**IF184966 - Big Data**

**Academic Year: 2022/2023**

Student’s ID : 5025201251

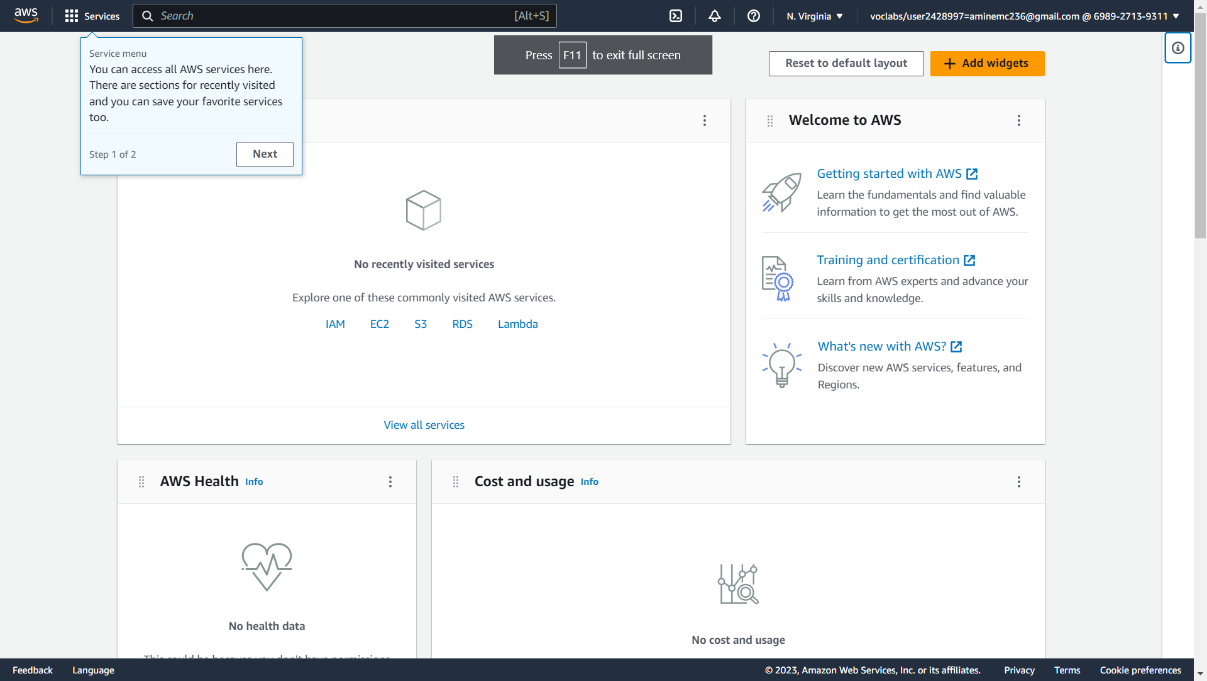
Student’s Name : Muhammad Amin­­

**Task 1**

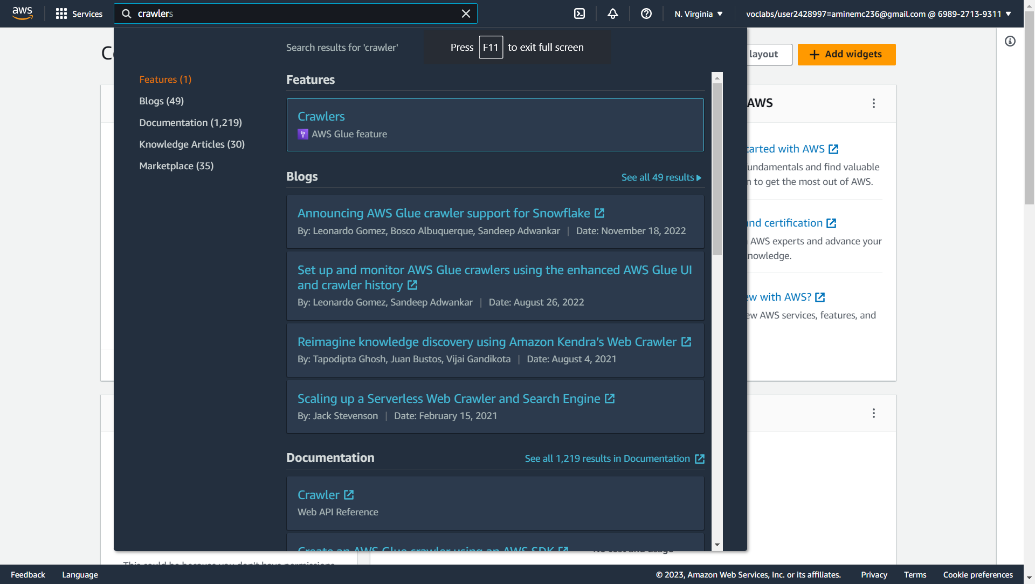
* Do the task on **AWS Academy Data Analytics [39663]: Lab 3 - Query data in Amazon S3 with Amazon Athena and AWS Glue**.
* Make documentation for each step and give a screenshot for each of them.

Every screenshot should display your account in the top right corner.

