LAMBTON COLLEGE



**A Report on**

[**Lab 4,5,6 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 4, 5, and 6

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

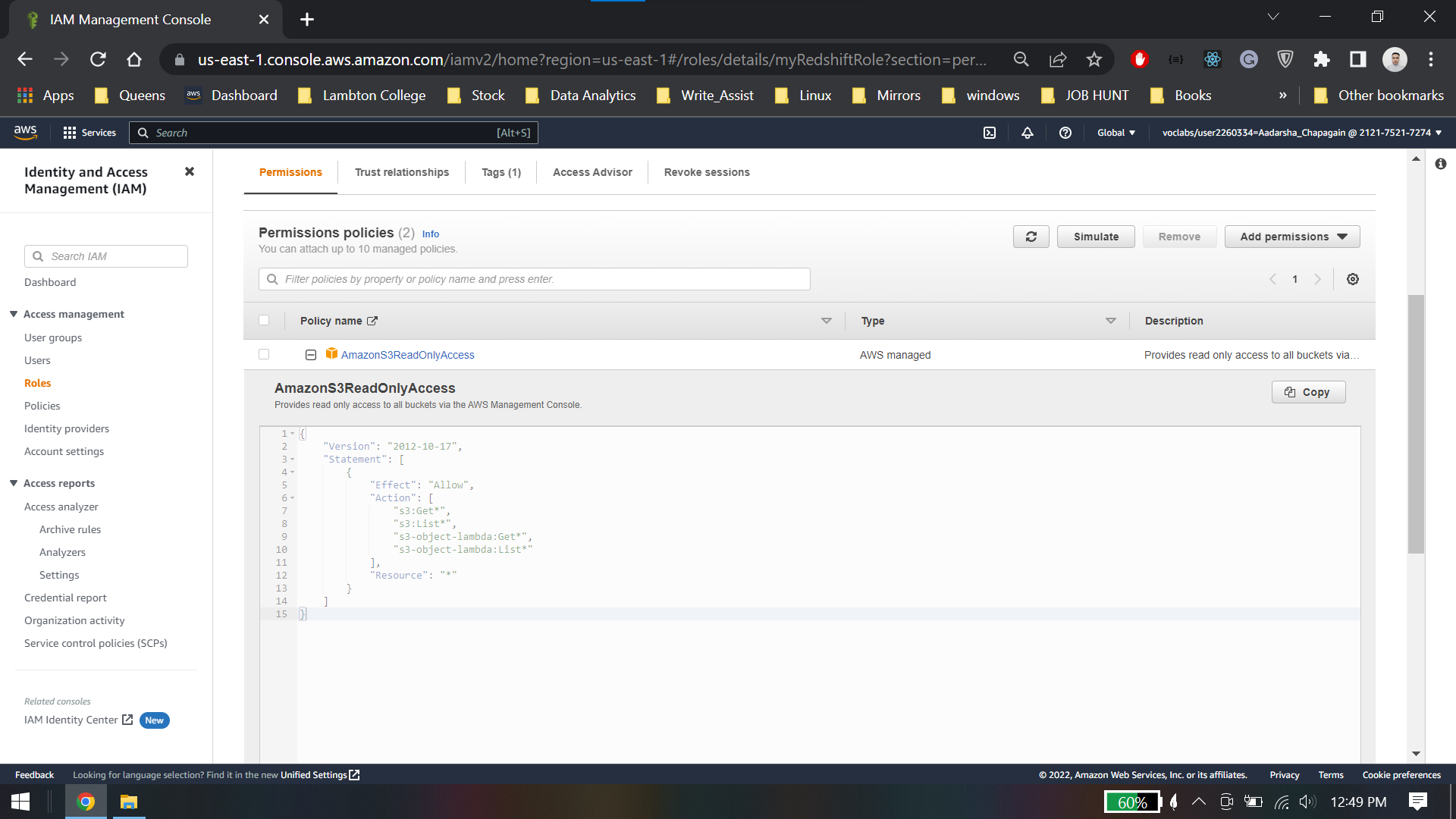
**Submission Date:**

27th  November 2022

# Lab4: Analyze Data with Amazon Redshift

## Task 1: Task 1: Review the security group for accessing the Amazon Redshift console

Myredshiftrole



## Task 2: Create and configure an Amazon Redshift cluster

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### Task 2.1: Create a security group for your cluster

