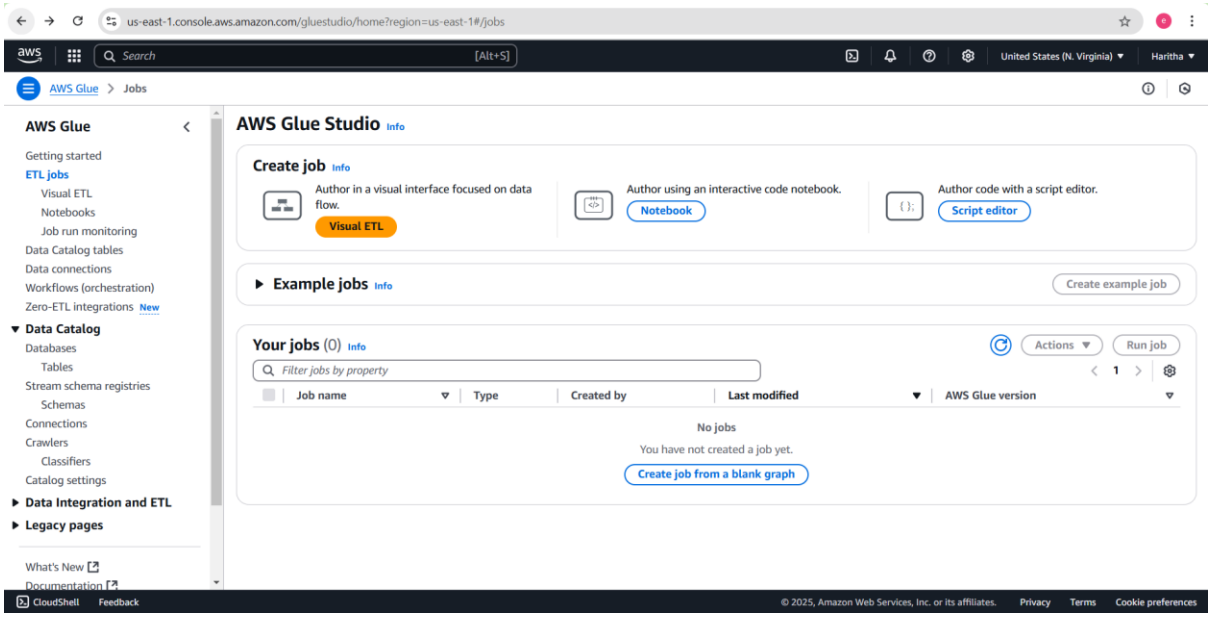

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon S3 > Buckets > glue-bucket28 > Create folder
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 11
Create folder Info
Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. Learn more
Your bucket policy might block folder creation
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the upload configuration to upload an empty folder and specify the appropriate settings.
Folder
Folder name
query_results
/
Folder names can't contain "/". See rules for naming
Server-side encryption Info
Server-side encryption protects data at rest.
The following encryption settings apply only to the folder object and not to sub-folder objects.
Server-side encryption
Don't specify an encryption key
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.
Specify an encryption key
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

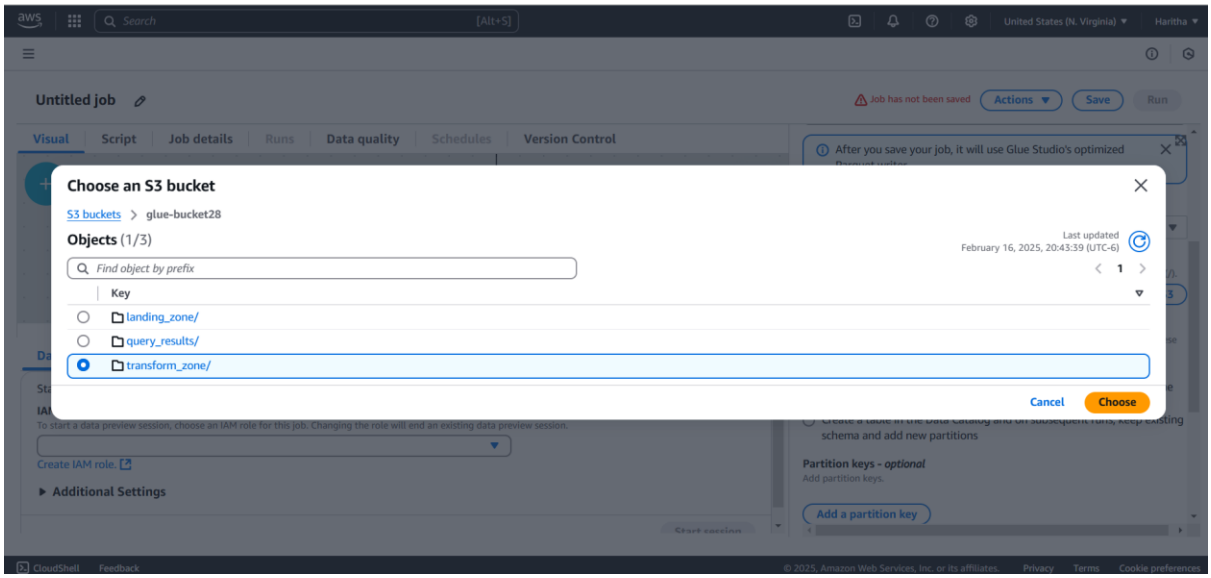

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon Athena > Query editor
Settings successfully updated.
Editor
Recent queries
Saved queries
Settings
Workgroup
primary
Query result and encryption settings
Manage
Query result location
s3://glue-bucket28/query_results/
Encrypt query results
-
Expected bucket owner
-
Assign bucket owner full control over query results
Turned off
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