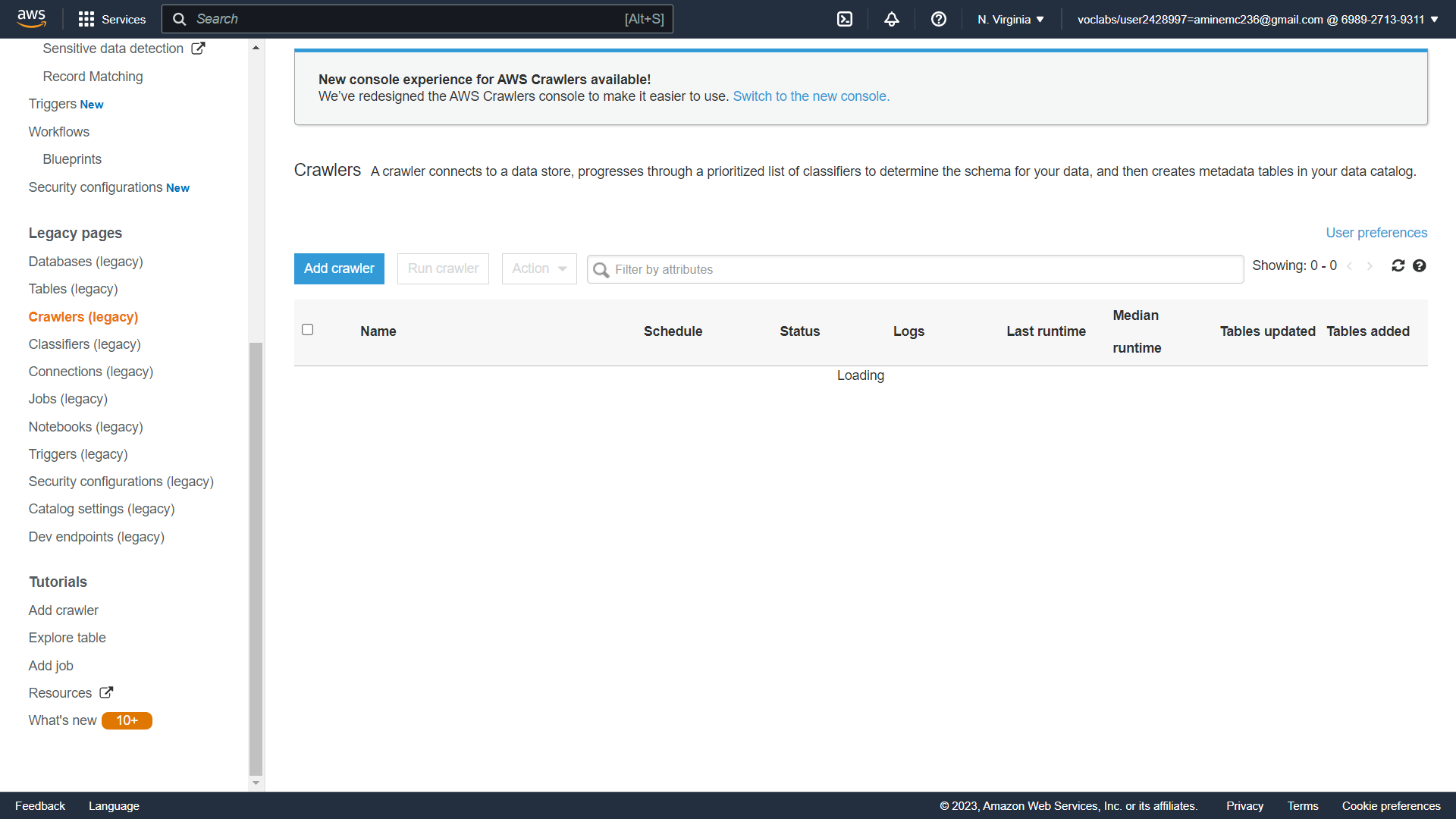
This is the dashboard



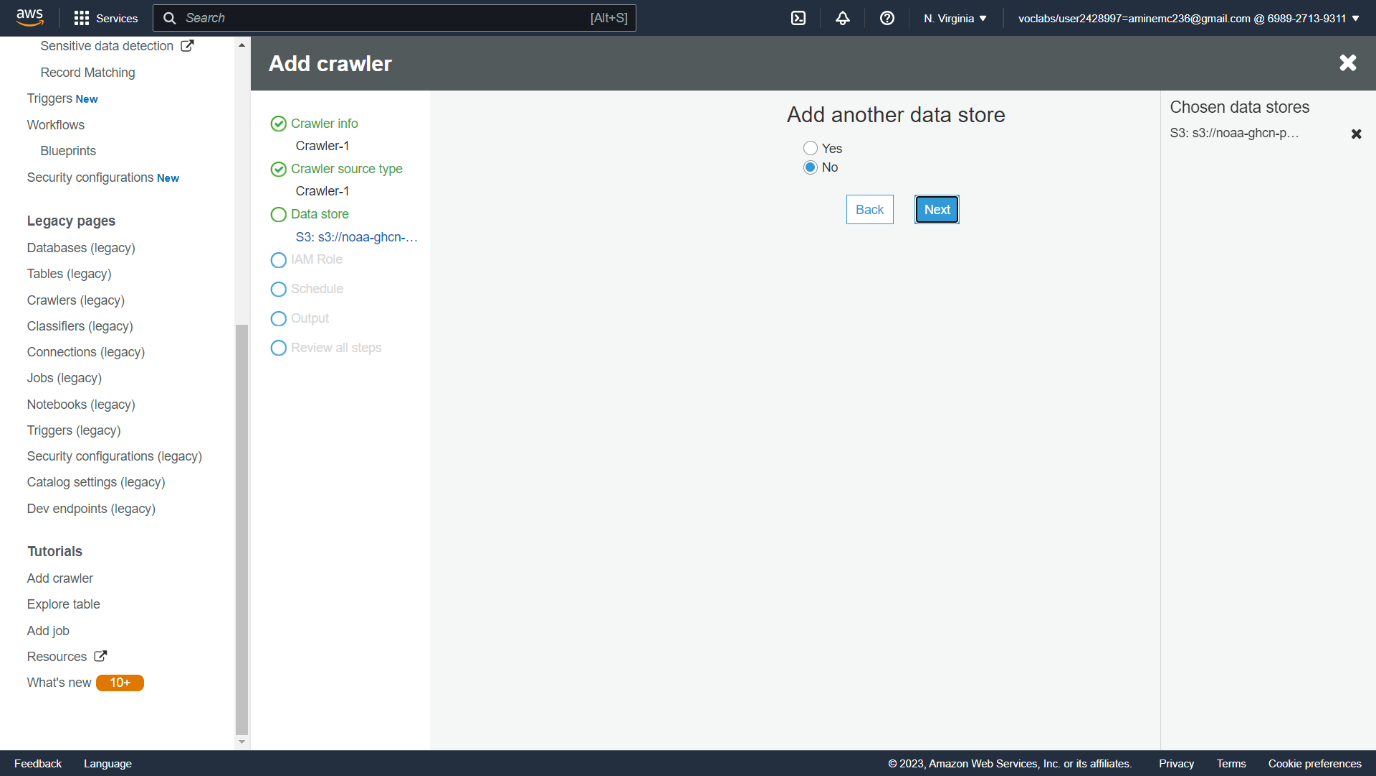
search crawlers



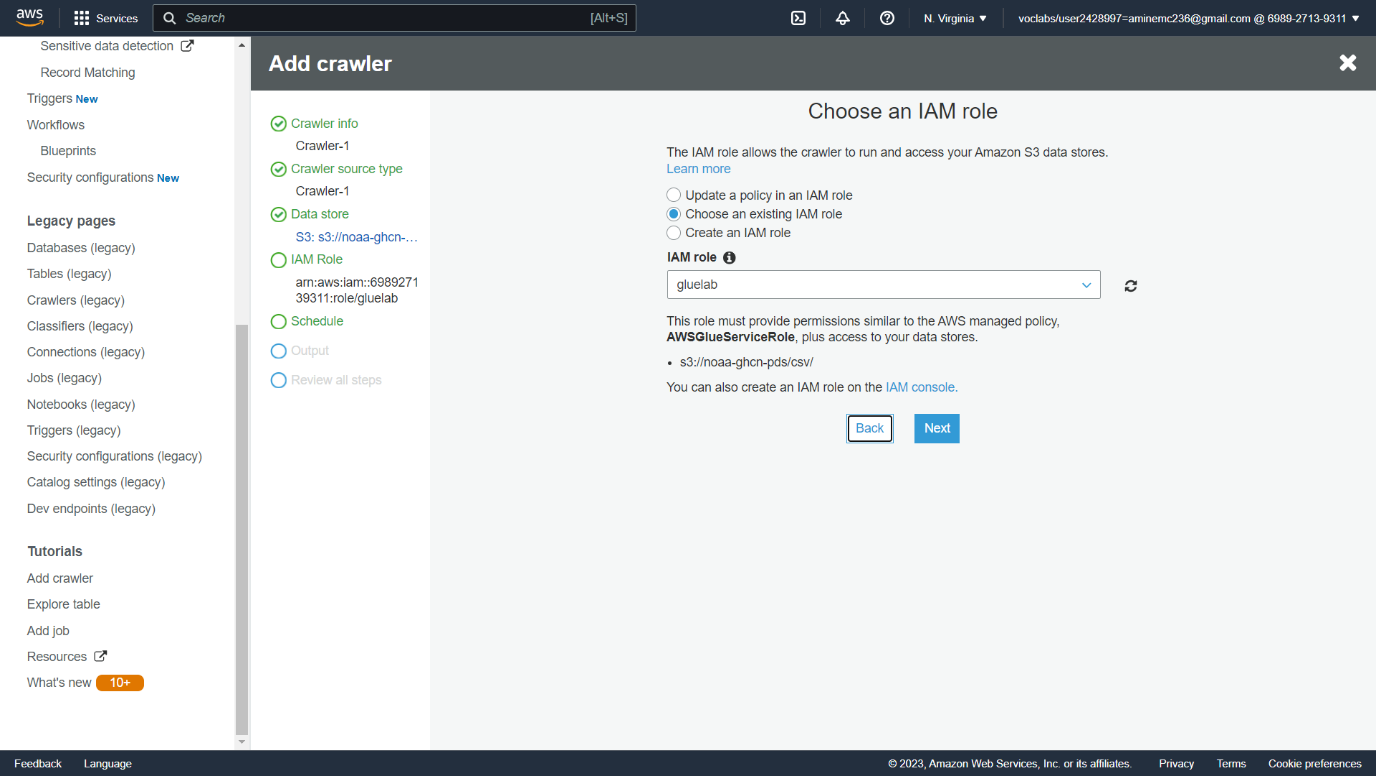
Press the add crawler button to add crawler



