**EX NO 7 QUERYING DATA IN S3 WITH AMAZON ATHENA**

**Aim**:

# To create a table and access the data by SQL Queries using Amazon Athena.

**YOU CAN USE AWS LEARNERS LAB**

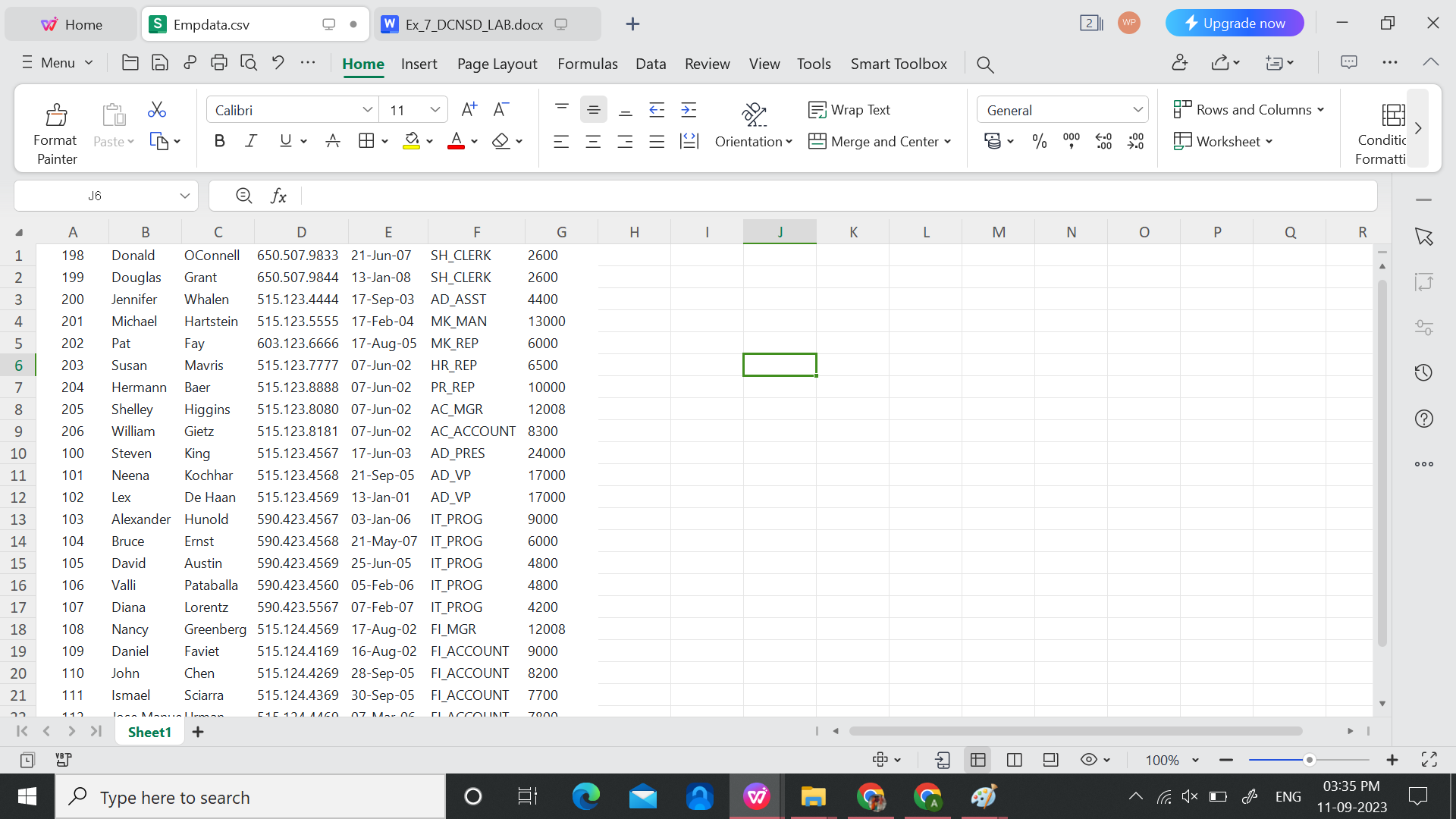
**Amazon Athena:**

* Amazon Athena is a serverless, interactive analytics service built on open-source frameworks, supporting open-table and file formats.
* Athena provides a simplified, flexible way to analyze petabytes of data where it lives.
* Analyze data or build applications from an Amazon Simple Storage Service (S3) data lake and 30 data sources, including on-premises data sources or other cloud systems using SQL or Python.
* Athena is built on open-source Trino and Presto engines and Apache Spark frameworks, with no provisioning or configuration effort required.

