

The screenshot shows the AWS Home page with a dark theme. On the left, there's a sidebar with links to Services, Features, Resources, Documentation, Knowledge articles, Marketplace, Blog posts, Events, and Tutorials. The main content area displays three service cards: S3 Scalable Storage in the Cloud, S3 Glacier Archive Storage in the Cloud, and AWS Snow Family Large Scale Data Transport. Below these are sections for Features, Imports from S3 (DynamoDB feature), Feature spotlight (S3 feature), and S3 Access Grants (S3 feature). To the right, there's a sidebar for creating applications, showing a search bar, a list of applications, and a button to 'Create application'.

The screenshot shows the Amazon S3 Buckets page. On the left, there's a sidebar with links to 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, and Block Public Access settings for this account. Under Storage Lens, there are links to Dashboards, Storage Lens groups, and AWS Organizations settings. The main content area shows an account snapshot updated every 24 hours, a table of General purpose buckets (3), and a 'Create bucket' button. The table lists three buckets: haritha-bucket, haritha28, and my-unique-bucket-1234566, with details like Region, IAM Access Analyzer, and Creation date.

The screenshot shows the Create bucket page. It starts with a 'General configuration' section where the AWS Region is set to US East (N. Virginia) us-east-1 and the Bucket type is set to General purpose. The bucket name is entered as glue-bucket28. There's a 'Copy settings from existing bucket - optional' section with a 'Choose bucket' button. The next section is 'Object Ownership', which controls ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). At the bottom, there are links to CloudShell and Feedback, and standard footer links for Privacy, Terms, and Cookie preferences.

Screenshot of the AWS IAM Roles page:

Identity and Access Management (IAM)

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.

Role name	Trusted entities	Last activity
AWSServiceRoleForOrganizations	AWS Service: organizations (Service-Linked Role)	-
AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role)	15 hours ago
AWSServiceRoleForSupport	AWS Service: support (Service-Linked Role)	-
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service-Linked Role)	-

Roles Anywhere Info

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>

Screenshot of the "Create role" wizard Step 2: Trusted entity type:

Step 2

- Add permissions
- 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](#)

Screenshot of the "Create role" wizard Step 2: Add permissions:

Step 1

- Select trusted entity

Step 2

- Add permissions
- 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

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 permit...
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 bu...
AmazonS3TablesReadOnlyAccess	AWS managed	Provides read only access to all S3 tabl...
AWSBackupServiceRolePolicyForS3Backup	AWS managed	Policy containing permissions necessar...

[CloudShell](#) [Feedback](#)

Add permissions

Permissions policies (2/1032) Info

Choose one or more policies to attach to your new role.

Filter by Type 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...
AWSSyncPushToCloudWatchLogs	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

Name, review, and create

Role details

Role name
Enter a meaningful name to identify this role.
glue-role2@

Description
Add a short explanation for this role.
Allows Glue to call AWS services on your behalf.

Step 1: Select trusted entities

Trust policy

```

1- [
2-   "Version": "2012-10-17",
3-   "Statement": [
4-     {
5-       "Effect": "Allow",
6-       "Principal": {
7-         "Service": "glue.amazonaws.com"
8-       },
9-       "Action": "sts:AssumeRole"
10-    }
11- ]
12- ]

```

Step 2: Add permissions

Permissions policy summary

Policy name	Type	Attached as
AmazonS3FullAccess	AWS managed	Permissions policy
CloudWatchLogsFullAccess	AWS managed	Permissions policy

Step 3: Add tags

us-east-1.console.aws.amazon.com/s3/buckets/glue-bucket28?region=us-east-1&bucketType=general&tab=objects

Amazon S3 > Buckets > glue-bucket28

glue-bucket28 Info

Objects | Metadata | Properties | Permissions | Metrics | Management | Access Points

Objects (0)

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

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

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 /

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

us-east-1.console.aws.amazon.com/s3/buckets/glue-bucket28

Successfully created folder "landing_zone".

glue-bucket28 Info

Objects | Metadata | Properties | Permissions | Metrics | Management | Access Points

Objects (1)

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	-	-	-

Screenshot of the AWS S3 console showing the 'landing_zone/' folder in the 'glue-bucket28' bucket. The 'Objects (0)' section is displayed, with a search bar and various actions like Copy S3 URI, Copy URL, Download, Open, Delete, Actions, Create folder, and Upload. A message states 'No objects' and 'You don't have any objects in this folder.' A prominent 'Upload' button is at the bottom.

