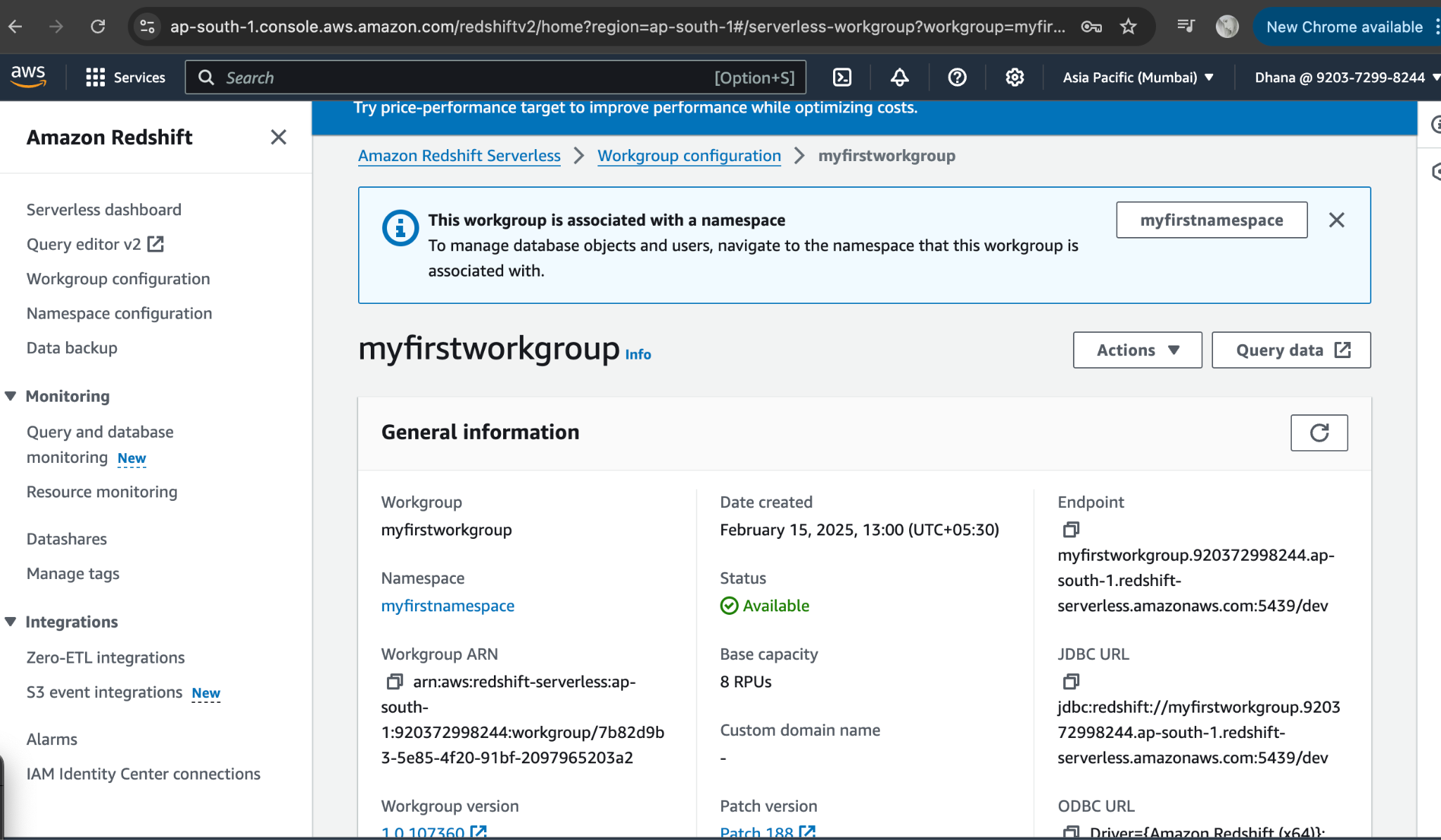
### **1. Set Up Amazon Redshift**

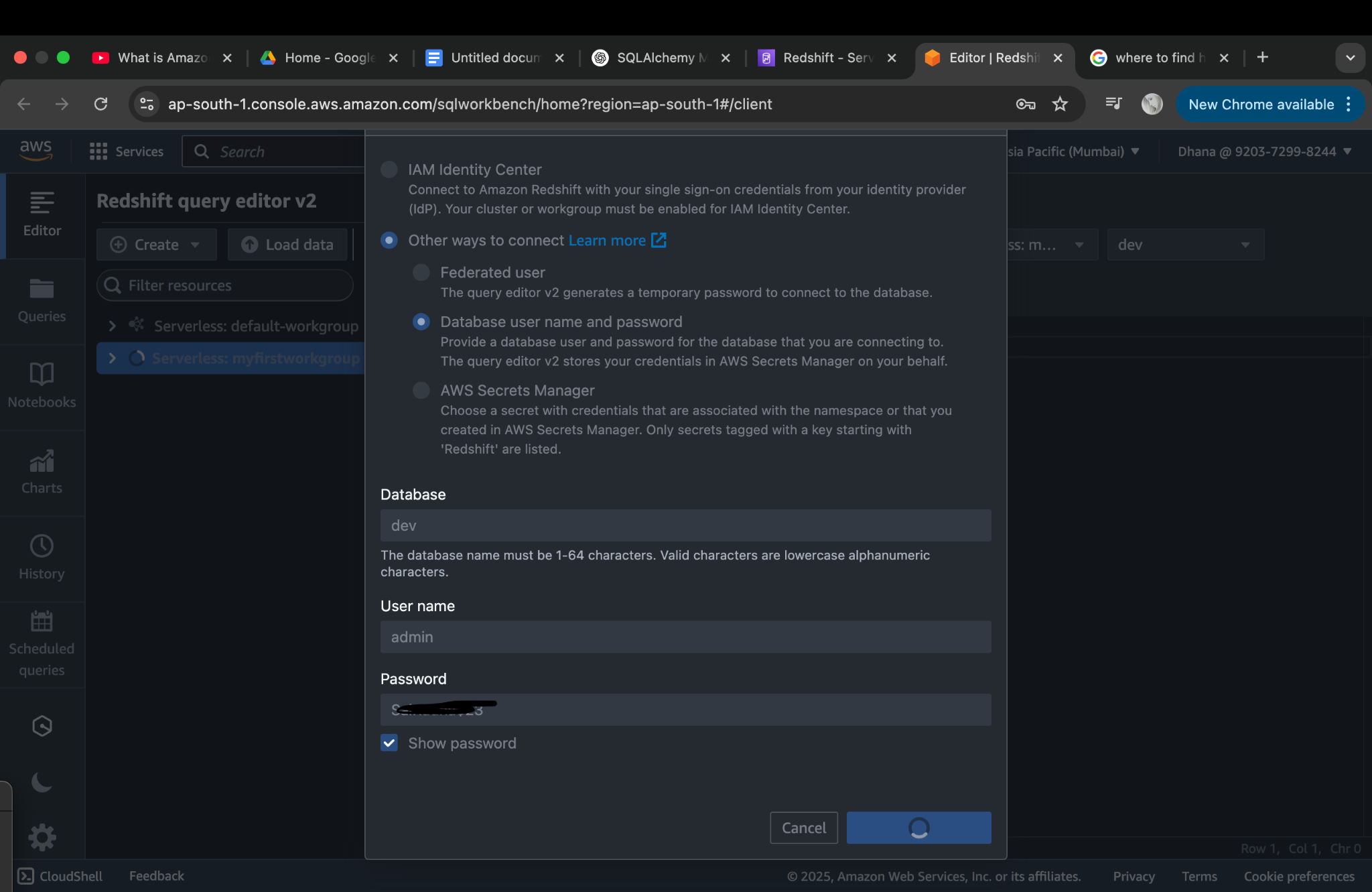
To begin with, you'll need an **Amazon Redshift cluster** where you'll store your data.