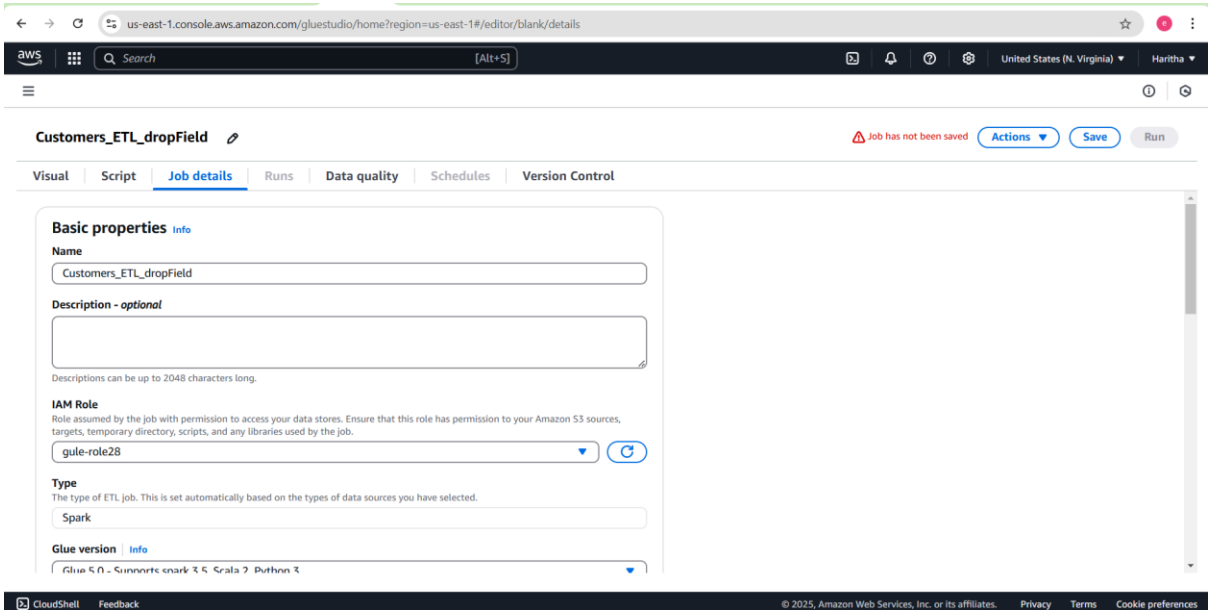

Search [Alt+S]
United States (N. Virginia)
Haritha
Amazon Athena > Query editor
Tables and views
Create
Filter tables and views
Tables (1)
customers
Partitioned
Views (0)
SQL
Ln 1, Col 1
Run again
Explain
Cancel
Clear
Create
Reuse query results
up to 60 minutes ago
Query results
Query stats
Completed
Time in queue: 65 ms
Run time: 1.155 sec
Data scanned: 3.89 KB
Results (10)
Copy
Download results CSV
Search rows
customerid gender age annual income (k$) spending score (1-100) load_date
1 1 Male 19 15 39 20230603
2 2 Male 21 15 81 20230603
3 3 Female 20 16 6 20230603
4 4 Female 23 16 77 20230603
5 5 Female 31 17 40 20230603
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