Screenshot of the AWS S3 console showing the 'Create folder' dialog for the 'landing_zone/' folder. It includes a note about bucket policy restrictions and a 'Your bucket policy might block folder creation' section. The 'Folder' section has a 'Folder name' input field containing 'customer'. The 'Server-side encryption' section notes that server-side encryption protects data at rest. A note states that encryption settings apply only to the folder object and not to sub-folder objects. The 'Server-side encryption' section shows two options: 'Don't specify an encryption key' (selected) and 'Specify an encryption key'. The selected option indicates default encryption is used.

Screenshot of the AWS S3 console showing the 'Create folder' dialog for the 'landing_zone/' folder. It includes a note about bucket policy restrictions and a 'Your bucket policy might block folder creation' section. The 'Folder' section has a 'Folder name' input field containing 'customer'. The 'Server-side encryption' section notes that server-side encryption protects data at rest. A note states that encryption settings apply only to the folder object and not to sub-folder objects. The 'Server-side encryption' section shows two options: 'Don't specify an encryption key' (selected) and 'Specify an encryption key'. The selected option indicates default encryption is used.

aws | Search [Alt+S] United States (N. Virginia) Haritha

Amazon S3 > Buckets > glue-bucket28 > landing_zone/ Copy S3 URI

landing_zone/

Objects Properties

Objects (1)

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	-	-	-

Actions Create folder Upload

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws | Search [Alt+S] United States (N. Virginia) Haritha

Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > customers/ Copy S3 URI

customers/

Objects Properties

Objects (0)

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

aws | 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 /
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.

aws | ⚙️ | Search [Alt+S] | United States (N. Virginia) | Haritha

Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > customers/

Successfully created folder "load_date=20230603".

customers/

Objects Properties

Objects (1)

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](#)

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 Microsoft Excel Comma Separate... 99 bytes

Yesterday

- VSCodeUserSetup-x64-1.97.2.exe Visual Studio Code Setup Microsoft Corporation
- cl-user_accessKeys.csv Microsoft Excel Comma Separate... 99 bytes
- 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

No files or folders You have not chosen any files or folders to upload.

Upload

Amazon S3 REST API. [Learn more](#)

Upload here, or choose Add files or Add folder.

Add files Add folder Remove

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.

cloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 Search [Alt+S] United States (N. Virginia) Haritha

Amazon S3 > Buckets > glue-bucket28 > landing_zone/ > customers/ > load_date=20230603/

load_date=20230603/ Copy S3 URI

Objects Properties

Objects (1)

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.csv	csv	February 16, 2025, 19:31:00 (UTC-06:00)	3.9 KB	Standard

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws Search [Alt+S] United States (N. Virginia) Haritha

AWS Glue > Databases

Databases (0)

A database is a set of associated table definitions, organized into a logical group.

Filter databases

Name	Description	Location URI	Created on (UTC)
No resources	No resources to display.		

Last updated (UTC) February 17, 2025 at 01:35:10 Edit Delete Add database

< 1 > ⚙️

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

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws Search [Alt+S] United States (N. Virginia) Haritha

AWS Glue > Databases > Add database

Create a database

Create a database in the AWS Glue Data Catalog.

Database details

Name Database name is required, in lowercase characters, and no longer than 255 characters.

Description - optional

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

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

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue > Databases

Databases (1)

A database is a set of associated table definitions, organized into a logical group.

Name	Description	Location URI
project_db		

Last updated (UTC) February 17, 2025 at 01:34:44 | Edit | Delete | Add database

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 CloudShell Feedback

AWS Glue > Databases > project_db

project_db

Announcing new optimization features for Apache Iceberg tables
Optimize storage for Apache Iceberg tables with automatic snapshot retention and orphan file deletion. Learn more

Database properties

Name	Description	Location	Created on (UTC)
project_db			February 17, 2025 at 01:34:41

Tables (0)

No available tables

Last updated (UTC) February 17, 2025 at 01:35:45 | Delete | Add tables using crawler | Add table

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 CloudShell Feedback

AWS Glue > Tables > Add table

Set table properties

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

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.