A screenshot of a computer

Description automatically generated

### Task 2.2: Configure your Amazon Redshift cluster

Graphical user interface, text, application, email

Description automatically generated

## Task 3: Load data into your Amazon Redshift cluster

### Task 3.1: Create the tables in the dev database

User table:

A screenshot of a computer

Description automatically generated

Date table

Graphical user interface, text, application

Description automatically generated

Sales table

Graphical user interface, text, application, Teams

Description automatically generated

### Task 3.2: Load data from Amazon S3

arn:aws:iam::212175217274:role/myRedshiftRole

Copy users table

Graphical user interface, application

Description automatically generated

Copy date table

Graphical user interface, text, application, email

Description automatically generated

Copy sales table

Graphical user interface, application

Description automatically generated

## Task 4: Query the data

Preview sales table

A screenshot of a computer

Description automatically generated

Number of item sold on particular date

Graphical user interface, application, Teams

Description automatically generated

Top 10 buyers by quantity

Graphical user interface, application, Teams

Description automatically generated

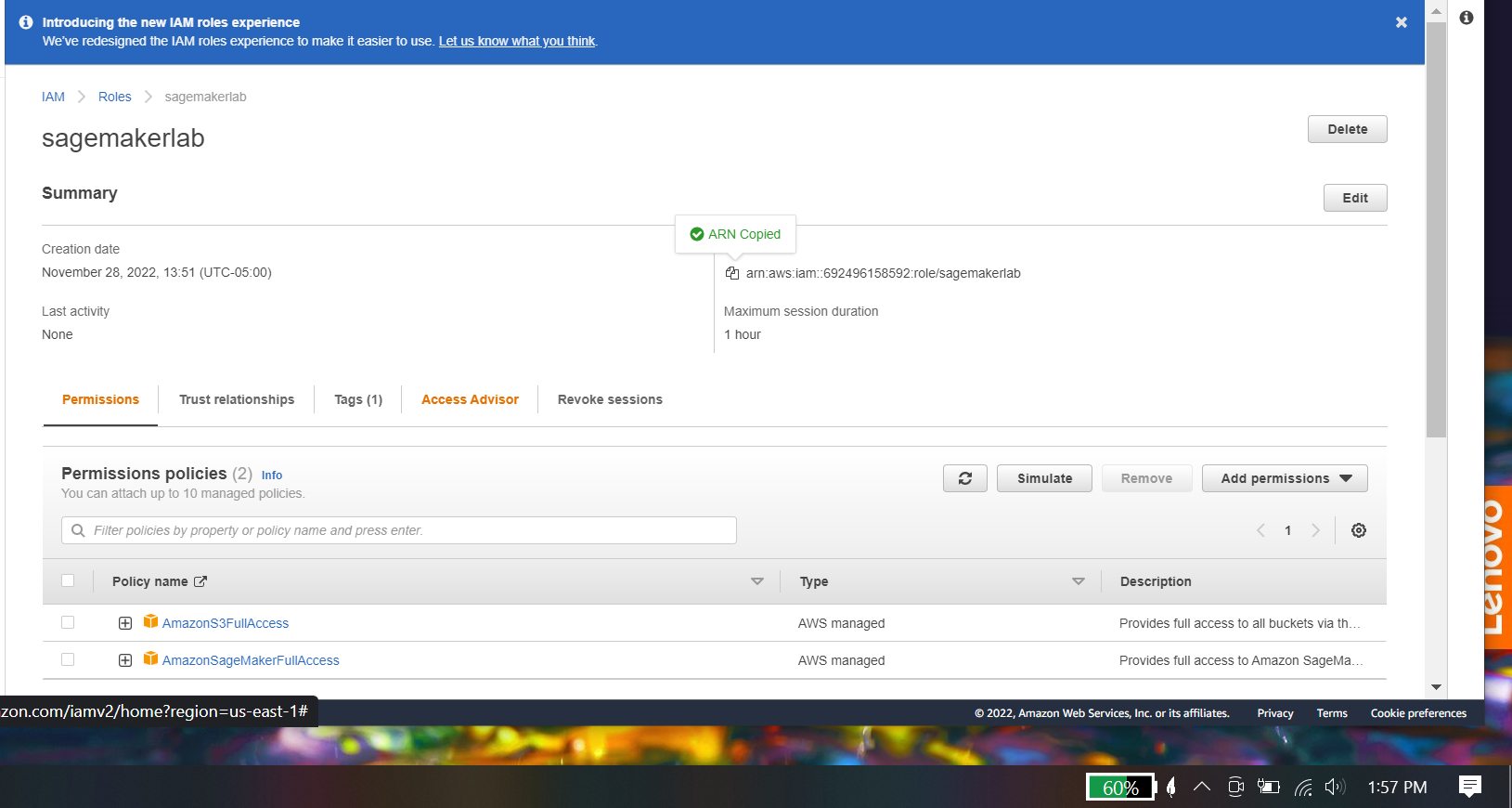
## Lab 4 Conclusion.

* Accessed Amazon Redshift in the AWS Management Console
* Created an Amazon Redshift cluster.
* Load data from Amazon Simple Storage Service (Amazon S3) into an Amazon Redshift table
* Queried data in Amazon Redshift

# Lab5: Analyze Data with Amazon Sagemaker, Jupyter Notebooks and Bokeh

## Task 1: Obtain the AWS Identity and Access Management (IAM) role

Sagemakerrolearn: arn:aws:iam::692496158592:role/sagemakerlab



## Task 2: Create a Jupyter notebook with Amazon SageMaker

A screenshot of a computer

Description automatically generated

## Task 3: Open your Jupyter notebook instance

## Task 4: Create visualizations with Bokeh

### Task 4.1: Create a line graph

Graphical user interface, text, application

Description automatically generated



### Task 4.2: Create a bar chart

Graphical user interface, application

Description automatically generated

### Task 4.3: Create a grouped bar chart

Graphical user interface, text, application

Description automatically generated

## Task 5: Create a visualization from a dataset

### Task 5.1: Download the data file

A screenshot of a computer

Description automatically generated with medium confidence

### Task 5.2: Create the notebook

Graphical user interface, text, application, email

Description automatically generated

Chart, scatter chart

Description automatically generated

## Lab 5 Conclusion

* Described Jupyter notebooks and the Bokeh visualization package.
* Created a Jupyter notebook with Amazon SageMaker.
* Imported data into a Jupyter notebook.
* Created a presentation with a Jupyter notebook.
* Visualized data with the open-source Bokeh Python package.

# Lab 6: Automate Loading Data with AWS Data Pipeline

## Task 1: Review the security group for accessing the Amazon Redshift console

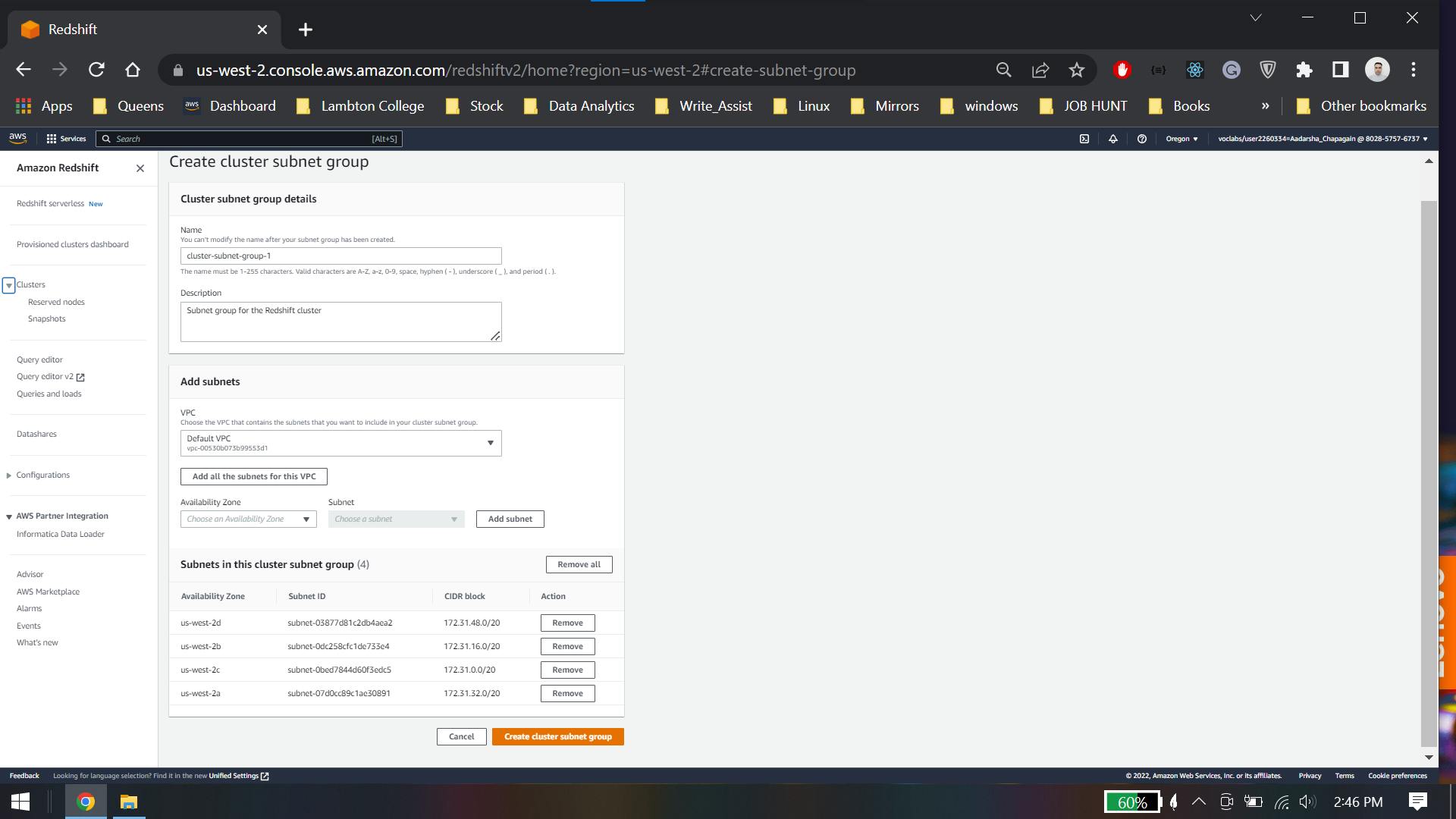
Arn: arn:aws:iam::802857576737:role/myRedshiftRole