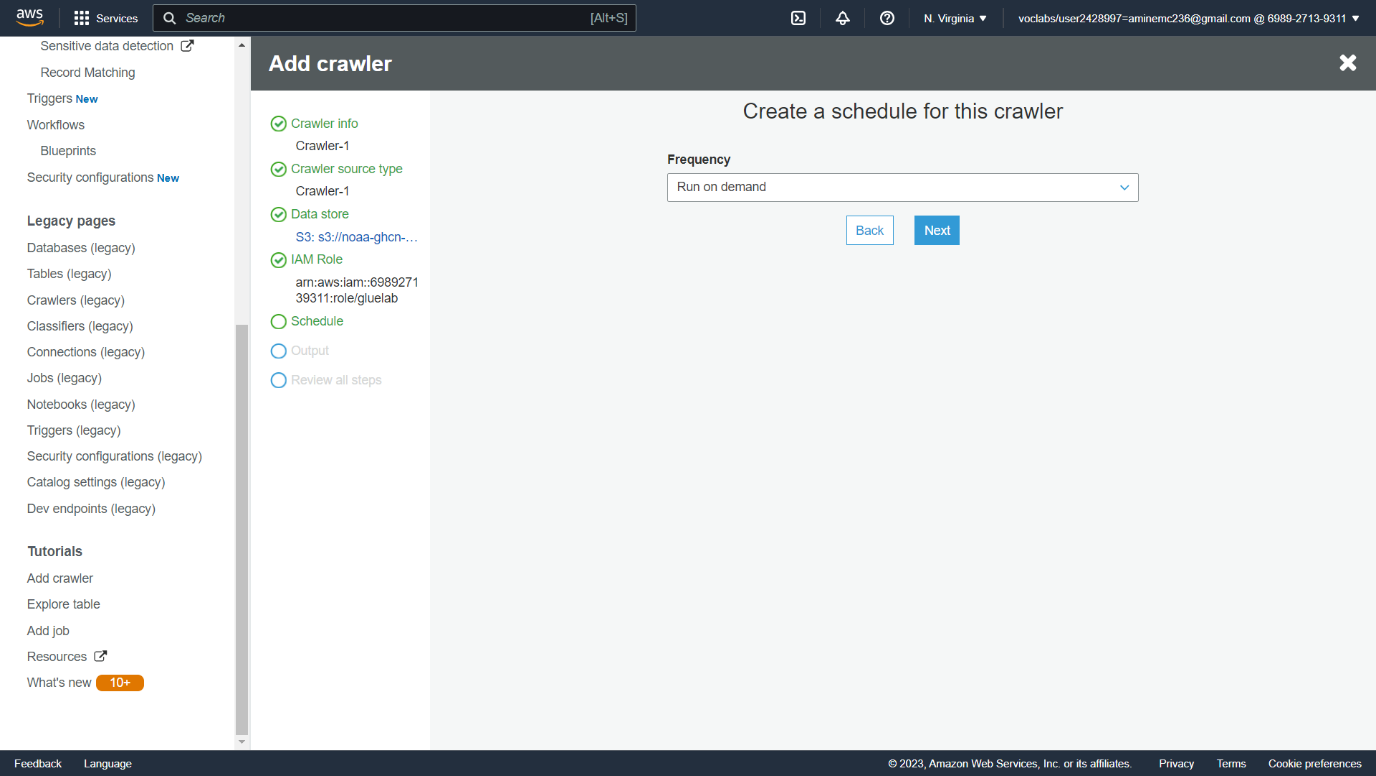
Choose No

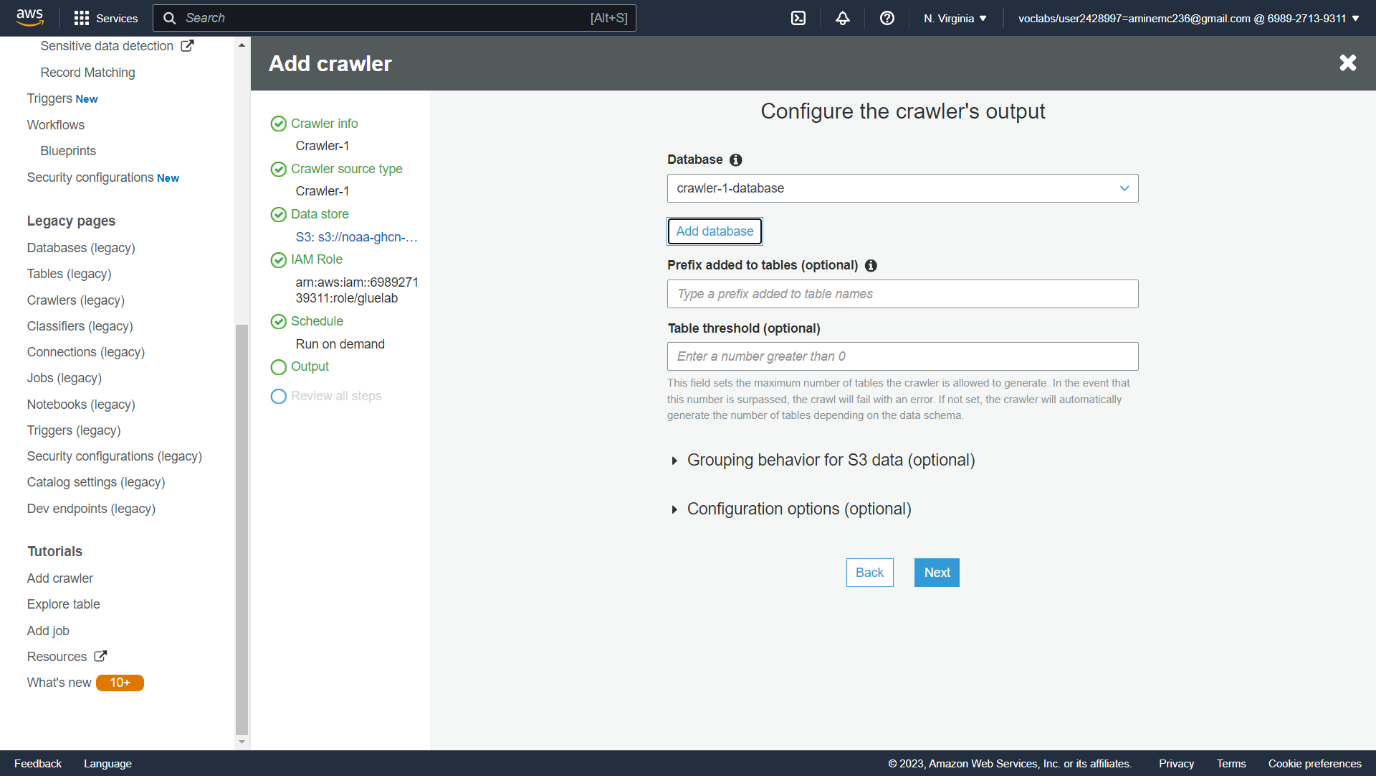


Then choose an existing IAM role and choose gluelab

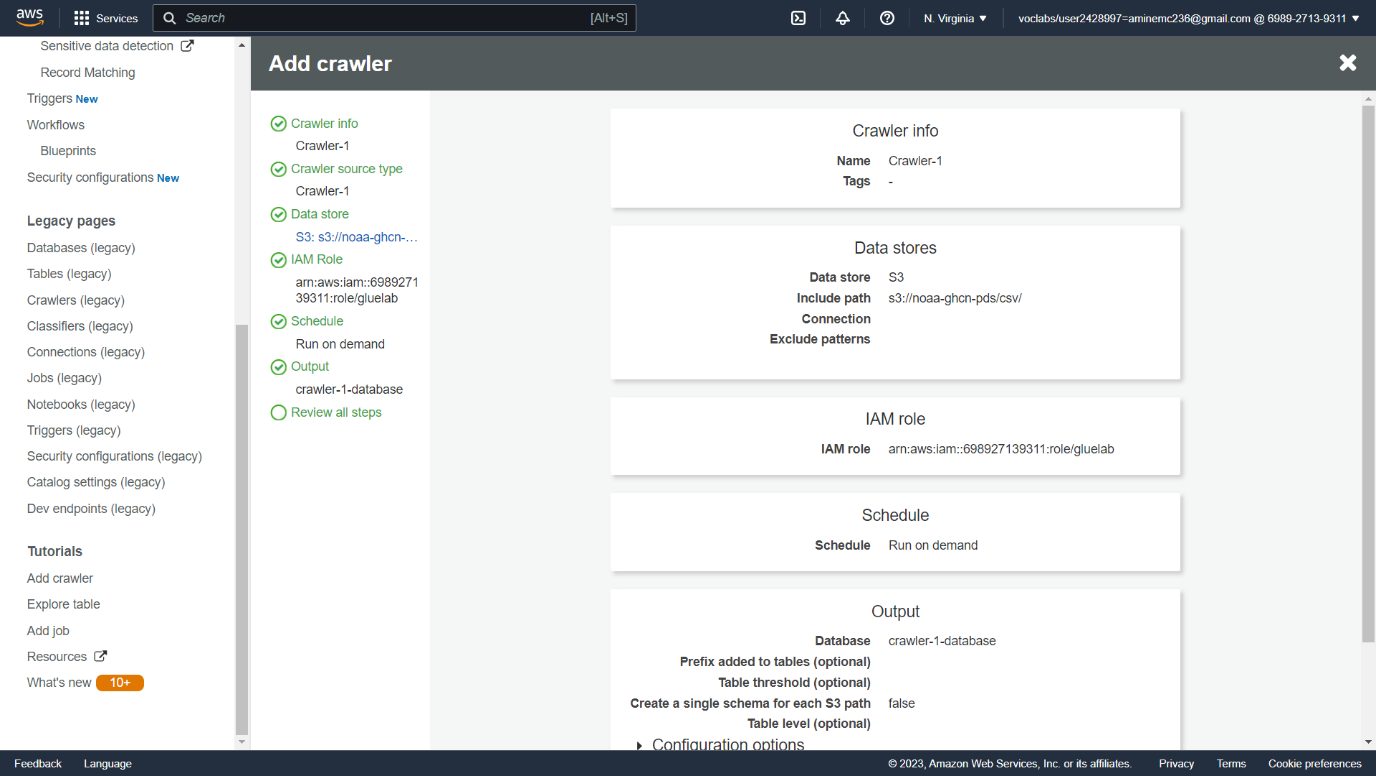


Just choose the default, Click next

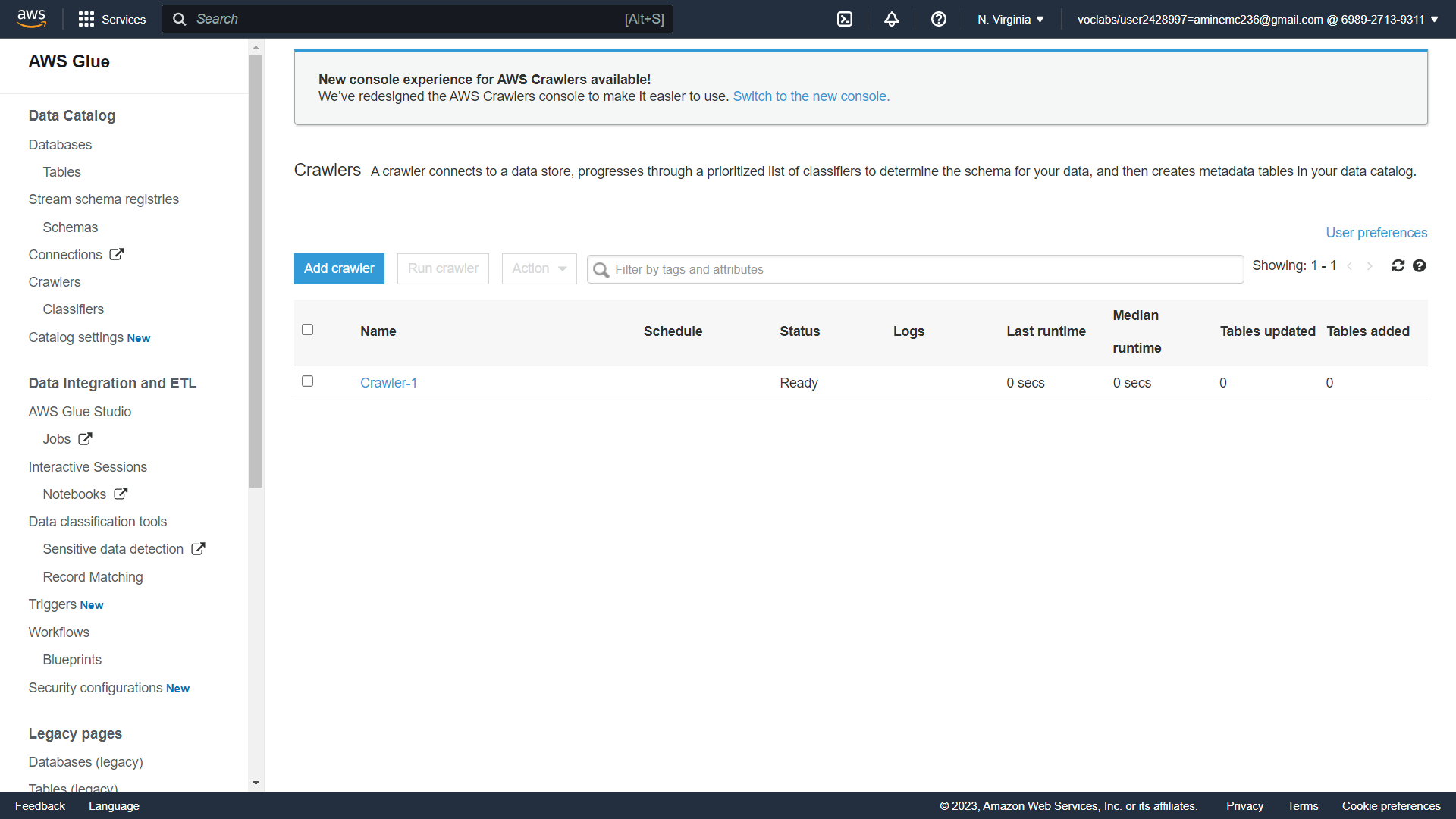


In the configuration, set a database name then click next 

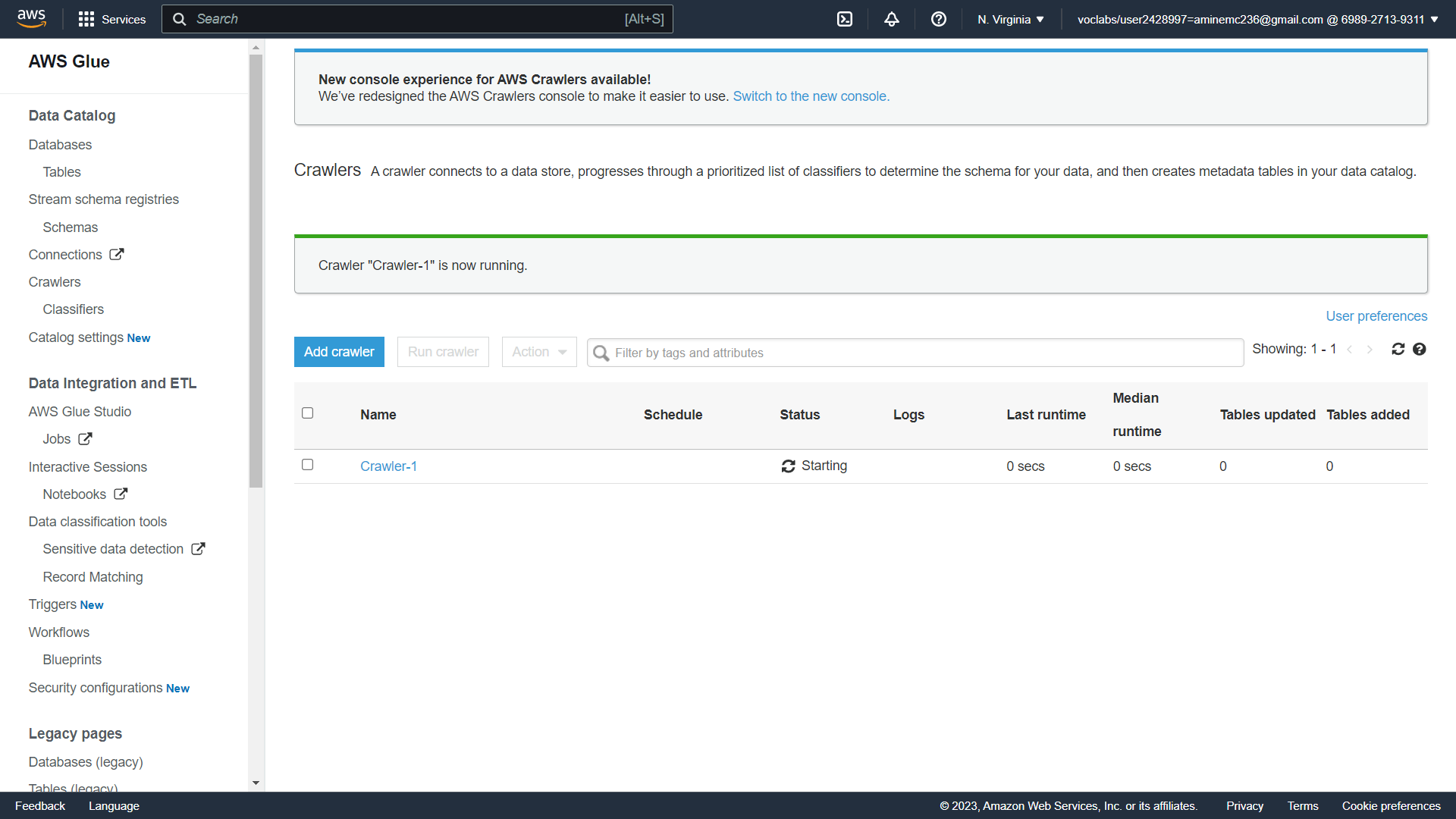
Review the crawler information



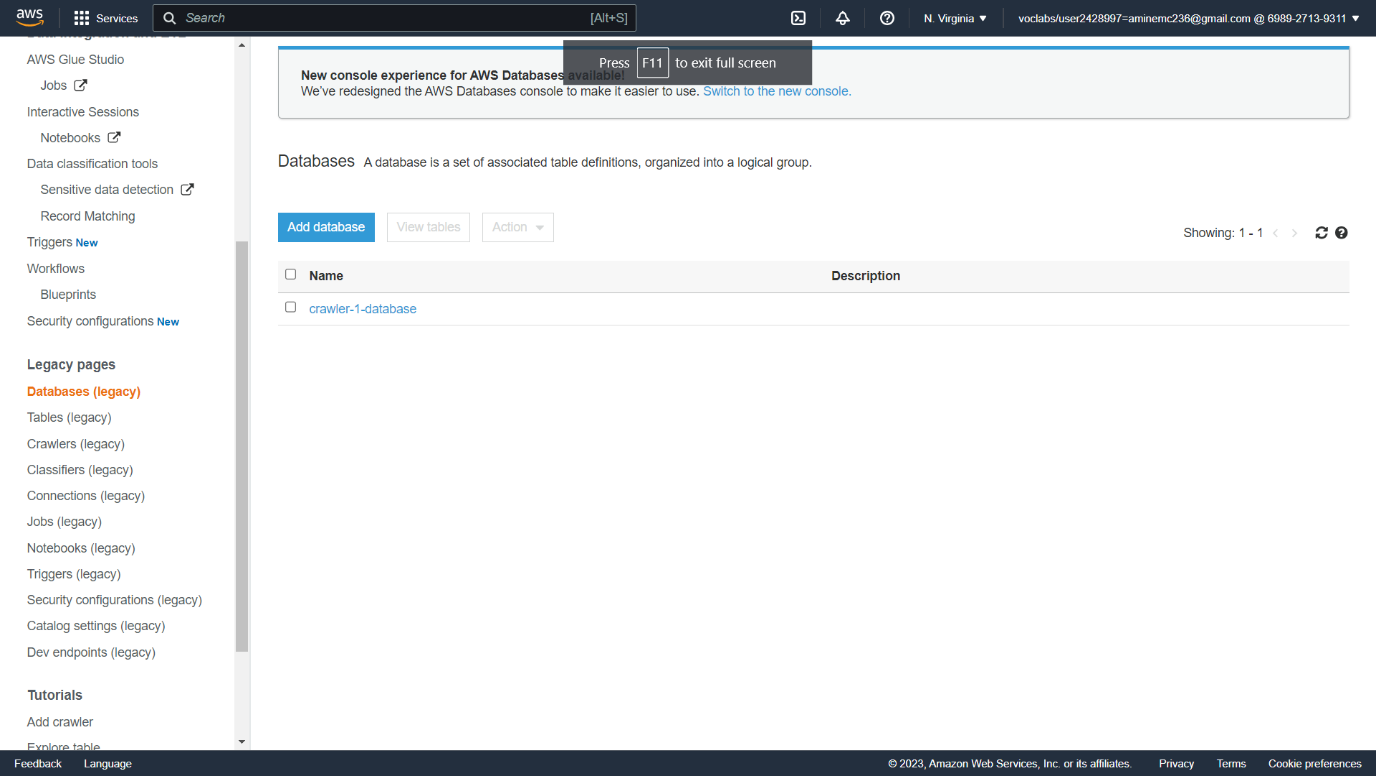
The crawler is ready to use



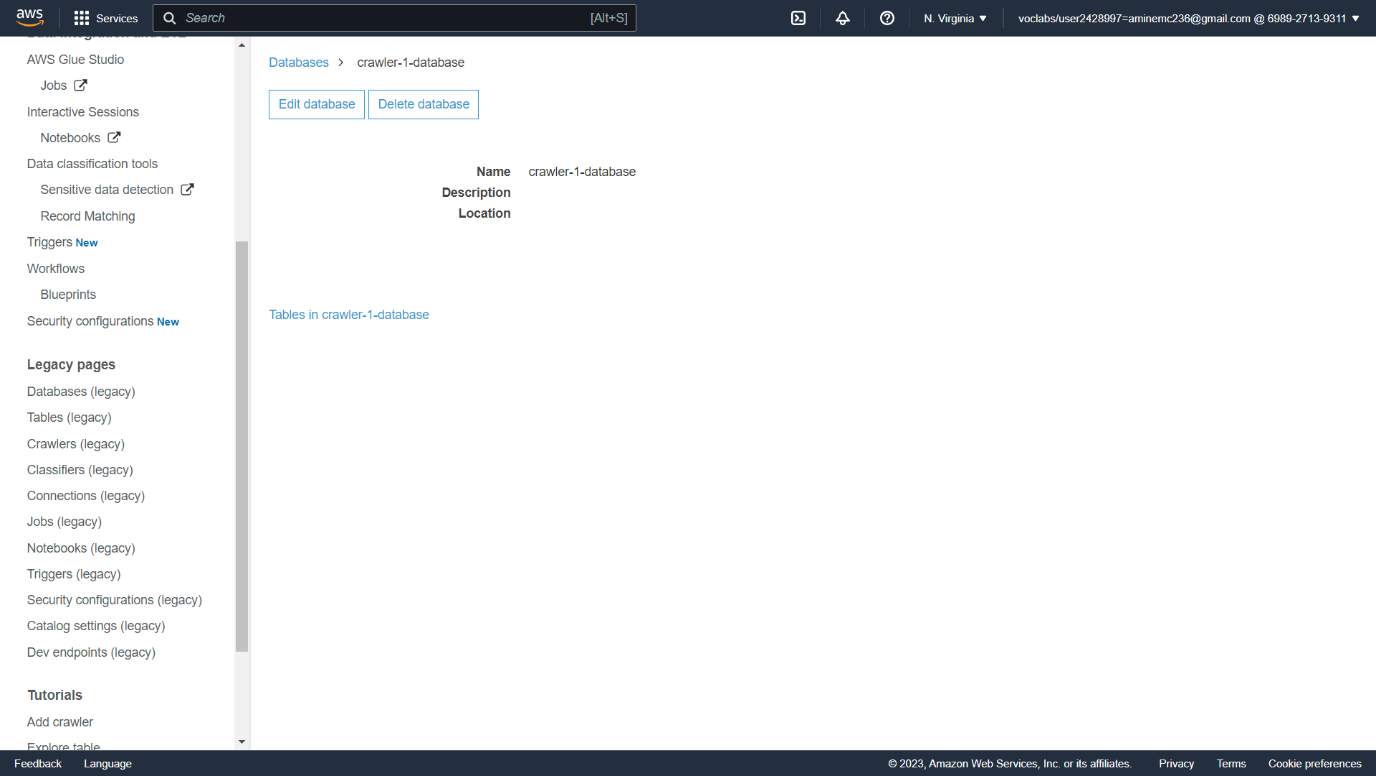
Run the Crawler

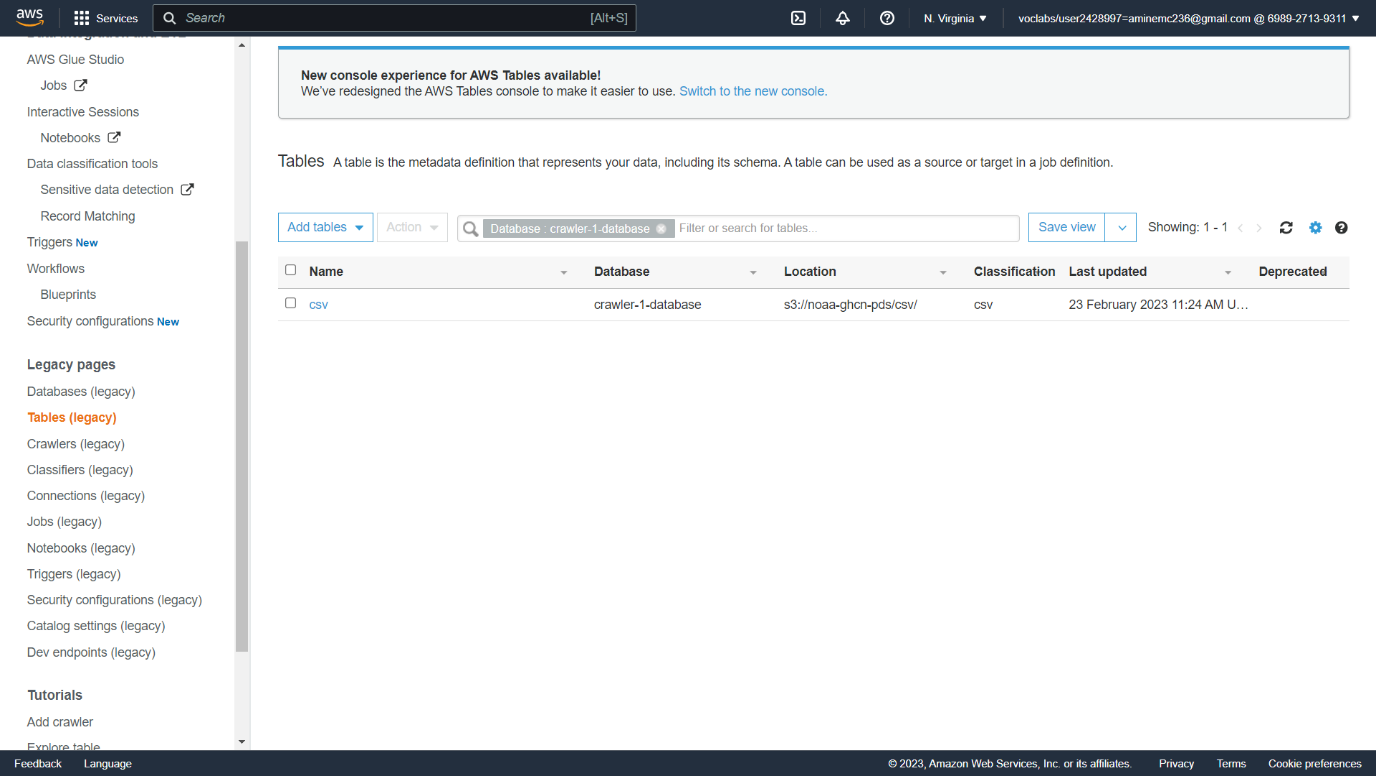


