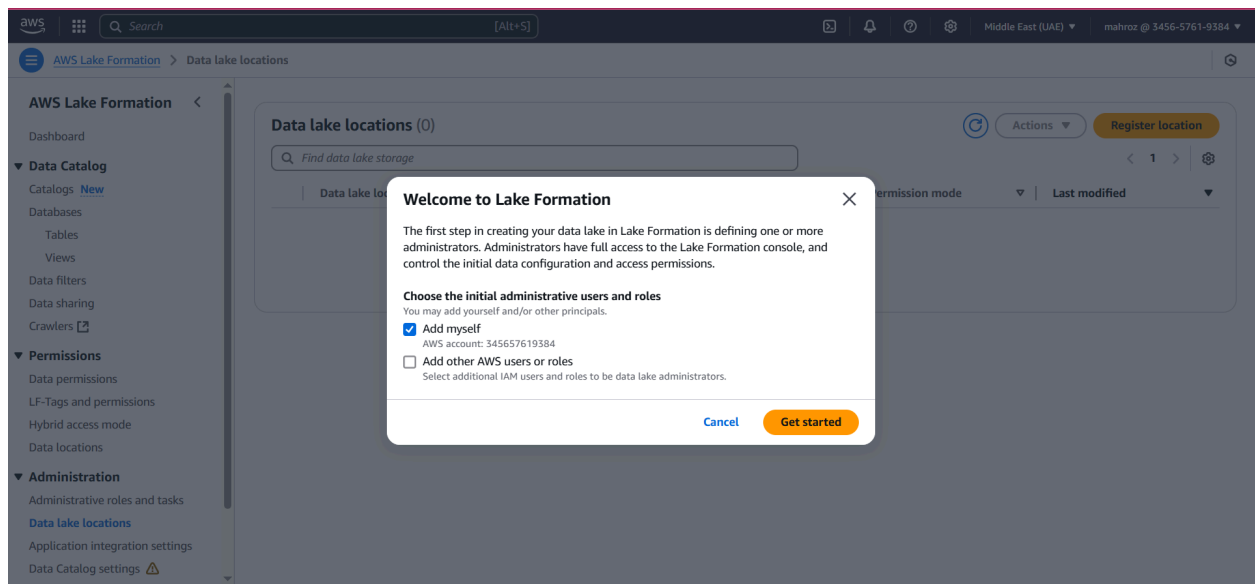


Data Governance using AWS

1. Enable Lake Formation & Register S3 Buckets



aws

Search

[Alt+S]

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Lake Formation

> Data lake locations

▼ Data Catalog

Catalogs [New](#)

Databases

Tables

Views

Data filters

Data sharing

Crawlers [↗](#)

▼ Permissions

Data permissions

LF-Tags and permissions

Hybrid access mode

Data locations

▼ Administration

Administrative roles and tasks

[Data lake locations](#)

Application integration settings

Data Catalog settings ⚠

▼ Ingestion

Blueprints

✔ Your data lake administrators have been updated successfully.

✕

Data lake locations (0)

🕒

Actions ▼

Register location

Find data lake storage

Data lake location ▼

IAM role ▼

Location Type ▼

Permission mode ▼

Last modified ▼

No Data lake storage

Register location

CloudShell

Feedback

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

Privacy

Terms

Cookie preferences

aws

Search

[Alt+S]

☰

Amazon S3

Buckets

🔍

🔔

⚙️

Middle East (UAE)

mahroz @ 3456-5761-9384

🟢

Successfully created bucket "my-middle-east-data-lake-bucket"

To upload files and folders, or to configure additional bucket settings, choose [View details](#).

View details

▶ Account snapshot - updated every 24 hours

All AWS Regions

View Storage Lens dashboard

Storage lens provides visibility into storage usage and activity trends. Metrics don't include directory buckets. [Learn more](#)

General purpose buckets

Directory buckets

General purpose buckets (1)

Info

All AWS Regions

🔄

Copy ARN

Empty

Delete

Create bucket

Buckets are containers for data stored in S3.

Find buckets by name

< 1 >

⚙️

Name	AWS Region	IAM Access Analyzer	Creation date
my-middle-east-data-lake-bucket	Middle East (UAE) me-central-1	View analyzer for me-central-1	May 23, 2025, 15:28:11 (UTC+04:00)

aws

Search

[Alt+S]

☰

Middle East (UAE)

mahroz @ 3456-5761-9384

🟢

Upload succeeded

For more information, see the [Files and folders](#) table.