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Screenshot of the AWS Glue Data Catalog interface showing the creation of a new table. A modal window titled "Choose S3 path" is open, displaying a list of four S3 buckets: "glue-bucket28", "haritha-bucket", "haritha28", and "my-unique-bucket-1234566". The "Classification" section shows "JSON" selected. The "Data format" section also shows "JSON" selected.

Screenshot of the AWS Glue Data Catalog interface showing the creation of a new table. A modal window titled "Choose S3 path" is open, displaying a list of objects under "glue-bucket28/landing_zone/". One object, "customers", is selected. The "Classification" section shows "JSON" selected. The "Data format" section also shows "JSON" selected.

Screenshot of the AWS Glue Data Catalog interface showing the creation of a new table. The "Data location is specified in" section indicates "my account". The "Include path" field contains "s3://glue-bucket28/landing_zone/custom". The "Data format" section shows "CSV" selected. The "Delimiter" field contains "Comma (,)".

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 [New](#)

Documentation [Feedback](#)

CloudShell Feedback

Search [Alt+S]

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.

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

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 [New](#)

Documentation [Feedback](#)

CloudShell Feedback

Search [Alt+S]

project_db

Last updated (UTC) February 17, 2025 at 01:43:21

Database properties

Name	Description	Location	Created on (UTC)
project_db	-	-	February 17, 2025 at 01:34:41

Tables (0)
View and manage all available tables.

Filter tables

Name	Database	Location	Classification	Deprecated	View data	Data quality	Column stats...
No available tables							

Last updated (UTC) February 17, 2025 at 01:43:23

Delete Add tables using crawler Add table

Cancel Edit Delete

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 [New](#)

Documentation [Feedback](#)

CloudShell Feedback

Search [Alt+S]

Set crawler properties

Crawler details [Info](#)

Name
Name can be up to 255 characters long. Some character set including control characters are prohibited.

Description - optional

Descriptions can be up to 2048 characters long.

Tags - optional
Use tags to organize and identify your resources.

Cancel Next

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 [New](#)

Documentation [New](#)

CloudShell Feedback

Search

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](#)

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](#)

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 [New](#)

Documentation [New](#)

CloudShell Feedback

Search

Step 1 Set crawler properties

Step 2 Choose data sources

Step 3 Configure security settings

Step 4 Set output and scheduling

Step 5 Review and create

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](#)

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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 [New](#)

Documentation [New](#)

CloudShell Feedback

Search

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.

Type	Data source	Parameters
<input type="radio"/> 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](#)

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue > Crawlers > Add crawler

Configure security settings

IAM role Info
Existing IAM role
glue-role28 [View](#) [Edit](#)

[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](#)

AWS Glue > Crawlers > Add crawler

Set output and scheduling

Output configuration Info
Target database
project_db [View](#) [Edit](#)

[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

us-east-1.console.aws.amazon.com/glue/home?region=us-east-1#/v2/data-catalog/crawlers/add

AWS Glue > Crawlers > Add crawler

Review and create

Step 1: Set crawler properties [Edit](#)

Set crawler properties

Name	Description	Tags
customers	-	-

Step 2: Choose data sources and classifiers [Edit](#)

Data sources (1) Info
The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
S3	s3://glue-bucket28/landing_zone/cust...	Recrawl all

Step 3: Configure security settings [Edit](#)

Configure security settings

IAM role	Security configuration	Lake Formation configuration
glue-role28	-	-

Step 4: Set output and scheduling [Edit](#)

Set output and scheduling

AWS Glue > Crawlers > customers

One crawler successfully created
The following crawler is now created: "customers"

Last updated (UTC) February 17, 2025 at 01:54:57

Run crawler Edit Delete

Crawler properties

Name	IAM role glue-role28	Database	State
customers	gule-role28	project_db	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.

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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	-	-	-	-	-

< 1 > | ⚙

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue > Crawlers > customers

customers

Last updated (UTC) February 17, 2025 at 01:54:57

Run crawler Edit Delete

Crawler properties

Name	IAM role	Database	State
customers	glue-role28	project_db	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.

