Metropolitan State University
ICS 432 - 01: Distributed and Cloud Computing
Fall 2021

Assignment 1: Working with Cloud Virtual Machines
Total points: 55
Out: Saturday, September 11, 2021
Due: 11:59 PM on Friday, October 1, 2021

The goal of this homework is practice using virtual machines in the cloud. There are two exercises:

1- Exercise 1: Hosting a web site on an AWS EC2 instance.
2- Exercise 2: Data Analysis using SQL on a GCP virtual machine.

IMPORTANT NOTES

- At various parts in this assignment, you are asked to take screenshots of your work. Open a word document and paste the screenshots in that document in the same order as mentioned in this assignment. Make sure to highlight the screenshot number. The assignment may also include some questions that requires text-based response so make sure to include answers to these questions in your submitted report.

- Make sure your screenshots are clear and easily readable. I will not give you points for a certain screenshot if I cannot read the text inside that screen shot.

- After you complete the two exercises, upload the word document to the designated D2L folder by 11:59 PM on Friday, October 1, 2021.

- To avoid wasting your AWS Educate and GCP credits, stop the virtual machine instances when you are not using them.

- Do not delete the virtual machine instances created in this homework until after the homework is graded.

- No submissions will be accepted after 11:59 PM on Monday, October 4.

- On Windows machines, you may consider using Snip & Sketch for screenshot handling.

Exercise 1: Hosting a web site on an AWS EC2 instance

To complete this exercise, you need to follow the following steps:

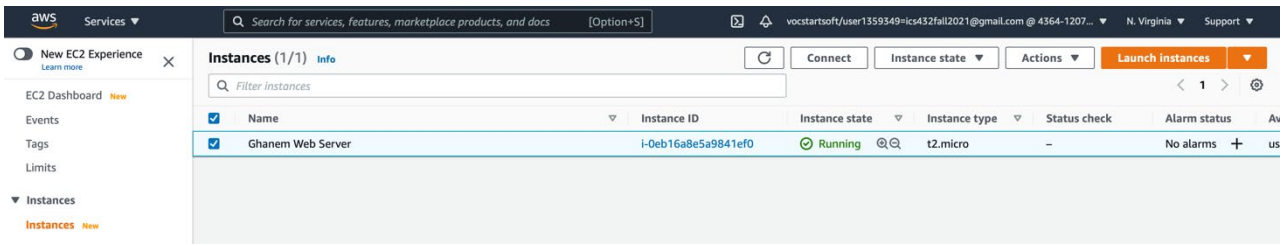
1. Step 1: Launch an EC2 instance.
2. Step 2: Connect to the EC2 instance using SSH.
3. Step 3: Install the nginx web server software on the EC2 instance.
4. Step 4: migrate an existing website to your instance.
5. Step 5: Change the security setting of the instance to allow public access.
6. Step 6: Host another simple web page on the web server.
7. Step 7: Stop the EC2 instance.
8. Step 8: Summarize your learning.

Note: To avoid wasting your AWS Educate credits, make sure to stop your EC2 instance when you are not using it. You can restart the instance to resume your work.

Step 1: Launch an EC2 instance

1- Log in to your AWS Educate account.
2- Create an EC2 instance with the following parameters:
   - **Amazon AMI:** Amazon Linux 2 AMI
   - **Instance type:** t2.micro
   - **Instance details:** keep all the defaults.
   - **Add storage:** keep defaults.
   - **Add tags:** keep defaults.
   - **Configure security group:** keep defaults.

3- When prompted for key pair, choose 'Create a new keypair' from the drop-down menu. Set key pair name to: ics432homework1keypair.
4- IMPORTANT: Click on Download Key Pair to download the file to your computer. A file called homework1keypair.pem will be downloaded and stored under the Downloads folder on your computer.
5- Launch the instance then open the EC2 dashboard and check the status of your instance.
6- While on the EC2 dashboard, click on the dash '–' under the Name column and enter the instance name as: <your-last-name> Web Server (for example, mine will be `Ghanem Web Server`)

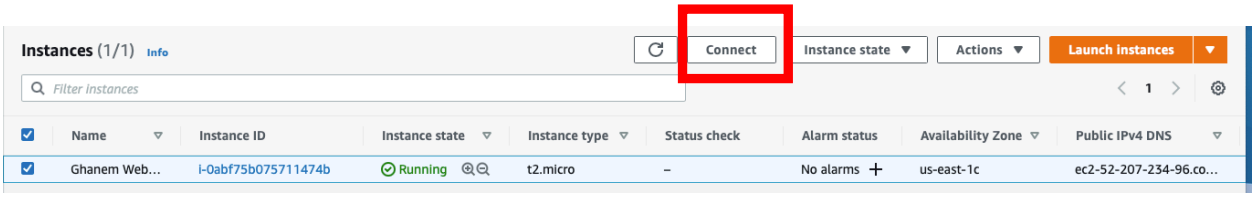


7- On your computer, create a folder called ics432-homework1. Copy the homework1keypair.pem file from your downloads folder to your ics432-homework1 folder.

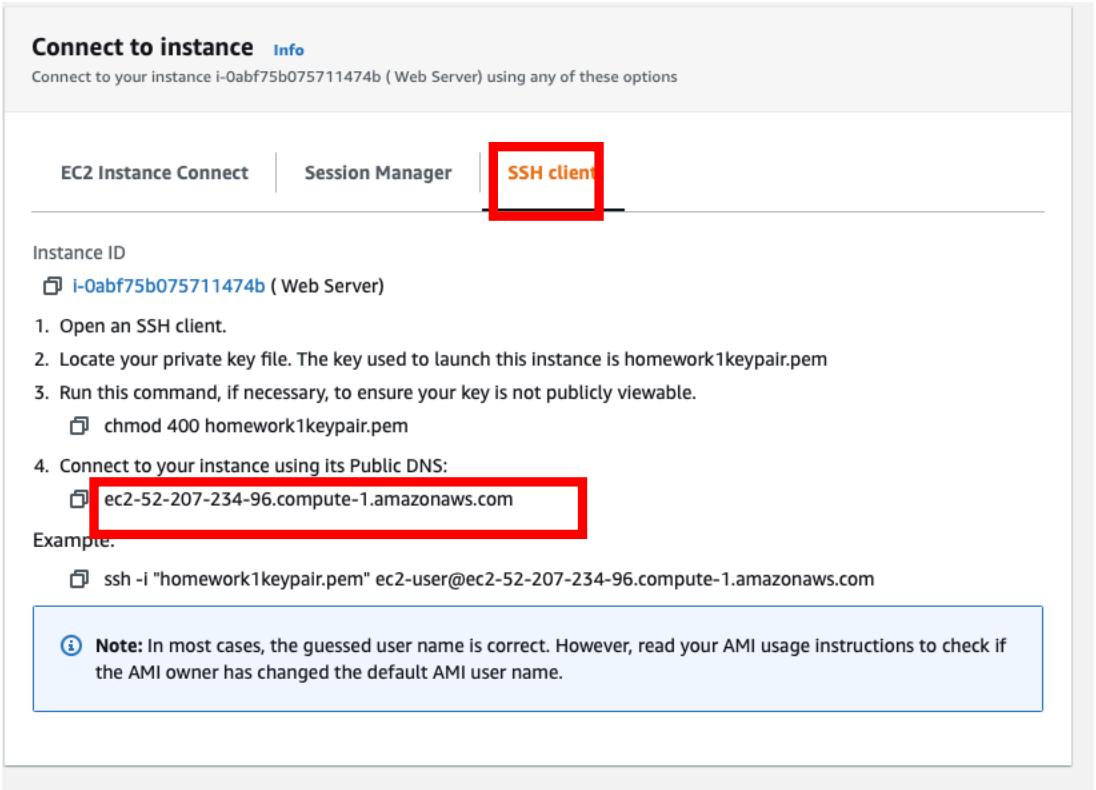
Homework report screenshot 1: Take a screen shot of EC2 dashboard showing your running instance. Your screenshot should include the top bar that shows your AWS Educate account it.

Step 2: Connect to the EC2 instance using SSH

1- From the instance Dashboard, click on Connect.



2- Select SSH client. A screen similar to the following screen is displayed. This screen gives the instance's DNS and gives you instructions on how to connect to the VM. The next two subsections explain in details how to establish the SSH connection using Mac and Windows computers respectively.



SSH connection to from a Mac Computer

1- Open a Terminal window.

2- Change directory to ics431-homework1 folder (where the key pair file is stored) using the command

```
cd ics432-homework1
```

3- As instructed in the above figure, run the following command, if necessary, to ensure your key is not publicly viewable.

```
chmod 400 ics432homework1keypair.pem
```

4- Connect to your AWS instance using the following command. Make sure to change the Public DNS part in the following command with the Public DNS of your instance.

```
ssh -i "ics432homework1keypair.pem" ec2-user@ec2-52-207-234-96.compute-1.amazonaws.com
```

5- If prompted to continue, type 'yes' and press enter.

Homework report screenshot 2: Take a screen shot of your terminal window after connecting to the instance.

SSH connection from a Windows Computer

You need to install PuTTy software to use to establish SSH connection to an EC2 instance. Refer to the file windows-ssh-filecopy on D2L for detailed instructions.

Homework report screenshot 2: take a screen shot of your terminal window after connecting to the EC2 instance.

Step 3: Install the nginx web server software on the VM

Run the following commands on the EC2 instance's command window:

1- install updates. You may receive a message that there are no updates available.

```
sudo yum update
```

2- Install the **nginx web server** on the instance by the running the following three commands in order

a. Run the following command to list all the packages that are available

```
amazon-linux-extras
```

b. In the displayed list, check the name of the current nginx package (for example nginx1). Run the following command to install nginx web server.

```
sudo amazon-linux-extras install nginx1
```

c. If you are prompted to enter y/d/N, enter Y and hit enter The output on your screen should show the successful installation of nginx.

Homework report screenshot 3: take a screenshot of the command window after successful installation of nginx. Note that you may need to scroll up to find the 'Complete!' notification. Make sure your screenshot is readable.

Step 4: Migrate an existing website to the EC2 instance

In this step, you will migrate a web site, called BallotOnline, to your EC2 instance and configure the nginx web server to serve it.

1- Clear the command window screen by running the following command:

```
clear
```

2- Create a directory for the BallotOnline website data

```
sudo mkdir -p /data/www
```

3- Download the website files from the legacy webserver. The files have been compressed into website.zip and the URL is www.ballotonline.biz/website.zip

```
cd /data/www
sudo wget www.ballotonline.biz/website.zip
```

4- Type `ls` to make sure the file is successfully downloaded.

5- unzip the file into the /data/www/BallotOnline directory.

```
sudo unzip website.zip -d /data/www/BallotOnline
```

6- Move inside the BallotOnline folder and list its contents.

```
cd BallotOnline
ls
```

Homework report screenshot 4: Take a screen shot of your shell screen with all the files listed under BallotOnline directory. There should be five files and one directory.

7- Make sure that the index.html file is downloaded to the `/data/www directory`:

```
cat BallotOnline/index.html
```

You should see text starting with <!DOCTYPE html>

Homework report screenshot 5: take a screen shot of the displayed index.html file.

8- Change the permissions on the index.html file to make it less restrictive

```
sudo chmod 755 BallotOnline/index.html
```

9- Replace the default nginx configuration file (located in /etc/nginx/nginx.conf) with the one that is provided within the zip file that you have downloaded and unzipped.

```
sudo mv /etc/nginx/nginx.conf /etc/nginx/nginx.conf.old

cat /etc/nginx/nginx.conf
```

Note: You should receive an error "No such file or directory"

Homework report screenshot 6: take a screen shot to show the error message.

10- copy the downloaded *nginx.conf* to the correct location.

```
sudo cp /data/www/BallotOnline/nginx.conf /etc/nginx/
```

11- You should now have a file named *nginx.conf* in the */etc/nginx* directory. To double-check, run the following command:

```
cat /etc/nginx/nginx.conf
```

You should see the content of *nginx.conf* from the */etc/nginx* directory.

Homework report screenshot 7: take a screen shot of the displayed nginx.conf file.

12- Start nginx by running the following command

```
sudo nginx
```

13- From your SSH session, test the web server by running the following command:

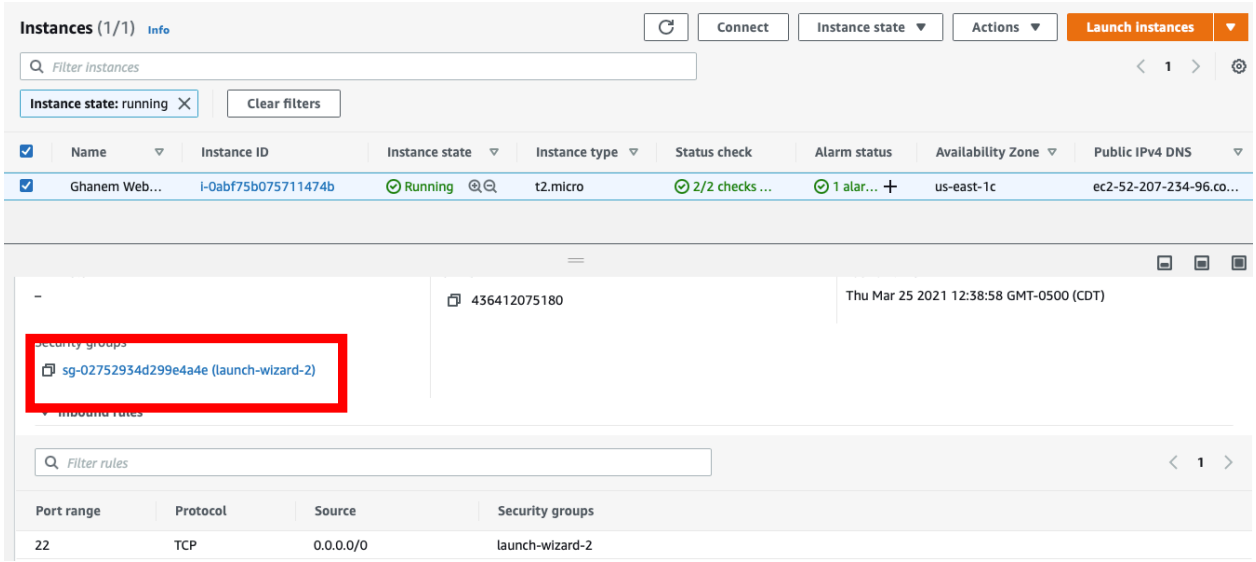
```
curl http://localhost
```

You should see the HTML file for BallotOnline. You have now confirmed that your basic nginx web server configuration is correct for your instance.

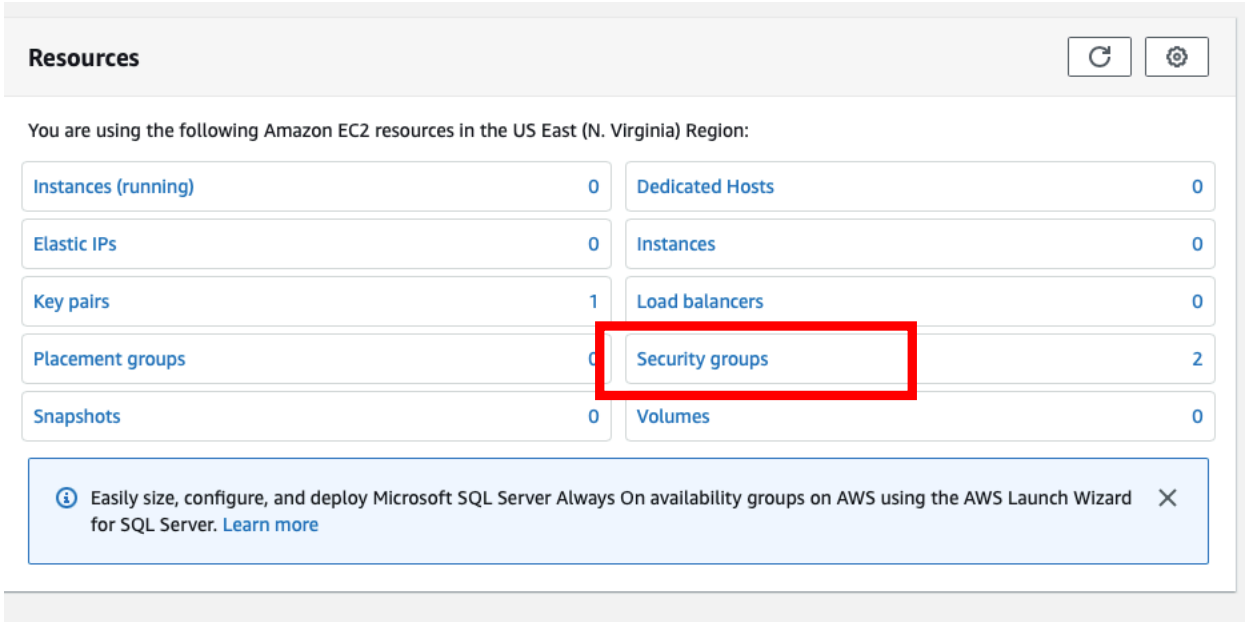
Homework report screenshot 8: Take a screen shot of your shell screen that shows the main HTML file of the BallotOnline web site.

Step 5: Change the security setting of the instance to allow public access

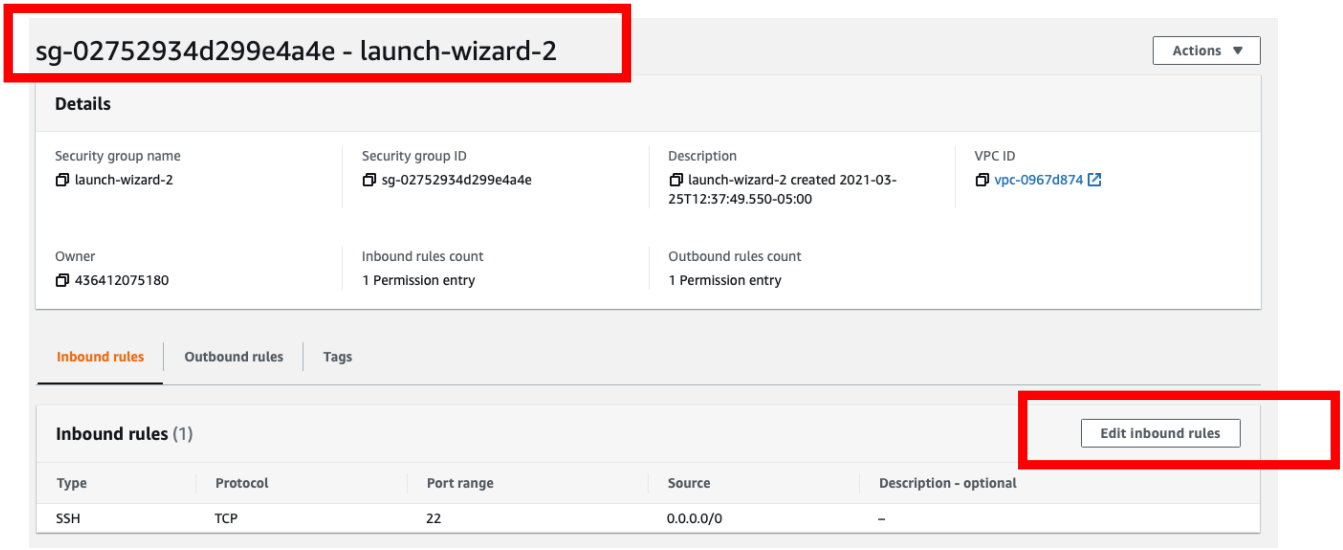
1- Check the name of the security group of your instance by going to instance information in the dashboard.



2- Go to the "Security Groups" menu in the EC2 Dashboard, and click on it.

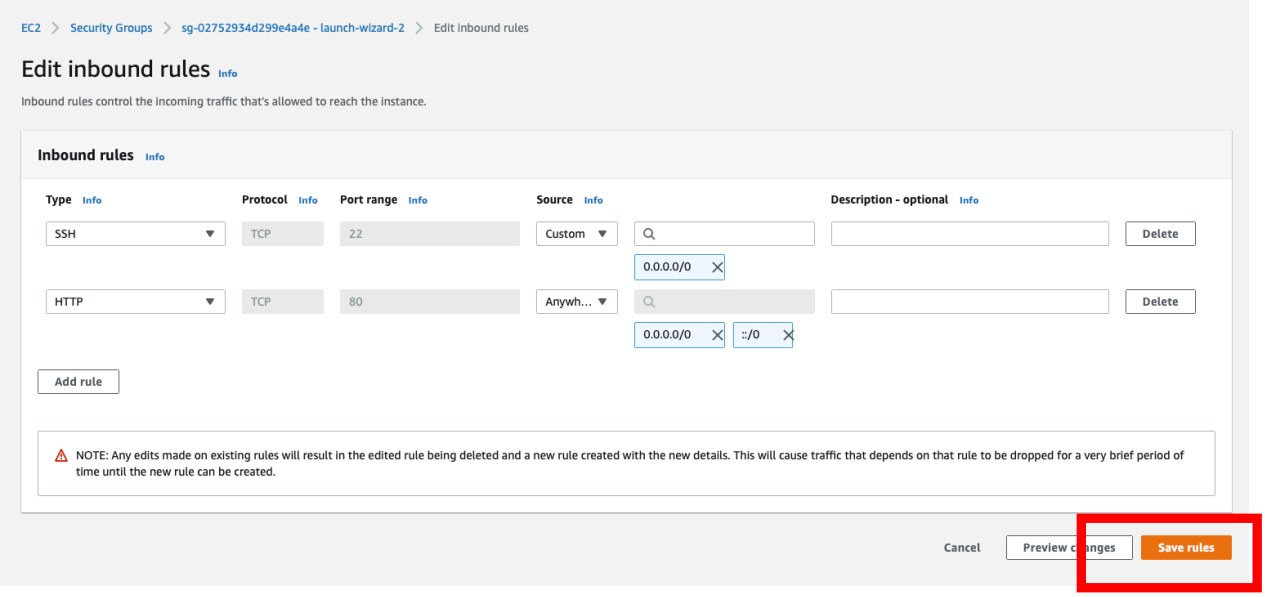


3- Choose the security group of your instance. For example, in my case, the name of the security group is 'launch-wizard-2'. Click on the security group name to open the security group dashboard.



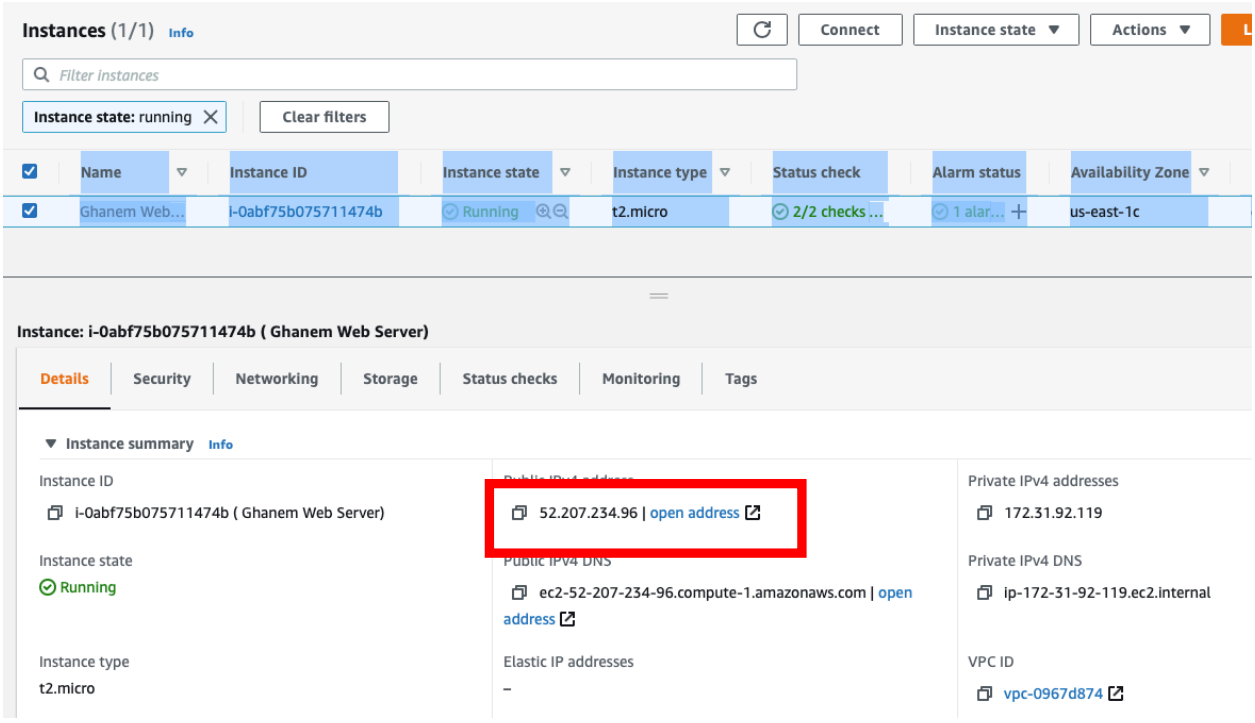
4- Click on 'Edit inbound rules'.

5- Click on 'Add rule' and add a new rule for HTTP with the following settings:
Protocol: TCP
Port Range: 80
Source:Anywhere



Homework report screenshot 9: Take a screen shot of the security group screen that shows the two rules. Make sure that the name of the security group is displayed.

6- Click on 'Save rules'.
7- Get the Public ID Address of your instance from the instance dashboard.



8- Access the web site from a browser by using the following URL:

http://NN.NN.NN.NN/index.html

by replacing NN.NN.NN.NN with the public DNS of your instance. For example, in my case, I will use the following url:

http://52.207.234.96/index.html

Homework report screenshot 10: Take a screen shot of the web site when it opens in your browser and include the screenshot in your homework report.

Step 6: Host another simple web page on the web server

1- Open a text editor (e.g., Notepad on Windows or TextEdit on Mac) and create a HTML file with the following text. Replace my name with your name and add your own message instead of 'I hope you enjoy it'.
2- Make sure to save the file under ics432-homework directory with extension 'html'. For example, my file name is 'ghanemindex.html'.

```
<!DOCTYPE html>
<html>
<body>

<h1>Thanaa Ghanem Page</h1>
<p>I hope you enjoy it</p>

</body>
</html>
```

3- Copy the html file to the EC2 instance as follows.

From a Mac computer

a- Open a terminal window and go to `ics432-homework1` directory.

b- Run the following command to copy the file to the instance. Note that you will need to use the homework1keypair.pem file and the machine's public DNS that you used in Step 1. If prompted to enter a password, enter your local machine's password. If you are asked 'do you want to continue', choose Yes.

```
sudo scp -r -i ics432homework1keypair.pem ghanemindex.html
ec2-user@ec2-52-207-234-96.compute-1.amazonaws.com:~/
```

```
Warning: Permanently added 'ec2-3-222-116-212.compute-1.amazonaws.com,3.222.116.
212' (ECDSA) to the list of known hosts.
ghanemindex.html                              100%  114     2.2KB/s   00:00
(base) FH0217ML1880742:ics432-homework1 xn7118mm$
```

From a Windows computer:

a- To copy files from a Windows machine to an AWS instance, you will use a software called WinSCP. Refer to the following link for more information about using WinSCP: https://asf.alaska.edu/how-to/data-recipes/moving-files-into-and-out-of-an-aws-ec2-instance-windows/

b- Refer to the document titled windows-ssh-filecopy on D2L for detailed steps.

4- After the file is copied to the EC2 instance, from the shell screen that is connected to the virtual machine, type the following commands to first move to the home directory and then list files to make sure the file is successfully copied.

```
cd ~/
ls
```

Homework report screen shot 11: take a screen shot of the shell screen to show the successful copy of your html file to the virtual machine.

5- To host the simple HTML page on nginx, copy the ghanemindex.html to the main webserver folder.

```
sudo cp ghanemindex.html /data/www/BallotOnline
```

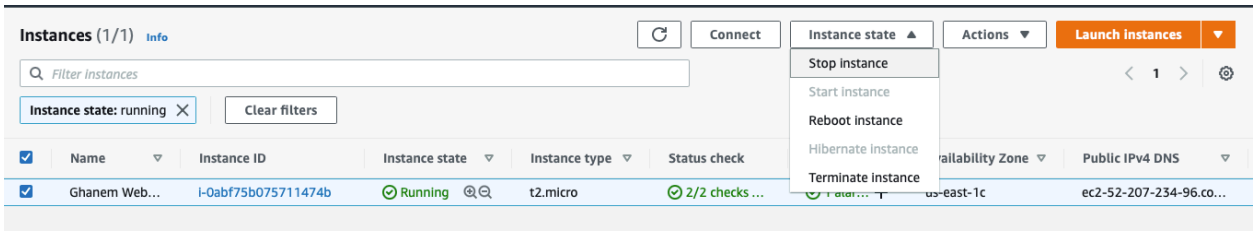
6- Now you can access new html from the browser using the instance's public IP address.

http://52.207.234.96/ghanemindex.html

Homework report screenshot 12: take a screen shot of the web browser page that shows the display of the new html page that contains your name.

Step 7: Stop your instance

Go back to your AWS EC2 dashboard and Stop your instance. Do not terminate the instance until I grade the homework.



Step 8: Summarize your learning

Write ~100-200 words to summarize your learning from this exercise. First, explain the most important things you learned from this exercise. Second, write about what parts of the exercise were straightforward and what challenges did you face.

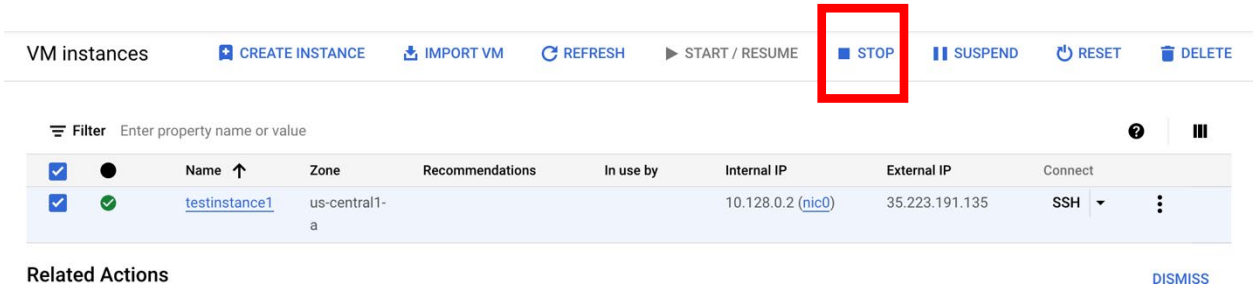
Exercise 2: Data Analysis using SQL on a GCP Virtual Machine

In this exercise, you will set up a MySQL DBMS on a virtual machine on Google Cloud Platform. Then, you will create a database write SQL queries to analyze a movie rating data set. These are the steps you will follow to complete the exercise:

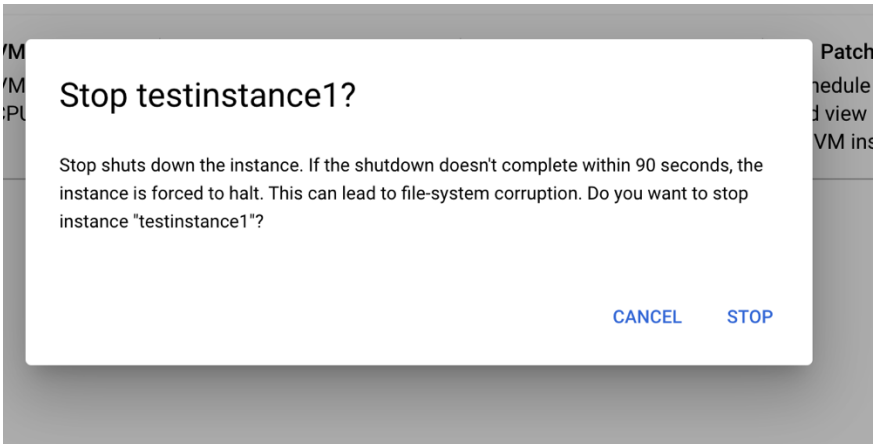
Step 1: Launch a VM instance on GCP
Step 2: Set up MySQL server on the VM.
Step 3: Create a database and fill tables with a data set about movie reviews.
Step 4: Write SQL queries to analyze a movie-rating data set.
Step 5: Stop the instance.
Step 6: Summarize your learning