✕

Upload: status

Close

🔍 After you navigate away from this page, the following information is no longer available.

Summary

Destination s3://my-middle-east-data-lake-bucket	Succeeded 🟢 1 file, 82.0 B (100.00%)	Failed 🔴 0 files, 0 B (0%)
---	---	-------------------------------

Files and folders

Configuration

Files and folders (1 total, 82.0 B)

Find by name

< 1 >

⚙️

Name	Folder	Type	Size	Status	Error
sales-data-csv.txt	-	text/plain	82.0 B	🟢 Succeeded	-

aws

Q s3

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Lake Formation

Data lake locations

Register location

Amazon S3 location

Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path
Choose an Amazon S3 path for your data lake.

Q s3://my-middle-east-data-lake-bucket

X

Browse

Review location permissions - strongly recommended
Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role
To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the `AWSServiceRoleForLakeFormationDataAccess` service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

☐ Enable Data Catalog Federation
Checking this box will allow Lake Formation to assume a role to access tables in a federated database.

Permission mode
Select the permission mode you want to use to manage access.

☒ Hybrid access mode
Lake Formation permissions can co-exist with IAM permission policies for AWS Glue and S3 actions to manage access.
[Learn more](#)

☐ Lake Formation
Only Lake Formation permissions are enforced.

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

aws

Q s3

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Lake Formation

Data lake locations

▼ Data Catalog

Catalogs [New](#)

Databases

Tables

Views

Data filters

Data sharing

Crawlers

▼ Permissions

Data permissions

LF-Tags and permissions

Hybrid access mode

Data locations

▼ Administration

Administrative roles and tasks

[Data lake locations](#)

Application integration settings

Data Catalog settings

▼ Ingestion

Blueprints

✔ Your Amazon S3 path: s3://my-middle-east-data-lake-bucket has been registered successfully. For information about how to set up hybrid access mode, see [Hybrid access mode](#).

Data lake locations (1/1)

Find data lake storage

Data lake location	IAM role	Location Type	Permission mode
<input checked="" type="radio"/> s3://my-middle-east-data-lake-buc...	AWSServiceRoleForLakeFormation...	Amazon S3	Hybrid access mode

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

aws

Search

[Alt+S]

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Glue

Crawlers

AWS Glue

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Data Integration and ETL

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

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 (0)

Info

Last updated (UTC)

May 23, 2025 at 11:33:24

Action

Run

Create crawler

View and manage all available crawlers.

Filter crawlers

< 1 >

Name

State

Schedule

Last run

Last run times...

Log

Table changes fr...

No resources

No resources to display.

CloudShell

Feedback

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

Privacy

Terms

Cookie preferences

aws

Search

[Alt+S]

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Glue

Crawlers

Add crawler

AWS Glue

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Data Integration and ETL

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

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

middle-east-data-lake-bucket-crawler

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

Create Database First from Data Lake console (recommended)

The screenshot shows the AWS Glue console interface for adding a new crawler. The left sidebar contains navigation links for AWS Glue, Getting started, ETL jobs, Visual ETL, Notebooks, Job run monitoring, Data Catalog tables, Data connections, Workflows (orchestration), Data Catalog, and Data Integration and ETL. The main content area is titled 'Choose data sources and classifiers' and includes a progress bar with five steps: Step 1 (Set crawler properties), Step 2 (Choose data sources and classifiers), Step 3 (Configure security settings), Step 4 (Set output and scheduling), and Step 5 (Review and create). The 'Data source configuration' section asks 'Is your data already mapped to Glue tables?' with two options: 'Not yet' (selected) and 'Yes'. Below this, the 'Data sources (0)' section shows a table with columns 'Type', 'Data source', and 'Parameters'. A message states 'You don't have any data sources.' with an 'Add a data source' button. The 'Custom classifiers - optional' section provides information about classifiers and a 'Next' button.

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
► **Data Catalog**
► **Data Integration and ETL**

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

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.
Edit Remove Add a data source

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

The screenshot shows the 'Add data source' dialog box. It has a close button (X) in the top right corner. The 'Data source' section has a dropdown menu with 'S3' selected. The 'Network connection - optional' section has a dropdown menu with a refresh icon. The 'Location of S3 data' section has two radio buttons: 'In this account' (selected) and 'In a different account'. The 'S3 path' section has a text input field with 's3://my-middle-east-data-lake-bucket', a 'View' button, and a 'Browse S3' button. The 'Subsequent crawler runs' section has three radio buttons: 'Crawl all sub-folders' (selected), 'Crawl new sub-folders only', and 'Crawl based on events'. The dialog box has 'Cancel' and 'Add an S3 data source' buttons at the bottom.