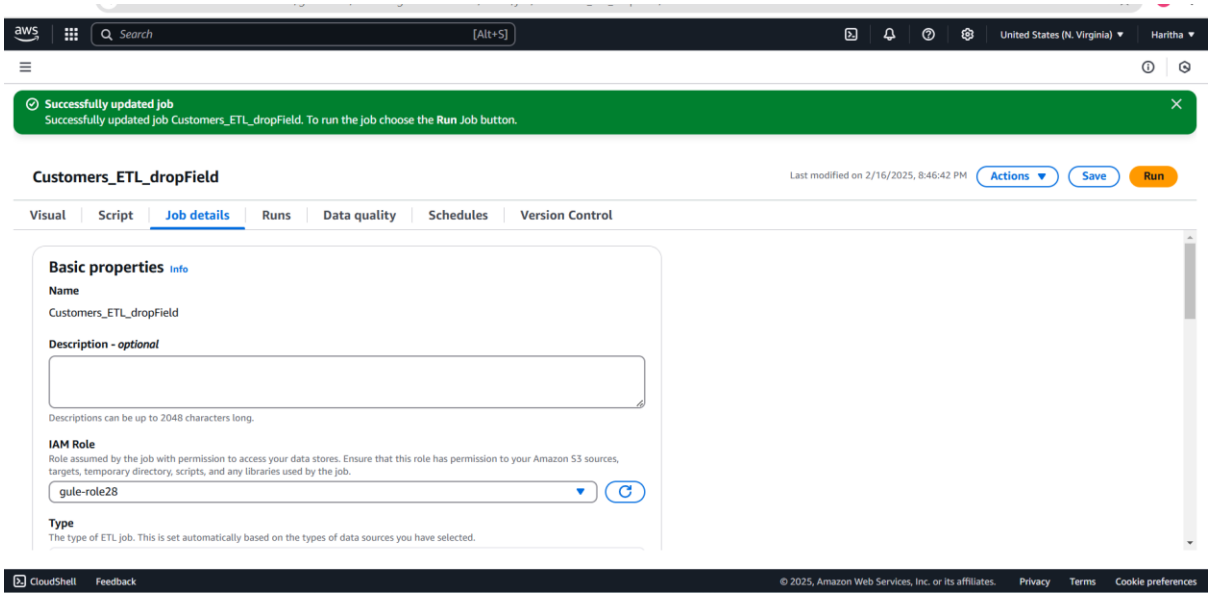

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Connectors
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
AWS Marketplace
Enable compact mode
Enable new navigation
Marketplace connectors
Subscribe to connectors from AWS partners to expand your data sources.
Go to AWS Marketplace
Custom connectors
Provide your own connector to expand your data sources. Creating custom connectors
Create custom connector
Connectors (0) Info
You can manage your connectors or use them to create connections.
Actions
Filter connections by property
Name
Status
Type
Last modified
No connectors
No connectors to display. You can create a custom connector or get a Marketplace connector.
Create custom connector
Connections (0) Info
You can manage your connections or use a connection in a job.
Actions
Create connection
Create job
Filter connections by property
Name
Status
Type
Last modified
Version
No connections
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

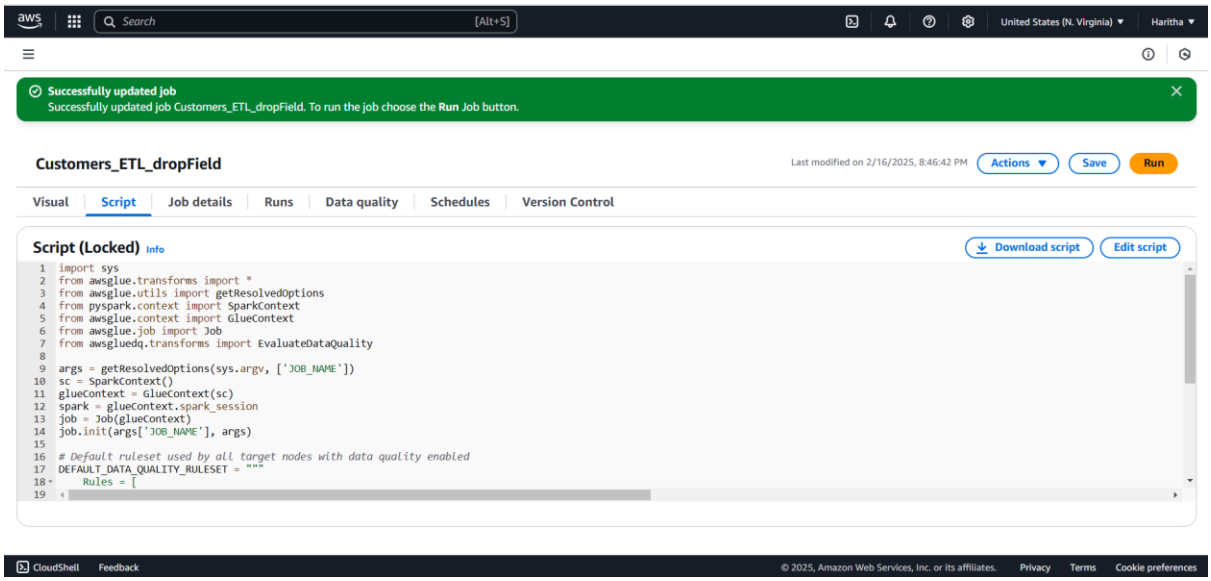

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Connectors > Create connection
Step 2
Configure connection
Step 3
Set properties
Step 4
Review and create
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
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
Data sources (72)
Find data sources
Grid
List
1 2 3 4 5
Adobe Analytics - new
Adobe Analytics is a business intelligence solution for combining and analyzing data from any digital point in the customer journey.
Amazon Aurora
Connect to Amazon Aurora.
Learn more
Amazon DocumentDB - new
Connect to Amazon DocumentDB.
Learn more
Amazon OpenSearch Service - new
Connect to Amazon OpenSearch Service.
Learn more
Amazon Redshift
Connect to Amazon Redshift.
Learn more
Apache Kafka
Connect to streaming data in Apache Kafka.
Learn more
Asana - new
Azure Cosmos - new
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