Important: Stopping and restarting the instance

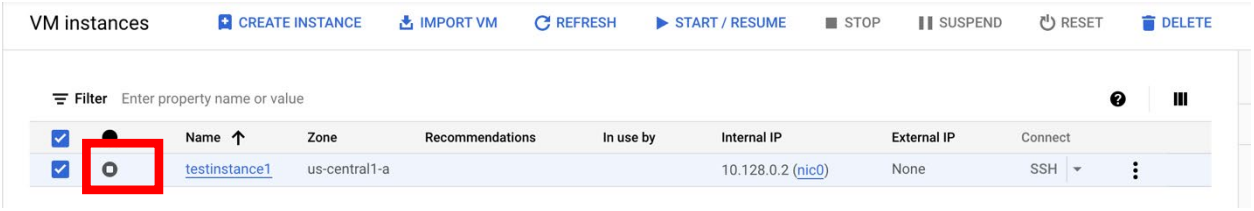
To avoid wasting your credits, make sure to stop the VM instance at any time when you are not working. You can stop the instance by going to the dashboard on GCP console and click on STOP.



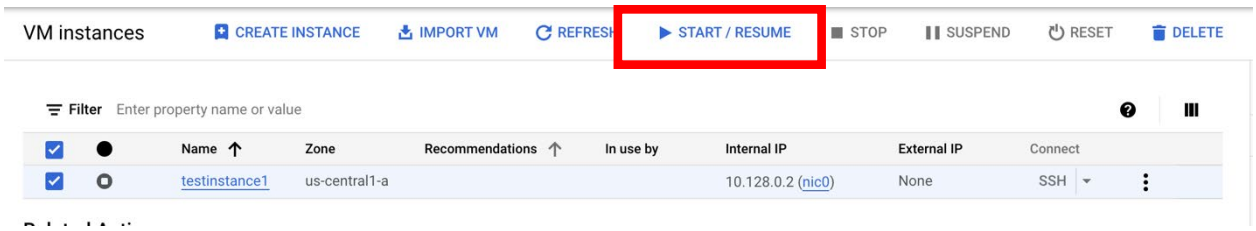
Confirm by clicking on STOP on the following screen.



Wait until you make sure the instance is stopped. The green circle to the left of the instance name is converted to black.

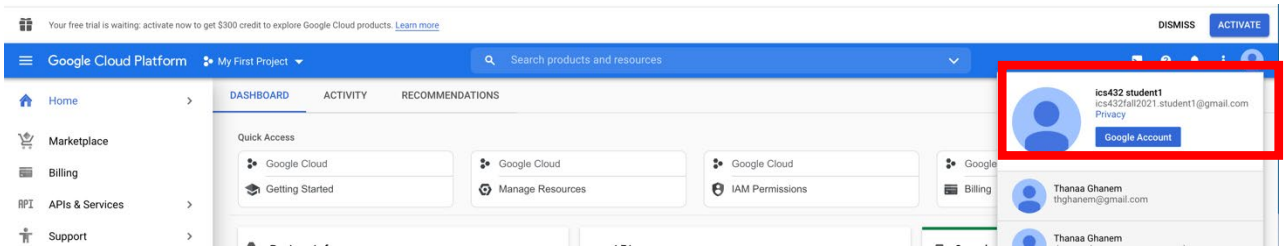


To continue working on your project, you can restart the instance by clicking on START/RESUME.

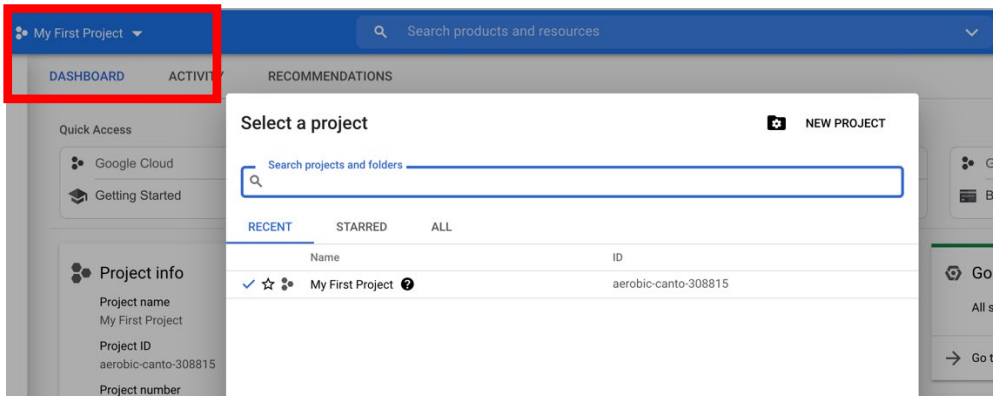


Step 1: Launch a VM instance on GCP

1- Log in to GCP console using the account you used to redeem the GCP coupon. To confirm that you are using the right account, click on the user icon of the top right corner of the screen

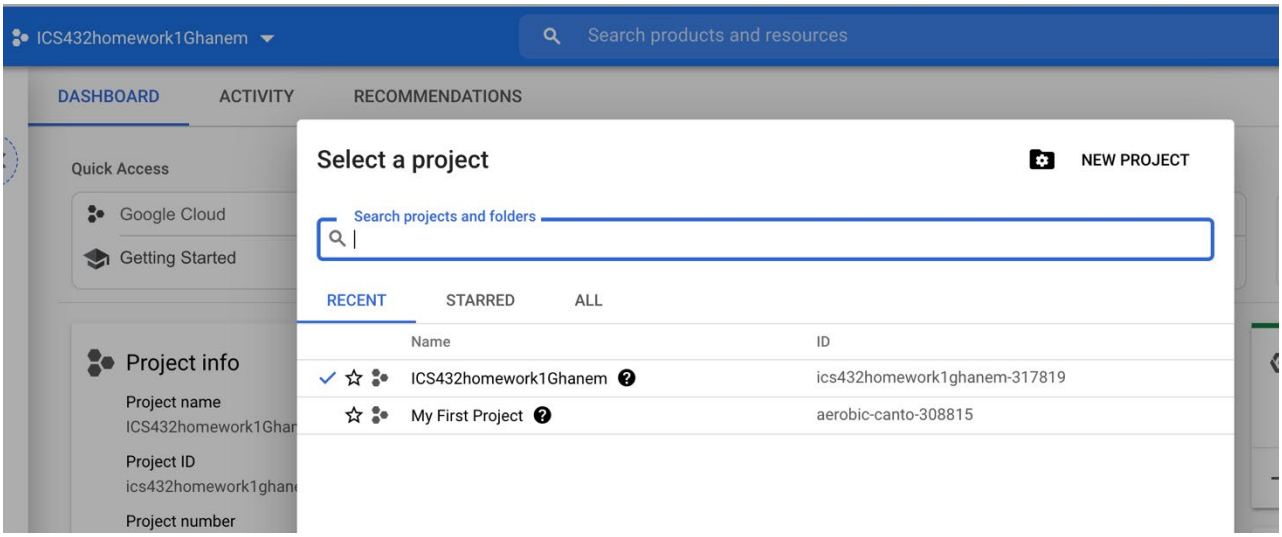


2- Click on the project drop down menu to open the project list.