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://my-middle-east-data-lake-bucket 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

aws

Search

[Alt+S]

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Glue

Databases

Add database

AWS Glue

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Data Integration and ETL

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

Create a database

Create a database in the AWS Glue Data Catalog.

Database details

Name

middle-east-data-lake-database

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.

Cancel

Create database

aws

Search

[Alt+S]

Middle East (UAE)

mahroz @ 3456-5761-9384

AWS Glue

Crawlers

Add crawler

AWS Glue

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Data Integration and ETL

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

Step 1

Set crawler properties

Step 2

Choose data sources and classifiers

Step 3

Configure security settings

Step 4

Set output and scheduling

Step 5

Review and create

Set output and scheduling

Output configuration

Target database

middle-east-data-lake-database

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

aws Search [Alt+S] Middle East (UAE) mahroz @ 3456-5761-9384

AWS Glue > Crawlers > middle-east-data-lake-bucket-crawler

One crawler successfully created
The following crawler is now created: "middle-east-data-lake-bucket-crawler"

middle-east-data-lake-bucket-crawler Last updated (UTC) May 23, 2025 at 11:47:42 [Run crawler](#) [Edit](#) [Delete](#)

Crawler properties

Name middle-east-data-lake-bucket-crawler	IAM role AWSGlueServiceRole-middle-east-data-lake-crawler	Database middle-east-data-lake-database	State READY
Description -	Security configuration -	Lake Formation configuration -	Table prefix -
Maximum table threshold -			

Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (0)
The list of crawler runs for this crawler.

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
------------------	----------------	-----------------------	--------	-----------	---------------

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

Notice permission mode.

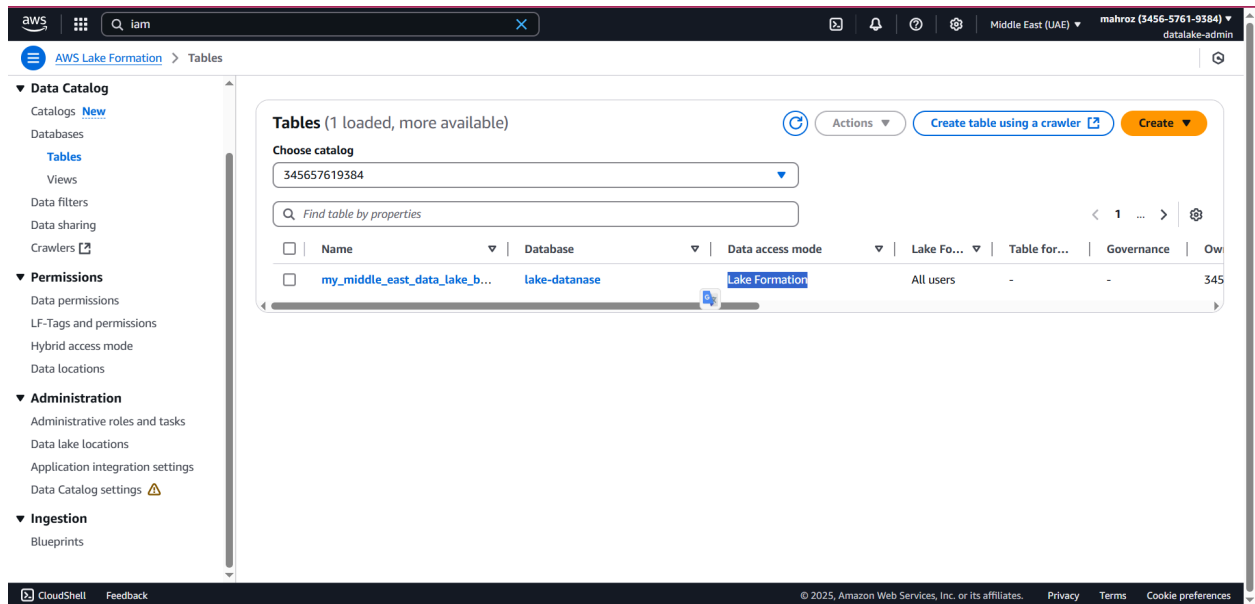
aws iam Middle East (UAE) mahroz (3456-5761-9384) datalake-admin