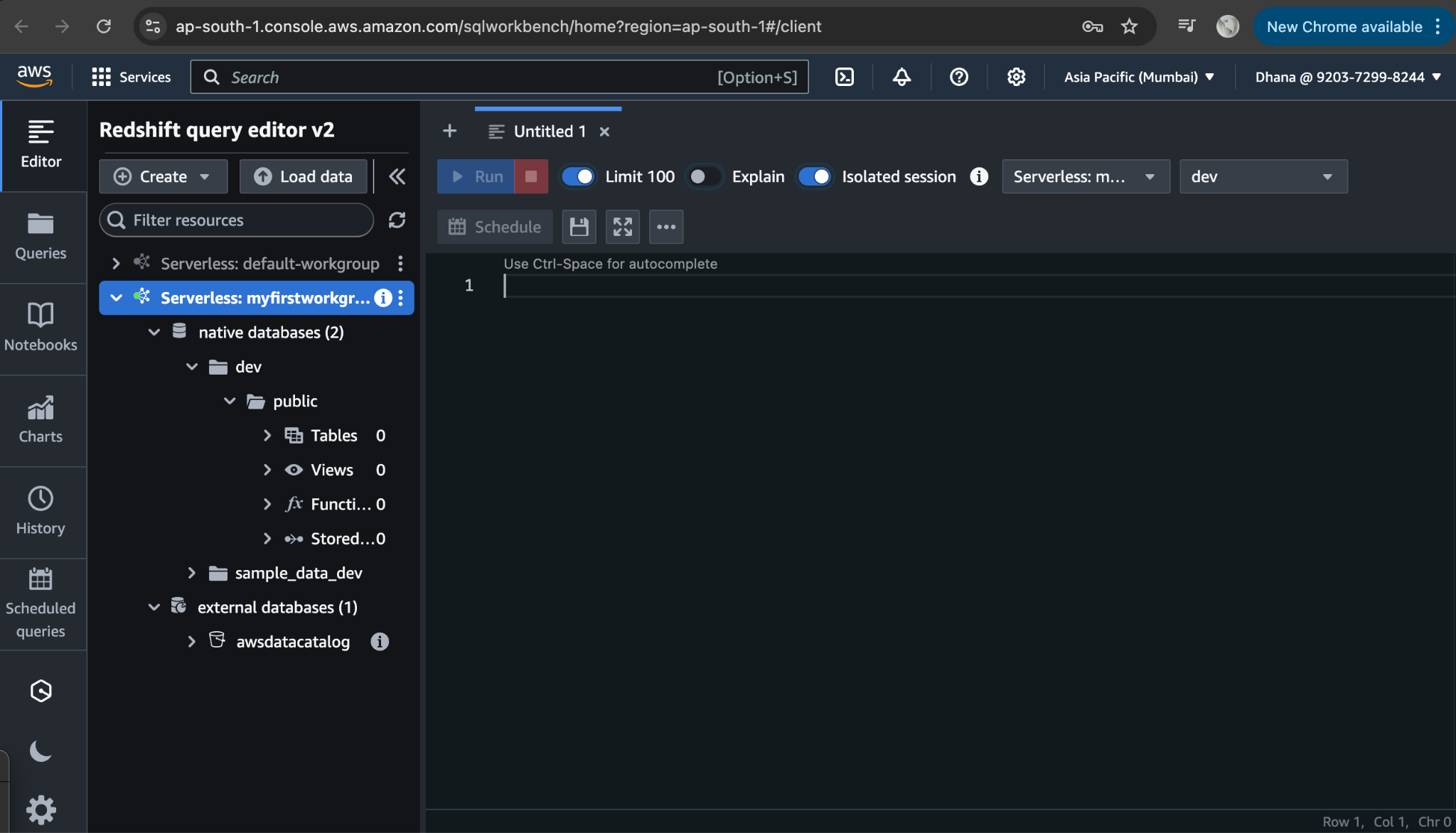
#### **Steps to set up Redshift:**

1. Go to **AWS Management Console**, navigate to **Redshift Serverless**, and create a new **Namespace and Workgroup** by choosing Base capacity and Maximum capacity.
2. Choose VPC, Subnets and Security Groups in which we want redshift to be created. Make sure the inbound rules of Security Group allows the traffic to Redshift with port 5439.
3. Define your **database** and create an **admin user**.
4. Associate an **IAM role** for Redshift to access other AWS resources.



6. Click on query data and login to WorkGroup by giving database name, user, password. You will be able to view Database, Schema, Tables.