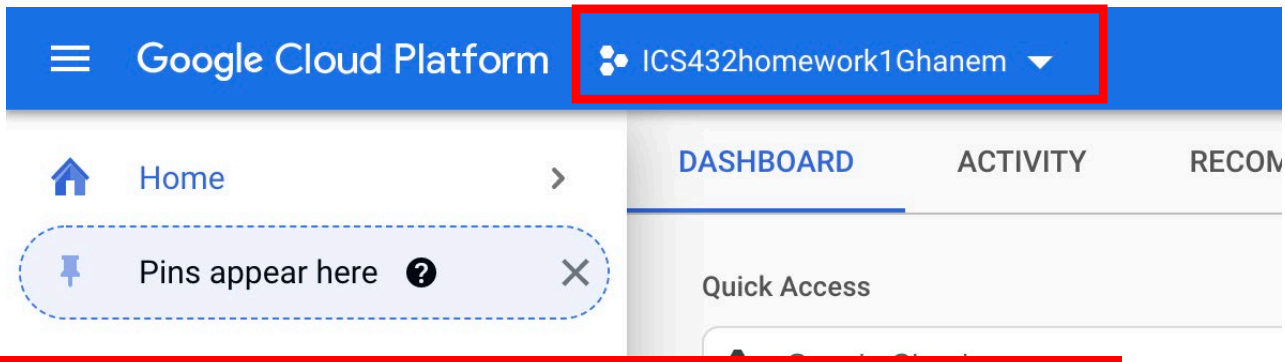


3- Click on NEW PROJECT and create a project with the name: `<yourlastname>ics432homework1`. For example, my project will be ghanemics432homework1.

4- Click again on the project drop-down list and click on the newly created project.

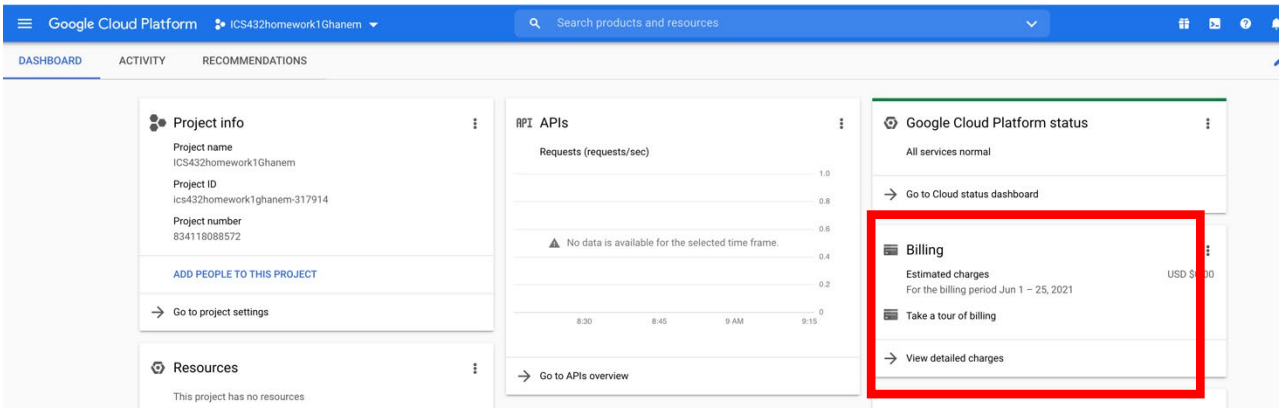


5- When you go back to the console menu, make sure the current project name is set to the project you created.

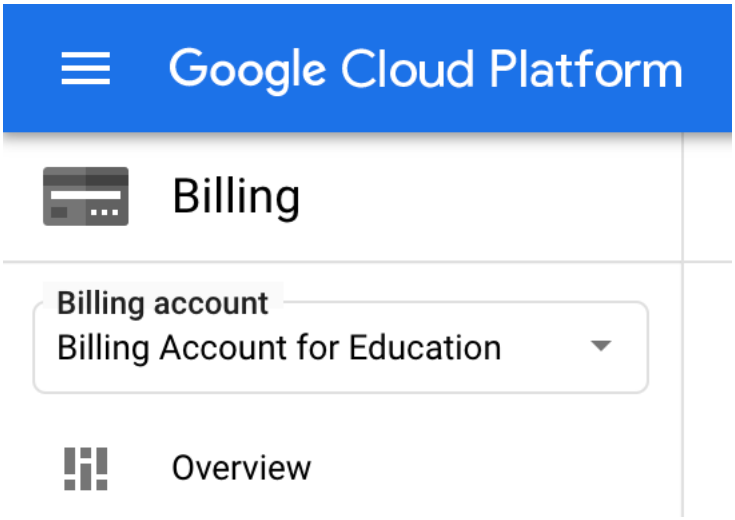


Homework report screenshot 13: take a screenshot of GCP console with your project name.

6- In project information page, click on Billing.



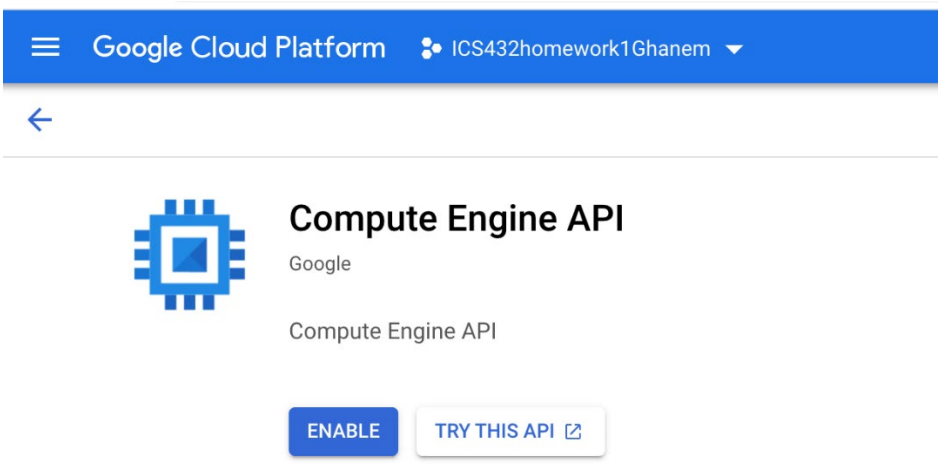
7- In the Billing page, confirm that the Account is Billing Account for Education.



8- From the left menu, click on Billing to go to your Account's billing account (which is the main billing account under which your project account falls) and make sure the Billing Account is set to": Billing Account for Education.

Homework report screenshot 14: take a screenshot to show the number of Credits in your account. This can be found in the bottom right corner of the page.

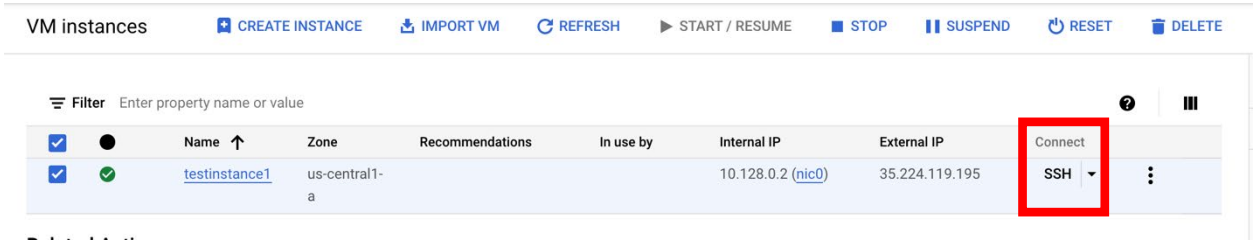
9- Click on Goolge Cloud Platform in the top left corner to go to your project. From the left menu, click on Compute→ Compute Engine→ VM instances.
10- If prompted to enable Compute Engine API, click on Enable.



11- Click on CREATE INSTANCE from the top menu bar and create a VM instance called <yourlastname>homework1mysql. Follow the steps as explained in Lab 1.

Homework report screenshot 15: take a screenshot to show the instances dashboard with your instance status as running.

12- Once the virtual machine is started, open an SSH connection to the instance:



Homework report screenshot 16: take a screenshot to show the SSH command window.Make sure your name and your machine name are clear in the screen shot.

