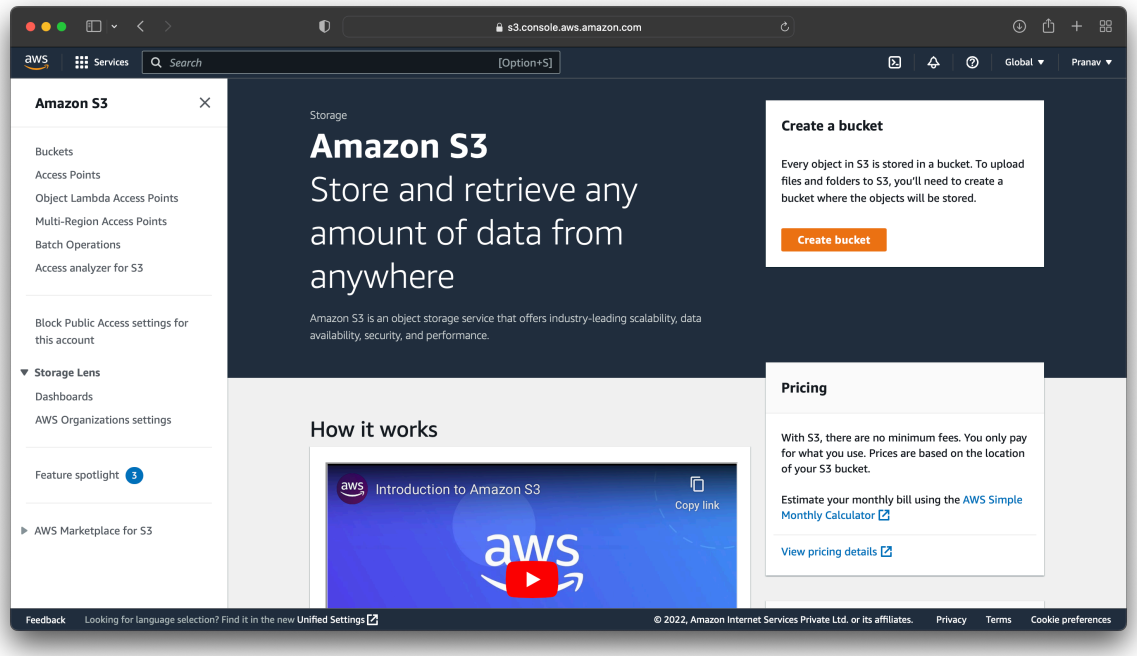
Experiment 6 - Querying Data in S3 with Amazon Athena

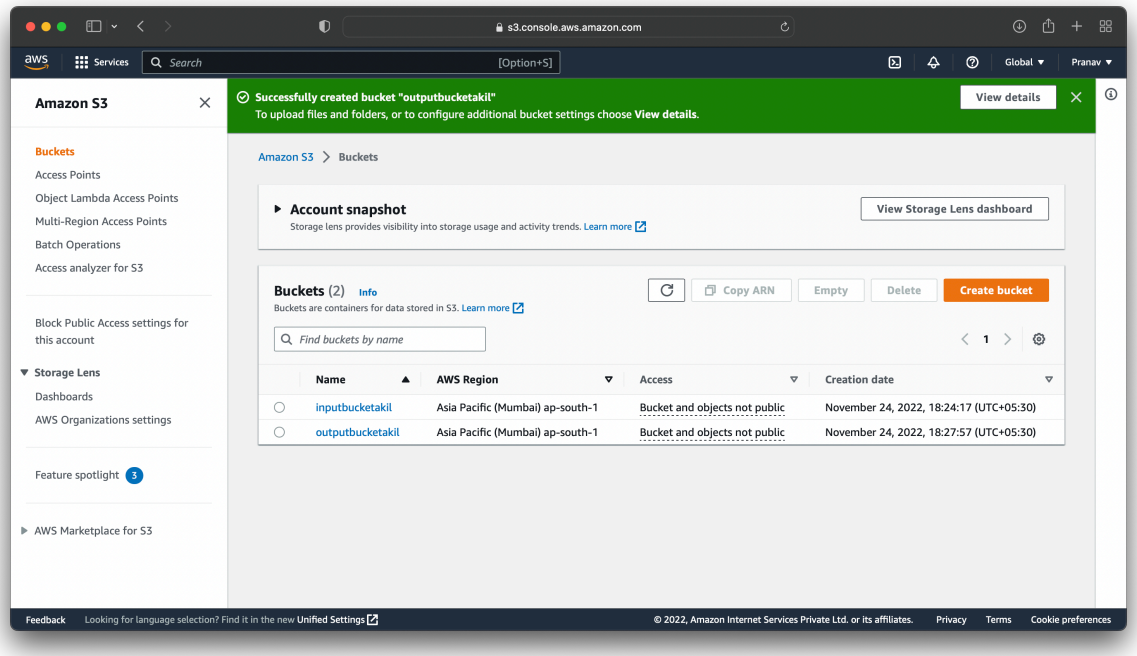
Aim: AWS Athena to query JSON/CSV files located in an s3 bucket