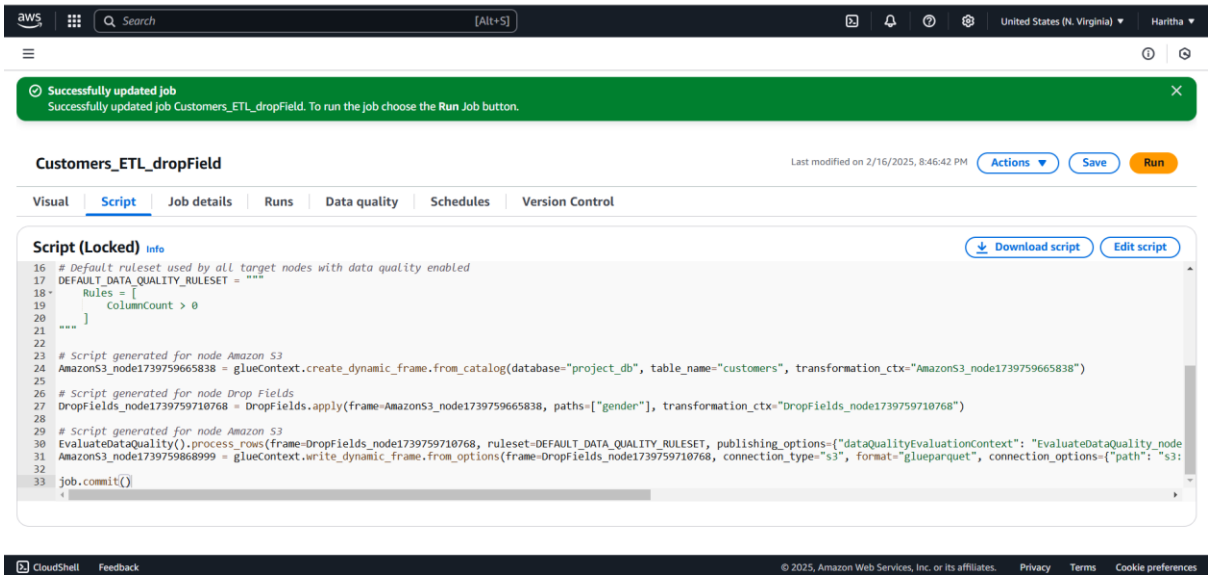

us-east-1.console.aws.amazon.com/gluestudio/home?region=us-east-1#/jobs
Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Jobs
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
What's New
Documentation
AWS Glue Studio Info
Create job Info
Author in a visual interface focused on data flow.
Visual ETL
Author using an interactive code notebook.
Notebook
Author code with a script editor.
Script editor
Example jobs Info
Create example job
Your jobs (0) Info
Actions
Run job
Filter jobs by property
Job name
Type
Created by
Last modified
AWS Glue version
No jobs
You have not created a job yet.
Create job from a blank graph
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

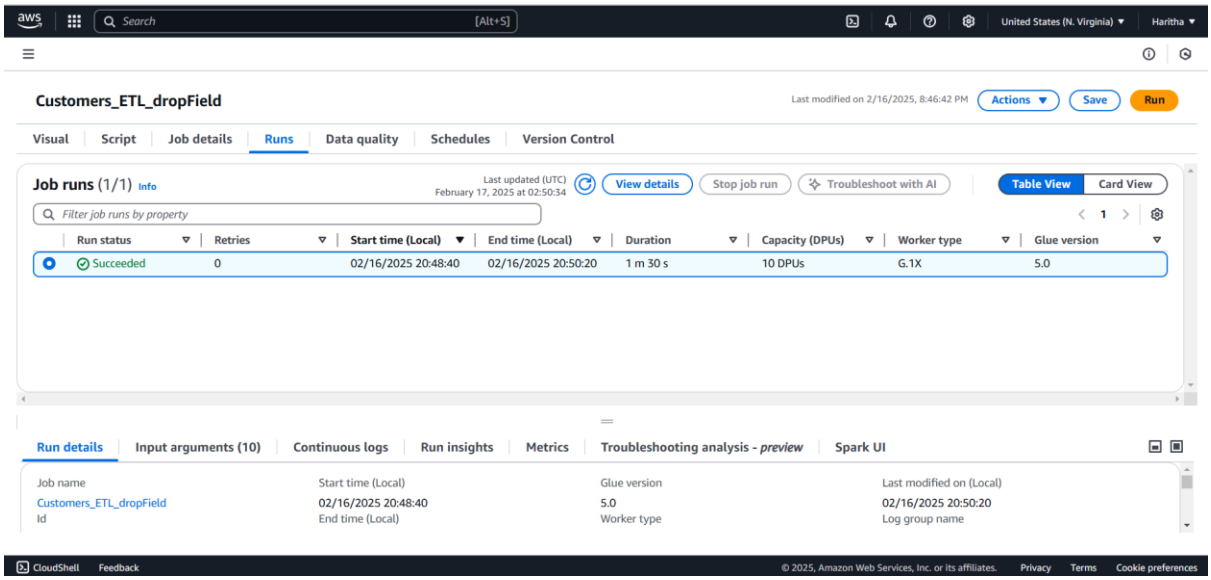

Untitled job
Job has not been saved
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Choose an S3 bucket
S3 buckets
glue-bucket28
Objects (1/3)
Find object by prefix
Last updated
February 16, 2025, 20:43:39 (UTC-6)
Key
landing_zone/
query_results/
transform_zone/
Cancel
Choose
Create IAM role.
Additional Settings
Partition keys - optional
Add partition keys.
Add a partition key