**CREATE A CSV FILE (empdata.csv)**

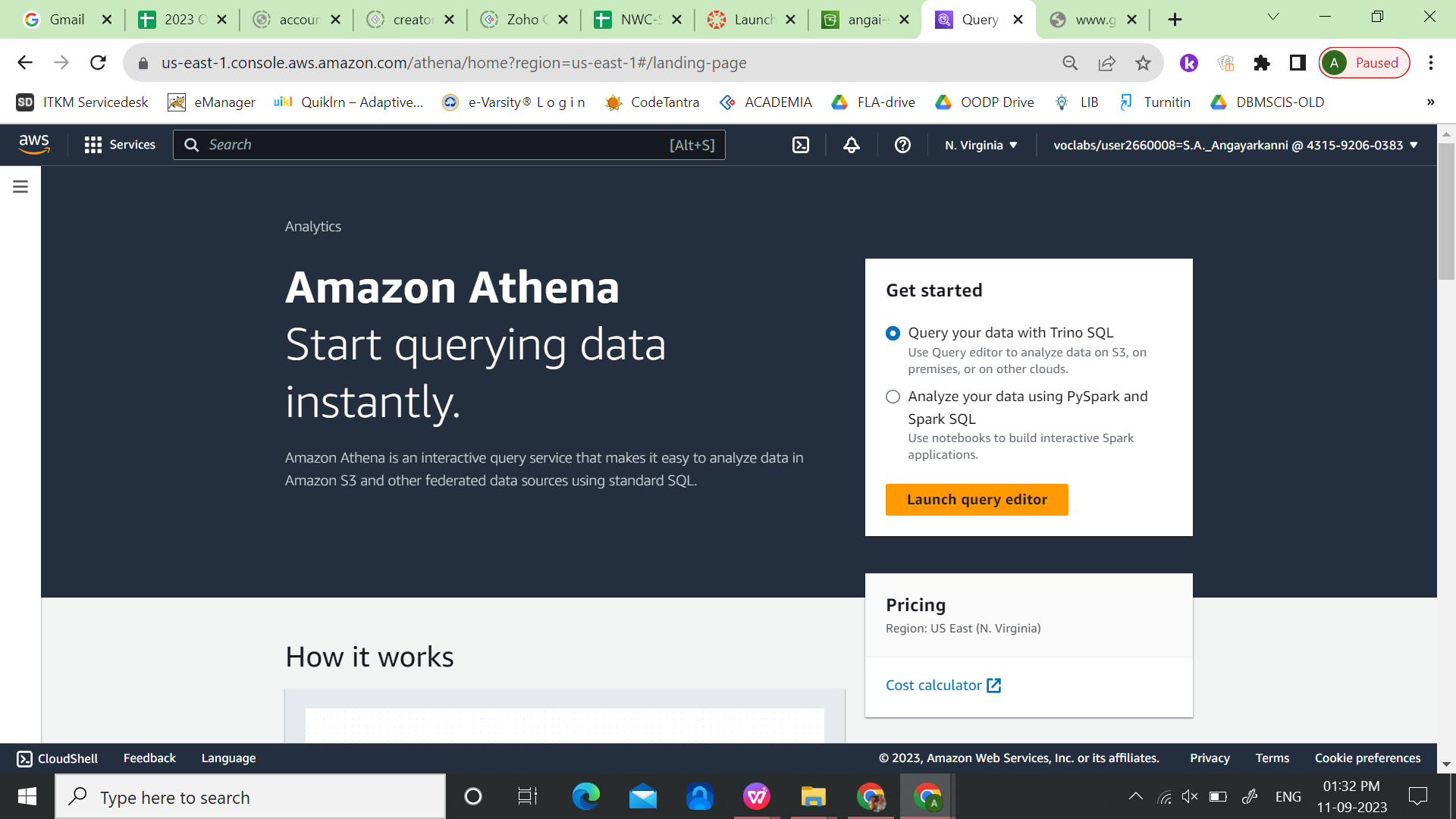
Use spreadsheet to create a new CSV file.

Sample data is shown below.