Graphical user interface, text, website

Description automatically generated

## Task 2: Create and configure an Amazon Redshift cluster

Create subnet cluster group



Create cluster

Graphical user interface, text, application

Description automatically generated

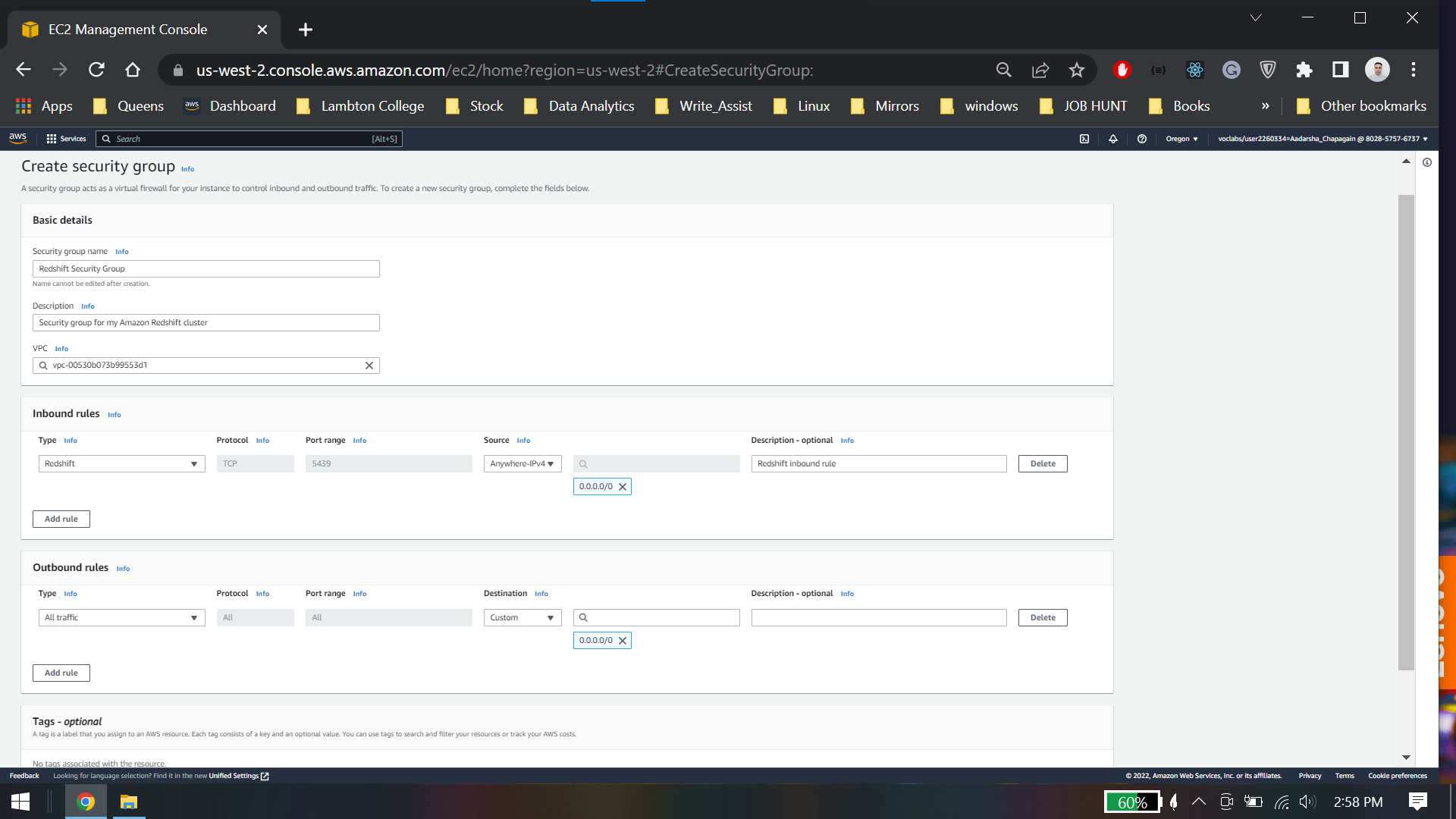
Graphical user interface, text, application, email