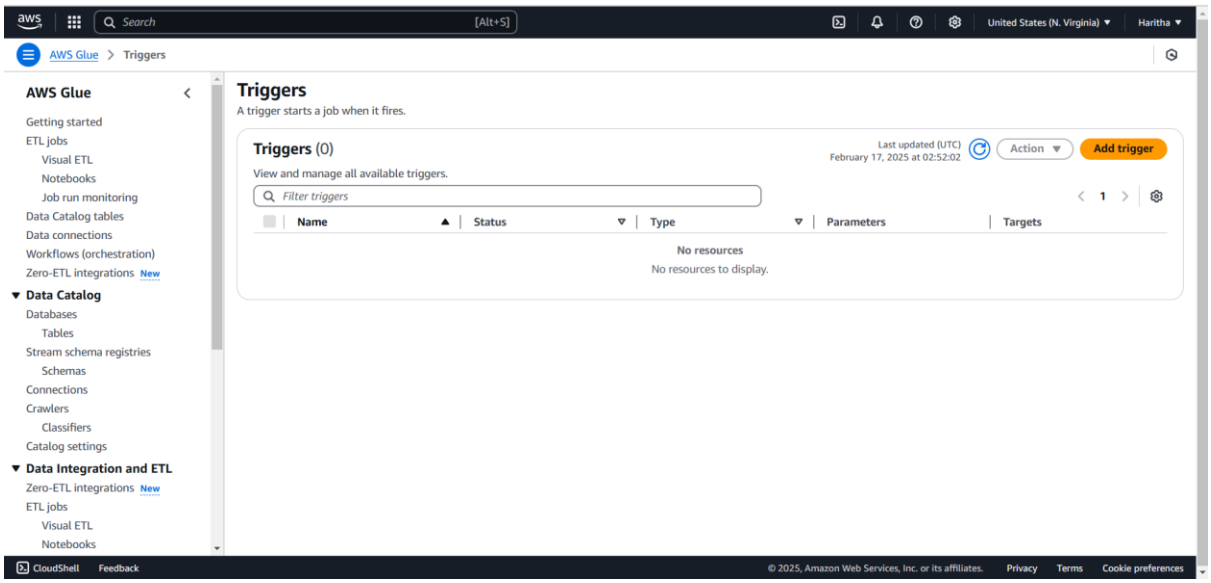

us-east-1.console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/details
Customers_ETL_dropField
Job has not been saved
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Basic properties
Info
Name
Customers_ETL_dropField
Description - optional
Descriptions can be up to 2048 characters long.
IAM Role
Role assumed by the job with permission to access your data stores. Ensure that this role has permission to your Amazon S3 sources, targets, temporary directory, scripts, and any libraries used by the job.
gule-role28
Type
The type of ETL job. This is set automatically based on the types of data sources you have selected.
Spark
Glue version
Info
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

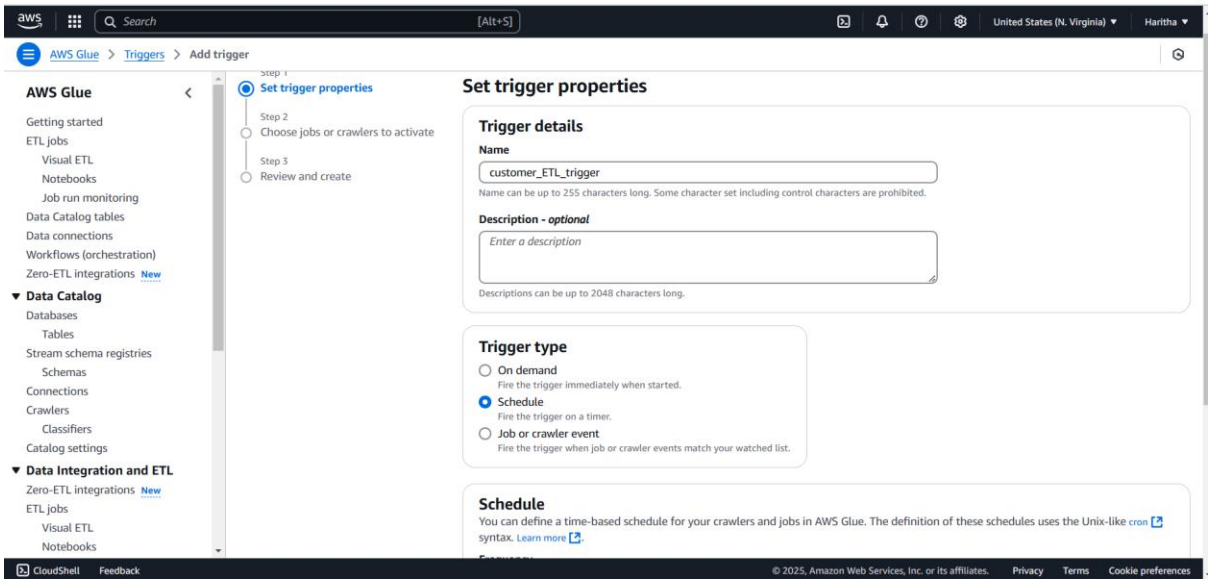

Successfully updated job
Successfully updated job Customers_ETL_dropField. To run the job choose the Run Job button.
Customers_ETL_dropField
Last modified on 2/16/2025, 8:46:42 PM
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Basic properties
Info
Name
Customers_ETL_dropField
Description - optional
Descriptions can be up to 2048 characters long.
IAM Role
Role assumed by the job with permission to access your data stores. Ensure that this role has permission to your Amazon S3 sources, targets, temporary directory, scripts, and any libraries used by the job.
gule-role28
Type
The type of ETL job. This is set automatically based on the types of data sources you have selected.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

