**IDS 410 – Business Database Technology**

**Fall 2022**

**AWS Lab 6**

**Submitted by:**

**Shubham Dwivedi** [**|sdwive4@uic.edu**](mailto:|sdwive4@uic.edu) **| 660328503**

**Review myRedshiftRole**

Graphical user interface, text, application, email

Description automatically generated

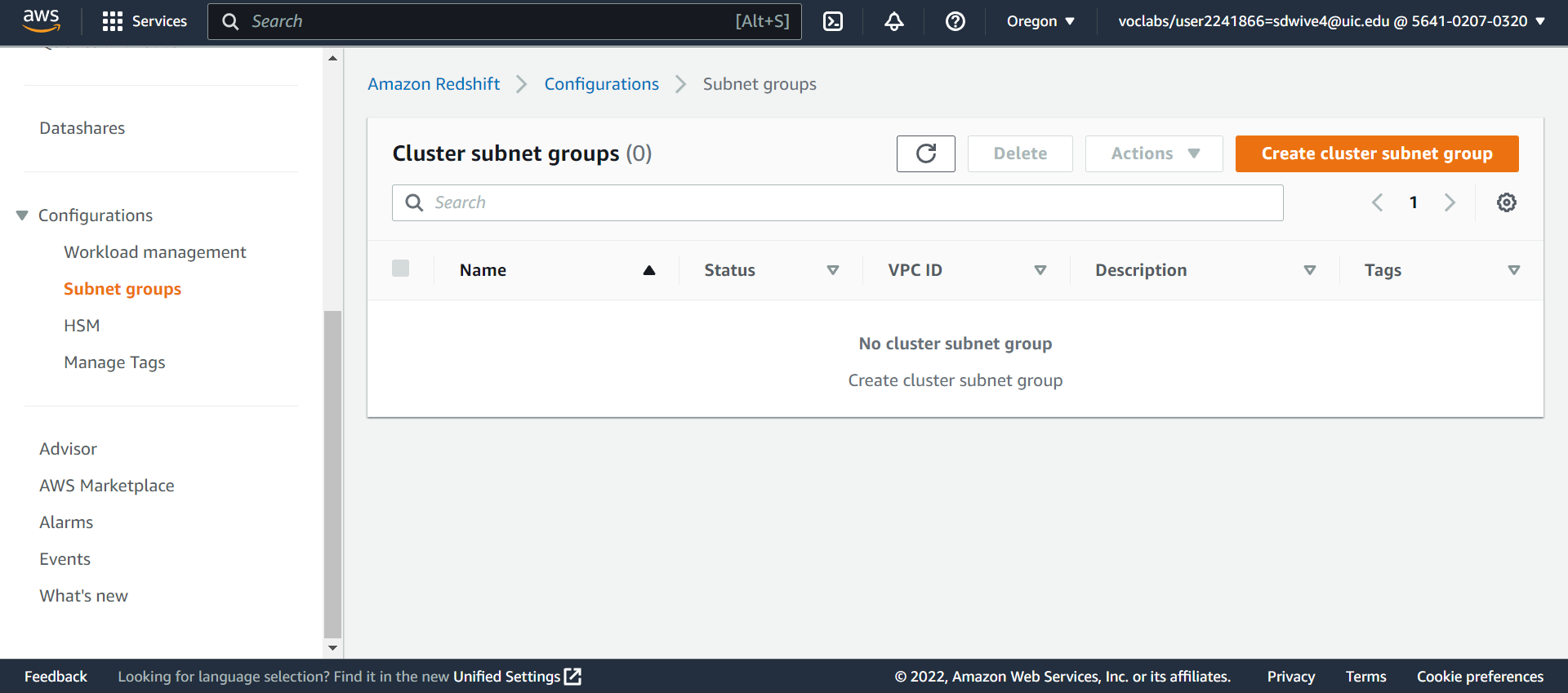
Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, email

Description automatically generated

**create a cluster for your data warehouse**



Graphical user interface, text, application, email

Description automatically generated

A screenshot of a computer

Description automatically generated

**Create a cluster**

Graphical user interface, text, application, email

Description automatically generated

**In the Cluster configuration pane, enter the following values:**

Graphical user interface, text, application

Description automatically generated

In the **Database configurations** pane, for **Admin user password**, enter Passw0rd1

Graphical user interface, text, application, email

Description automatically generated

Choose **Associate IAM role**

Graphical user interface, application

Description automatically generated

In the **Additional configurations** section, turn off **Use defaults**.

Graphical user interface, text, application, email

Description automatically generated

For the **Publicly accessible** setting, choose **Enable**.

Graphical user interface, text, application, email

Description automatically generated

**Cluster creation I process and created successfully**

Graphical user interface, text, application, website

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**Creating security groups in EC2**

A screenshot of a computer

Description automatically generated

Graphical user interface, text, application, Teams

Description automatically generated

**Adding Properties to Inbound Rules**

Graphical user interface, application

Description automatically generated

**Security Group successfully created**

Graphical user interface, text, application

Description automatically generated

**Go to Redshift and choose the cluster that you have created. In the properties tab edit the Network and security settings**

Graphical user interface, text, application, email

Description automatically generated

Under **VPC security groups**, choose the **Redshift Security Group** that you created.