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

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue > Crawlers > customers

customers

Last updated (UTC) February 17, 2025 at 01:54:57

Crawler properties

Name	customers	IAM role	glue-role28	Database	project_db	State	READY
Description	-	Security configuration	-	Lake Formation configuration	-	Table prefix	-
Maximum table threshold	-						

Advanced settings

Crawler runs (2)

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	-	-

Documentation **Feedback**

AWS Glue > Tables

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)

Name	Database	Location	Classification	Deprecated	View data	Data quality	Column stats...
customers	project_db	s3://glue-bucket28/	CSV	-	Table data	View data quality	View statistics

Documentation **Feedback**

AWS Glue > Databases > project_db

project_db

Last updated (UTC) February 17, 2025 at 02:10:49

Database properties

Name	project_db	Description	-	Location	-	Created on (UTC)	February 17, 2025 at 01:34:41
------	------------	-------------	---	----------	---	------------------	-------------------------------

Tables (1)

Name	Database	Location	Classification	Deprecated	View data	Data quality	Column stats...
customers	project_db	s3://glue-bucket28/	CSV	-	Table data	View data quality	View statistics

Documentation **Feedback**

AWS Glue > Tables > customers

customers

Last updated (UTC) February 17, 2025 at 02:11:59 Version 1 (Current version) Actions

Table details

Name	customers	Classification	CSV
Database	project_db	Location	s3://glue-bucket28/landing_zone/customers/
Description	-	Connection	-
Last updated	February 17, 2025 at 02:04:54	Advanced properties	

Schema | Partitions | Indexes | Column statistics - new

Schema (6)

View and manage the table schema.

#	Column name	Data type	Partition key	Comment
1	customerid	bigint		
2	gender	string		
3	age	bigint		
4	annual income (k\$)	bigint		
5	spending score (1-100)	bigint		
6	load_date	string (Partitioned)		

Edit schema as JSON | Edit schema

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws Search [Alt+S] United States (N. Virginia) Haritha

Amazon Athena > Query editor

Catalog

Tables and views

Tables (1)

- customers

Views (0)

SQL Ln 1, Col 1

Run Query Preview Table Generate table DDL Load partitions Insert Insert into editor

No results Run a query to view results

Reuse query results up to 60 minutes ago

Copy Download results CSV

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws 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

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

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Screenshot of the Amazon Athena Settings page. A green success message at the top says "Settings successfully updated." The "Settings" tab is selected. The "Query result location" is set to "s3://glue-bucket28/query_results/" with a "Manage" button. The "Encrypt query results" section shows "Expected bucket owner" as "-" and "Assign bucket owner full control over query results" as "Turned off". The "Workgroup" dropdown is set to "primary".

Screenshot of the Amazon Athena Query editor. The left sidebar shows "Tables and views" with one table "customers" and no views. The main area displays a SQL query "SQL Ln 1, Col 1" with options to "Run again", "Explain", "Cancel", "Clear", and "Create". Below the query is a "Query results" table showing 10 rows of customer data. The columns are: #, customerid, gender, age, annual income (k\$), spending score (1-100), and load_date. The data is as follows:

#	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
6	6	Female	22	17	76	20230603

Screenshot of the AWS Glue Connectors page. The left sidebar includes sections for "Data Catalog", "Data Integration and ETL", and "Legacy pages". The main area has two main sections: "Marketplace connectors" (with a "Go to AWS Marketplace" button) and "Custom connectors" (with a "Create custom connector" button). Below these are two tables: "Connectors (0) info" and "Connections (0) info", both showing no results.

AWS Glue > Connectors > Create connection

Step 2: Configure connection

Step 3: Set properties

Step 4: Review and create

Data sources (72)

Find data sources

Grid List

1 2 3 4 5 >

Adobe Analytics - new

Amazon Aurora

Amazon DocumentDB - new

Amazon OpenSearch Service - new

Amazon Redshift

Apache Kafka

Asana - new

Azure Cosmos - new

Learn more

CloudShell Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue > Jobs

AWS Glue Studio

Create job

Author in a visual interface focused on data flow.

Author using an interactive code notebook.

Author code with a script editor.