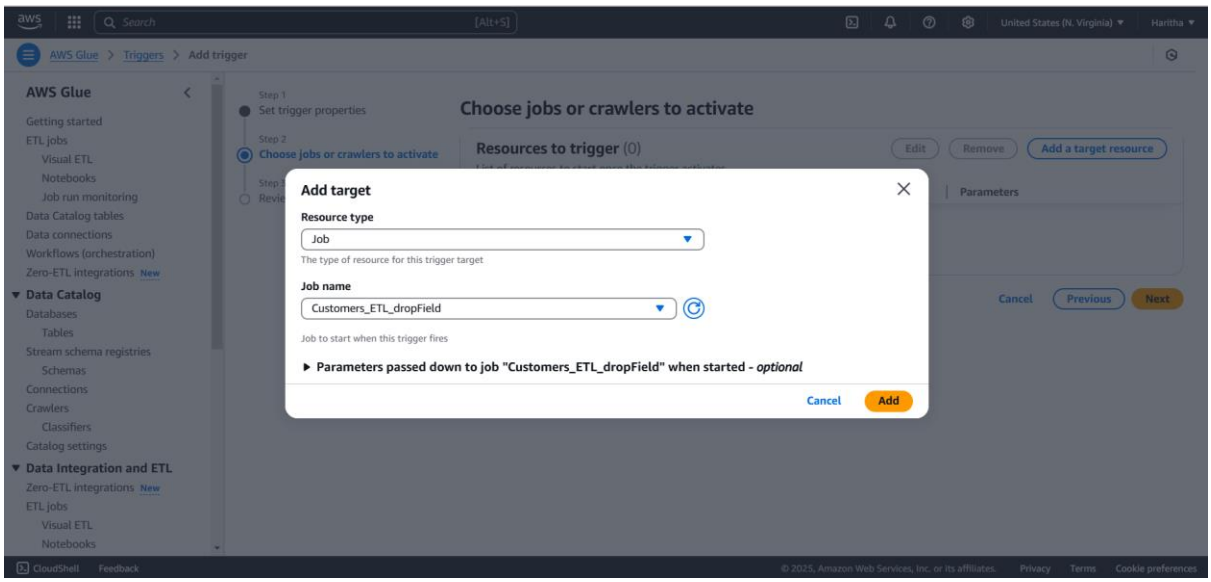

Search [Alt+S]
United States (N. Virginia)
Haritha
Successfully updated job
Successfully updated job Customers_ETL_dropField. To run the job choose the Run Job button.
Customers_ETL_dropField
Last modified on 2/16/2025, 8:46:42 PM
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Script (Locked) Info
Download script
Edit script
import sys
from awsglue.transforms import *
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job
from awsgluedq.transforms import EvaluateDataQuality
args = getResolvedOptions(sys.argv, ['JOB_NAME'])
sc = SparkContext()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
job.init(args['JOB_NAME'], args)
# Default ruleset used by all target nodes with data quality enabled
DEFAULT_DATA_QUALITY_RULESET = """
Rules = [
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

