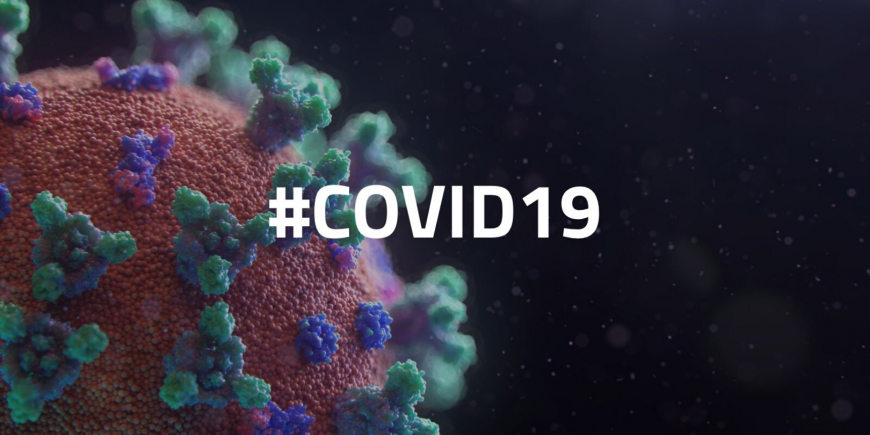
**AWS Covid 19 Project**

This is one of the easiest AWS projects you can practice. You will learn to examine the covid-19 dataset using AWS Athena, a serverless SQL query engine. The Covid-19 database comprises Covid-related timestamps, posts, and comments. The project will teach you how to leverage Amazon Glue to create tables and offers you a chance to explore various Athena joins. In the AWS Glue Data Catalog, you will use Python to create tables using crawlers. Working on this project will also help you understand how AWS Athena pricing varies depending on the file size. Additionally, this project involves a few other services such as Amazon S3, Amazon CloudWatch, etc. AWS S3 allows you to store the dataset (CSV file) in S3 buckets for further processing, and CloudWatch keeps track of your data's log files and lets you analyze them as needed.



Following steps need to follow to Create Covid 19 AWS project using Amazon Services.

1. Collect or Download Covid 19 dataset from any official, GitHub, Kaggle, or any other data source.
2. Use Amazon S3 to Store the data.
3. Create Crawler using Amazon Glue for data in S3.

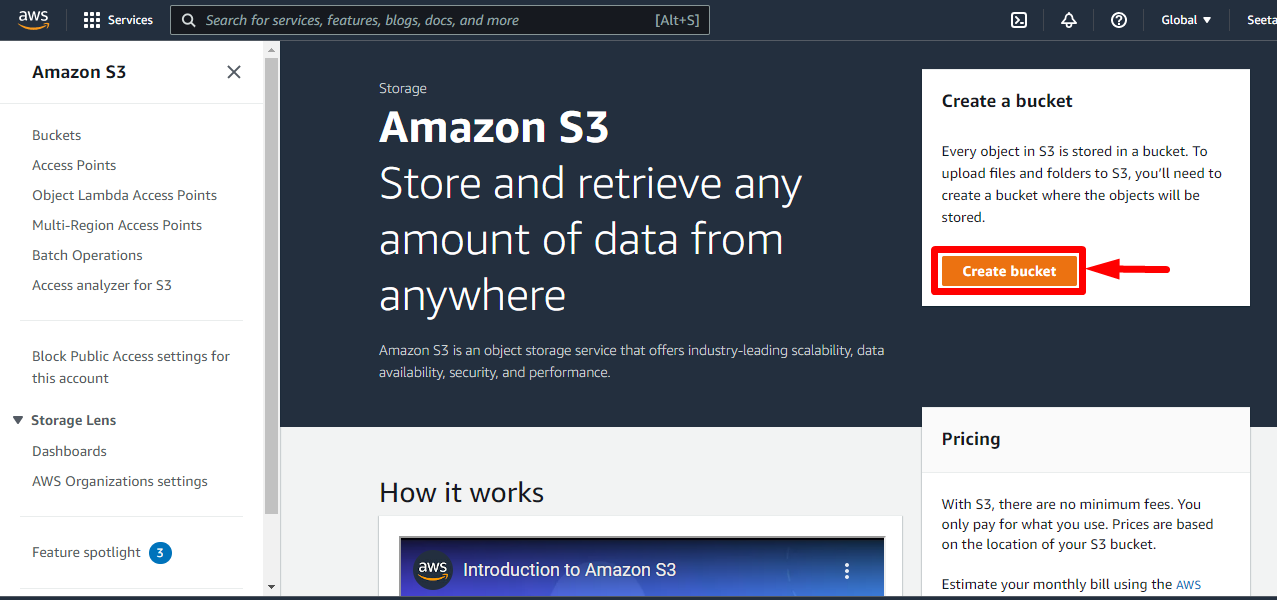
**Step 1**

Download data of Covid 19 from available data source. The link is provided below just check if you want to download data.

<https://github.com/SSSirsat/AWS/tree/main/Covid19%20Project/Covid19_Data>

Upload CSV data into S3 Bucket

1. Create Bucket in Amazon S3



1. You can see the Bucket is Created in Asia Pacific (Mumbai) ap-south-01 AWS Region.

Here we have given “ covid-19-project” bucket name to our bucket.

Graphical user interface, text, application

Description automatically generated

1. Go inside that Bucket and Upload Covid 19 Data file that we have downloaded from above mentioned Git Hub link.

Graphical user interface, text, application

Description automatically generated

Inside this folder you can see uploaded csv file in folder.

Some data in .gz file and some data is in csv or JSON file.