Description automatically generated

A screenshot of a computer

Description automatically generated

### Task 2.1: Create a security group for your cluster



### Task 2.2: Configure your Amazon Redshift cluster

A screenshot of a computer

Description automatically generated

### Task 2.3: Capture JDBC connection information

Graphical user interface, application

Description automatically generated

### Task 2.4: Create an S3 bucket to store log files

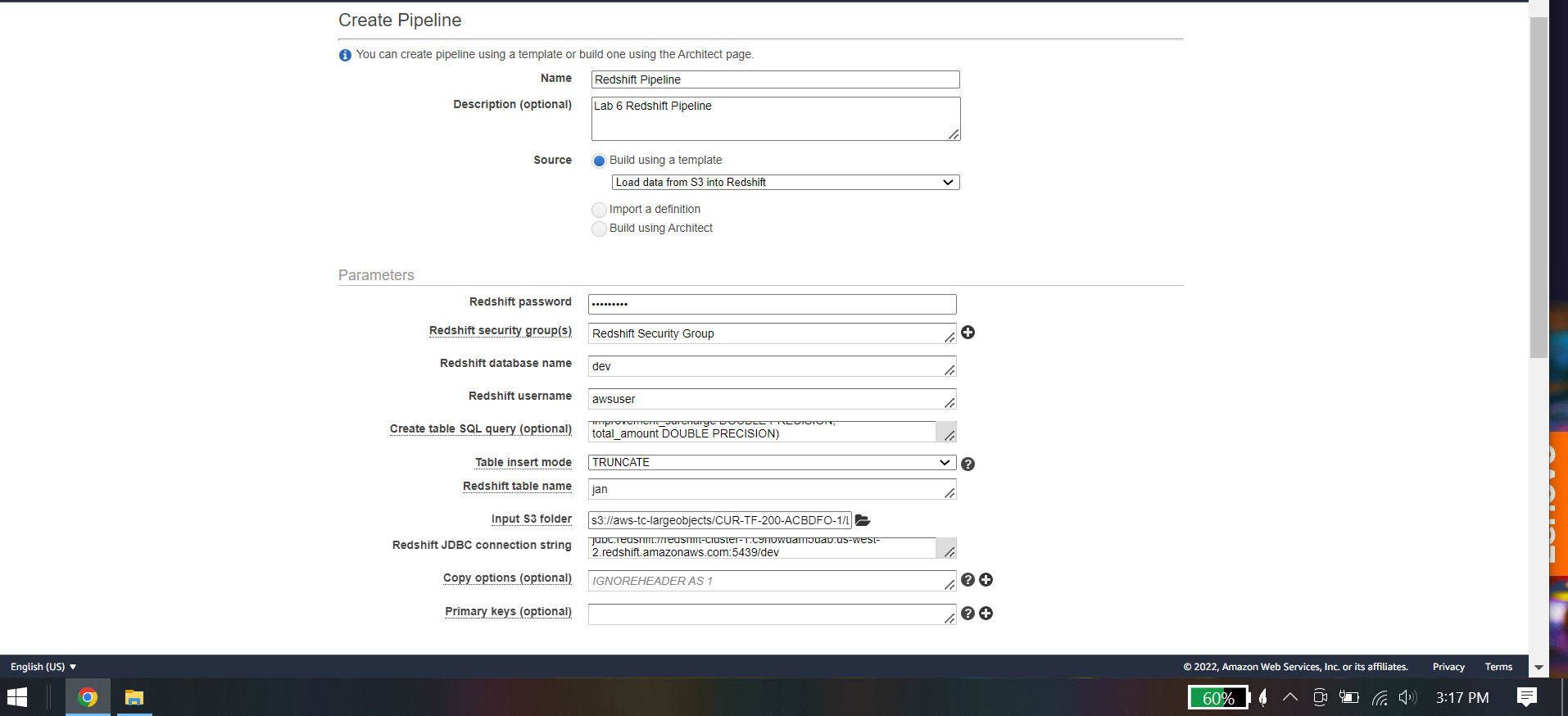
A screenshot of a computer

Description automatically generated

Bucketname:logbucketgroupa

Jdbc Connection: jdbc:redshift://redshift-cluster-1.c9howuam5uab.us-west-2.redshift.amazonaws.com:5439/dev