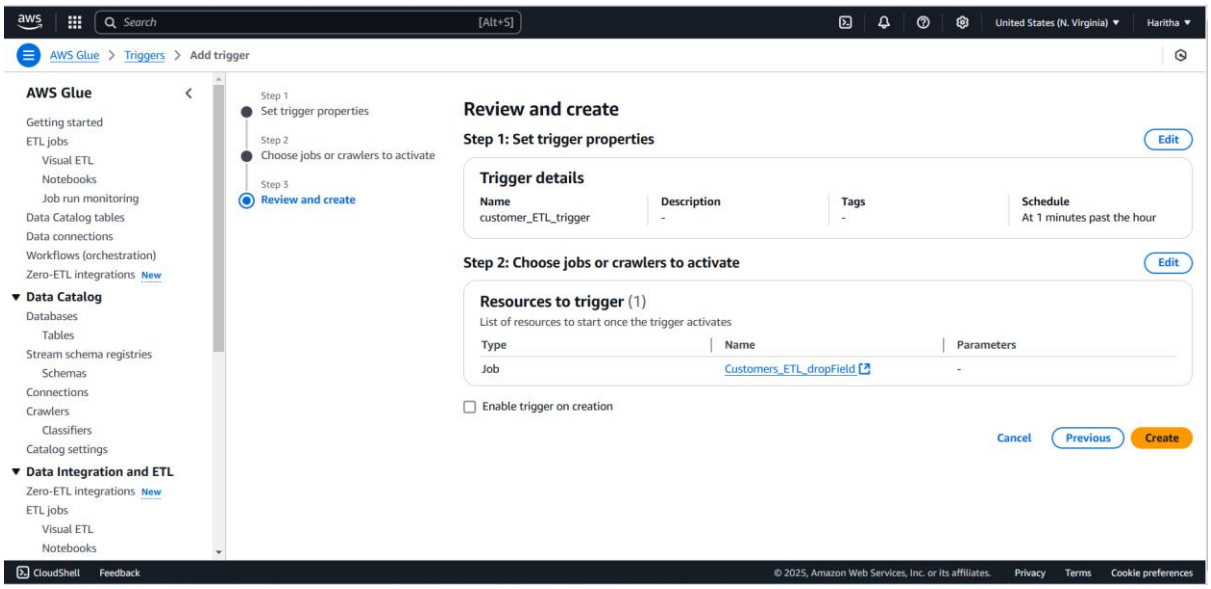

Search [Alt+S]
United States (N. Virginia)
Haritha
Successfully updated job
Successfully updated job Customers_ETL_dropField. To run the job choose the Run Job button.
Customers_ETL_dropField
Last modified on 2/16/2025, 8:46:42 PM
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Script (Locked) Info
Download script
Edit script
# Default ruleset used by all target nodes with data quality enabled
DEFAULT_DATA_QUALITY_RULESET = """
Rules = [
ColumnCount > 0
]
"""
# Script generated for node Amazon S3
AmazonS3_node1739759665838 = glueContext.create_dynamic_frame.from_catalog(database="project_db", table_name="customers", transformation_ctx="AmazonS3_node1739759665838")
# Script generated for node Drop Fields
DropFields_node1739759710768 = DropFields.apply(frame=AmazonS3_node1739759665838, paths=["gender"], transformation_ctx="DropFields_node1739759710768")
# Script generated for node Amazon S3
EvaluateDataQuality().process_rows(frame=DropFields_node1739759710768, ruleset=DEFAULT_DATA_QUALITY_RULESET, publishing_options={"dataQualityEvaluationContext": "EvaluateDataQuality_node
AmazonS3_node1739759868999 = glueContext.write_dynamic_frame.from_options(frame=DropFields_node1739759710768, connection_type="s3", format="glueparquet", connection_options={"path": "s3:
job.commit()
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

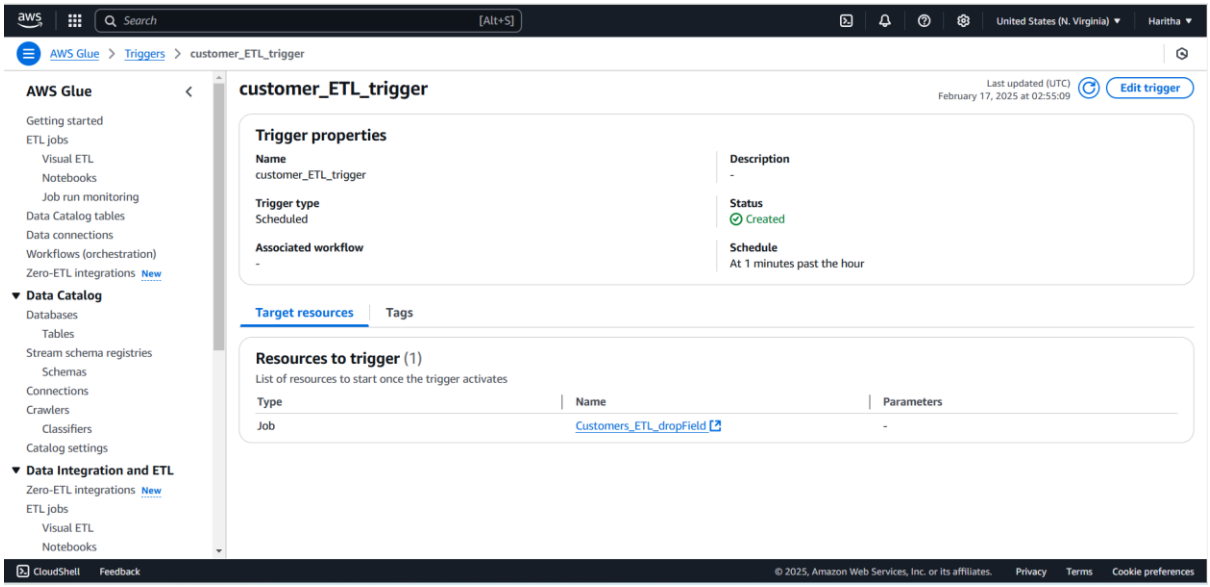

Search [Alt+S]
United States (N. Virginia)
Haritha
Customers_ETL_dropField
Last modified on 2/16/2025, 8:46:42 PM
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Job runs (1/1) Info
Last updated (UTC)
February 17, 2025 at 02:50:34
View details
Stop job run
Troubleshoot with AI
Table View
Card View
Filter job runs by property
Run status
Retries
Start time (Local)
End time (Local)
Duration
Capacity (DPUs)
Worker type
Glue version
Succeeded
0
02/16/2025 20:48:40
02/16/2025 20:50:20
1 m 30 s
10 DPUs
G.1X
5.0
Run details
Input arguments (10)
Continuous logs
Run insights
Metrics
Troubleshooting analysis - preview
Spark UI
Job name
Customers_ETL_dropField
Id
Start time (Local)
02/16/2025 20:48:40
End time (Local)
Glue version
5.0
Worker type
Last modified on (Local)
02/16/2025 20:50:20
Log group name
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