Step 2: Set up MySQL server on the VM

1- On the command window, run the following command in the SSH window to install the *wget* tool which is a tool that can be used to download files from different locations on the internet. When prompted to enter [y/n], enter y.

```
sudo apt install wget
```

2- Follow the following steps set up MySQL. Refer to the following link for explanation of the steps as needed. https://www.digitalocean.com/community/tutorials/how-to-install-the-latest-mysql-on-debian-10

a. Run the following command in the SSH window to update the *apt package management system* which is a library that we are going to use to download packages from the internet.

```
sudo apt update

sudo apt install gnupg
```

b. move to the `/tmp` directory

```
cd /tmp.
```

c. Download MySQL package

```
wget https://dev.mysql.com/get/mysql-apt-config_0.8.13-1_all.deb
```

d. List files in the current directory

```
ls
```

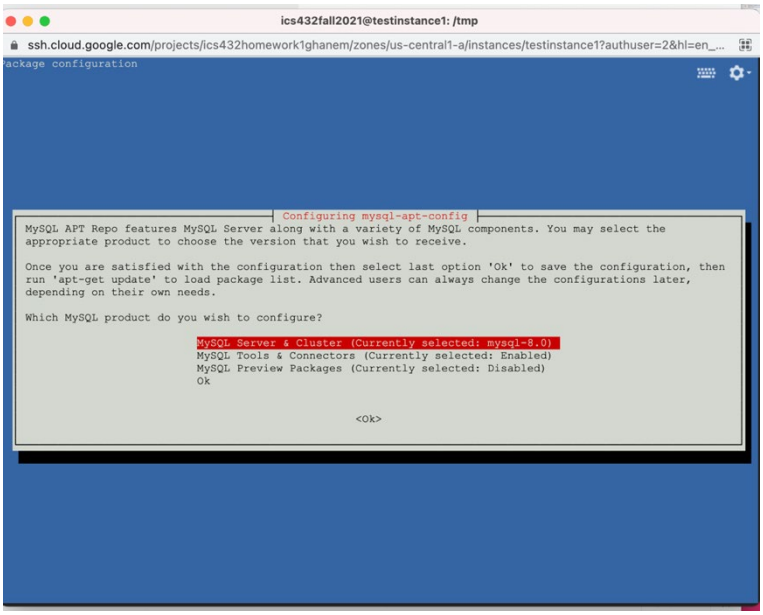
If the download is successful, `ls` should display mysql-apt-config_0.8.13-1_all.deb as follows

Homework report screenshot 17: take a screenshot to show the command window after running `ls`. Make sure the command prompts (the includes your username and instance name) are clear and readable in the screen shot.

e. Set up MySQL server by running the following command.

```
sudo dpkg -i mysql-apt-config*
```

f. In the following screen, use the down arrow to navigate to the `Ok` menu option and hit `ENTER`.



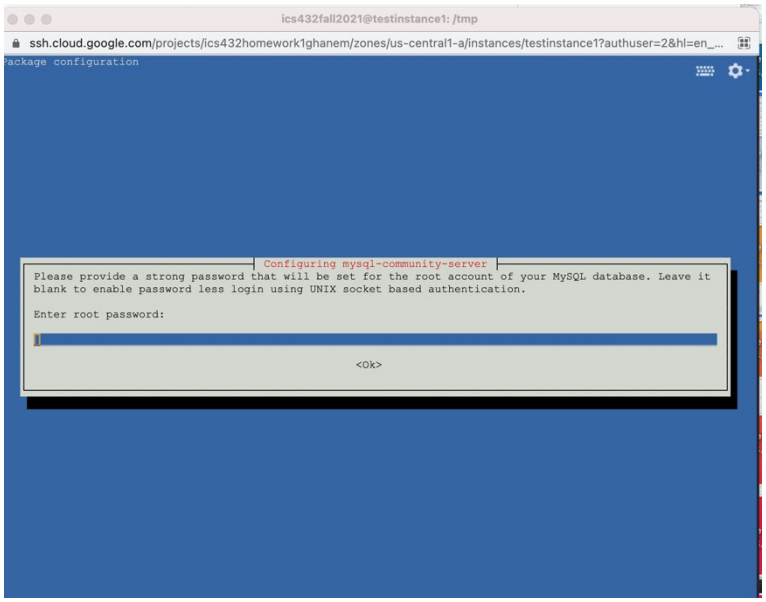
g. Refresh your `apt` package cache to make the new software packages available:

```
sudo apt update
```

h. Having added the repository and with our package cache freshly updated, we can now use `apt` to install the latest MySQL server package. When prompted, enter Y.

```
sudo apt install mysql-server
```

i. You will be asked to set a **root** password during the configuration phase of the installation. Although you can enter any password, I suggest that you enter the simple password root. Click ENTER. You will be asked to enter the password again. Enter the same password root.

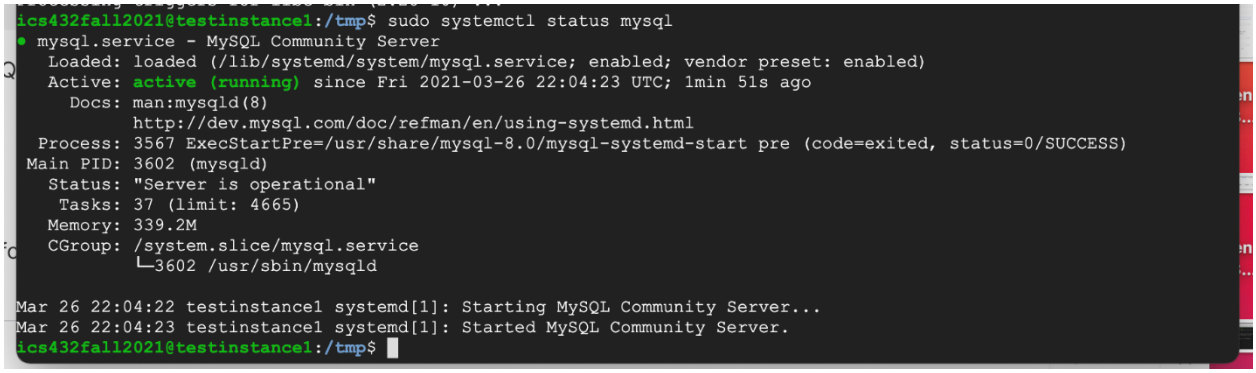


j. Hit Enter to accept the default Encryption method.
k. Wait until the installation is complete.

l. MySQL should be running now. Let's check it using the following command:

```
sudo systemctl status mysql
```

You should see a screen similar to the following screen with the Active: active (running) line means MySQL is installed and running.



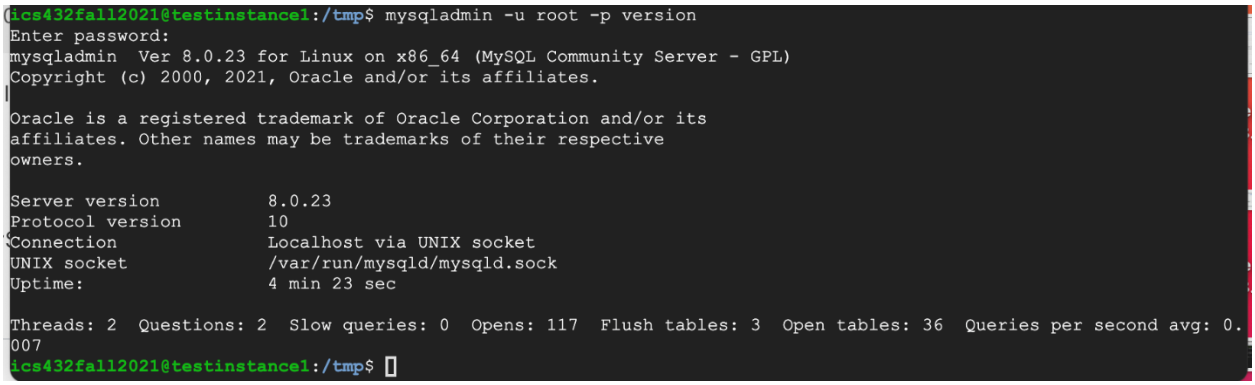
Homework report screenshot 18: take a screenshot to show the command window with the status of MySQL. Make sure the command prompts (the includes your username and instance name) are clear in the screen shot.

m. Test the MySQL installation using the mysqladmin command as follows. This command will connect to the server and output some version and status information:

```
mysqladmin -u root -p version
```

The `-u` root portion tells mysqladmin to log in as the MySQL root user, `-p` instructs the client to ask for a password, and version is the actual command we want to run.

n. When prompted for the password, enter root. You should receive a screen similar to the following screen.



o. Connect to MySQL server using the following command. Enter password root.

```
mysql -u root -p
```

You should see the `mysql>` prompt and this can be used to start working with the database management system.