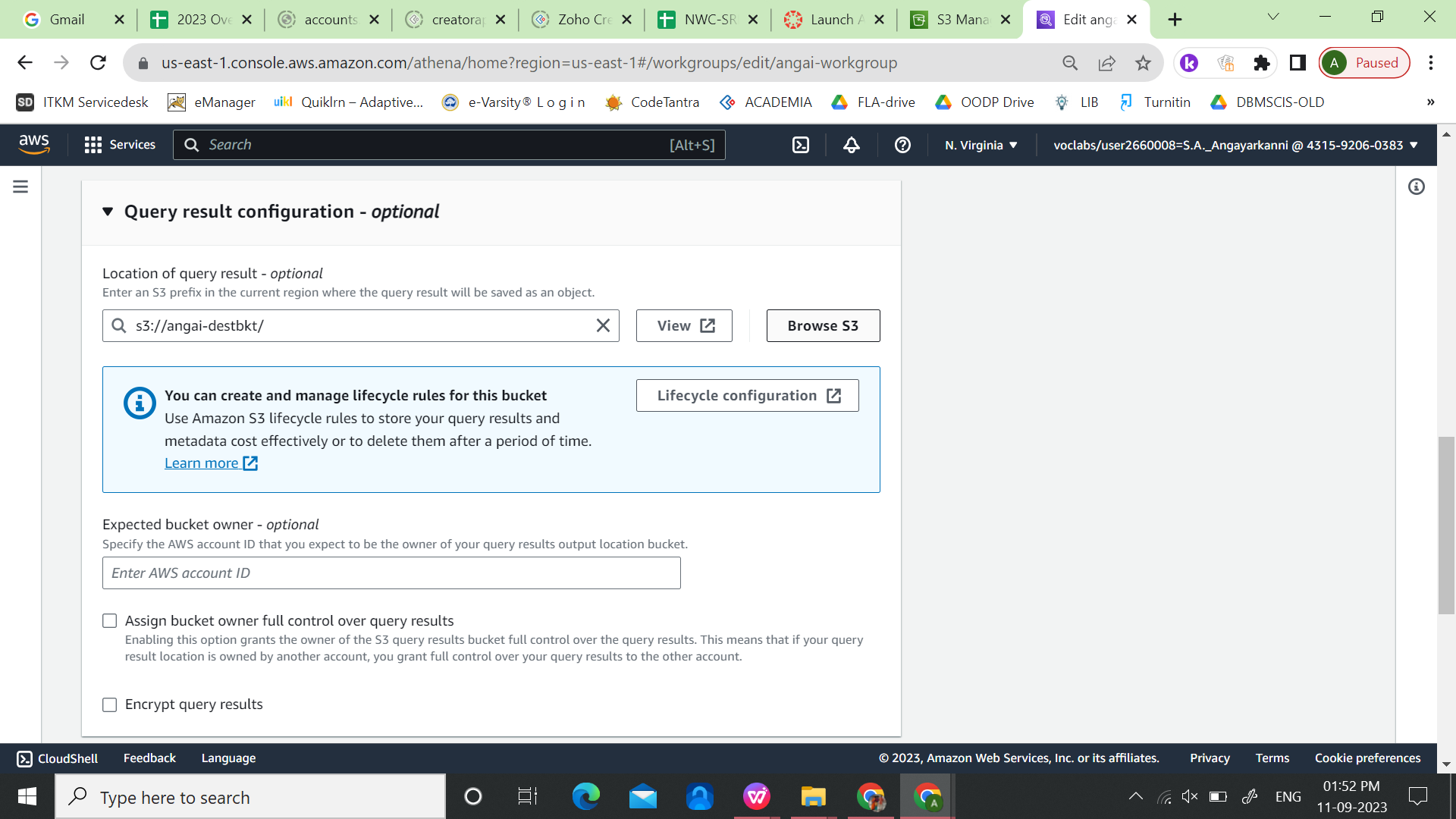
**CREATE TABLE FORM EXISTING CSV FILE**

1. Open the [Amazon S3 console](https://s3.console.aws.amazon.com/s3" \t "https://repost.aws/knowledge-center/_blank).
2. Choose Create two Bucket.
3. Choose a [DNS-compliant name](https://docs.aws.amazon.com/AmazonS3/latest/userguide/bucketnamingrules.html" \t "https://repost.aws/knowledge-center/_blank) for your new bucket (bucket-src).
4. Click the created bucket and Upload a new file [We can use any csv to upload existing data into the database, here ‘empdata.csv’ is taken]
5. Choose Create Bucket.
6. Choose a [DNS-compliant name](https://docs.aws.amazon.com/AmazonS3/latest/userguide/bucketnamingrules.html" \t "https://repost.aws/knowledge-center/_blank) for your new bucket (bucket-dest).