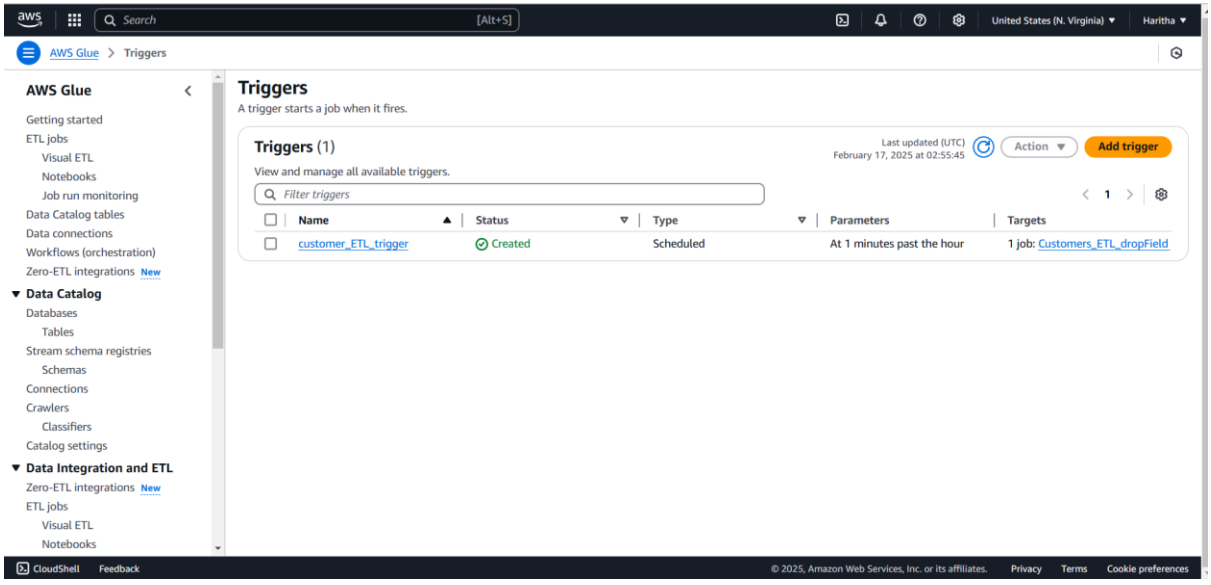

Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Triggers
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Zero-ETL integrations New
ETL jobs
Visual ETL
Notebooks
Triggers
A trigger starts a job when it fires.
Triggers (0)
Last updated (UTC)
February 17, 2025 at 02:52:02
Action
Add trigger
View and manage all available triggers.
Filter triggers
Name
Status
Type
Parameters
Targets
No resources
No resources to display.
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Triggers > Add trigger
AWS Glue
Step 1
Set trigger properties
Step 2
Choose jobs or crawlers to activate
Step 3
Review and create
Set trigger properties
Trigger details
Name
customer_ETL_trigger
Name can be up to 255 characters long. Some character set including control characters are prohibited.
Description - optional
Enter a description
Descriptions can be up to 2048 characters long.
Trigger type
On demand
Fire the trigger immediately when started.
Schedule
Fire the trigger on a timer.
Job or crawler event
Fire the trigger when job or crawler events match your watched list.
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
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Triggers > Add trigger
AWS Glue
Step 1
Set trigger properties
Step 2
Choose jobs or crawlers to activate
Choose jobs or crawlers to activate
Resources to trigger (0)
Edit
Remove
Add a target resource
Parameters
Cancel
Previous
Next
Add target
Resource type
Job
The type of resource for this trigger target
Job name
Customers_ETL_dropField
Job to start when this trigger fires
Parameters passed down to job "Customers_ETL_dropField" when started - optional
Cancel
Add
CloudShell
Feedback


Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Triggers > Add trigger
AWS Glue
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
Zero-ETL integrations New
ETL jobs
Visual ETL
Notebooks
Step 1
Set trigger properties
Step 2
Choose jobs or crawlers to activate
Step 3
Review and create
Review and create
Step 1: Set trigger properties
Edit
Trigger details
Name
customer_ETL_trigger
Description
-
Tags
-
Schedule
At 1 minutes past the hour
Step 2: Choose jobs or crawlers to activate
Edit
Resources to trigger (1)
List of resources to start once the trigger activates
Type
Name
Parameters
Job
Customers_ETL_dropField
-
Enable trigger on creation
Cancel
Previous
Create
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Triggers > customer_ETL_trigger
AWS Glue
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
Zero-ETL integrations New
ETL jobs
Visual ETL
Notebooks
customer_ETL_trigger
Last updated (UTC)
February 17, 2025 at 02:55:09
Edit trigger
Trigger properties
Name
customer_ETL_trigger
Description
-
Trigger type
Scheduled
Status
Created
Associated workflow
-
Schedule
At 1 minutes past the hour
Target resources
Tags
Resources to trigger (1)
List of resources to start once the trigger activates
Type
Name
Parameters
Job
Customers_ETL_dropField
-
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Search [Alt+S]
United States (N. Virginia)
Haritha
AWS Glue > Triggers
AWS Glue
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
Zero-ETL integrations New
ETL jobs
Visual ETL
Notebooks
Triggers
A trigger starts a job when it fires.
Triggers (1)
Last updated (UTC)
February 17, 2025 at 02:55:45
Action
Add trigger
View and manage all available triggers.
Filter triggers
1
Name
Status
Type
Parameters
Targets
customer_ETL_trigger
Created
Scheduled
At 1 minutes past the hour
1 job: Customers_ETL_dropField
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences