LAMBTON COLLEGE



**A Report on**

[**Lab 1,2,3 on** [**AWS Academy Data Analytics**](https://moodle.queenscollege.ca/moodle/mod/lesson/view.php?id=436567)**]**

121 Brunel Rd, Mississauga

ON L4Z 3E9

A Group assignment with screenshots of Lab 1, 2, and 3

on Aws academy

Big Data Analytics DSMM

**Under the supervision**

**of**

**Professor Teresa Zhu**

**Submitted BY:**

Aadarsha Chapagain (C0825975)

Roshan Acharya (C0831342)

Anjana Kuriakose (C0829580)

Onyinye Mbanefo (C0831578)

**Submitted To:**

Lambton College

Professor Teresa Zhu

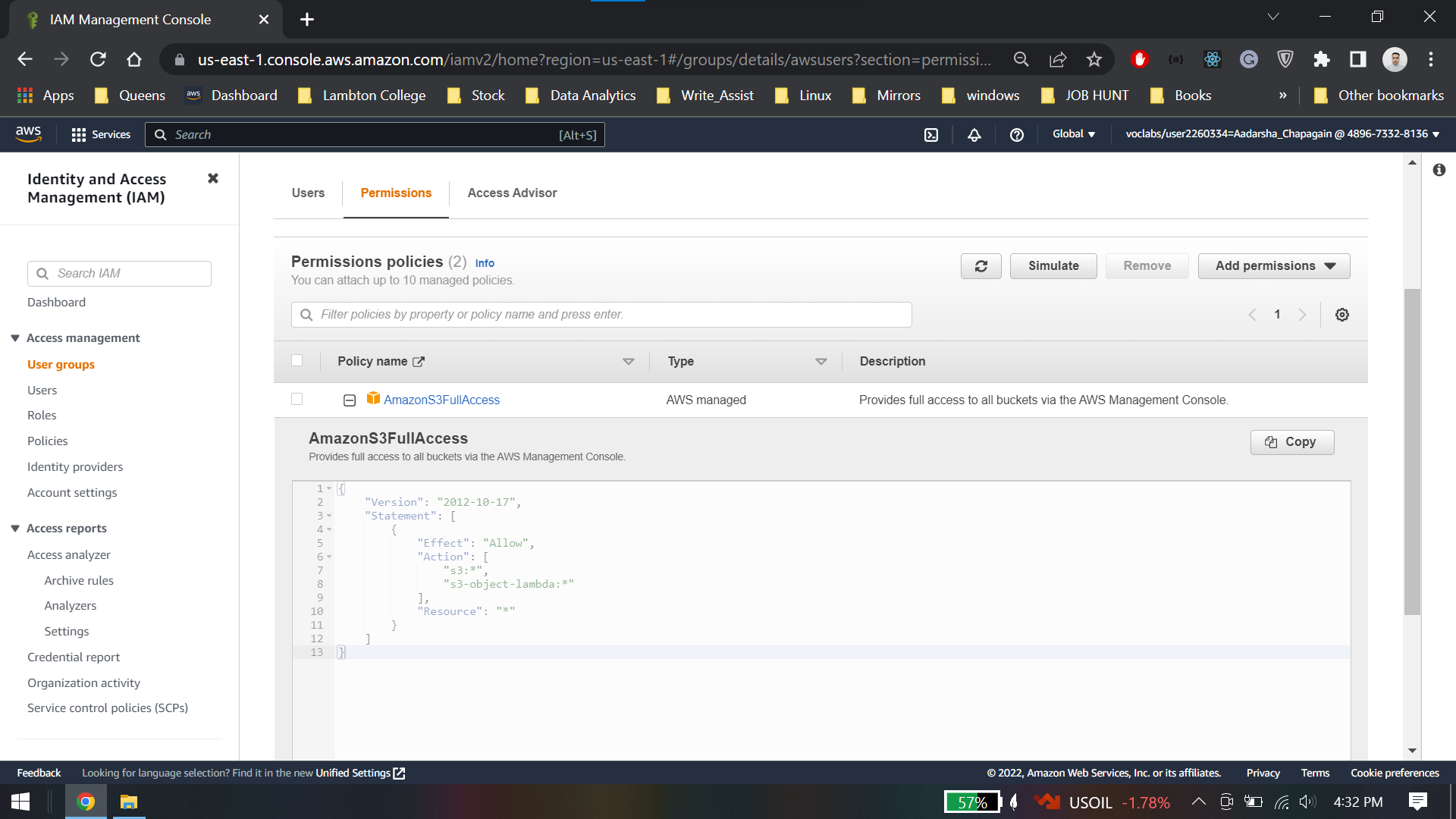
**Submission Date:**

27th  November 2022

# Lab1: Store data in Amazon S3

## Task 1: Create an IAM user account

### Task 1.1: Review users and group permissions in the IAM console



Task 1.2: Add awsuser to the awsusers group

A screenshot of a computer

Description automatically generated

AccountId: 489673328136

## Task 2: Load data into Amazon S3

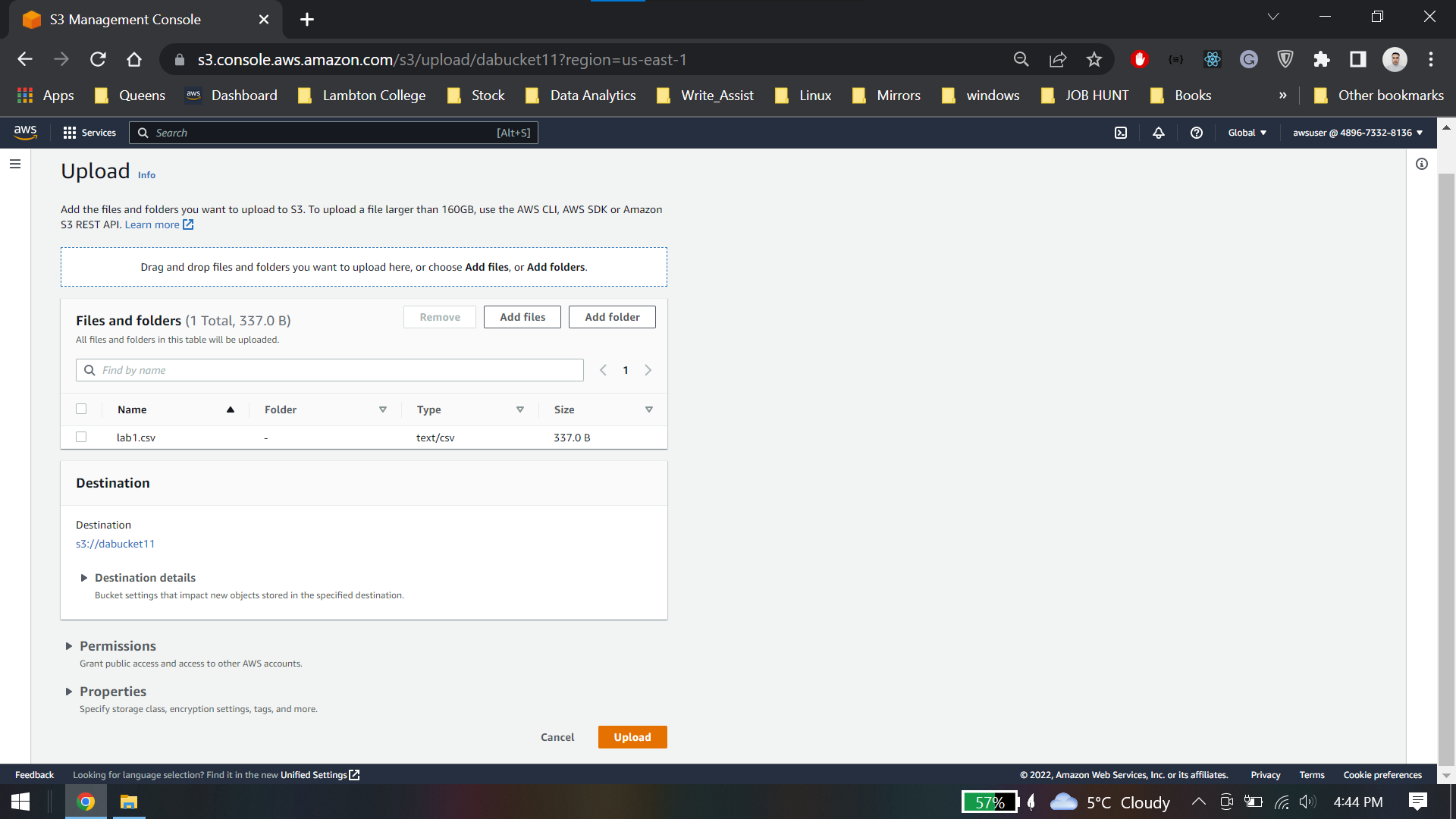
### Task 2.1: Create an S3 bucket

Bucket name: dabucket11

A screenshot of a computer

Description automatically generated

### Task 2.2: Upload an object



### Task 2.3: Query the object you uploaded

A screenshot of a computer

Description automatically generated

### Task 2.4: Change the encryption properties and storage type

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

### Task 2.5: Upload a compressed file



Compressed file can be queried in the same way as a non-compressed file.

A screenshot of a computer

Description automatically generated

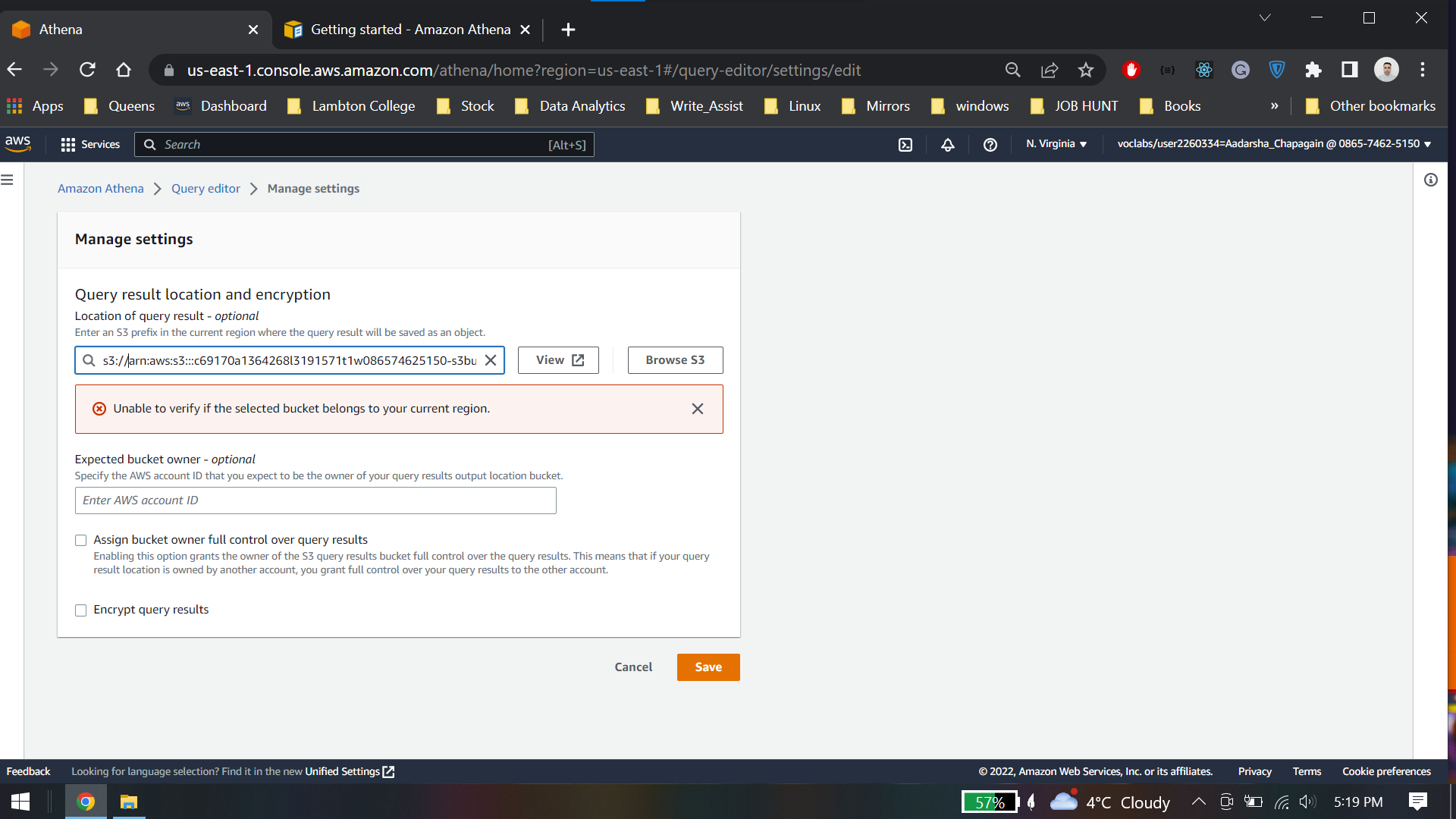
## Lab 1 Conclusion.