**2.** **Set Up ORM Integration**

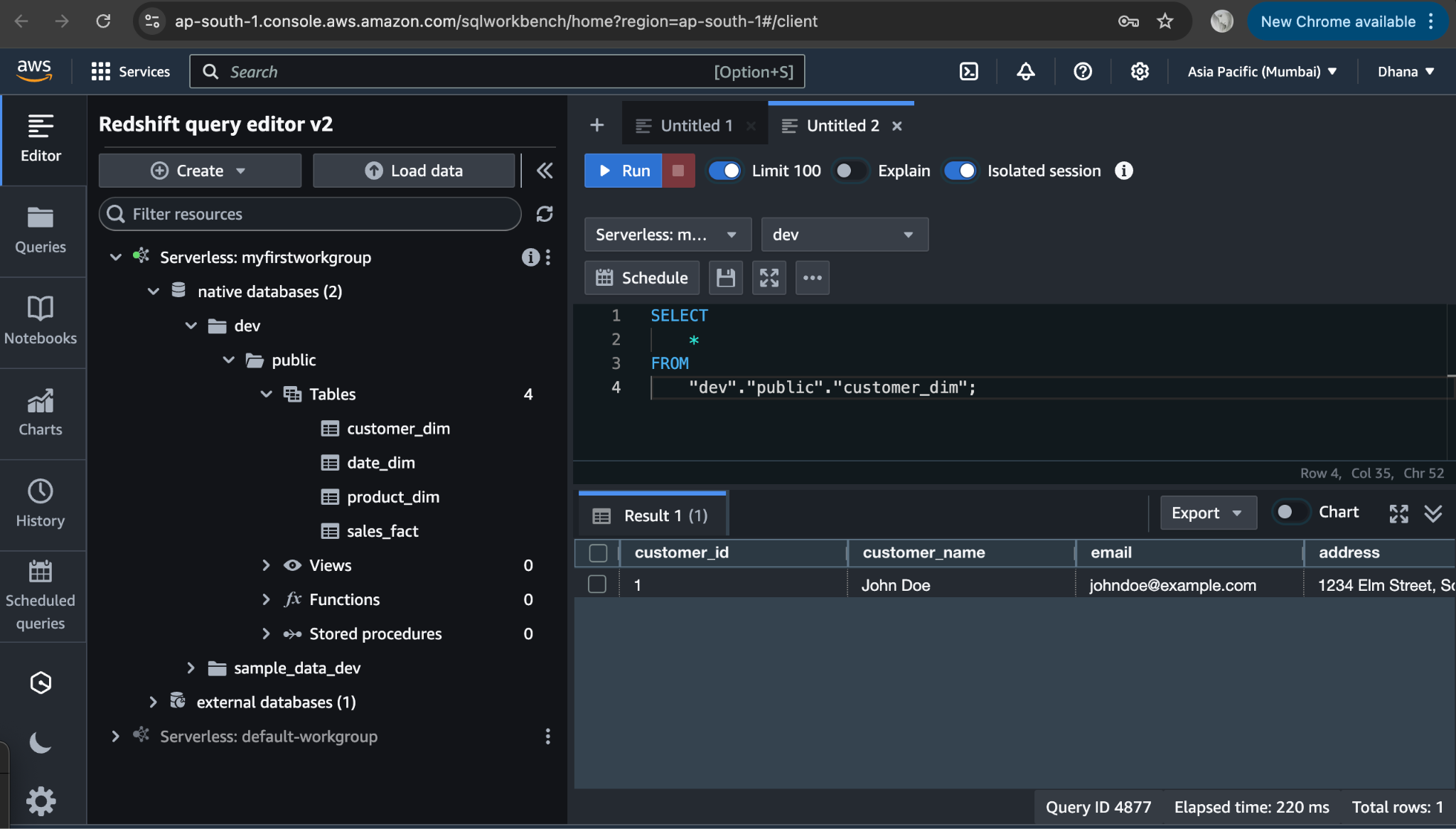
You will use an **ORM** to interact with Redshift(abstracting raw SQL queries). In Python, you can use **SQLAlchemy** to achieve this. You can create a python script for Creating a SQLAlchemy Engine and Session to interact with Redshift, to define Models via ORM Classes, to perform CRUD operations and to query the data.