Graphical user interface, text, application, email

Description automatically generated

**Choose the cluster that you created and from the general information copy the JDBC URL.**

Graphical user interface, text, application

Description automatically generated

**Create a S3 bucket**

Graphical user interface, website

Description automatically generated

**Bucket Created**

Graphical user interface, text, application, email, website

Description automatically generated

**To create a data pipeline. Select Data Pipeline from the All services Menu**

Graphical user interface, text, application

Description automatically generated

**Add the below details for creating the data pipeline.**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

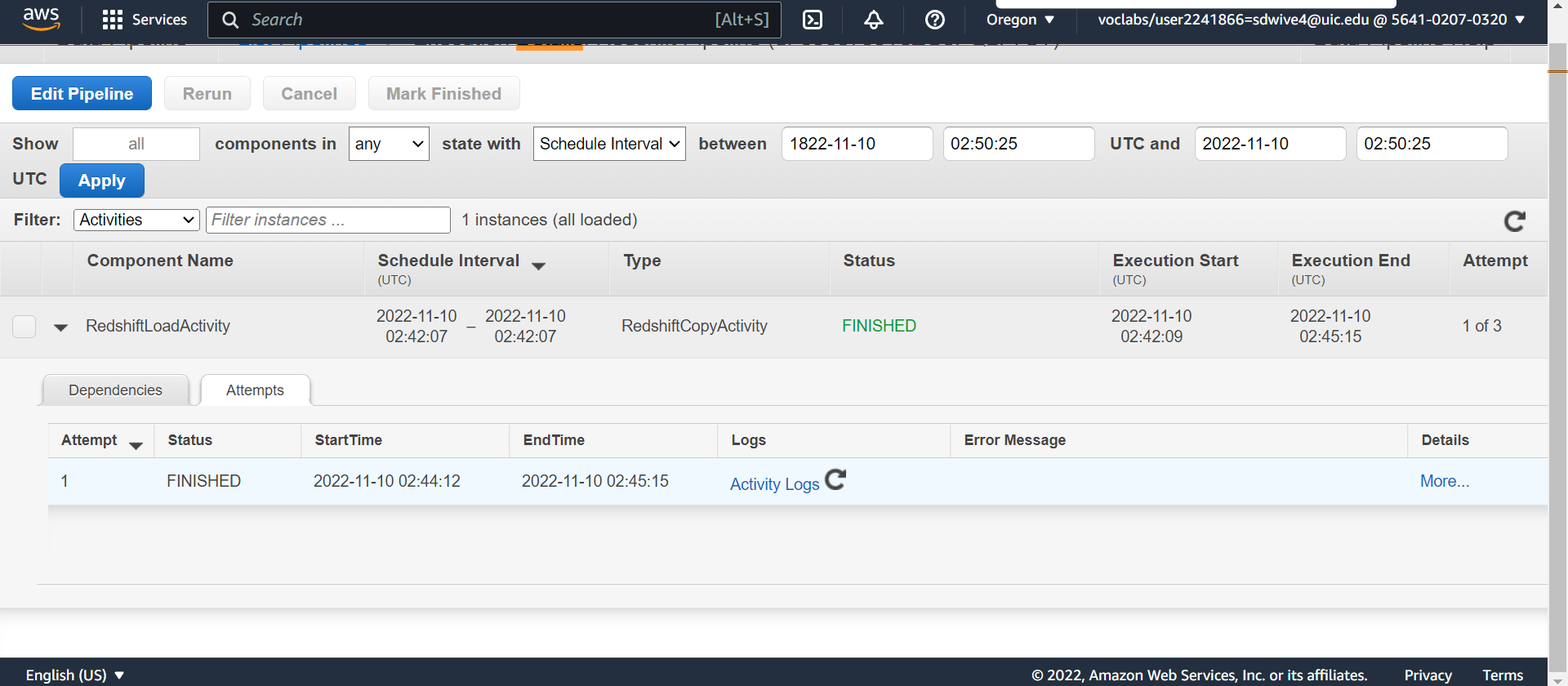
Description automatically generated

**To monitor pipelines, select the data pipeline created from the List Pipelines**

Graphical user interface, text, application, email

Description automatically generated

**To Check Logs, select Activity Logs**



**Log Files**

Text

Description automatically generated

**Query the Amazon Redshift database**

**On the AWS Management Console, on the Services menu, choose Amazon Redshift. In the navigation pane, choose CLUSTERS**

**From the clusters, open query editor**

Graphical user interface, text, application

Description automatically generated

**Choose Connect to database and enter the following values: Database name: dev Database user: awsuser**

Graphical user interface, text, application, email

Description automatically generated

**From the Select schema dropdown list on the left, choose the public schema. In the Resources pane, choose the jan table. Choose the three dots next to the jan table and choose Preview data.**

Graphical user interface, application

Description automatically generated

**After the query runs, you see the first 10 rows for the January data. You might need to scroll down to see all the data.**