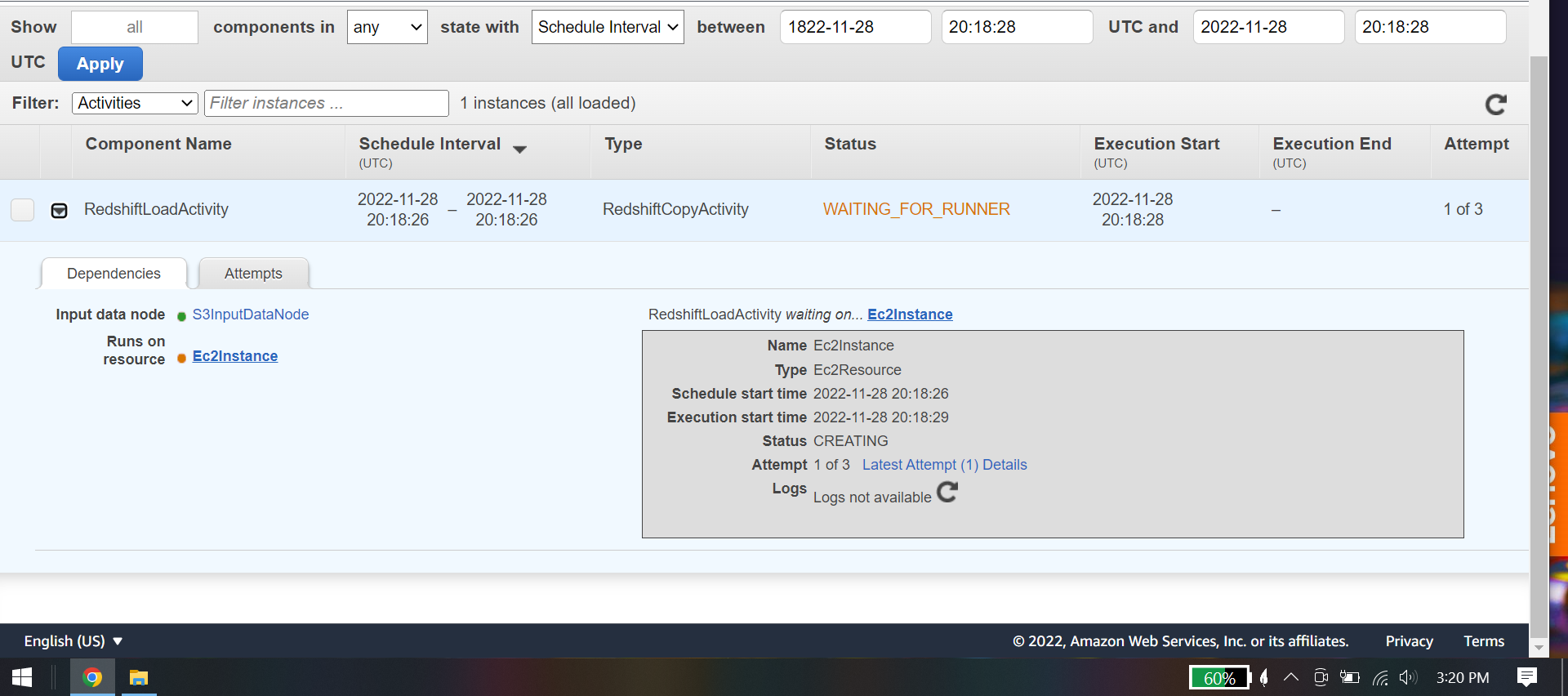
## Task 3: Create the pipeline



Graphical user interface, text, application

Description automatically generated

## Task 4: Monitor your pipeline



## Task 5: View log files

A screenshot of a computer

Description automatically generated

## Task 6: Query the Amazon Redshift database

Connect to redshift using username and password

A screenshot of a computer

Description automatically generated

## Lab 6 Conclusion

* Accessed AWS Data Pipeline in the AWS Management Console.
* Created a data pipeline.
* Load data from Amazon S3 into Amazon Redshift with a data pipeline.
* Troubleshoot a data pipeline.
* Export data from Amazon Redshift to a Jupyter notebook