Go to databse (legacy) to see the database

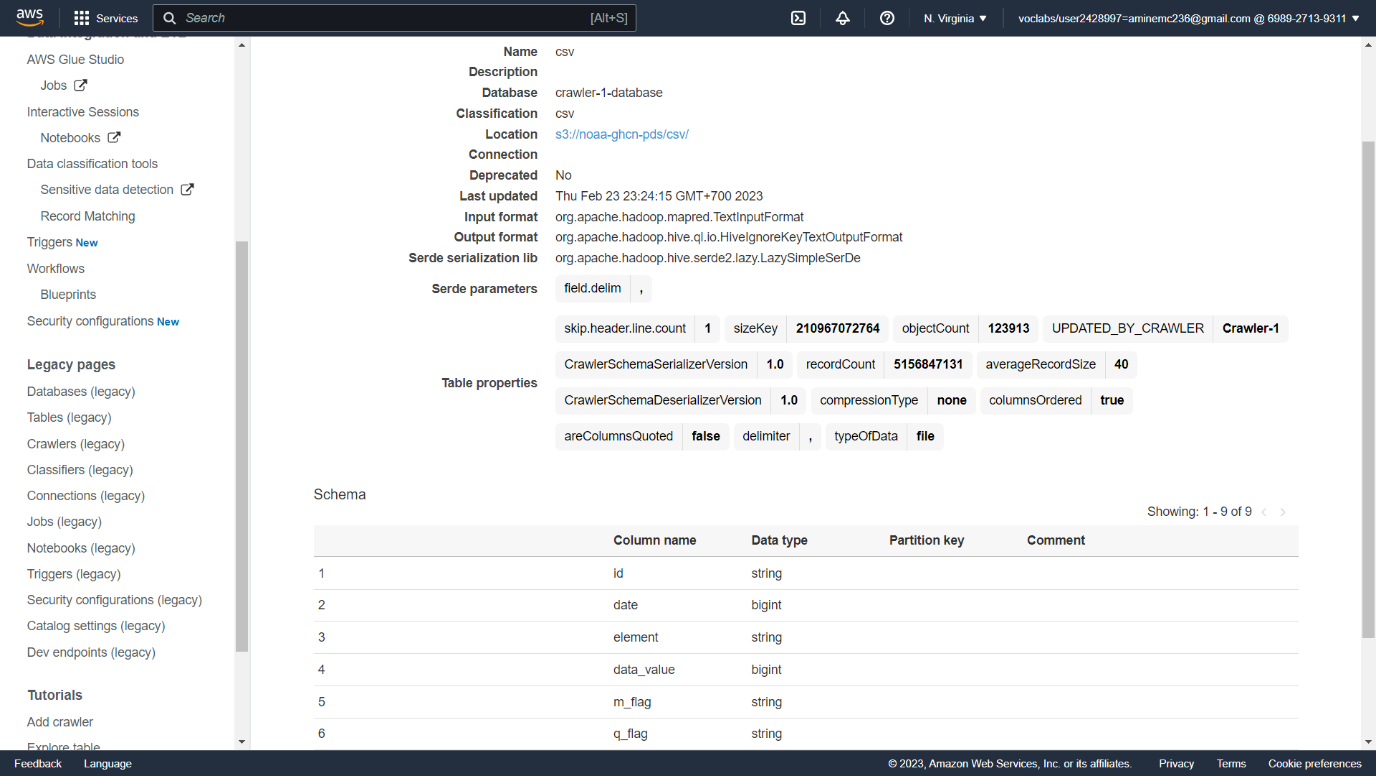


After that try to edit the database

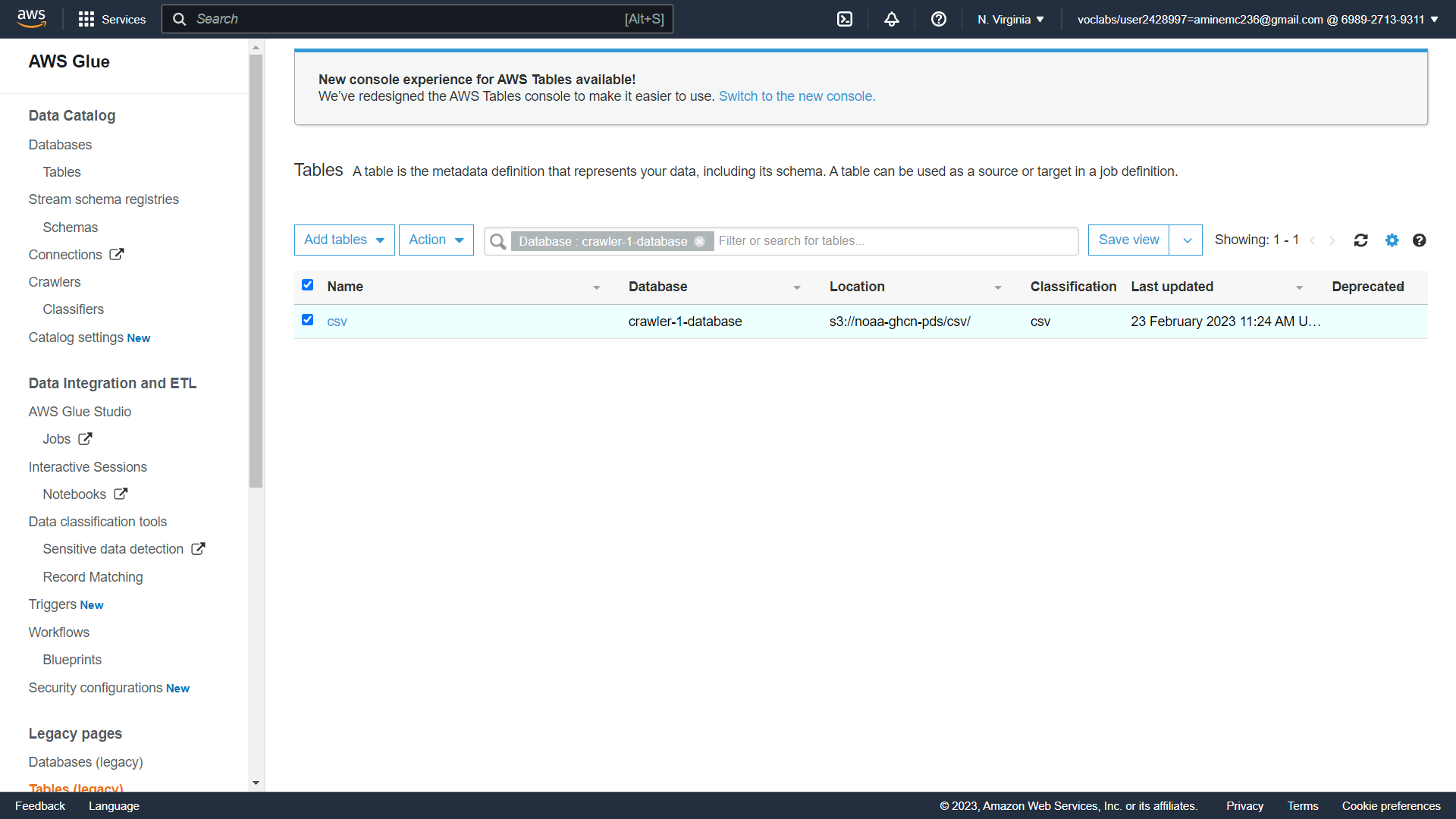