Visual ETL Notebook Script editor

Example jobs

Your jobs (0)

Filter jobs by property

Actions Run job

1 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

Key

landing_zone/

query_results/

transform_zone/

Last updated February 16, 2025, 20:43:39 (UTC-6)

Create a table in the Data Catalog and on subsequent runs, keep existing schema and add new partitions

Partition keys - optional Add partition keys

Cancel Choose

Create IAM role

Additional Settings

Start session

Add a partition key

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#/editor/blank/details

aws Search [Alt+S] United States (N. Virginia) Haritha

Customers_ETL_dropField Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Basic properties Info

Name Customers_ETL_dropField

Description - optional

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.

Type The type of ETL job. This is set automatically based on the types of data sources you have selected.

Spark

Glue version Info

Glue 5.0 - Supports spark 3.5, Scala 2, Python 3

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 Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Basic properties Info

Name Customers_ETL_dropField

Description - optional

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.

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

Successfully updated job Successfully updated job Customers_ETL_dropField. To run the job choose the Run Job button.

Customers_ETL_dropField Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Script (Locked) Info

Download script Edit script

```
1 import sys
2 from awsglue.transforms import *
3 from awsglue.utils import getResolvedOptions
4 from pyspark.context import SparkContext
5 from awsglue.context import GlueContext
6 from awsglue.job import Job
7 from awsgluedq.transforms import EvaluateDataQuality
8
9 args = getResolvedOptions(sys.argv, ['JOB_NAME'])
10 sc = SparkContext()
11 glueContext = GlueContext(sc)
12 spark = glueContext.spark_session
13 job = Job(glueContext)
14 job.init(args['JOB_NAME'], args)
15
16 # Default ruleset used by all target nodes with data quality enabled
17 DEFAULT_DATA_QUALITY_RULESET = """
18 *     Rules = [
19 *
```

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

```

16 # Default ruleset used by all target nodes with data quality enabled
17 DEFAULT_DATA_QUALITY_RULESET = """
18     Rules = [
19         ColumnCount > 0
20     ]
21 """
22
23 # Script generated for node Amazon S3
24 AmazonS3_node1739759665838 = glueContext.create_dynamic_frame.from_catalog(database="project_db", table_name="customers", transformation_ctx="AmazonS3_node1739759665838")
25
26 # Script generated for node Drop Fields
27 DropFields_node1739759710768 = DropFields.apply(frame=AmazonS3_node1739759665838, paths=["gender"], transformation_ctx="DropFields_node1739759710768")
28
29 # Script generated for node Amazon S3
30 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://project-db/customers.parquet", "partitionKeys": ["year", "month"]})"})
31
32 job.commit()
33

```

[Download script](#) [Edit script](#)

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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](#)

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

[Table View](#) [Card View](#)

Run details [Input arguments \(10\)](#) [Continuous logs](#) [Run insights](#) [Metrics](#) [Troubleshooting analysis - preview](#) [Spark UI](#)

Job name Customers_ETL_dropField	Start time (Local) 02/16/2025 20:48:40	Glue version 5.0	Last modified on (Local) 02/16/2025 20:50:20
Id	End time (Local)	Worker type	Log group name

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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
- 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](#)

Name	Status	Type	Parameters	Targets
No resources	No resources to display.			

Screenshot of the AWS Glue Trigger creation process - Step 1: Set trigger properties.

Trigger details

- Name:** customer_ETL_trigger
- Description - optional:** Enter a description

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.

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Screenshot of the AWS Glue Trigger creation process - Step 2: Choose jobs or crawlers to activate.

Resources to trigger (0)

Add target

Resource type: Job

Job name: Customers_ETL_dropField

Parameters passed down to job "Customers_ETL_dropField" when started - optional

Cancel Previous Next

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Screenshot of the AWS Glue Trigger creation process - Step 3: Review and create.

Step 1: Set trigger properties

Trigger details

- Name:** customer_ETL_trigger
- Description:** -
- Schedule:** At 1 minutes past the hour

Step 2: Choose jobs or crawlers to activate

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

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

S | Search [Alt+S] | United States (N. Virginia) | Haritha

AWS Glue > Triggers > customer_ETL_trigger

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
Associated workflow	-	Schedule

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

S | Search [Alt+S] | United States (N. Virginia) | Haritha

AWS Glue > Triggers

Triggers

A trigger starts a job when it fires.

Triggers (1)

Last updated (UTC) February 17, 2025 at 02:55:45 | Action | Add trigger

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