CS 79B Final Project - Part 2

1. Document and describe the dataset you selected.

Dataset: Department of Veterans Affairs COVID-19 National Summary:

The Department of Veterans Affairs (VA) is the largest integrated health system in the United States with more than nine million enrolled Veterans and over six million Veterans receiving health care each year. VA employs nearly 380,000 individuals including more than 350,000 professionals within the Veterans Health Administration.

This report summarizes what we know about the status of COVID-19 patients who have been tested or treated at VA facilities.

VA COVID-19 Cumulative Cases:

Running total of all patients tested or treated at a VA facility for known or probable COVID-19. This includes Veterans, employees, and non-Veterans.

Active Cases:

Patients tested or treated at a VA facility for known or probable COVID-19 who have neither dead nor reached convalescent status.

Convalescent Cases:

Patients tested or treated at a VA facility for known or probable COVID-19 who are either post-hospital discharge, or 14 days after their last positive test, whichever comes later.

Known Deaths:

All deaths among VA patients known to have tested positive for COVID-19. "Inpatient" indicates the death occurred in a VA hospital. "Known other" indicates the death was reported to VA but occurred elsewhere.

Categories of Cases:

We group Active and Convalescent Cases into the following categories: Veteran, Employee, Veteran-Employee and All Other. "All Other" includes civilians admitted to VA hospitals as humanitarian cases, Tricare patients, Active Duty Military, and other groups.

Vaccinations:

All the vaccinations administered by VA – Pfizer, Moderna, and Janssen - are tallied in this report. Each of these options have been deemed to be effective and safe, and have an Emergency Use Authorization (EUA) from the Food and Drug Administration (FDA).

The tables break the numbers down as follows:

1. Who we vaccinated*: Veteran, VA Employee, or Federal Partner**.
2. How many of those received a full vaccination course (2 doses of either the Pfizer or Moderna vaccine, or a single dose of the Janssen vaccine).

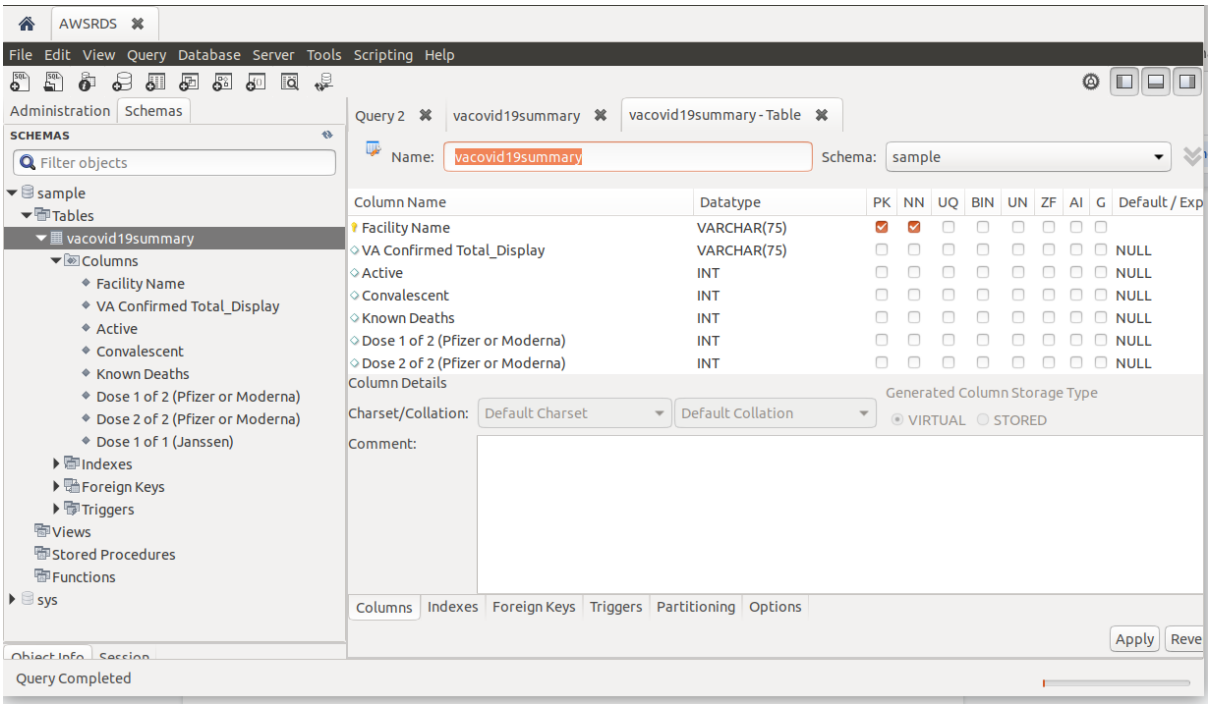
3. The number of total doses administered, as follows:
    1. “Dose 1 of 2” refers to the initial dose of a 2-dose vaccine series, that is, Pfizer or Moderna.
    2. “Dose 2 of 2” refers to the final dose of a 2-dose series.
    3. “Dose 1 of 1” refers to the single dose needed for the Janssen vaccine.
4. Cumulative totals are provided by the facility. Totals for smaller sites from a multi-division facility are included in the overall totals.
5. If a particular subcategory has a count of less than 10, it is included in the total for the next higher category. That is done to protect patient privacy.