Graphical user interface, text, application, email

Description automatically generated

**Step 2**

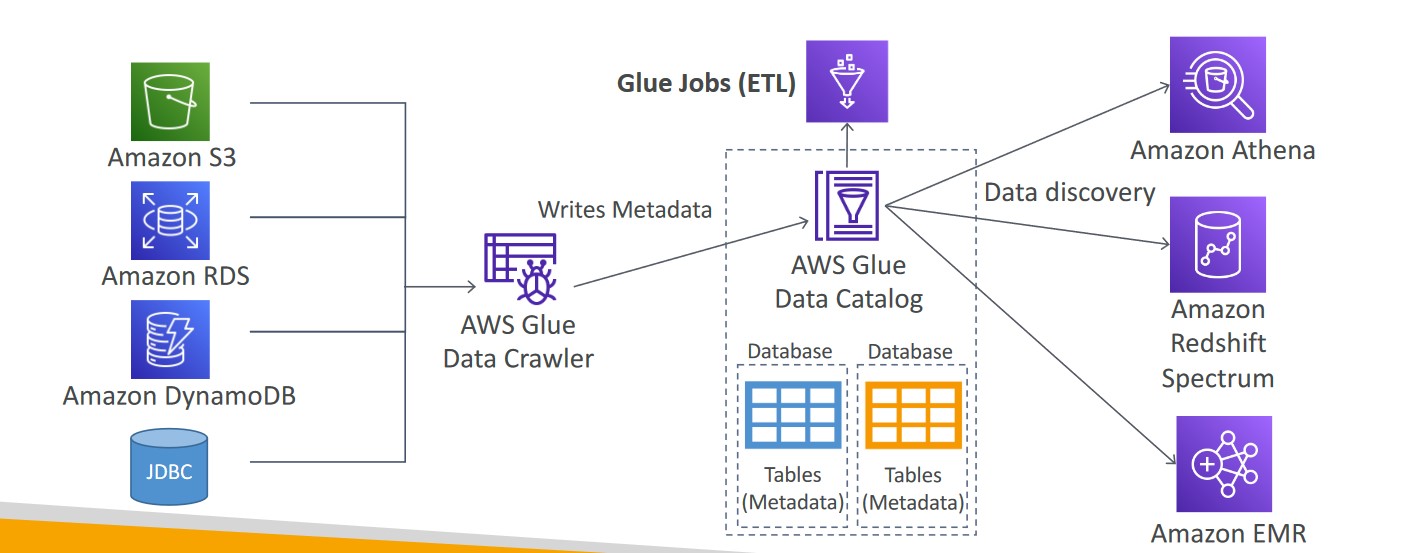
**AWS Glue**

* Managed extract, transform, and load (ETL) service
* Useful to prepare and transform data for analytics
* Fully serverless service

[](https://ind01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fraw.githubusercontent.com%2Fsudheera96%2FOrchestrate-Redshift-ETL-using-AWS-Glue-and-Step-Functions%2Fmain%2Fimages%2Faws%2520glue.jpg&data=05%7C01%7Cshubham.s%40bourntec.com%7C1543d5a94aa648796b9508da75f2f7bf%7Ce047295072a7456f9ef123c644042bfd%7C0%7C0%7C637951983652716827%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=ZC1MM3qENOdMKuBVKFZyFUhEVhbJ4cHYe%2BbZiFo4bno%3D&reserved=0)

**Glue Data Catalog**

* Glue Data Catalog: catalog of datasets

[](https://ind01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fraw.githubusercontent.com%2Fsudheera96%2FOrchestrate-Redshift-ETL-using-AWS-Glue-and-Step-Functions%2Fmain%2Fimages%2Faws%2520data%2520catlog.jpg&data=05%7C01%7Cshubham.s%40bourntec.com%7C1543d5a94aa648796b9508da75f2f7bf%7Ce047295072a7456f9ef123c644042bfd%7C0%7C0%7C637951983652716827%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=J%2F2m0f6weG5tOYhXQNyicRbJJSIihZPPzYGvYRbHc1g%3D&reserved=0)

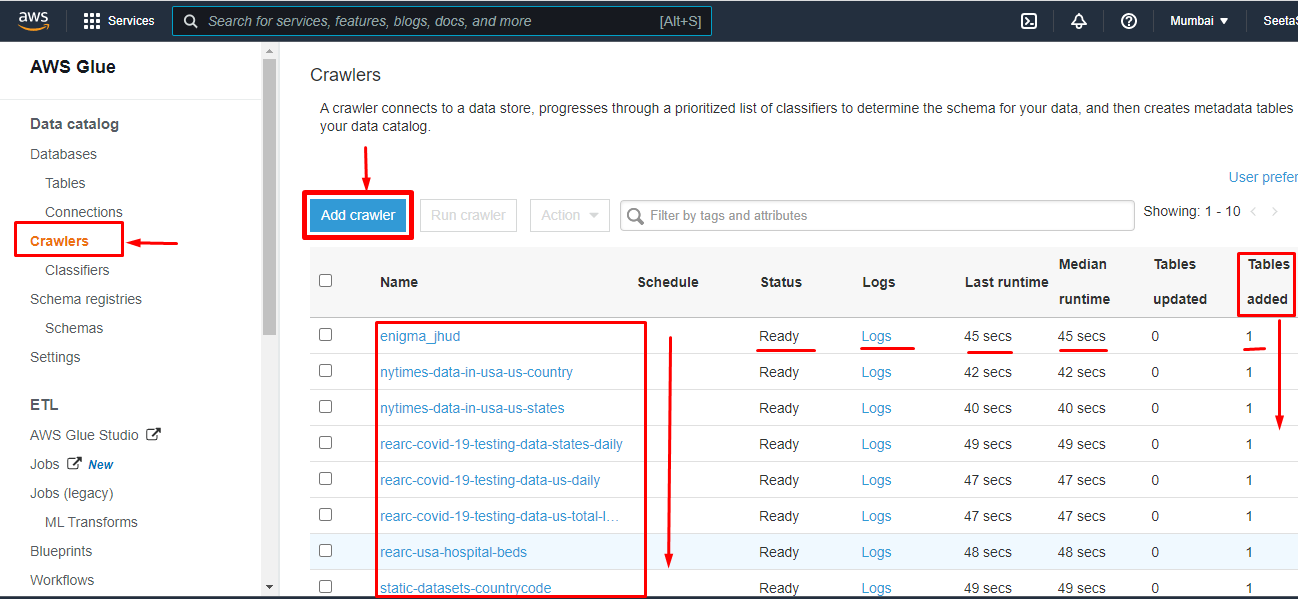
**Glue Job Creation**

1. Go to search >> glue >> connection >> Add connection (has to connect with redshift) >> name as "myredshiftconnection" >> connection type :AWS Redshift >> connection access >> cluster details (as only one cluster is running we add that cluster) >> review the steps >> finish
2. Now select the created connection >> test connection >> IAM role ( we have to give the glue to access the redshift) >> create new IAM role >> create role >> search glue >> select AWS glue service role (Policy for AWS Glue service role which allows access to related services including EC2, S3, and Cloudwatch Logs) >> Along with it create other policy for connecting to s3 and connecting to secret manager >> copy and paste the provide policy of glue with providing secret manager ARN connection id >> next tag >> next review >> my glue policy >> create policy
3. Refresh policy >> select created policy >> aws glue service role >> next >> name: mygluerole >> create
4. come back to the attaching IAM role to glue >> refresh >> choose mygluerole >> testconnection
5. Crete glue job (as we having table and scriptsa ready) >> click on job (legacy)
6. After Uploading data into S3 Bucket, we need to create a Crawler using Amazon Glue.

Just open new tab with Amazon Glue service. And click on Add new crawler

Then add new crawler for separate folder present in S3 Bucket.

You can see the tables is added into the AWS Glue and we can use this table for further use.