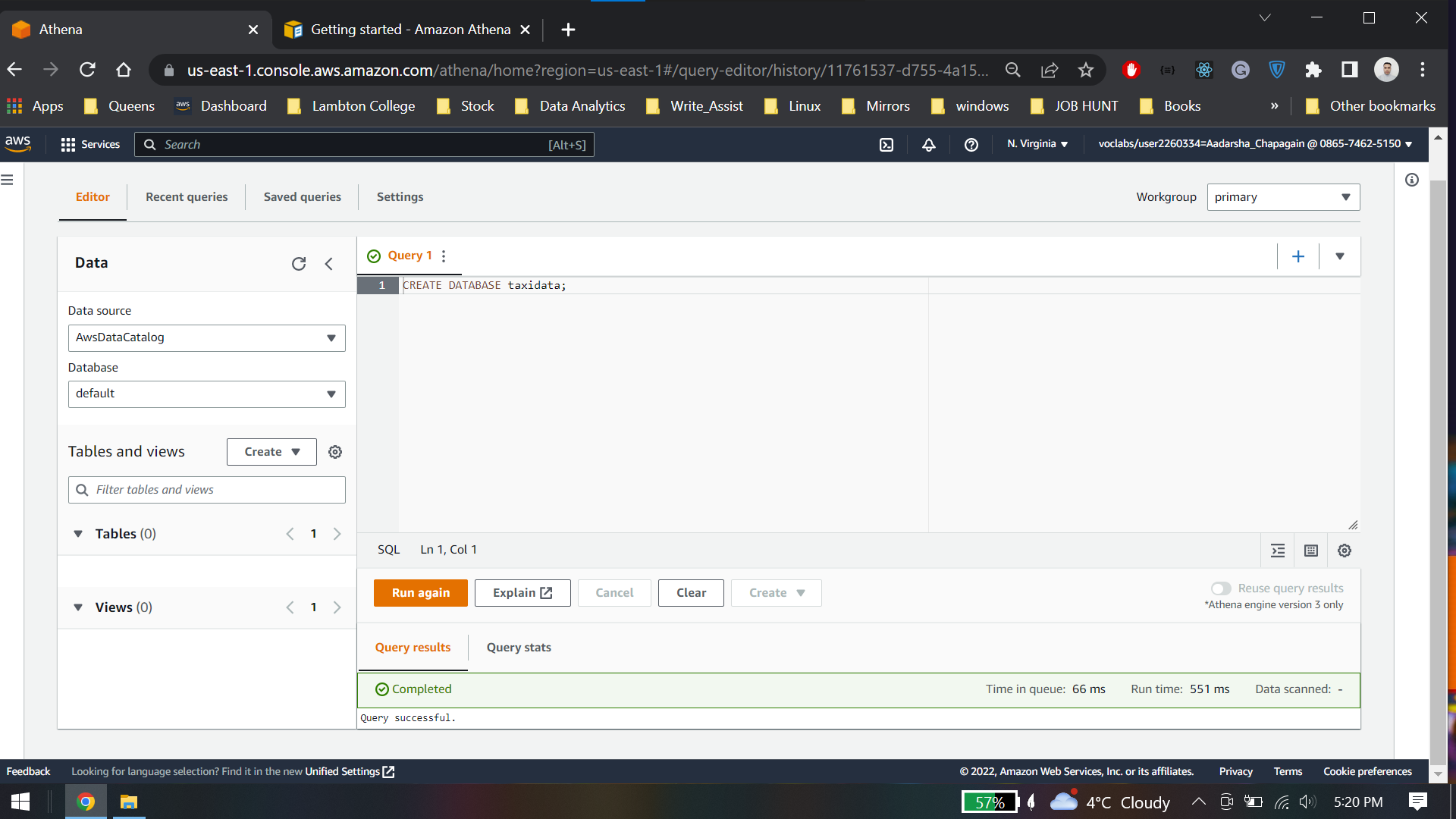
* Access Amazon S3 in the AWS Management Console
* Secure an S3 bucket with IAM
* Create a bucket with Amazon S3
* Load data into an S3 bucket
* Query an S3 bucket