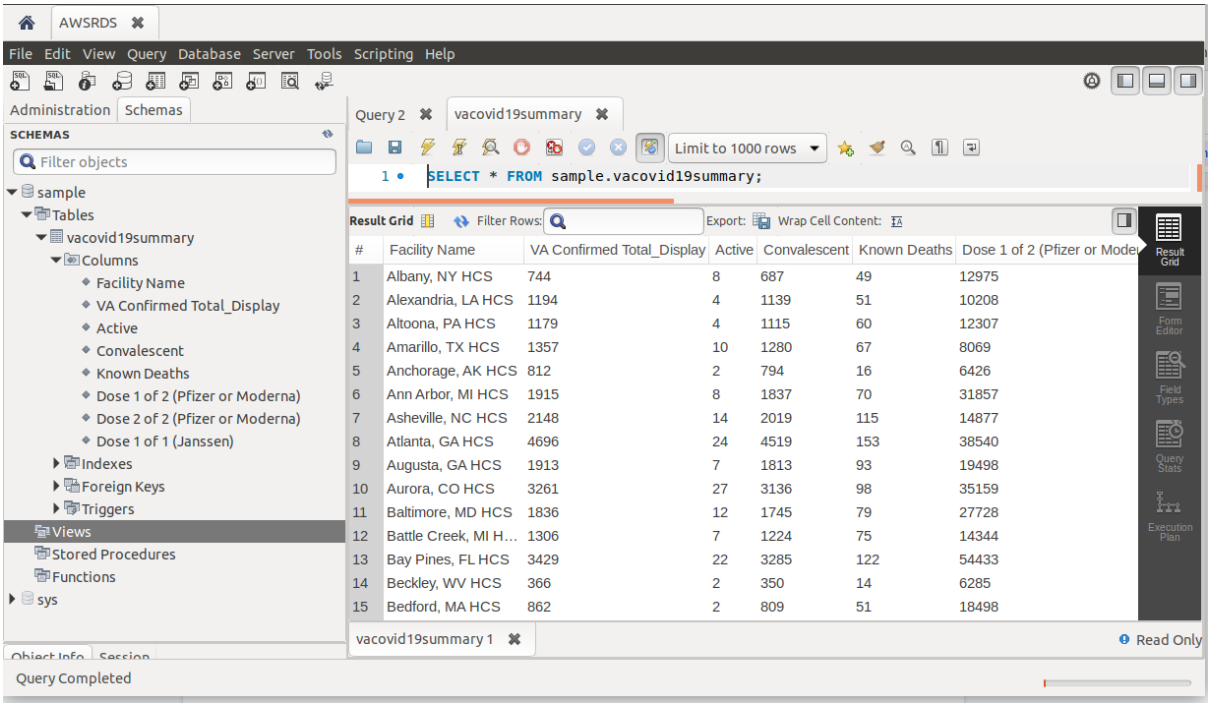
*Veterans that undergo vaccination because they are a VA employee or Federal Partner are counted under those latter categories, and not included in the “Veteran” category. That is to prevent double-counting.