AWS Lake Formation > Data lake locations

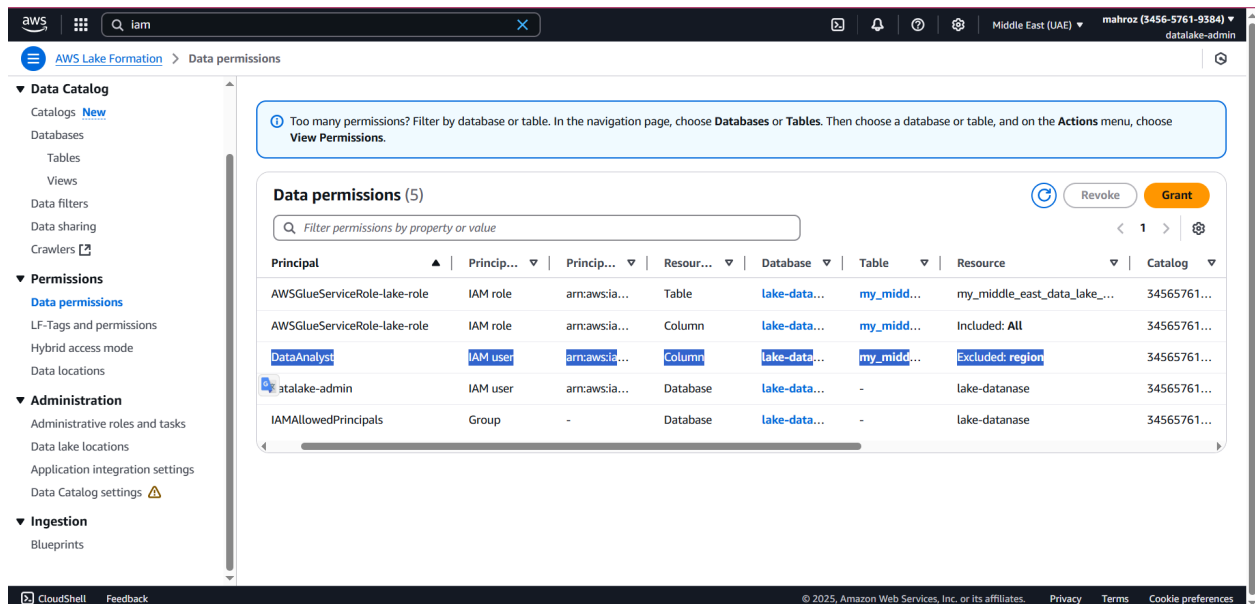
Data lake locations (1/1) [Actions](#) [Register location](#)

Data lake location	IAM role	Location Type	Permission mode	Last modified
s3://my-middle-east-data...	AWSServiceRoleForLakeFo...	Amazon S3	Lake Formation	May 23, 2025 at 6:27 PM ...

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



We are giving all column access to dataanalyst user except region column.



See 'DataAnalyst' can not see region column

The screenshot shows the Amazon Athena Query Editor interface. On the left, the 'Database' is set to 'lake-datanase'. The 'Tables and views' section shows 'Tables (0)' and 'Views (0)'. The main area displays a query result for a completed query. The query statistics show: Time in queue: 63 ms, Run time: 695 ms, Data scanned: 0.08 KB. The results table has 1 row and 4 columns: #, date, product, and amount. The data row shows: 1, 2023-01-01, Laptop, 2500. The user 'DataAnalyst' is logged in, and the 'region' column is not visible in the results.

#	date	product	amount
1	2023-01-01	Laptop	2500

Where datalake-admin user can see all columns

The screenshot shows the Amazon Athena Query Editor interface for the user 'datalake-admin'. The 'Database' is set to 'lake-datanase'. The 'Tables and views' section shows 'Tables (1)' with 'my_middle_east_data_lake_bucket' selected. The main area displays a query result for a completed query. The query statistics show: Time in queue: 69 ms, Run time: 690 ms, Data scanned: 0.08 KB. The results table has 1 row and 5 columns: #, date, region, product, and amount. The data row shows: 1, 2023-01-01, North, Laptop, 2500. The user 'datalake-admin' is logged in, and the 'region' column is visible in the results.

#	date	region	product	amount
1	2023-01-01	North	Laptop	2500