Homework report screenshot 18: take a screenshot to show the command window after connecting to MySQL. Make sure the command prompts (the includes your user name and instance name) are clear in the screen shot.

Step 3: Create a database and fill tables with data

1- If you stopped your VM for any reason, when you ready to continue working on your project, restart the VM and connect to it using SSH.

2- Start MySQL using the following command. The --local-infile=1 is needed to be able to load tables from local data files.

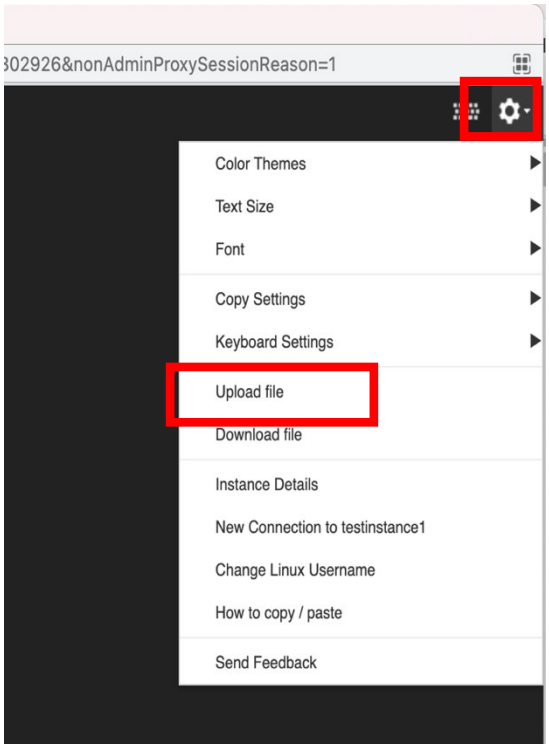
```
mysql --local-infile=1 -u root -p
```

3- Create a new database

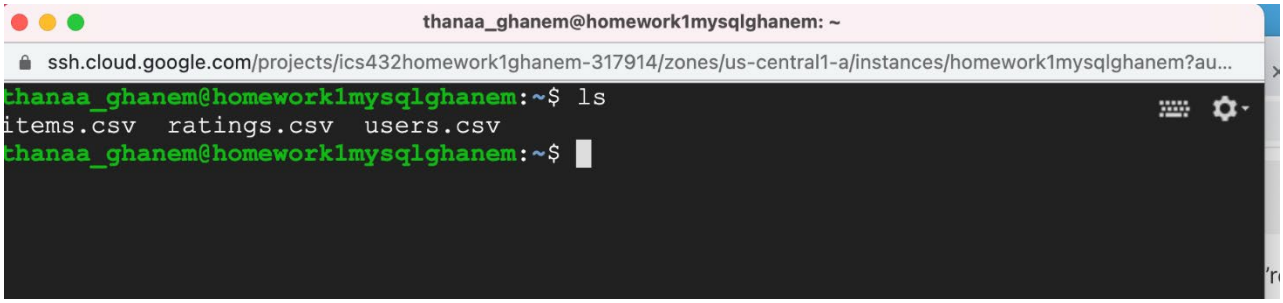
```
CREATE DATABASE movie_ratings_database;
```

4- You are given three data files (users.csv, ratings.csv, and items.csv). You need first to copy the data files to the VM and then you will upload these files to MySQL tables.

5- To copy files to the instance, from the SSH window, click on the icon at the upper right corner as shown in the following figure and then click in 'Upload Files'. Browse to the location where you have the data files and choose the files to be uploaded.



6- In the SSH window, if you still under the `mysql>` prompt, you can exit mysql by typing `exit`. In the SSH window, if you are still under `/tmp` directory, type `cd ~/` to go back to the home directory. Once you are in the home directory, type `ls` to see the list of files.



Homework report screenshot 19: take a screenshot to show the command window with the list of data files. Make sure the command prompts (the includes your user name and instance name) are clear in the screen shot.

7- Use the following command to log back to MYSQL.

```
mysql --local-infile=1 -u root -p
```

8- Create tables on MySQL using SQL commands. For example, the following command is used to create the 'location' table.

```
CREATE TABLE users(
     userid int primary key,
     age int,
     gender varchar(5),
     occupation varchar(100),
     zipcode varchar(10));
```

9- Load data from a local file to a MySQL table using the load command as follows.

```
LOAD DATA LOCAL INFILE 'users.csv'
INTO TABLE users
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n';
```

10- If you get error that the local file cannot be uploaded, you need to run the following command. (https://stackoverflow.com/questions/59993844/error-loading-local-data-is-disabled-this-must-be-enabled-on-both-the-client).

```
SET GLOBAL local_infile=1;
```

Run the load data command again.

11- Run the following SQL query to make sure the data is successfully loaded to the table.

```
select * from users;
```

Homework report screenshot 20: take a screenshot to show the command window with the output (or part of) of the select query. Make sure the command prompts (the includes your username and instance name) are clear in the screen shot.

Homework report: how many rows are there in the users table.

12- The data to be used in this assignment is provided to you in a zipped file called assignment-01-movieratingdata.zip. When you unzip the file, you will get the following 3 files (users.csv,items.csv, and ratings.csv). Open each file and familiarize yourself with the data format. The data in these files is interpreted as follows:

users – primary key (user id)
user id | age | gender | occupation | zip code
items – primary key (item id)
movie id | movie title | release date | video release date|
IMDb URL | unknown | Action | Adventure | Animation |

Children | Comedy | Crime | Documentary | Drama |Fantasy|
FilmNoir | Horror | Musical | Mystery | Romance | SciFi |
Thriller | War | Western |
ratings – primary key (user id, item id), user id references users (user id) and item id references (items item id)
user id (integer) | movie id (integer) | rating (integer) | timestamp (varchar (100))

13- Create two tables (called `ratings` and `items`) and load them with data from the csv files.

Homework report: Include SQL queries that you used to create and load the tables.

Homework report screenshot 21: run a SQL query to display rows from the `ratings` table and take a screen shot of the command window with the output (or part of) of the select query. Make sure the command prompts (the includes your user name and instance name) are clear in the screen shot.

Homework report: how many rows are there in the items table.

Homework report screenshot 22: run a SQL query to display rows from the `items` table and take a screen shot of the command window with the output (or part of) of the select query. Make sure the command prompts (the includes your user name and instance name) are clear in the screen shot.

Homework report: how many rows are there in the ratings table.

Step 4: Movie Rating Data Analysis using SQL

Write SQL queries to find answers to the following questions.

Homework report and screenshots 23-33: for each one of the following questions, include the SQL query and a screen shot to show the output for each of the following questions.

1- Find how many ratings are there for each of the following values (5,4,3,2,and 1). Draw a histogram to show these values

2- Find how many distinct zip codes are there.
Note that the zipcode field is defined as string with length 100 so you may encounter issues while working with this attribute in queries. One way to resolve these issues is to use `replace(zipcode,'\r','')` instead of `zipcode` in your queries. Or you can write an `update` query to change the values of zip codes to `replace(zipcode,'\r','')`.

3- Find how many users are there from each zip code.

4- What is the age and occupation of the user who gave the maximum number of reviews? (you can use more than one query).

5- What are the top five zip codes in terms of number of users? How many users are there in each one of these zip codes.

6- Find how many different occupations are there.

7- What are the five top occupations in terms of the number of users? How many users are there in each one of these top five occupations?

8- What are the top five movie ids in terms of the number of ratings.

9- Find the top ten movie ids in terms of average rating? What are the titles for these movies. You can use more than one query.

Step 5: Stop the VM instance

To avoid wasting your credits, make sure to stop the VM instance after you complete the assignment.

Do not delete the GCP project until I grade this homework.

After you receive your grade for this assignment, you can terminate the project.

Step 6: Summarize your learning

Write ~100-200 words to discuss what you learned from this exercise. Explain the most important things you learned? What are the challenges that you faced?