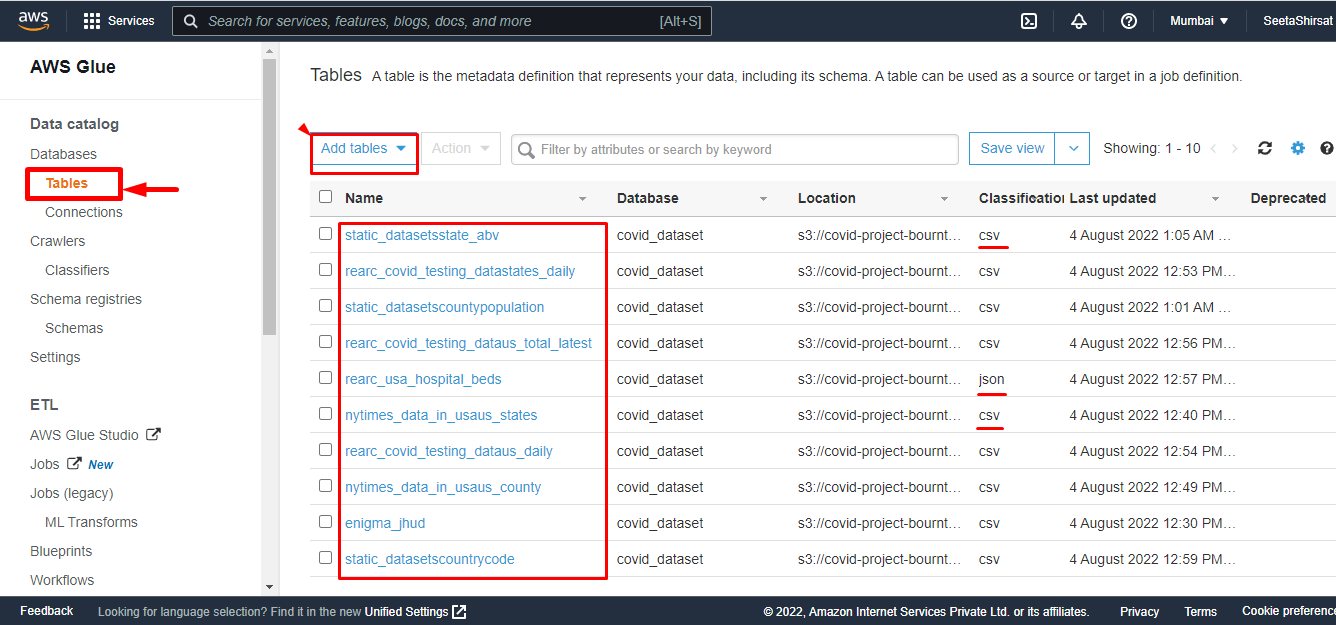
1. Now goto Tables and click on Add Tables

Then click on Add table using a crawler

Graphical user interface, application, Word

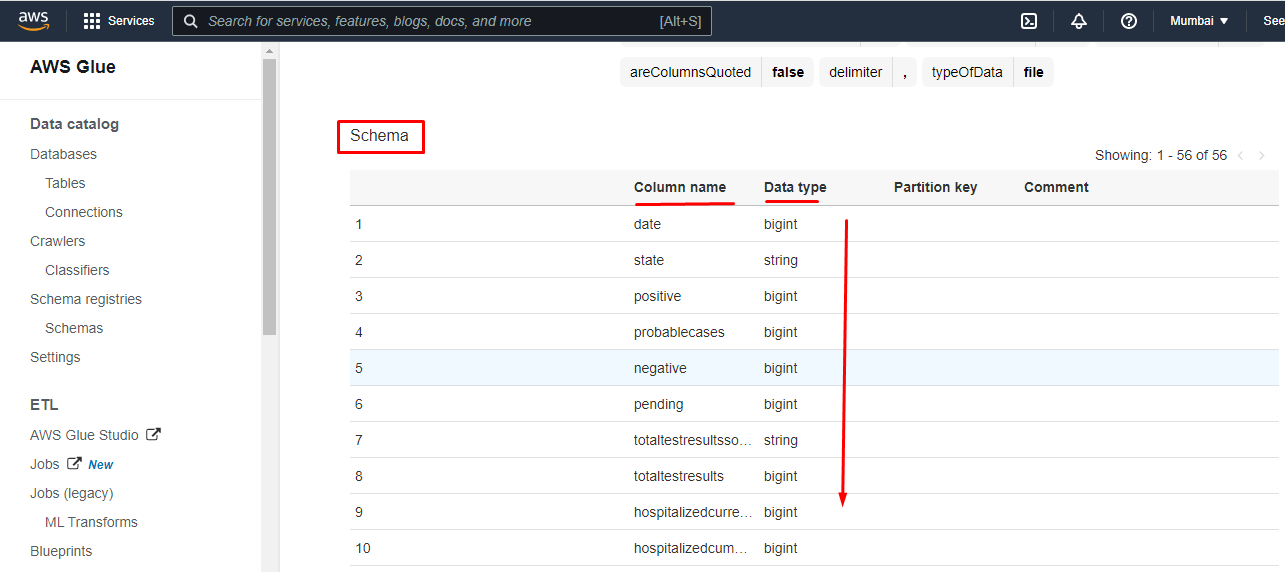
Description automatically generated

Add all table separately one by one and now we have a list of tables in.



By clicking on one of the Table name you can see the Schema of the table.

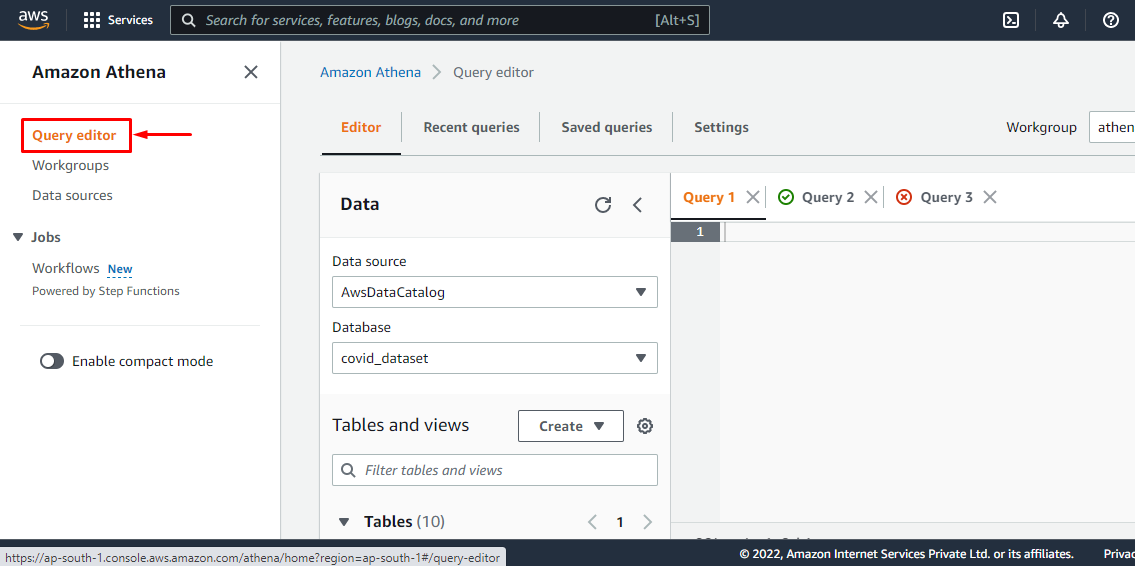
Also you can see columns , data type and other details of table in schema.



