PRADYOTH SINGENAHALLI PRABHU

02071847

[psingenahalliprabhu@umassd.edu](mailto:psingenahalliprabhu@umassd.edu)

CIS 602: Big Data – Homework 2

**Task 1: Creating and querying an AWS Glue database and table in Athena**

## 1.

A screenshot of a computer

Description automatically generated

2.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

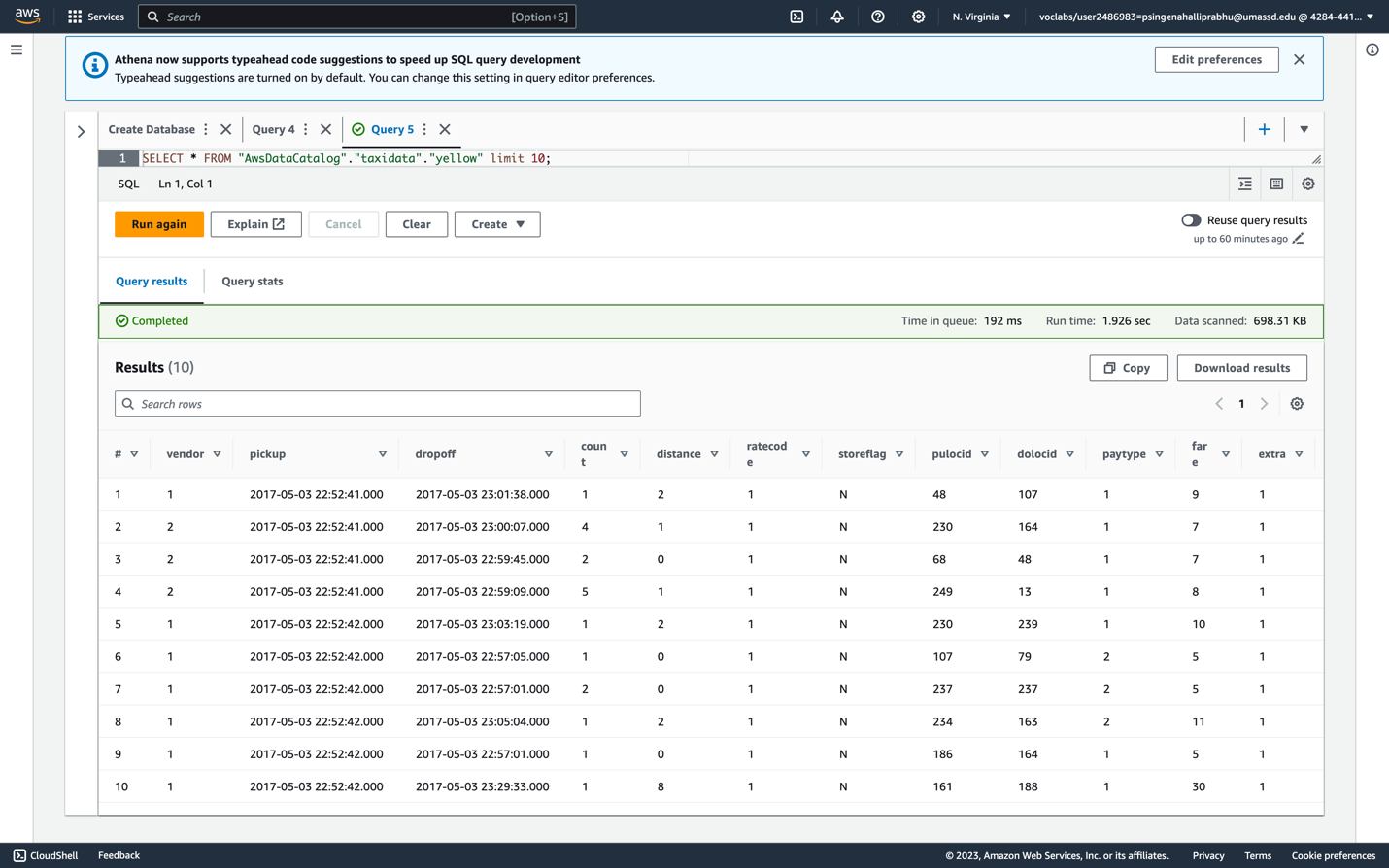
3. Table Created

A screenshot of a computer

Description automatically generated

4.Preview Data

Query runned: SELECT \* FROM "AwsDataCatalog"."taxidata"."yellow" limit 10;



**Task 2: Uploading sample data to an S3 bucket**

**Task 3: Querying the data**

**Task 4: Modifying an object's encryption properties and storage type**

**Task 5: Compressing and querying the dataset**

**Conclusion Remarks:**

In conclusion, this report provides a summary of our project's key achievements, highlighting the important lessons learned and tasks completed. Firstly, we became proficient in accessing AWS Cloud9, a valuable skill for modern technology environments. We automated S3 bucket creation using YAML scripts and AWS CLI commands, enhancing the efficiency of our infrastructure setup. We ensured stack reliability by verifying its status through AWS CLI.

We also learnt how to delete the stack using AWS CLI. With the ability to remove stacks through the command-line interface, we not only ensured the elimination of unnecessary resources but also demonstrated our proficiency in managing cloud-based systems effectively. We acquired essential data acquisition skills by downloading files using wget. Subsequently, we smoothly uploaded these files to S3 buckets, ensuring data accessibility and integrity.  
  
Our proficiency with AWS CLI allowed us to list files in S3 buckets efficiently, simplifying data management. Complex data queries were executed using S3 Select, including tasks like retrieving specific data fields and limiting results.

Data security was addressed through encryption using 'Intelligent-Tiering'. This proactive measure played a pivotal role in safeguarding sensitive information and ensuring the confidentiality and integrity of our data assets. By employing this encryption method, we not only protected our data from unauthorized access but also optimized our data storage costs. The intelligent tiering system dynamically moves data between different storage classes, such as Standard and Glacier, depending on its access patterns. This means that frequently accessed data is kept in more expensive but faster storage, while less frequently accessed data is moved to more cost-effective storage options.

Furthermore, we optimized data storage and transfer by compressing files with 'gzip' and successfully uploaded compressed gzip files to S3 buckets. Lastly, we demonstrated versatility by running S3 queries on compressed data.