Procedure:

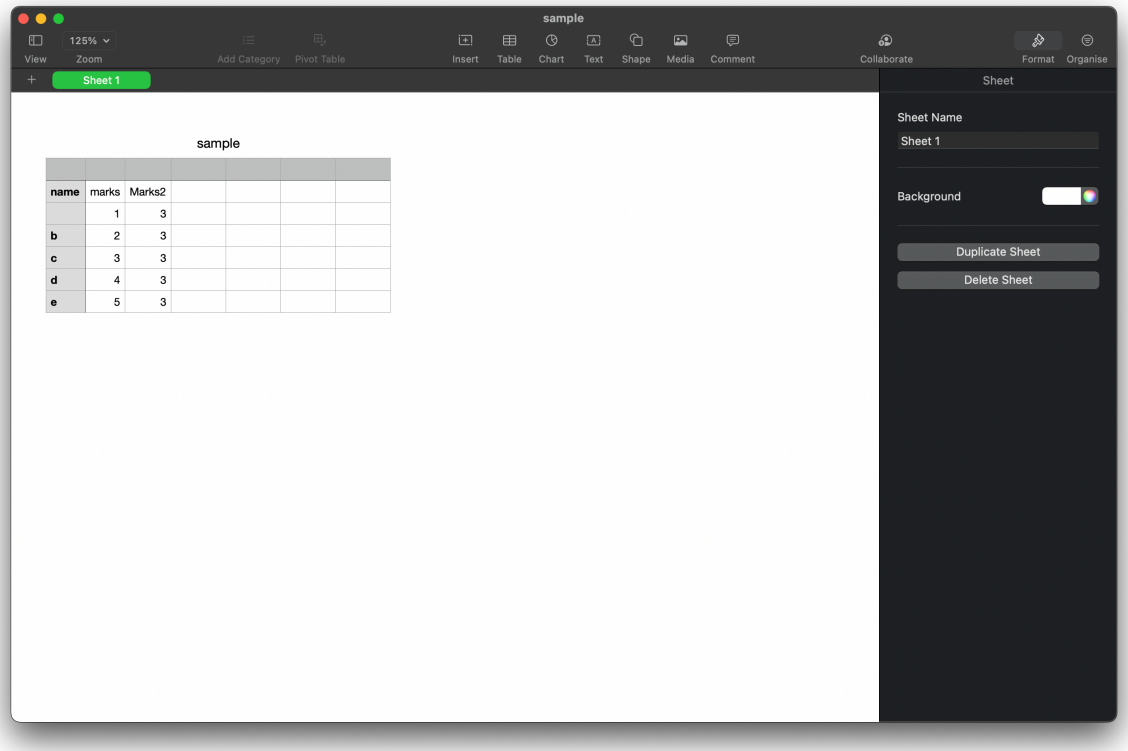
1. Firstly, open the AWS console homepage on browser (https://aws.amazon.com/console/) and select S3.