**Step 3**

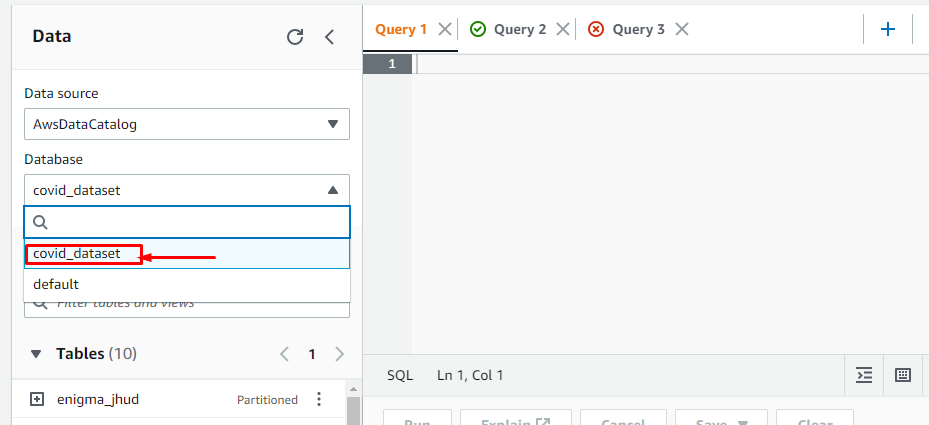
In step 3 you need to open amazon Athena.

Click on query editor it will open and show you the dashboard in which you need to select the Data source and database



You need to select the database which will directly show into the database drop up bar.

Select that table.

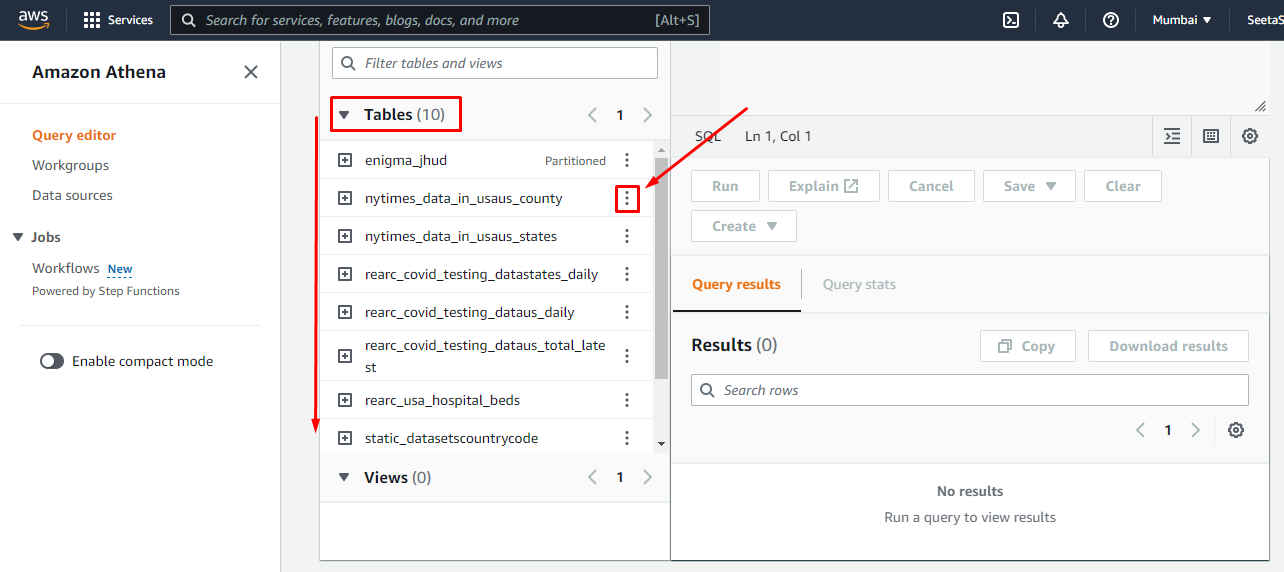


Graphical user interface, text, application

Description automatically generated

It will show you all the data tables in Table

Click any of the three dots and open the



Click on the Preview Table

Graphical user interface, text, application

Description automatically generated

It will open one Query editor and write one query automatically to show the 10 data records present in the table.