# Lab 2: Query Data in Amazon Athena

## Task 1 : Query Data in Amazon Athena

Bucket ARN: arn:aws:s3:::c69170a1364268l3191571t1w086574625150-s3bucket-q3gdz12cykyj





A screenshot of a computer

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated

## Task 2: Optimize the database

### Task 2.1: Create a table for the January 2017 data

A screenshot of a computer

Description automatically generated

### Task 2.2: Run a query using the data that is not divided into buckets

A screenshot of a computer

Description automatically generated

### Task 2.3: Run a query using the data that is divided into buckets for each month

A screenshot of a computer

Description automatically generated

Following results was found

* No buckets:
  + Total data scanned: 9.32 GB
* Buckets:
  + Total data scanned: 815 MB

Task 2.4: Query partitioned data

### Task 2.4.1: Partition the data

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Yellow table:

Run time: 7.19 seconds & Data scanned: 9.32 GB

Credit card table:

Run time: 3.32 seconds & Data scanned: 71.8 MB

## Task 3: Create and query views

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A picture containing chart

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Lab2 Conclusion

* Accessed Athena in the AWS Management Console
* Created tables and define data types
* Queried data in Amazon S3 from Athena
* Optimized queries with partitioning

# Lab 3: Query Data in Amazon S3 with Amazon Athena and AWS Glue

## Task 1: Create a crawler for the GHCN-D dataset

Create a Crawler

A screenshot of a computer

Description automatically generated

Choose Iam role

A screenshot of a computer

Description automatically generated

Create a database

A screenshot of a computer

Description automatically generated

Output Configuration

A screenshot of a computer

Description automatically generated

### Task 1.1: Run the crawler

Graphical user interface, text, application

Description automatically generated

### Task 1.2: Review the metadata created by AWS Glue

A screenshot of a computer

Description automatically generated

### Task 1.3: Edit the schema

A screenshot of a computer

Description automatically generated

## Task 2: Query the table using the AWS Glue Data Catalog

A screenshot of a computer

Description automatically generated

### Task 2.1: Create a table for data after 1950

Create a bucket in same region

A screenshot of a computer

Description automatically generated

Create a table specifying the bucket location

Graphical user interface, text, application, email

Description automatically generated

Preview table

A screenshot of a computer

Description automatically generated

### Task 2.2: Run a query from the selected data

Create a view

Graphical user interface, text, application, Word

Description automatically generated

Preview the data

Graphical user interface, application, Teams

Description automatically generated

Average maximum temperature from 1950 to 2018

Graphical user interface, text, application, email

Description automatically generated

Results

A screenshot of a computer

Description automatically generated

## Lab 3 Conclusion

* Accessed AWS Glue in the AWS Management Console
* Created a crawler with AWS Glue
* Created tables and a schema with AWS Glue
* Queried data in Amazon Simple Storage Service (Amazon S3) from Amazon Athena with the AWS Glue data catalog