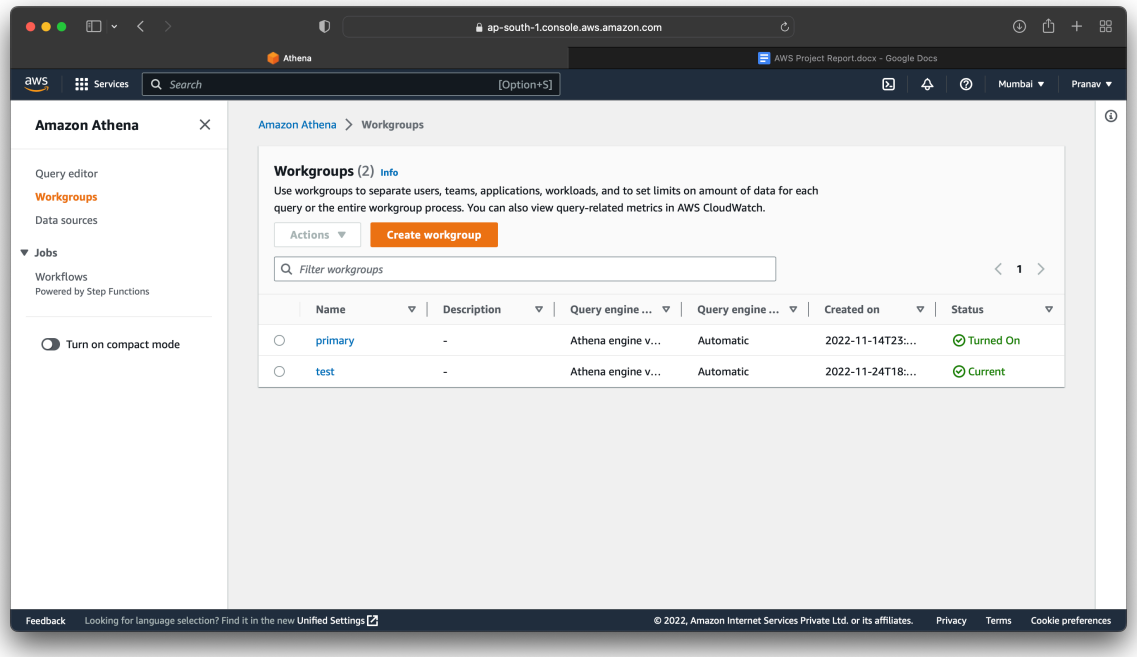
2. Create two buckets, one bucket for input data file and another for output result of the query.



3. Upload a sample dataset (json, csv, tsv, etc) file in your AWS S3 bucket.

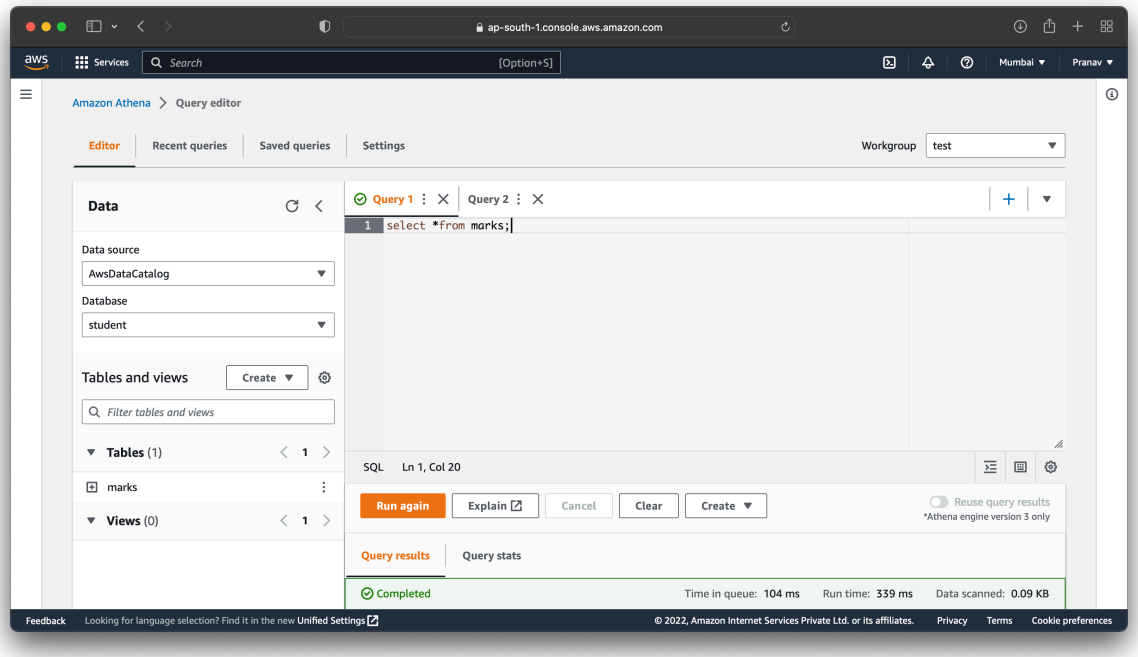


4. Go to AWS Athena.
5. Firstly, create a workgroup (workgroup is nothing but a kind of a container where our athena service stores the temporary data).