**Install SQLAlchemy and psycopg2 (PostgreSQL adapter):**  
pip install sqlalchemy psycopg2

**Python Script Used:**

RedshiftORM.py

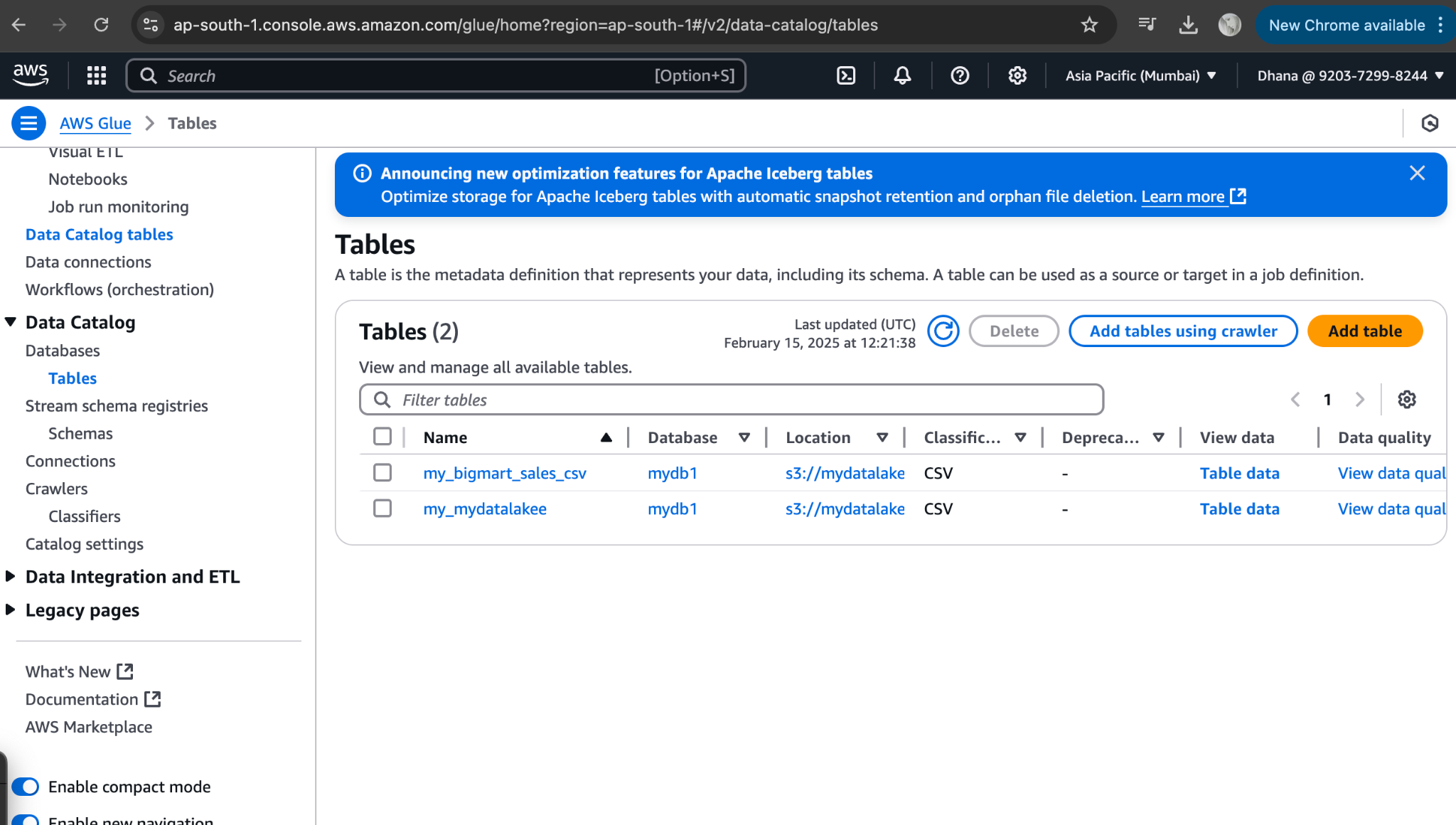
After Executing above Python script, we should be able to see the tables that has been created in Redshift cluster.



**3. Redshift Spectrum for querying data in S3 without loading data to Redshift.**

Upload files in the S3 folder..

Create Data Catalog in AWS Glue using crawlers.



Create External Schema to connect to Redshift Spectrum using below Query. The query should look as shown in screenshot.

