In order to use glue we need an S3 bucket

1. Click create S3 bucket
2. Name it and create it

Create an IAM role

1. Click on roles (left ribbon)
2. Select AWS service and search for glue below.
3. Click on next and add “AmazonS3FullAccess”, “AWSGlueConsoleFullAccess” and “CloudWatchLogsFullAccess” as role policy
4. Click next and name the Role and click on create role at the bottom.

Create an S3 Folder in the bucket

1. Name it
2. Click next
3. If you need extra subfolders, you can add it.
4. You can upload data into these folders.

After uploading the file, we can create a database in glue

1. Click on databases in glue at the left ribbon
2. Click on add databases
3. Enter the name of the database
4. Click create database

You can add tables into these databases

1. Open the DB and click “add table”
2. Add the name of the table
3. Find the object and select the S3 Folder
4. Select the data format
5. Click on next
6. Chose between crawler or data source.

Alternative solution for crawler:

1. Click on databases on the left ribbon
2. Select our DB and click on “add tables using crawler”
3. Enter the name of crawler and click next
4. Click add data source
5. Click browse s3 and pick up the relevant folder
6. Click add
7. Click next.
8. Pick the IAM role we specified before
9. Click next and chose a target database, which we previously created.
10. Set the frequency of the crawler below.
11. Click create crawler in the next window.
12. If You picked “On demand” at the frequencies, click run crawler.

Crawler bugfix

1. If the crawler stopped working click on it.
2. Pick the failed entry and click view cloudwatch logs.
3. It specifies the type of error.

If the crawler finished successfully, click on tables in the left ribbon below databases. You can check the first table there.

If You want to query the database, pick Athena in AWS. If You are using it first, you have to edit it. Click on browse S3 and select a folder in the bucket for the query. Click on save.

If You want to create custom data source connection, click on connections on the left ribbon. Here You can select the data source you want to connect to. Here You can add the credentials in order to successfully establish the connection. You can even connect to different S3 buckets here.

After that, you can start an ETL job. Click on create an ETL job on the left ribbon. Here You can pick the type of ETL job.

Visual ETL: You can pick the source and the target too.

You can set triggers for the job by selecting “Triggers” function on the left ribbon. Click on Add Trigger, where You can set the type and frequency.