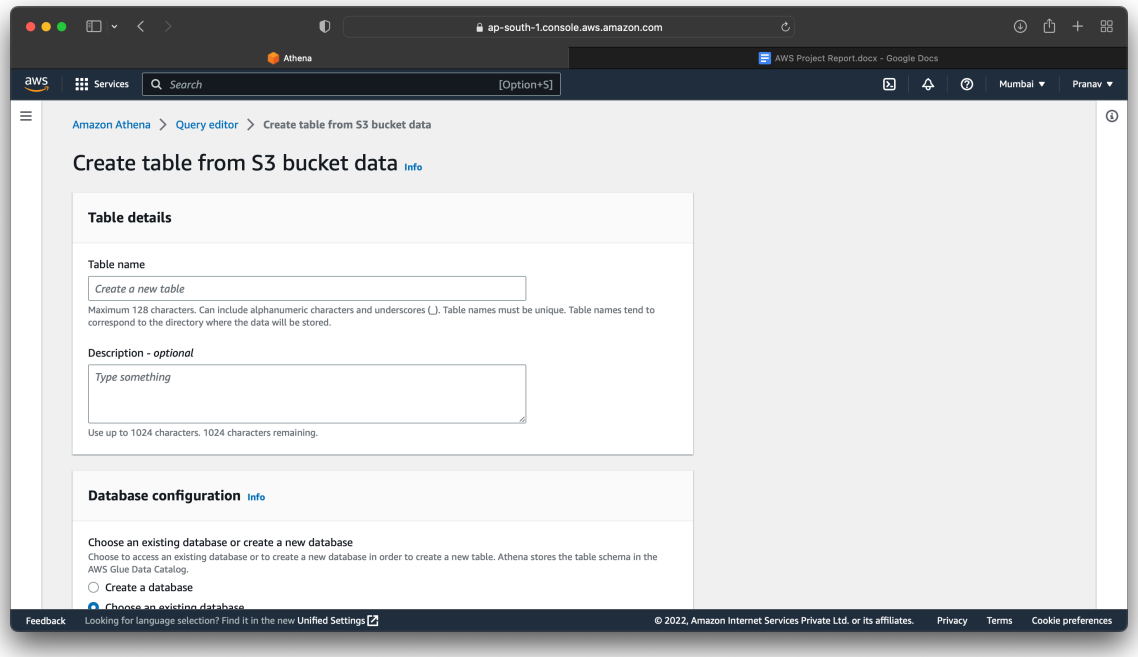


6. Give workgroup name, description, query result location, data usage limit.

7. Go to query editor panel, then go the settings, switch to your custom-made workgroup from the primary.
   There are two ways to query s3 dataset –
   - Using aws glue crawler
   - Select S3 Bucket