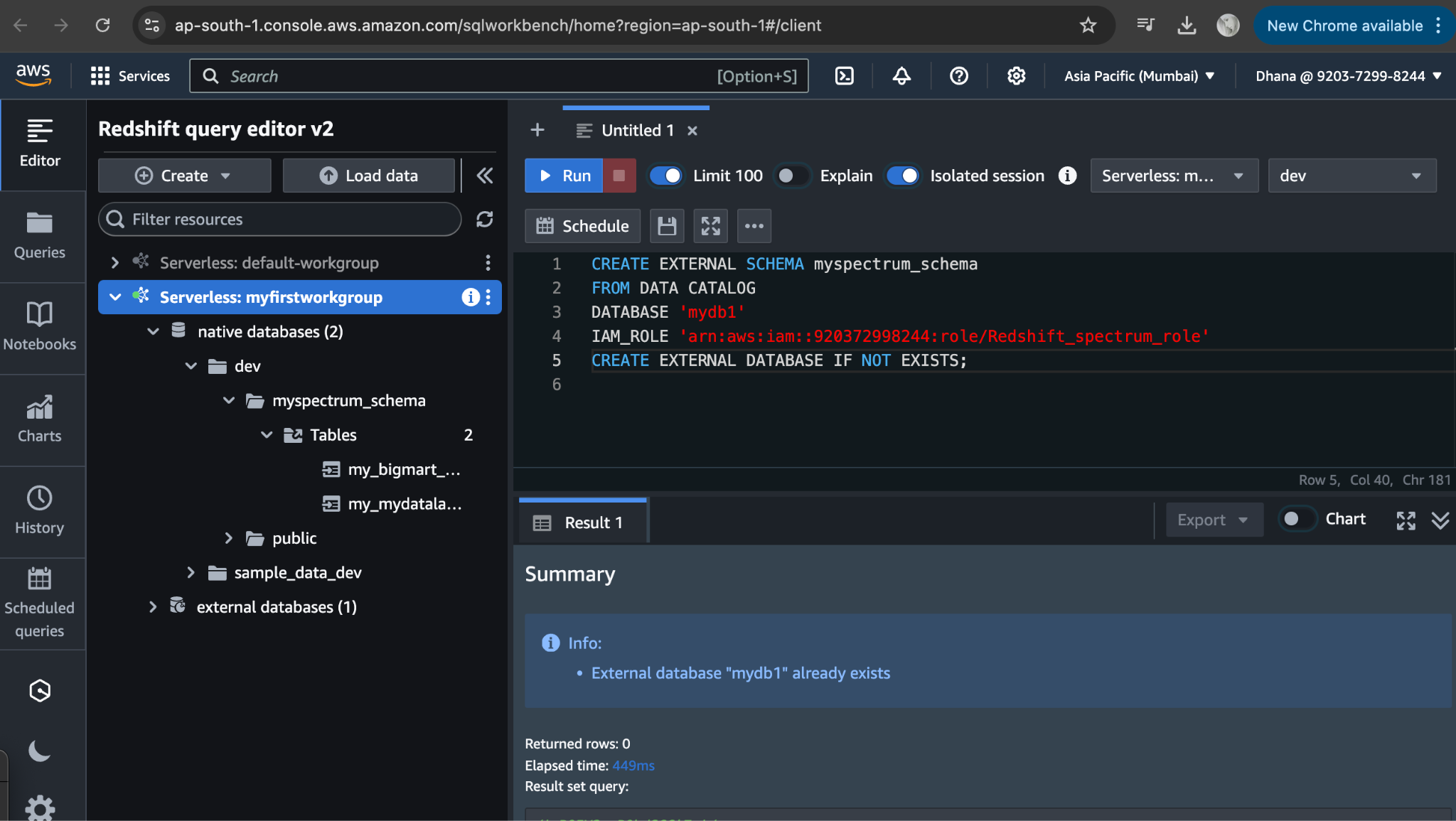
CREATE EXTERNAL SCHEMA spectrum\_schema

FROM DATA CATALOG

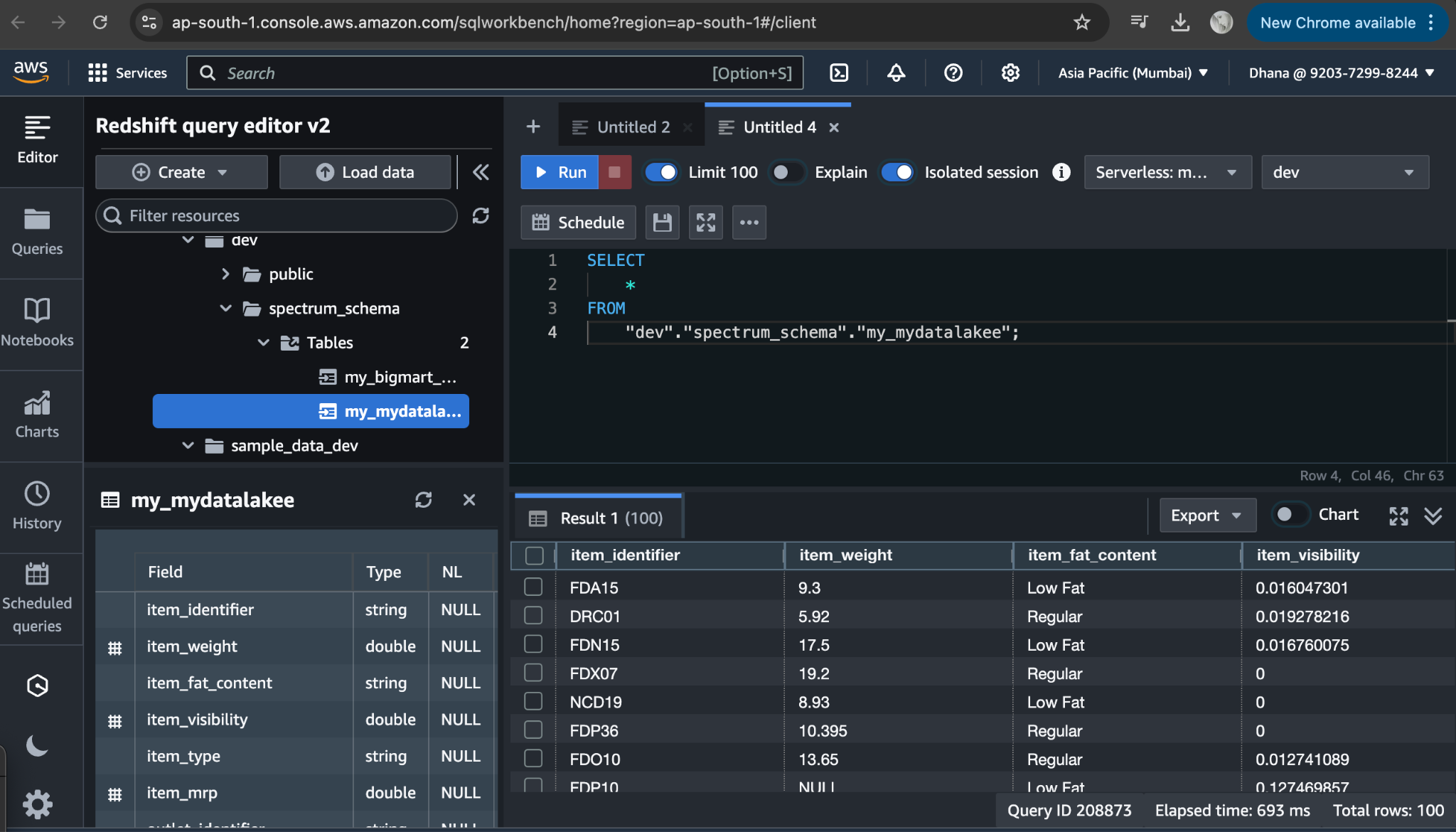
DATABASE 'your\_datacatalog\_database'

IAM\_ROLE 'arn:aws:iam::your-aws-account-id:role/RedshiftSpectrumRole'

CREATE EXTERNAL DATABASE IF NOT EXISTS;



When you create an external schema in Redshift linked to the Glue Data Catalog, the tables do not need to be physically created in Redshift. Instead, Redshift queries the metadata (table structure, column names, types, etc.) from the Glue Data Catalog, but the actual data remains in S3.



You can also create external tables using query example given below.

CREATE EXTERNAL TABLE spectrum\_schema.retail\_data (

Date VARCHAR(50),

Product\_ID VARCHAR(50),

Product\_Name VARCHAR(50),

Category VARCHAR(50),

Units\_Sold INT,

Unit\_Price DECIMAL(10,2),

Total\_Sales DECIMAL(10,2),

Store\_Location VARCHAR(50),

Payment\_Method VARCHAR(50)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LOCATION 's3://mydatalakee/retaildata.csv' TABLE PROPERTIES ('skip.header.line.count'='1');