Graphical user interface, text, application

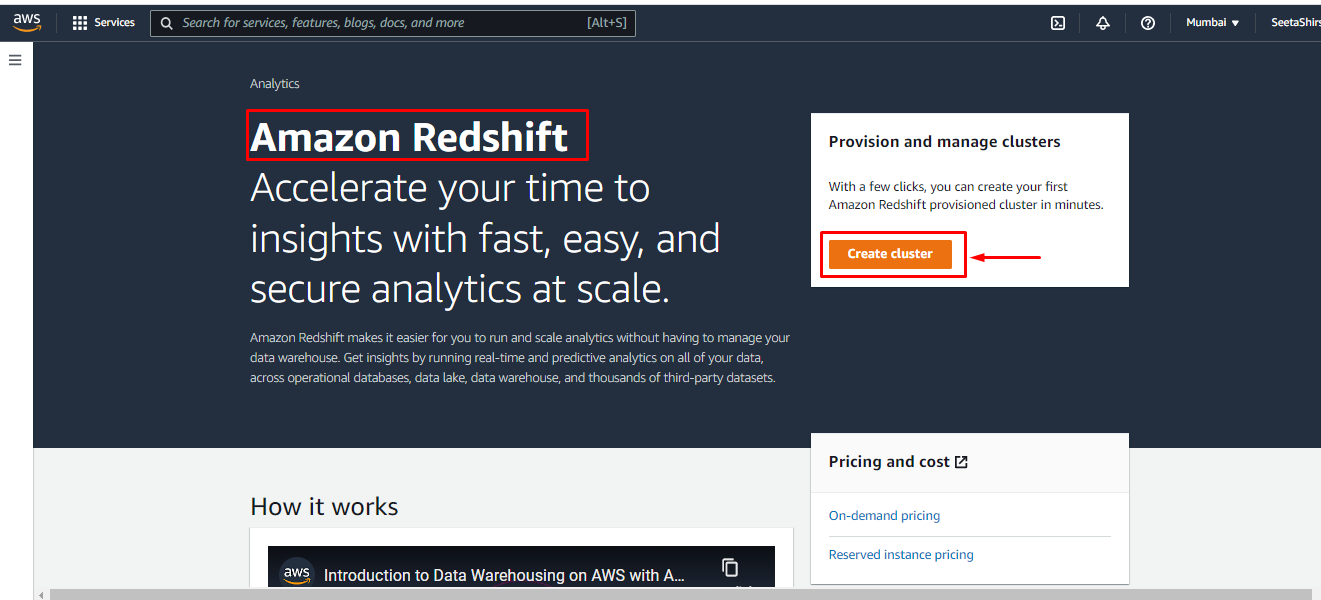
Description automatically generated

Graphical user interface, application, table

Description automatically generated

**Step 4**

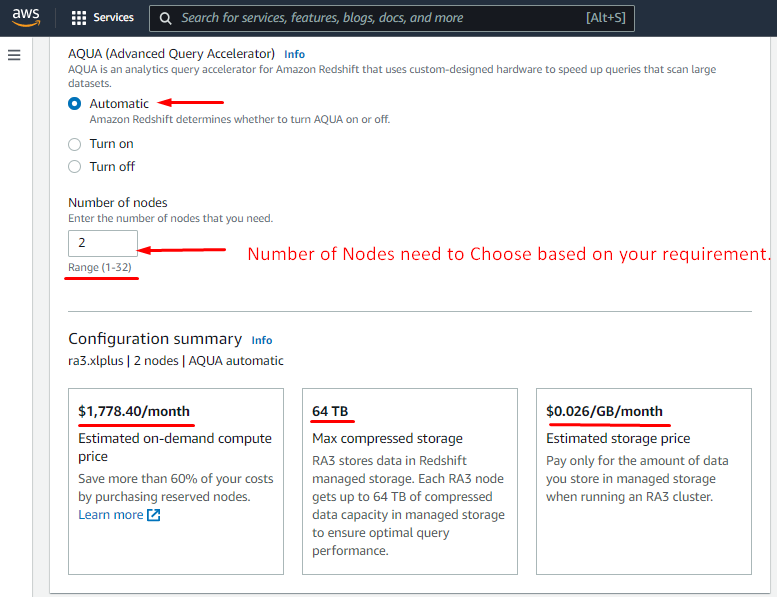
Create Amazon Redshift Cluster

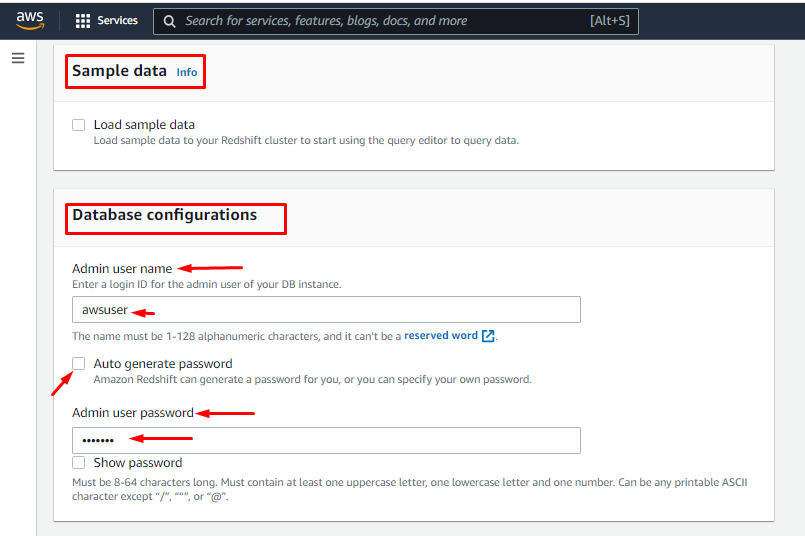


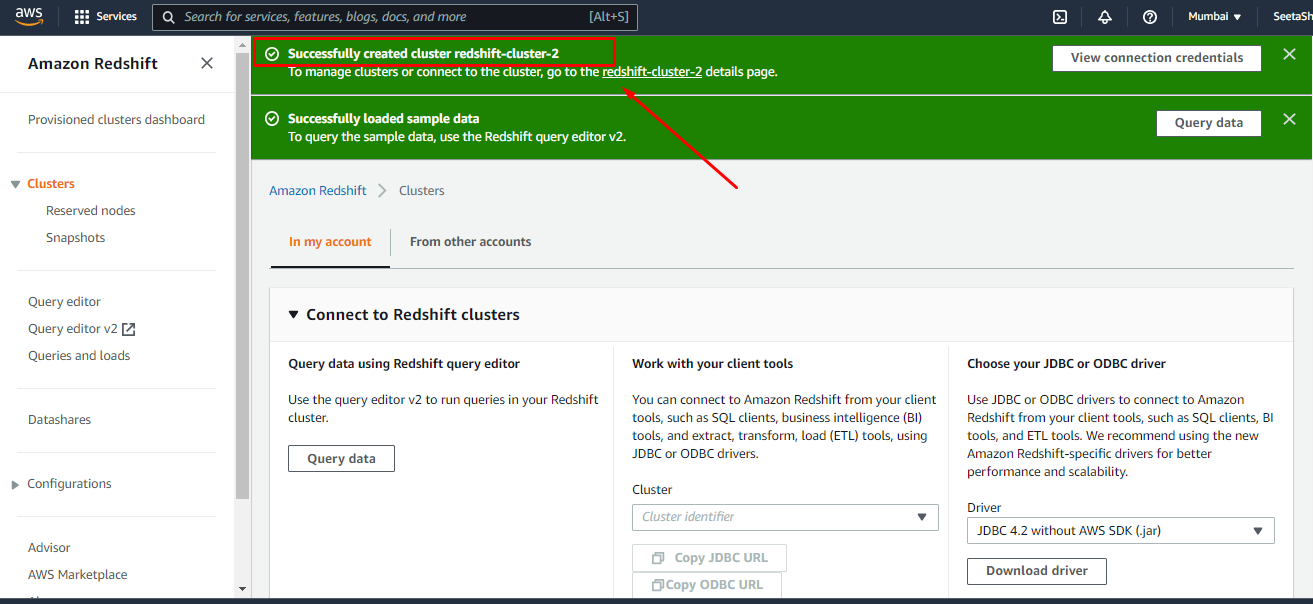
Graphical user interface, text, application