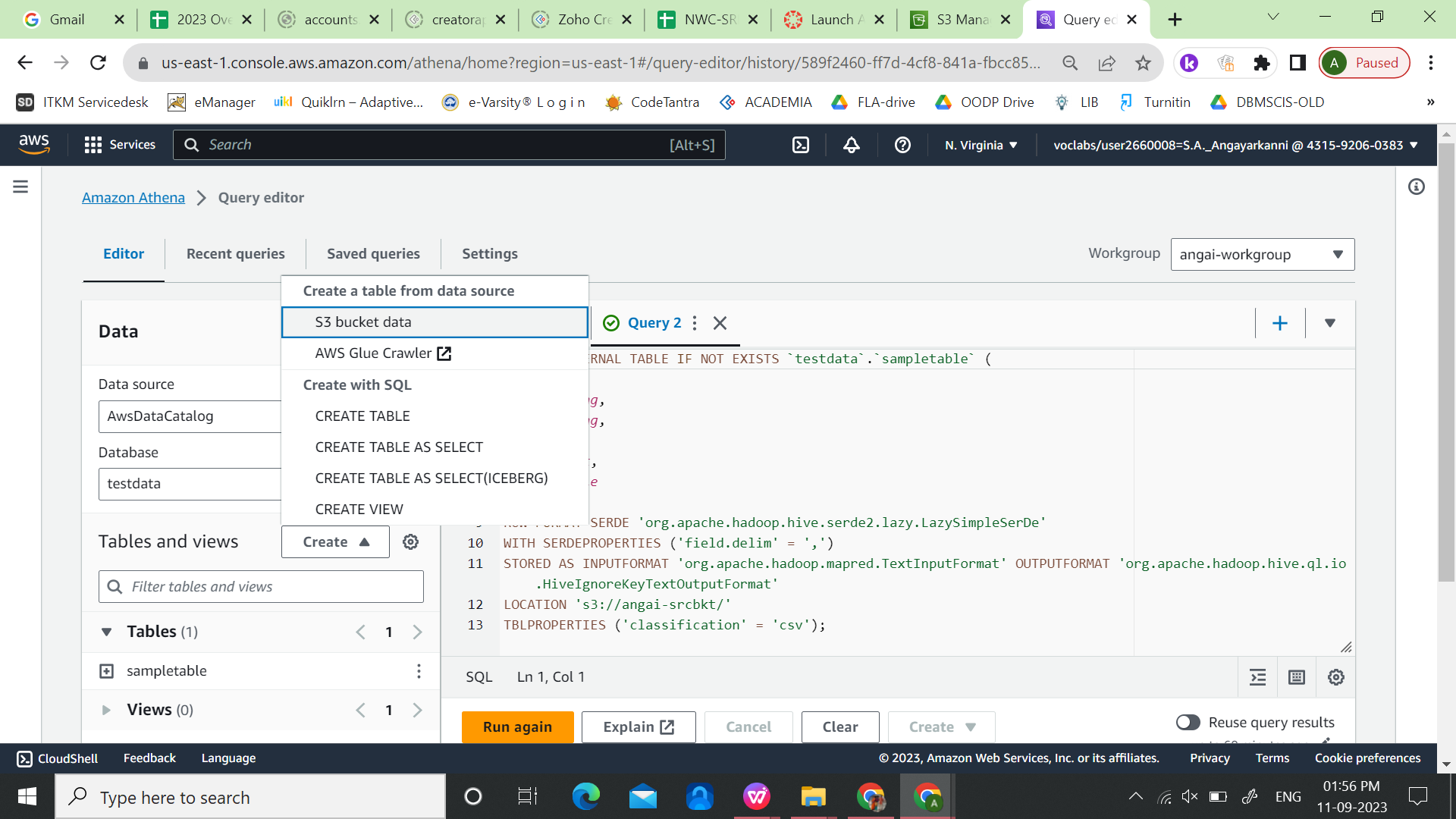


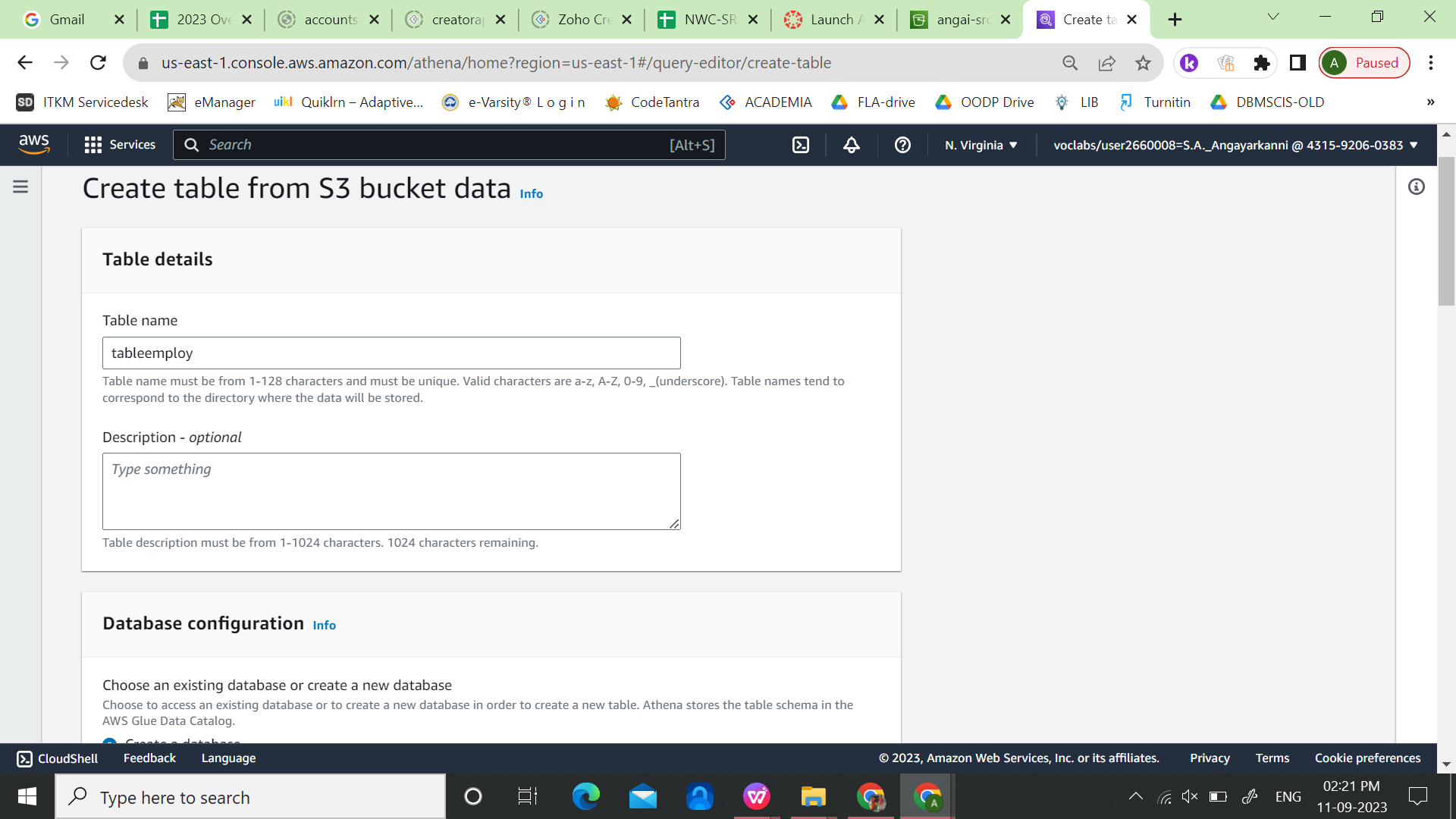
1. Open the Amazon Athena console.
2. Select Workgroup --> Create Workgroup (here, I am giving ‘angai-workgroup’

