



## Edit crawler



## Add information about your crawler



Crawler info

s3 specific bucket  
crawler

Data store

S3: s3://spark23mar...



IAM Role



Schedule

Run on demand



Output

testdb3



Review all steps

## Crawler name

s3 specific bucket crawler

▸ Tags, description, security configuration, and classifiers (optional)

▼ Grouping behavior for S3 data (optional)

☒ Create a single schema for each S3 path

By default, when a crawler defines tables for data stored in S3, it considers both data compatibility and schema similarity. Select this check box to group compatible schemas into a single table definition across all S3 objects under the provided include path. Other criteria will still be considered to determine proper grouping. [Learn more](#)

Next



Services

Resource Groups



sahil dadia

Ireland

Support

## Edit crawler



## Crawler info

s3 specific bucket  
crawler

## Data store

S3: s3://spark23mar...

## IAM Role

## Schedule

Run on demand

## Output

testdb3

## Review all steps

## Add a data store

## Choose a data store

S3

## Crawl data in

☒ Specified path

## Include path

s3://spark23march2019

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

## ▼ Exclude patterns (optional)

## Exclude patterns

\*.parquet/\*

glob pattern

The exclude pattern is relative to the include path. Objects that match the exclude pattern are not crawled. For example, with include path `s3://mybucket/` and exclude pattern, `mydir/**`, then all objects in the include path below the `mydir` directory are skipped. In this example, any object whose path matches `s3://mybucket/mydir/**` is not crawled. For more information about patterns, see [Cataloging Tables with a Crawler](#).

Back

Next

## Chosen data stores

S3: s3://spark23mar...





Services

Resource Groups



sahil dadia

Ireland

Support

## Edit crawler



Crawler info

s3 specific bucket  
crawler

Data store

S3: s3://spark23mar...



IAM Role



Schedule

Run on demand



Output

testdb3



Review all steps

## Add another data store

☐ Yes☒ No

Back

Next

## Chosen data stores

S3: s3://spark23mar...





## Edit crawler



## Crawler info

s3 specific bucket crawler

## Data store

S3: s3://spark23mar...

## IAM Role

## Schedule

Run on demand

## Output

testdb3

## Review all steps

## Choose an IAM role

The IAM role allows the crawler to run and access your Amazon S3 data stores. [Learn more](#)

- ☐ Update a policy in an IAM role
- ☒ Choose an existing IAM role
- ☐ Create an IAM role

## IAM role ⓘ

glue\_role



This role must provide permissions similar to the AWS managed policy, **AWSGlueServiceRole**, plus access to your data stores.

- s3://spark23march2019

You can also create an IAM role on the [IAM console](#).

Back

Next

[Edit crawler](#)

## Create a schedule for this crawler

👉 Crawler info

s3 specific bucket  
crawler

✔ Data store

S3: s3://spark23mar...

✔ IAM Role

```
arn:aws:iam::637994078207:role/glue_role
```

☐ Schedule

Run on demand

✔ Output

testdb3

✔ Review all steps

Frequency

Run on demand

[Back](#)

Next



Services

Resource Groups



sahil dadia

Ireland

Support

## Edit crawler



## Configure the crawler's output

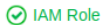


Crawler info

s3 specific bucket  
crawler

Data store

S3: s3://spark23mar...



IAM Role

arn:aws:iam::6379940  
78207:role/glue\_role

Schedule

Run on demand



Output

testdb3



Review all steps

## Database ⓘ

testdb3

Add database

## Prefix added to tables (optional) ⓘ

Type a prefix added to table names

► Configuration options (optional)

Back

Next



Services

Resource Groups



sahil dadia

Ireland

Support

## Edit crawler



## Crawler info

s3 specific bucket  
crawler

## Data store

S3: s3://spark23mar...

## IAM Role

arn:aws:iam::6379940  
78207:role/glue\_role

## Schedule

Run on demand

## Output

testdb3

## Review all steps

## Crawler info

Name s3 specific bucket crawler

Tags -

Create a single schema for each S3 path true

## Data stores

Data store S3

Include path s3://spark23march2019

Exclude patterns \*.parquet/\*

## IAM role

IAM role arn:aws:iam::637994078207:role/glue\_role

## Schedule

Schedule Run on demand

## Output

Database testdb3

Prefix added to tables (optional)

► Configuration options

Back

Finish



Services

Resource Groups



sahil dadia

Ireland

Support

## AWS Glue

## Data catalog

Databases

Tables

Connections

Crawlers

Classifiers

Settings

ETL

Jobs

Triggers

Dev endpoints

Notebooks

Security

Security configurations

Tutorials

Add crawler

Explore table

Add job

Resources

What's new

**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.

[User preferences](#)

Add crawler

Run crawler

Action

Filter by tags and attributes

Showing: 1 - 1

<input type="checkbox"/>	Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
<input type="checkbox"/>	<a href="#">s3 specific bucket crawler</a>		Ready	<a href="#">Logs</a>	1 min	1 min	0	3





## AWS Glue

## Data catalog

## Databases

## Tables

## Connections

## Crawlers

## Classifiers

## Settings

## ETL

## Jobs

## Triggers

## Dev endpoints

## Notebooks

## Security

## Security configurations

## Tutorials

## Add crawler

## Explore table

## Add job

## Resources

## What's new

**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.

Add tables

Action

Database : testdb3 Filter or search for tables...

Save view



Showing: 1 - 3 &lt; &gt; Refresh Settings Help

<input type="checkbox"/>	Name	Database	Location	Classification	Last updated	Deprecated
<input type="checkbox"/>	cfs_2012_pumf_csv	testdb3	s3://spark23march2019/cfs_2012_pumf.csv	csv	25 March 2019 8:23 AM UTC	
<input type="checkbox"/>	fire_dept_calls_parquet	testdb3	s3://spark23march2019/fire_dept_calls.parquet/	parquet	25 March 2019 8:23 AM UTC	
<input type="checkbox"/>	fireservice_data_for_ml_parquet	testdb3	s3://spark23march2019/fireservice_data_for_ML.parquet/	parquet	25 March 2019 8:23 AM UTC	