Choose csv tables

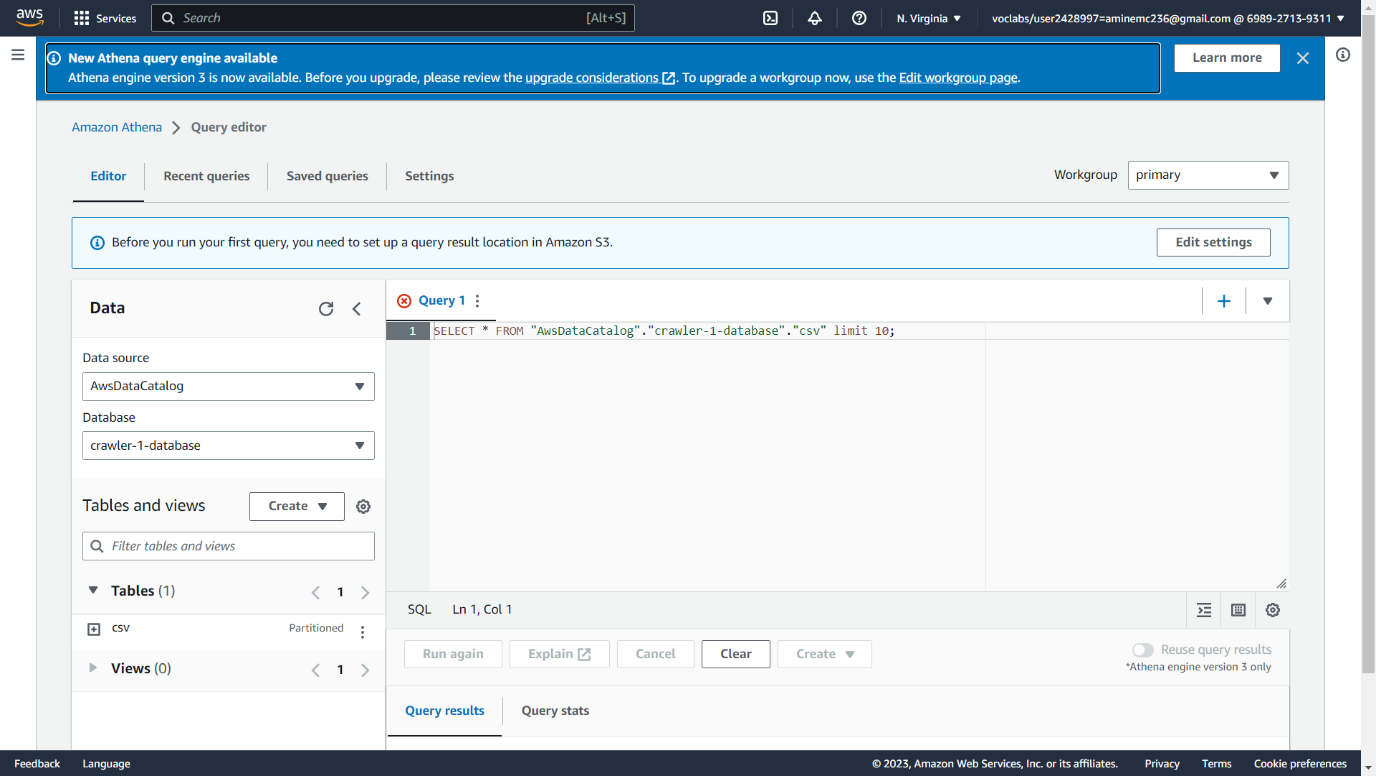
Review the metadata of the table



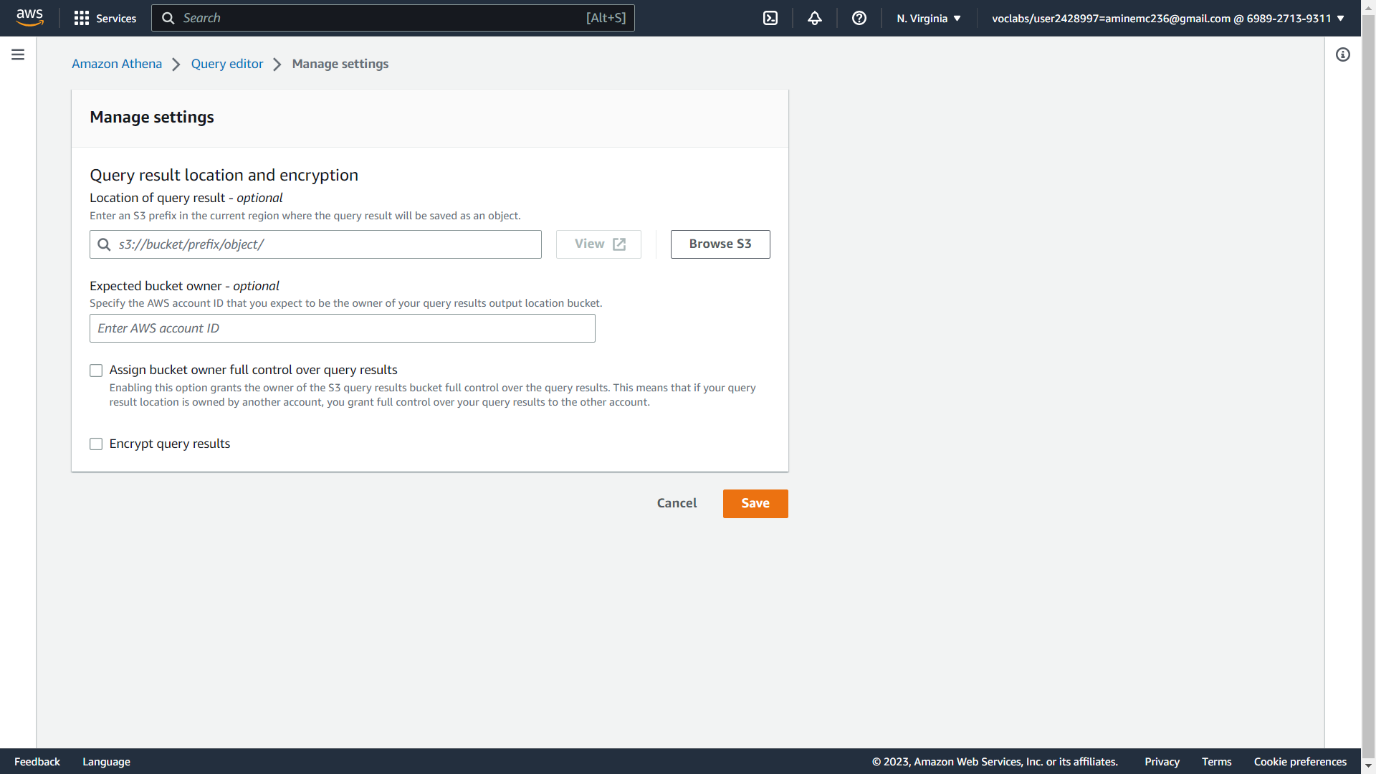
In table list, choose csv and click action to view data