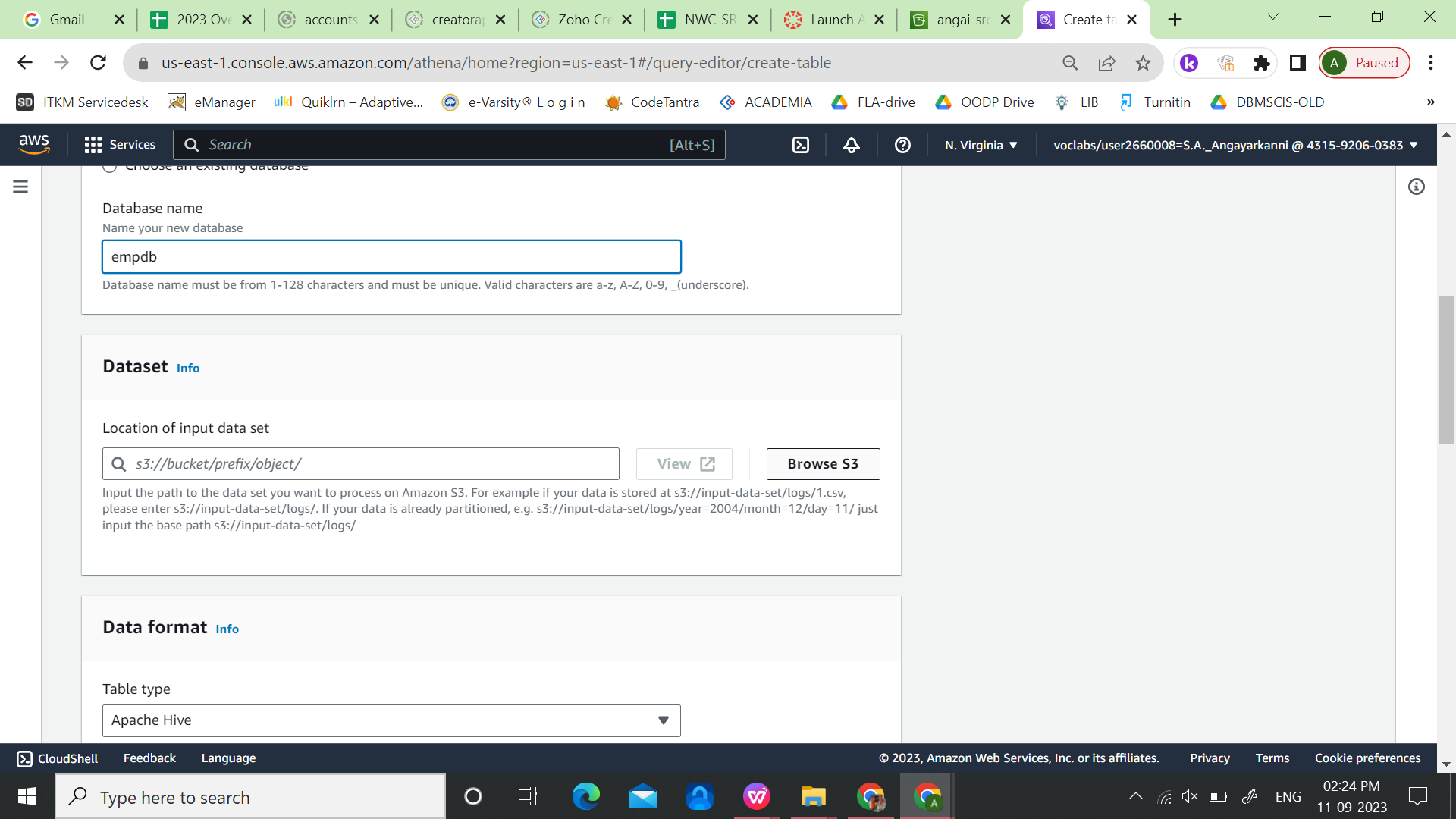
Under ‘Query result configuration’, browse and select the destination bucket