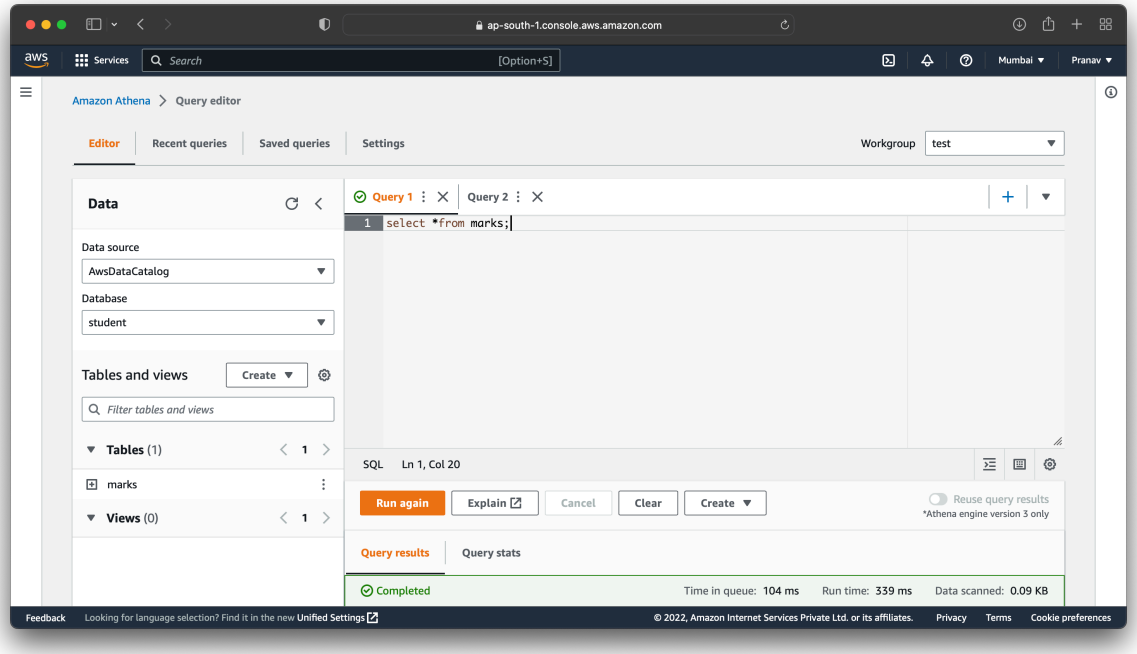


8. Create a table name and the database.

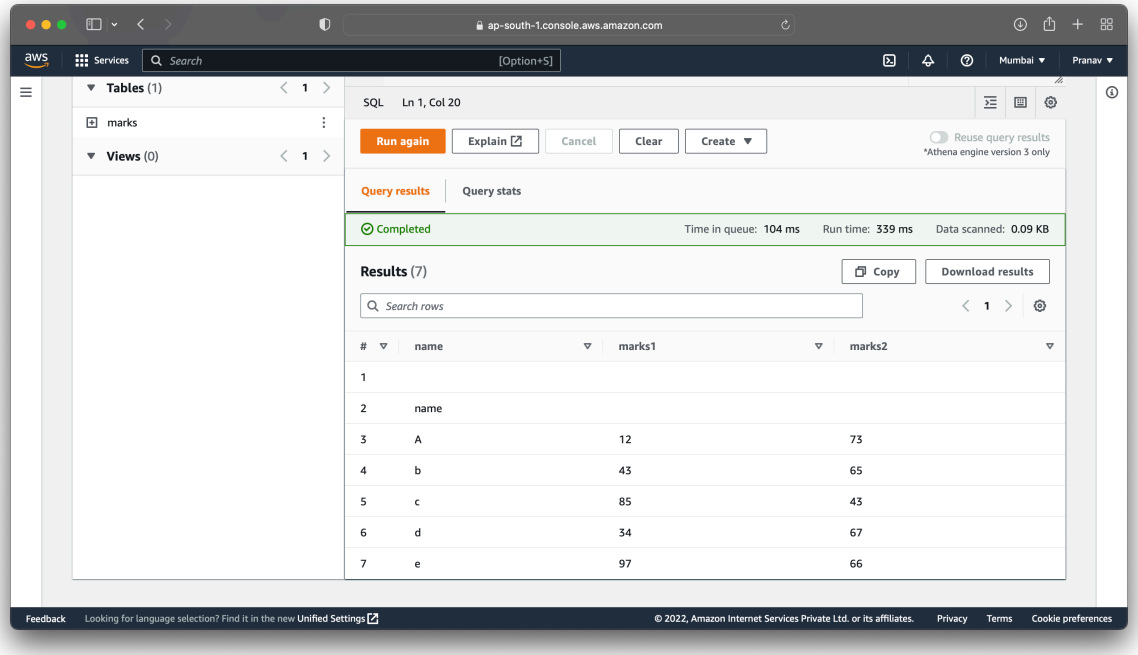


9. Select the input and output source with respect to the respective buckets created earlier.

10. Now go to the Query Editor and select the created database.



11. Run the desired queries and verify the output.



Result:

We have successfully used AWS Athena to query JSON/CSV files located in an s3 bucket by setting up an Athena Database and Table using AWS Glue's Crawler.