Graphical user interface, application, table

Description automatically generated

**Challenge Question**

**Create new Pipleline for Feb**

**Create Table SQL Query**

create table feb

(vendorid bigint,

pickup\_datetime VARCHAR,

dropoff\_datetime VARCHAR,

passenger\_count bigint,

trip\_distance DOUBLE PRECISION,

ratecodeid bigint,

store\_and\_fwd\_flag VARCHAR,

pulocationid bigint,

dolocationid bigint,

payment\_type bigint,

fare\_amount DOUBLE PRECISION,

extra DOUBLE PRECISION,

mta\_tax DOUBLE PRECISION,

tip\_amount DOUBLE PRECISION,

tolls\_amount DOUBLE PRECISION,

improvement\_surcharge DOUBLE PRECISION,

total\_amount DOUBLE PRECISION)

**Below are the screenshots for the Pipeline Creation properties**

Graphical user interface, text, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

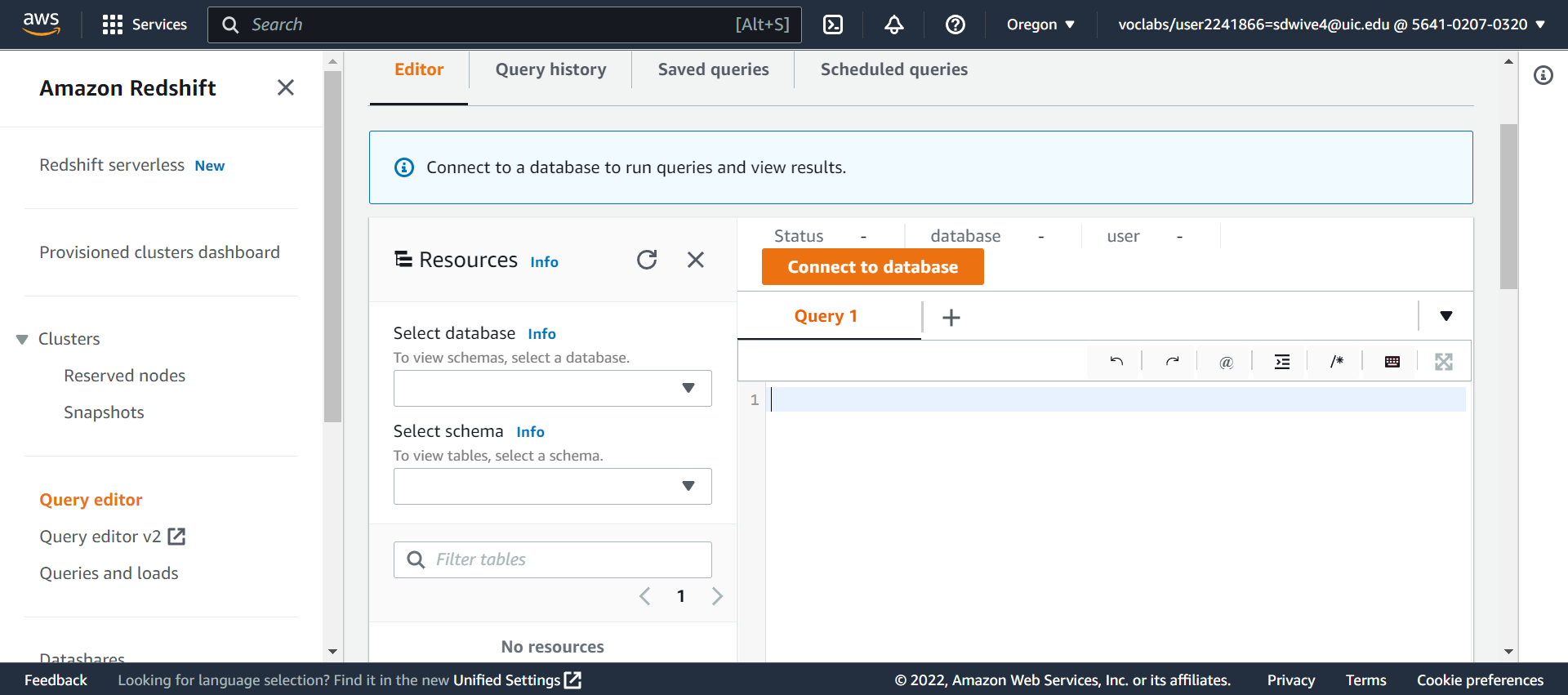
Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated

Query for most common pickup locations in Redshift



Connect to Database

Graphical user interface, text, application

Description automatically generated

**Query to find most common pickup locations in January**

SELECT pulocationid

FROM jan

GROUP BY pulocationid

ORDER BY COUNT(\*) DESC

LIMIT 1;

Graphical user interface, text, application, email

Description automatically generated

**Result: 79**

**Query to find most common pickup locations in February**

SELECT pulocationid

FROM feb

GROUP BY pulocationid

ORDER BY COUNT(\*) DESC

LIMIT 1;

Graphical user interface, text, application, email

Description automatically generated

**Result: 161**