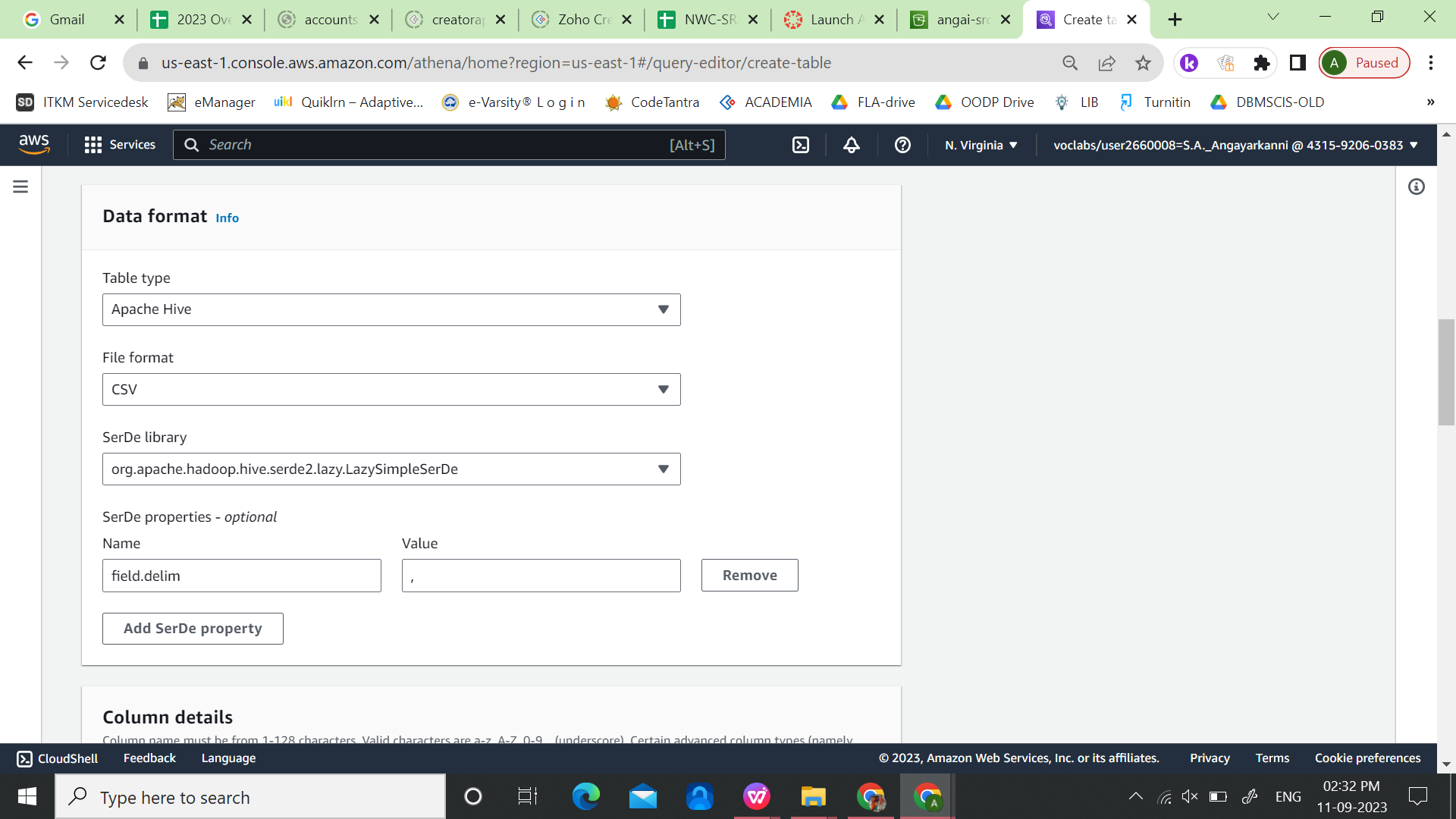
Select Query Editor. Click Create --> Create a table from data source --> s3 bucket data