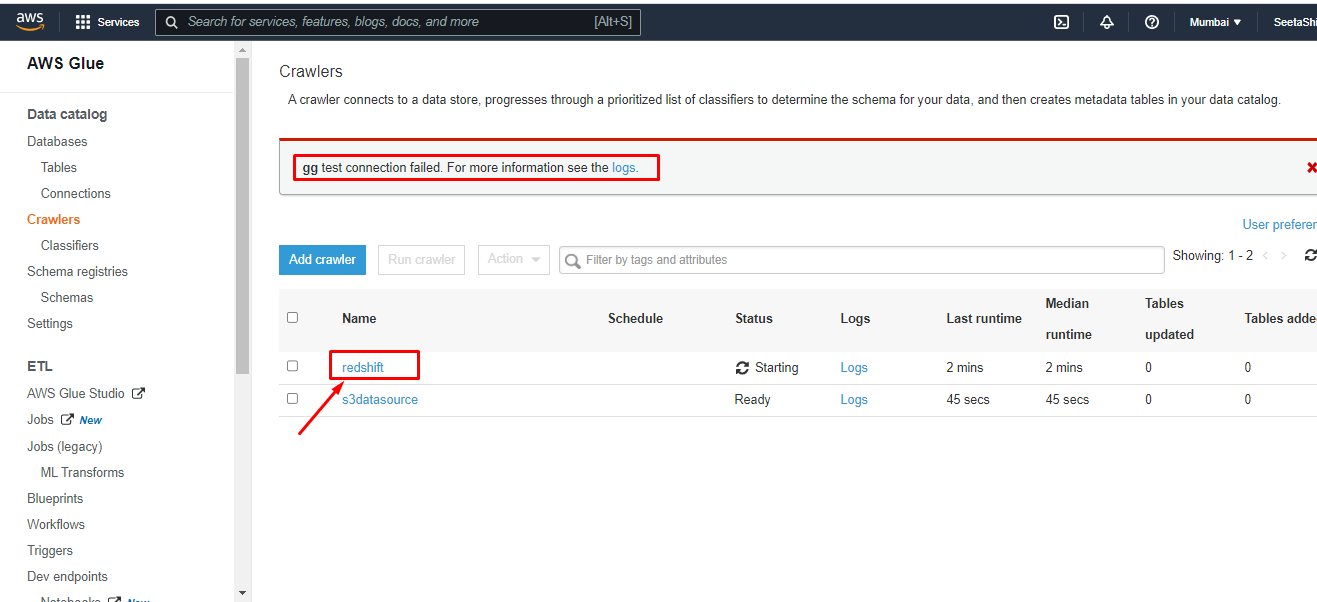
Description automatically generated

Amazon Redshift Cluster









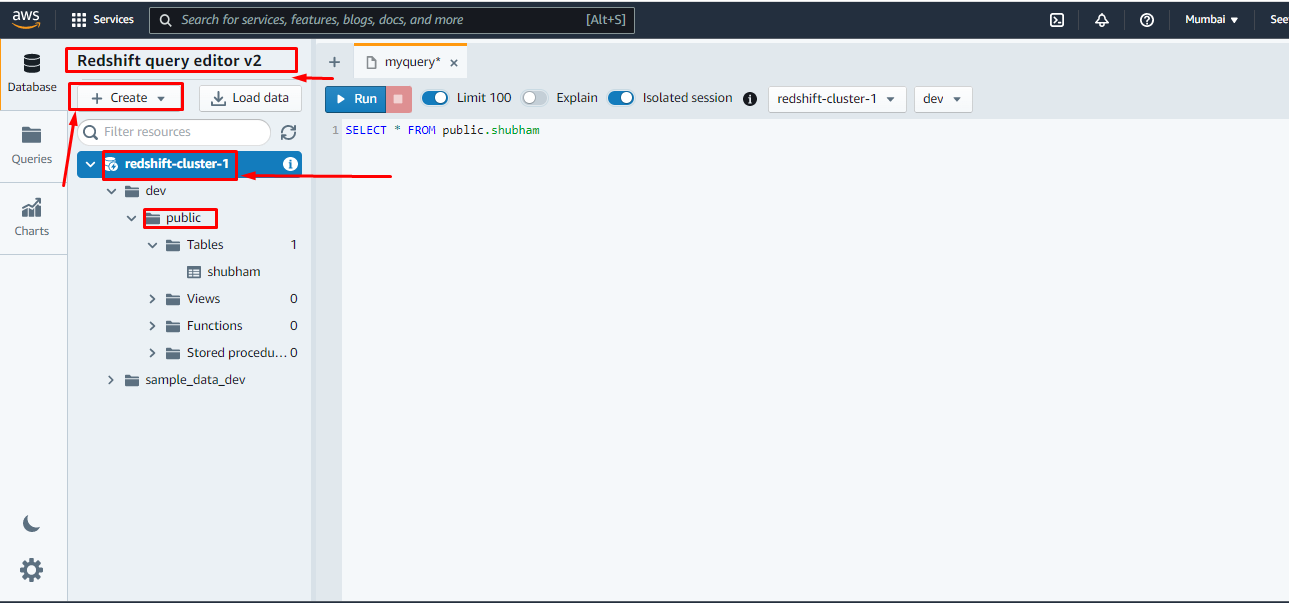
Open Query Editor 2 in your redshift query editor

Create one database that is “dev”

Create Schema that is “public”

Create Table that is “covid19”

In that create add all field in table.



Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

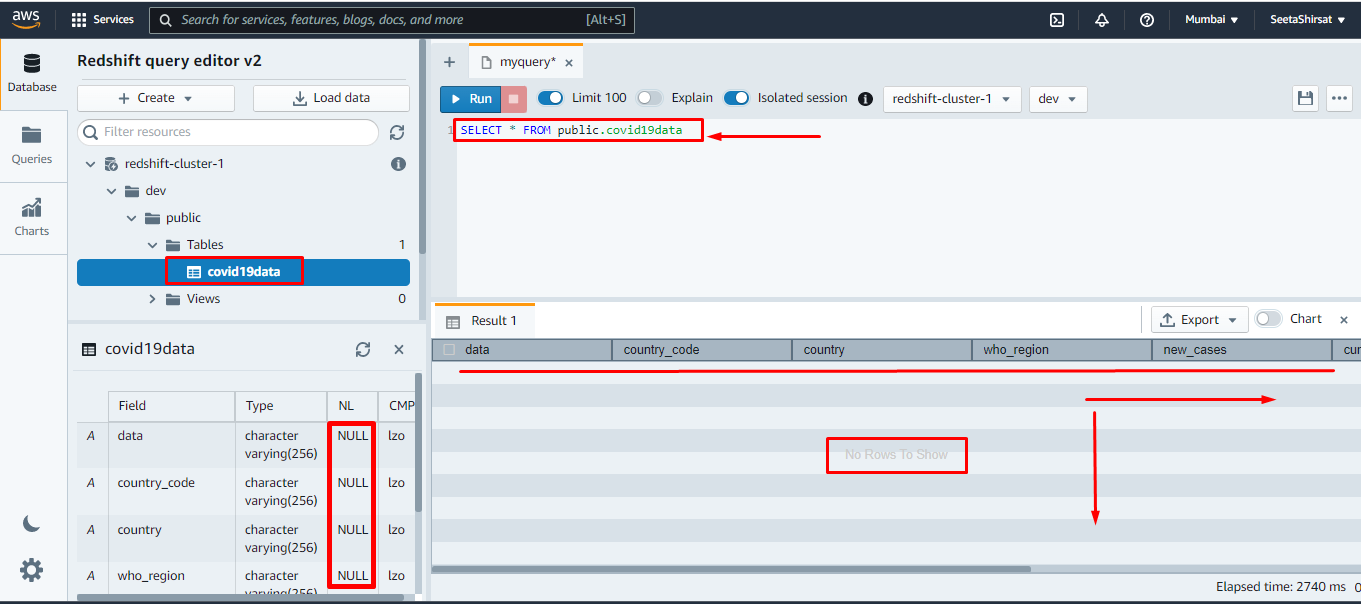
Description automatically generated

Graphical user interface, table

Description automatically generated

Graphical user interface, text, application

Description automatically generated



Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text

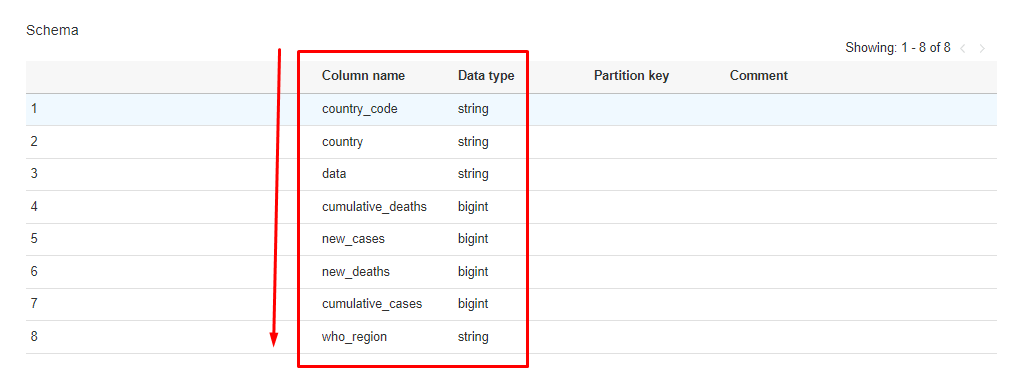
Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