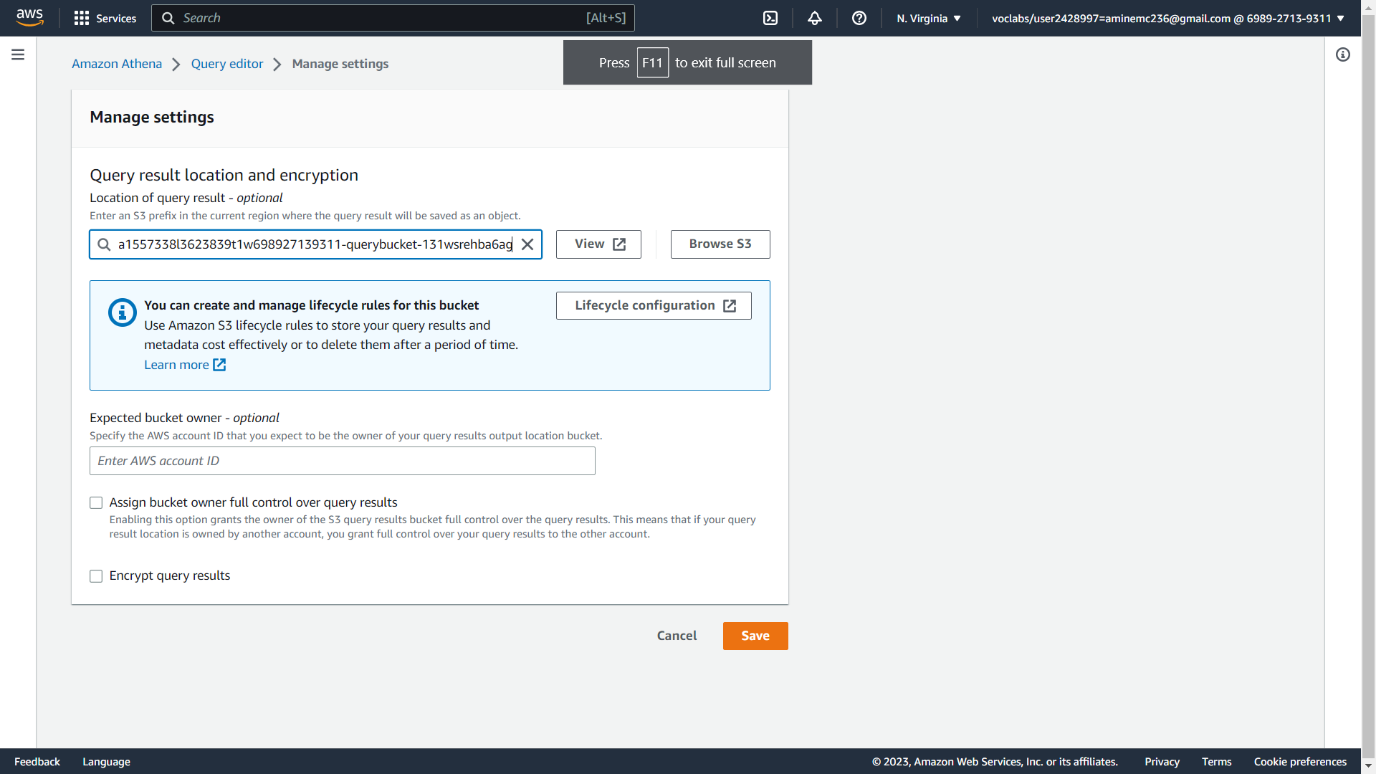


Now we are inside of the athena, before running a query, we need to set where to save the result (called bucket) just click edit settings

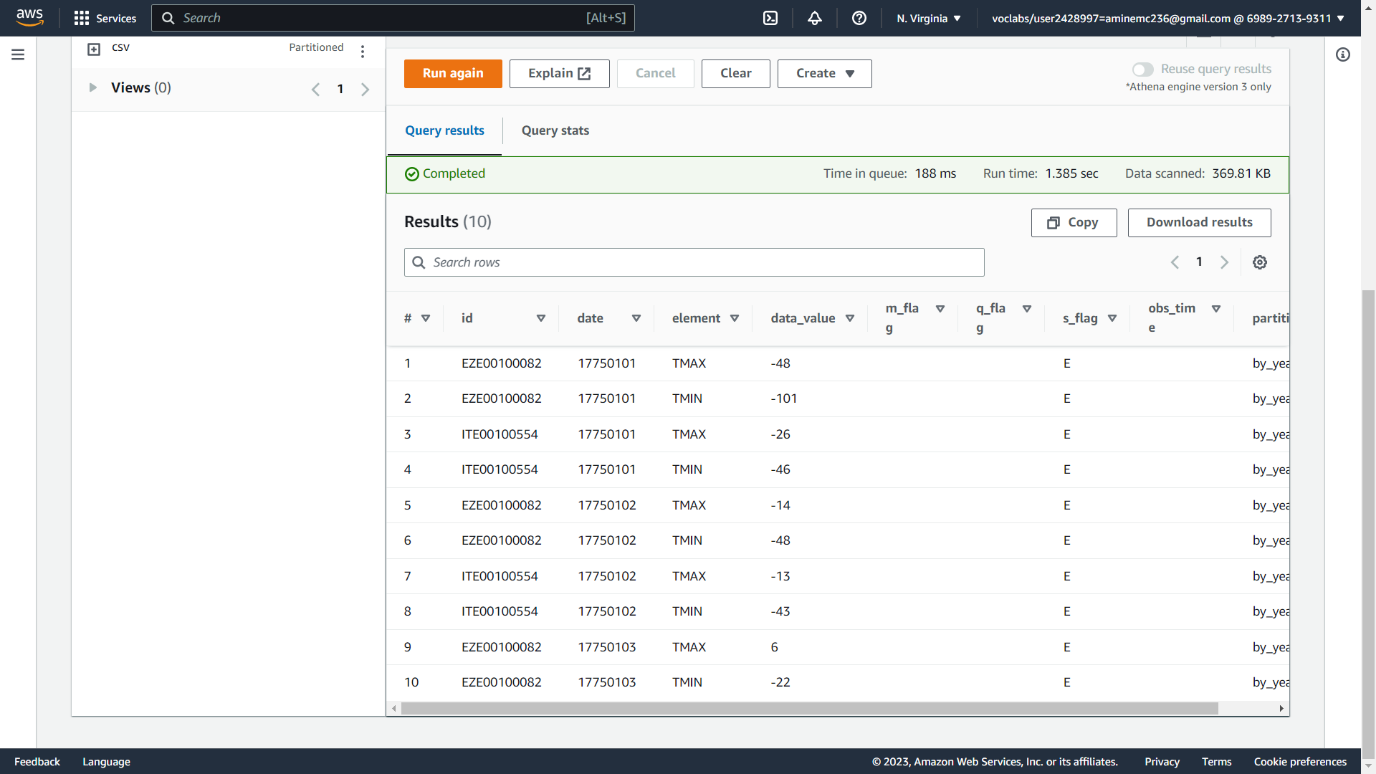


Choose a bucket





Run the query



**Task 2 (Map Reduce using Python)**

* Redo the programming example from here:
  + <https://icaml.org/canon/basics/mapreduce_wordcout_python.html>
  + <https://colab.research.google.com/drive/1blwHxoV55wHoI1bqj2VTOJkZy5lyz_eI?usp=sharing>
* Upload/commit your code on your GitHub.
* Paste the GitHub link here: …

<https://github.com/Aminemcc/Big-Data/tree/master/Week%202>

Collect this template as PDF file.