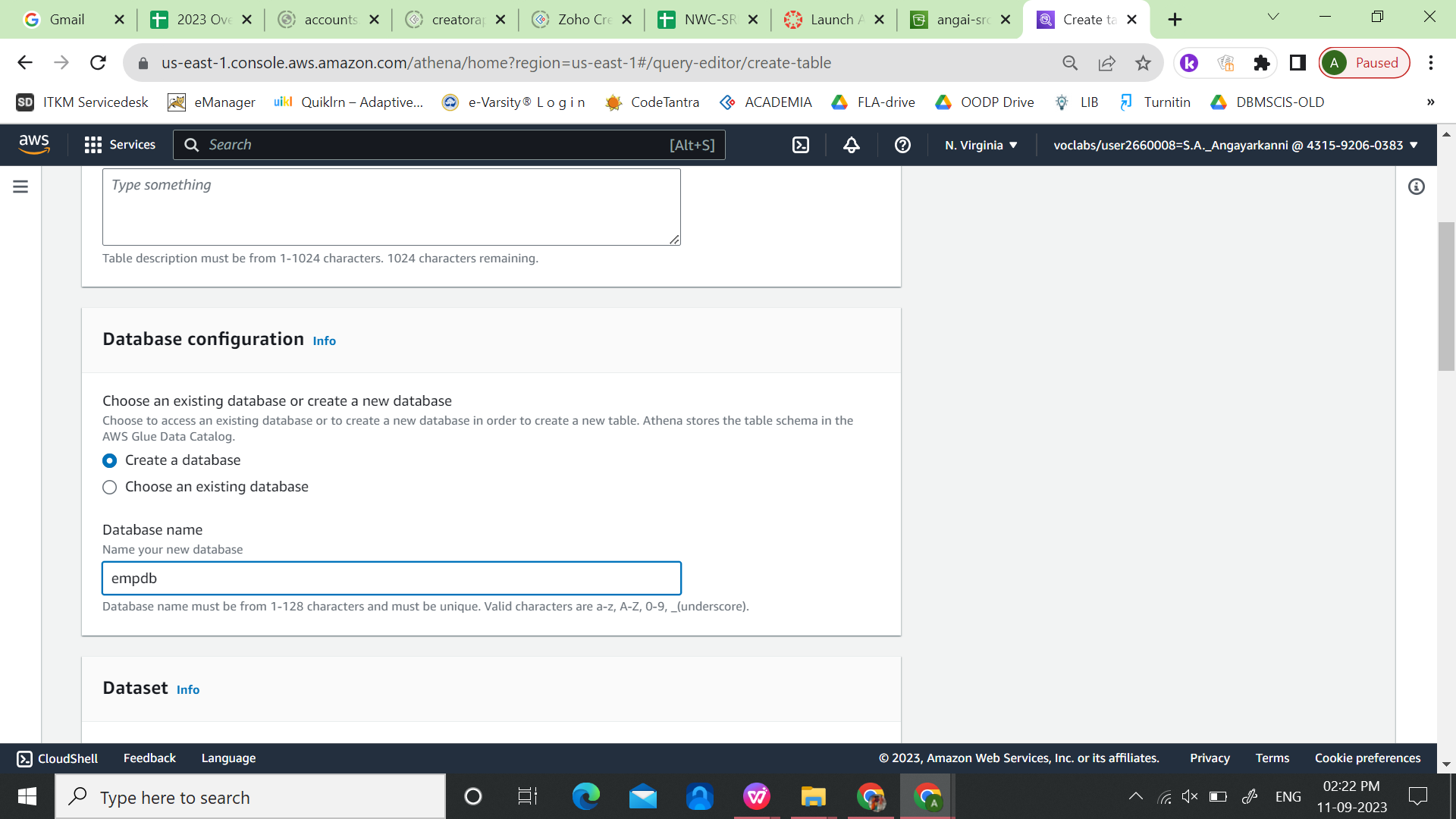
Select ‘create a database’ and give tale name as ‘empdb’.



Under dataset, For the location of input data set, click ‘Browse S3’ and select the csv file uploaded inside the source bucket.