**Federal Partners include Front Line Staff and First Responders that work in other agencies and were directed to receive their vaccination at a VA facility.

2. Create either an RDS or DynamoDB solution to store records from this dataset.

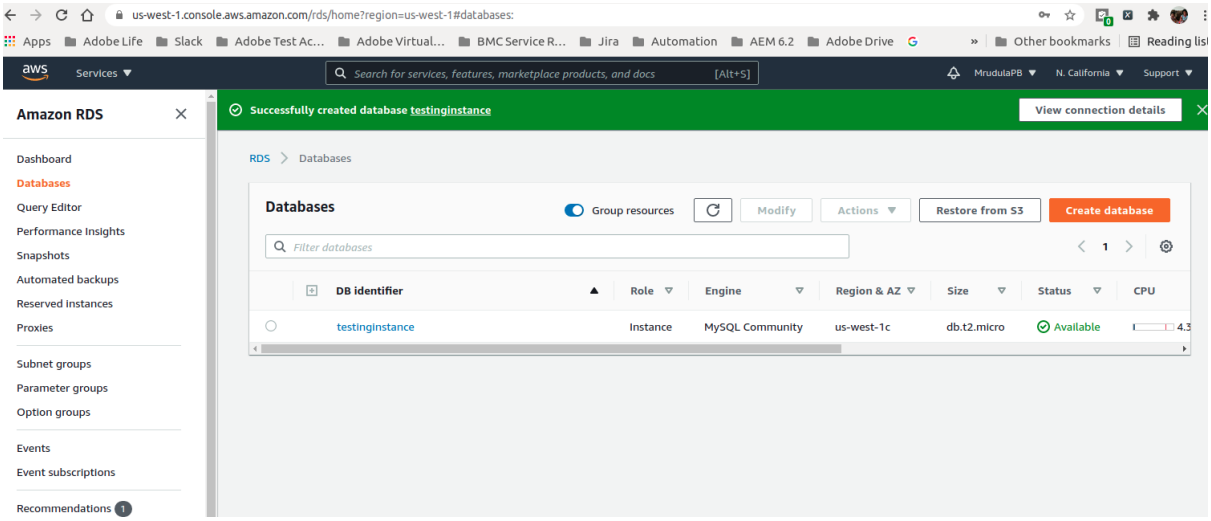
Creating RDS solution:

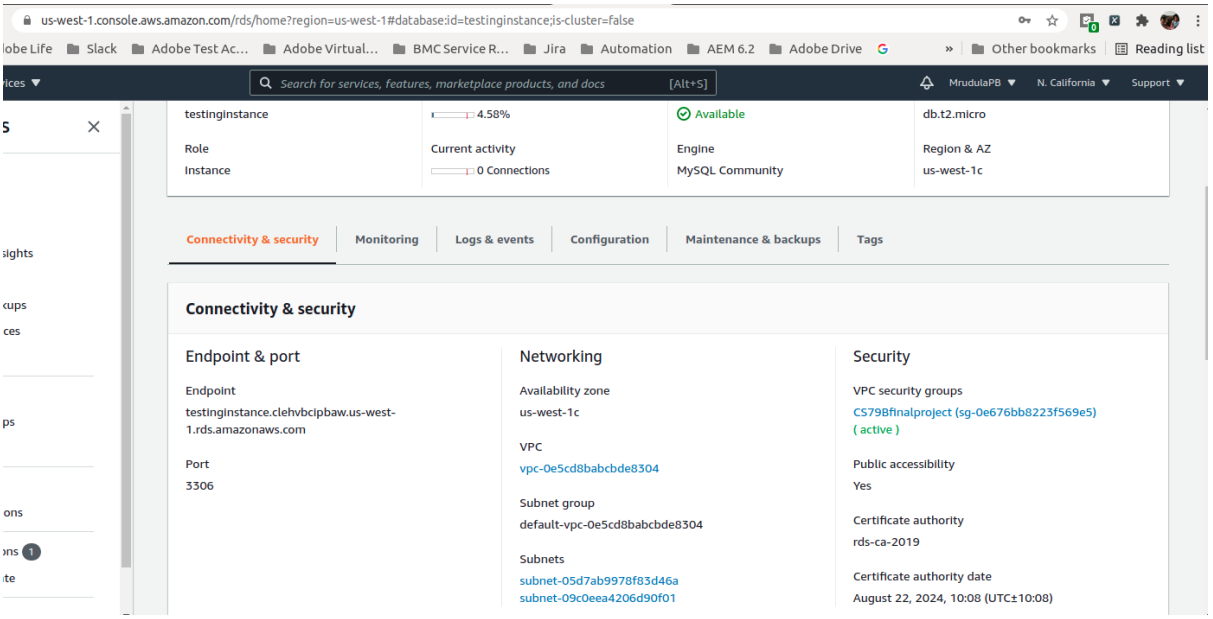




3. Document the structure of the storage solution you defined in AWS.

Table creation in MYSQL Workbench:



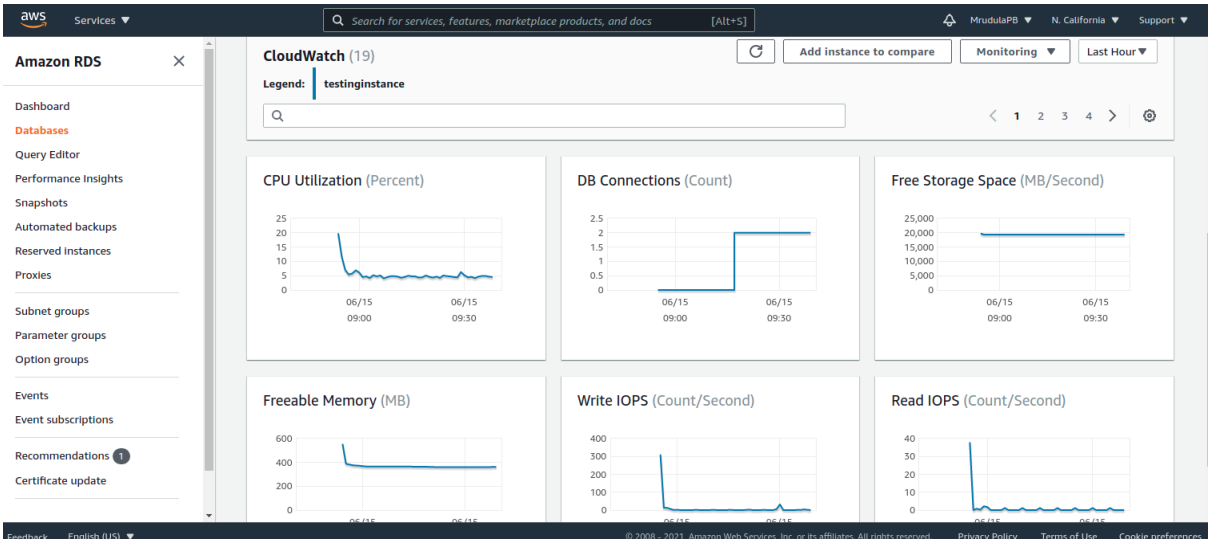


4. Upload various records into your storage solution that you downloaded from the government dataset.

https://www.accesstocare.va.gov/Healthcare/COVID19NationalSummary

5. Please take a screenshot showing me either your mySQLWorkbench or the table scan from the Items tab of the DynamoDB table you created to store this dataset.

I do not have an Items tab in RDS. Hence showing MOnitoring via CloudWatch in RDS:



6. Describe why you chose the AWS service you did for the data you are trying to store.

Amazon RDS allows us to use the AWS Management Console or a simple set of web services APIs to create, delete and modify relational database instances (DB Instances). You can also control access and security for your instance(s) and manage your database backups and snapshots.

It takes only a few clicks in the AWS Management Console to launch and connect to a production-ready MySQL database in minutes. Amazon RDS for MySQL database instances are pre-configured with parameters and settings for the server type you have selected. Database parameter groups provide granular control and fine-tuning of your MySQL database.

I had to create the RDS Database in AWS and connect to it via MySQL Workbench. I created tables and uploaded the data via SQL queries and .CSV files.