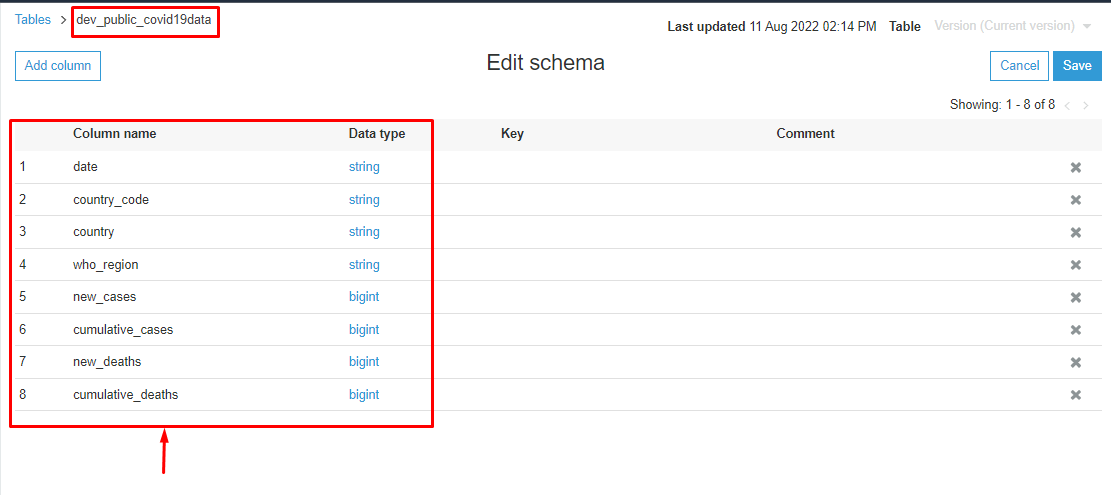


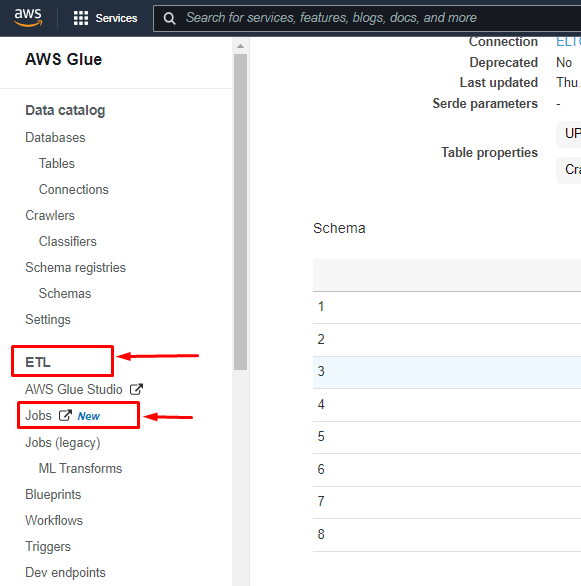
Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated





Amazon Job Creation

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface

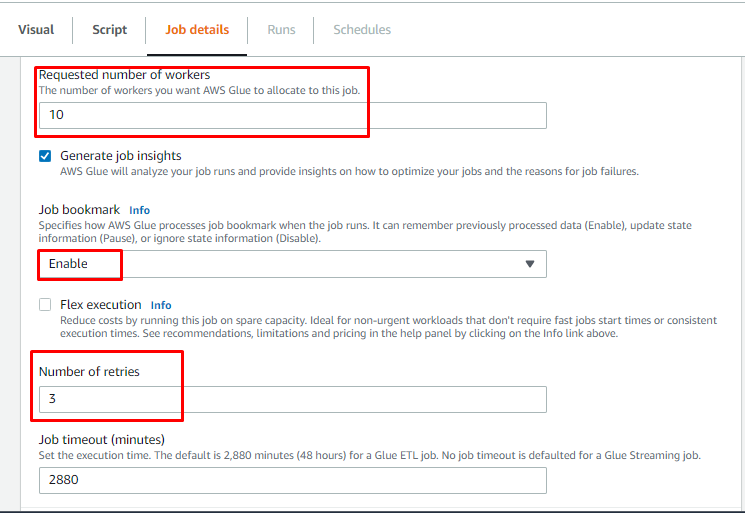
Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated



Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

A screenshot of a computer

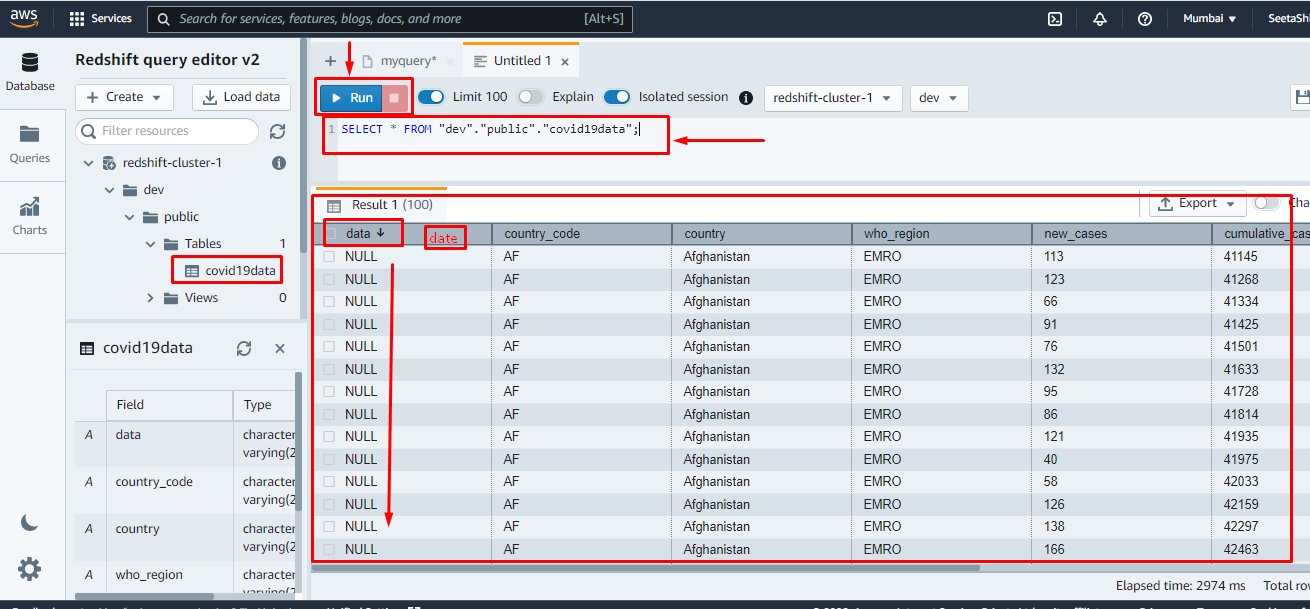
Description automatically generated

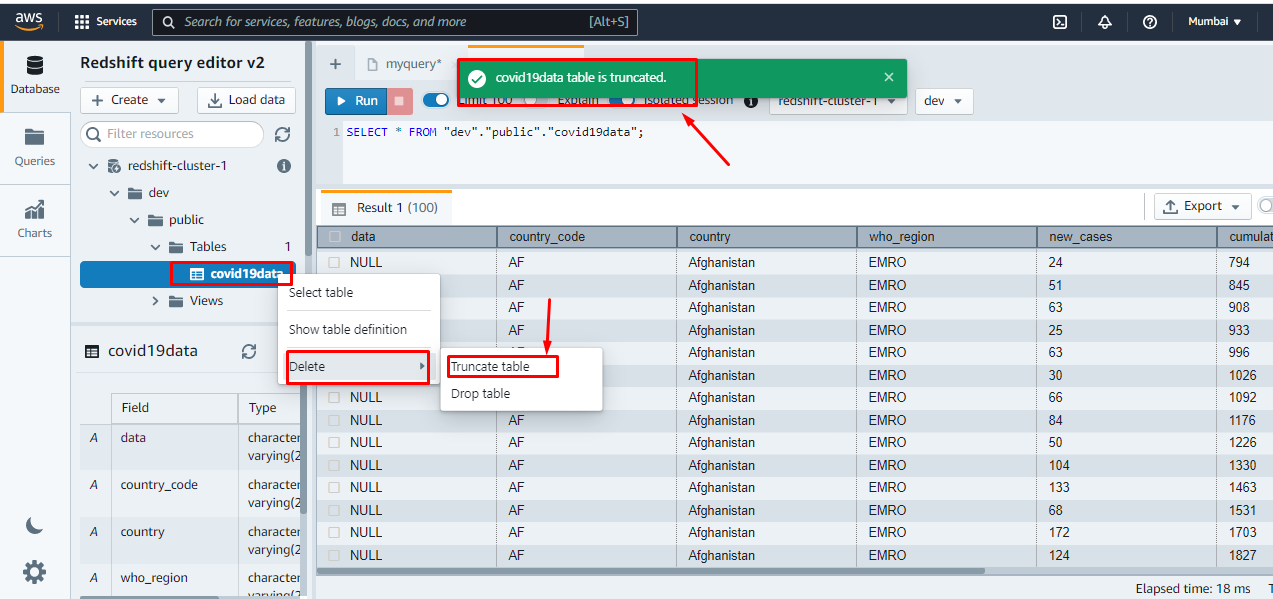
Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated



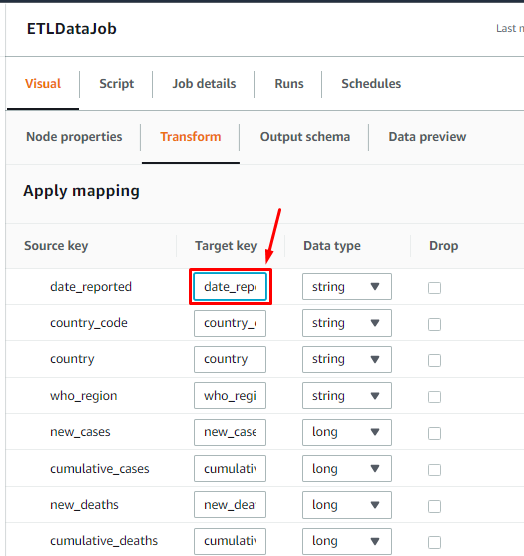


Graphical user interface, text, application, email

Description automatically generated

Chart

Description automatically generated with low confidence



Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

A screenshot of a computer

Description automatically generated

Analyse Data

Graphical user interface, text

Description automatically generated with medium confidence

Graphical user interface, application, Teams

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

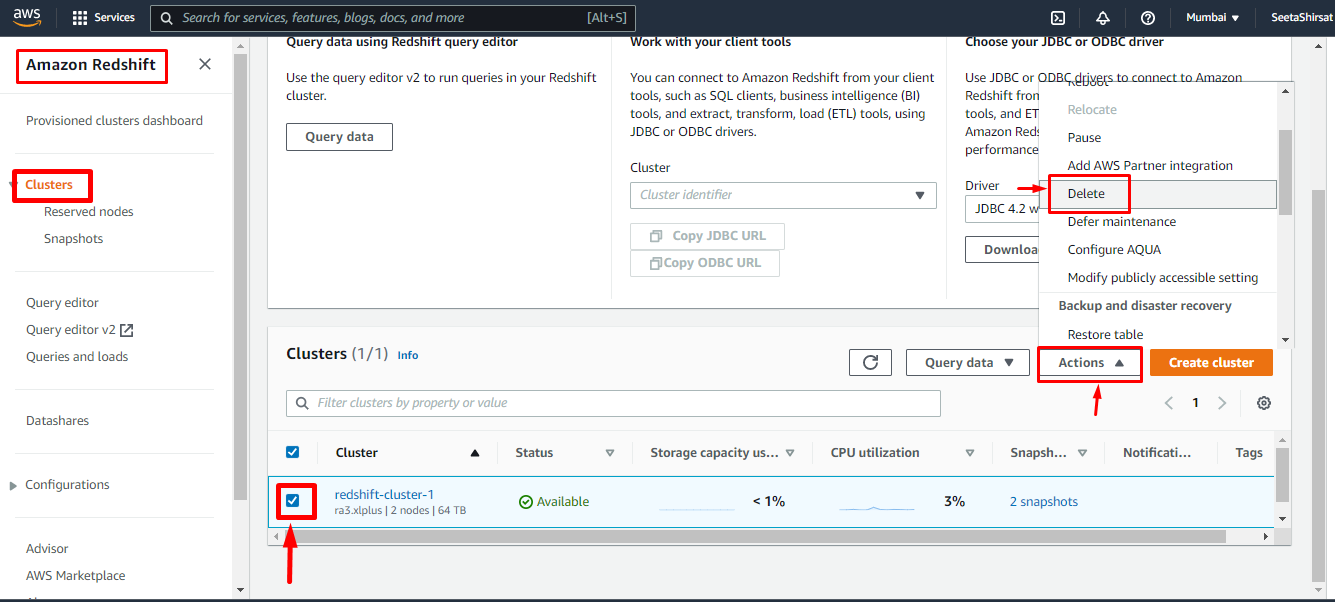
Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application

Description automatically generated

Deleting Cluster



Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, diagram

Description automatically generated