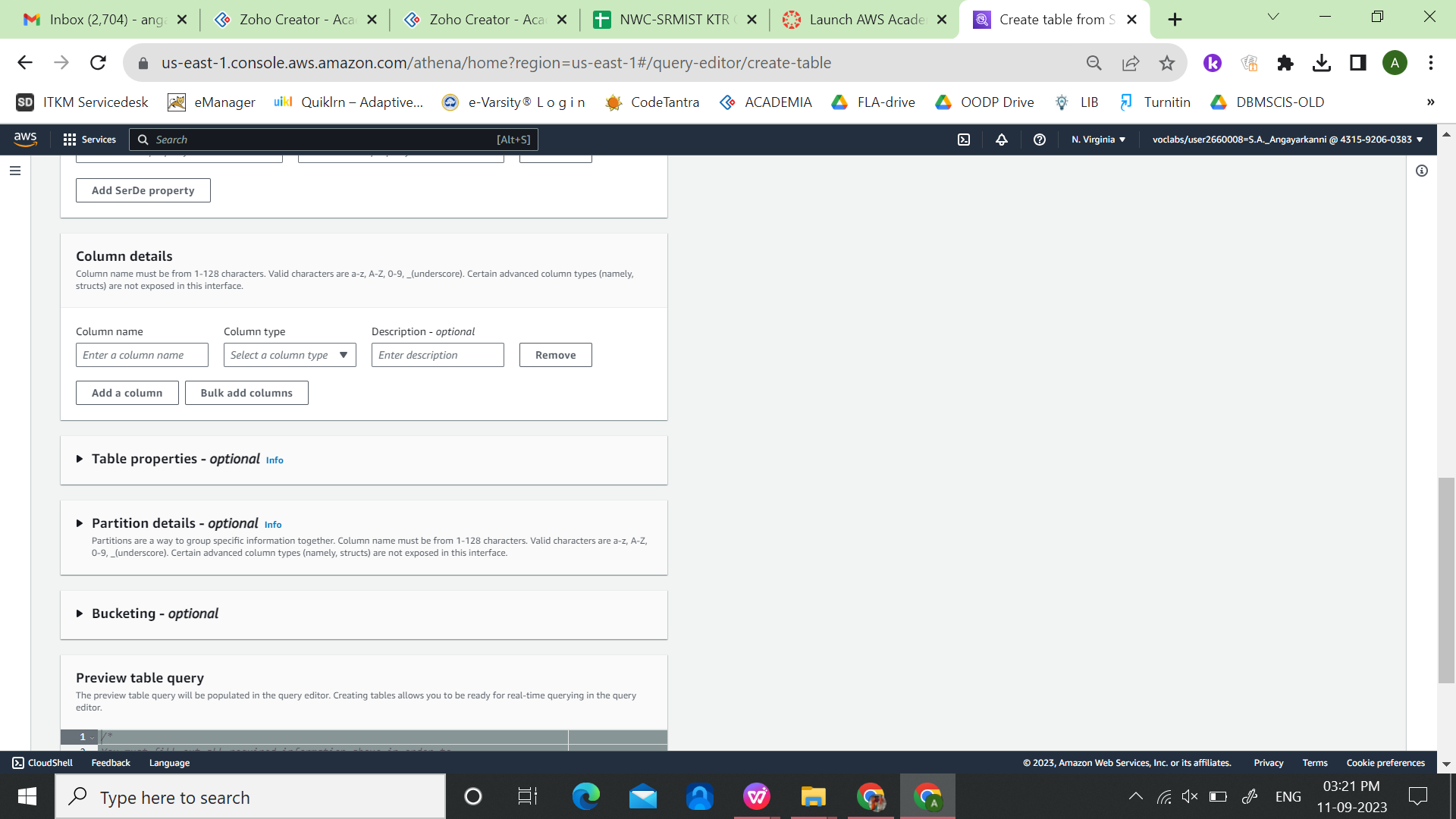


Select the File format as ‘CSV’



Under Column details

Add column by specifying name and datatype of that column.



Column Name Datatype

EID int

EMPFNAME string

EMPLNAME string

IPADDR string

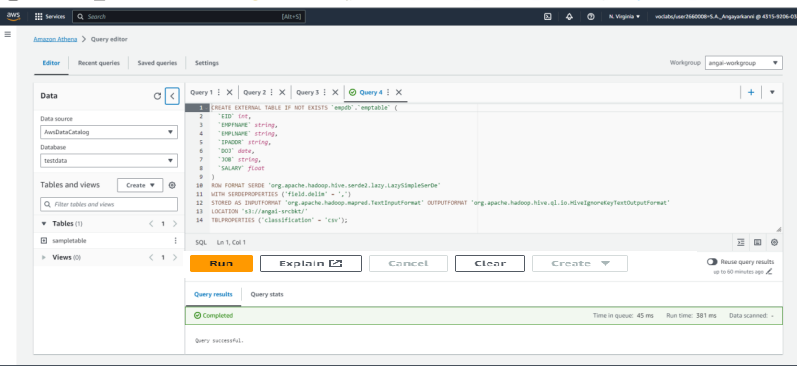
DOJ date

JOB string

SALARY float

Click ‘Create Table’.

DDL command to create the table is now avaiable in the query editor.



Click the ‘Run’ button to execute the SQL query. Now the table is created and a ‘completed’ message is shown.

Select query can be used to display the table values.

ADd new Query in the query editor, and type

select \* from "empdb"."emptable" limit 10;

Execution of the above query displays the first 10 rows of the employee table, which was created from CSV file.

