ADTA 5900/5550.701: Deep Learning with Big Data

Thuan L Nguyen, PhD

Assignment 1

1. Overview

The rise of cloud computing has been a facilitator to the emergence of big data. Cloud computing is the commodification of computing time and data storage using standardized technologies.

Big data is a term to describe large volumes of data that can be both structured and unstructured. These enormous volumes of data overwhelm the digital world every second. However, it is not the amount of data that is important. It is what we can do with the data that matters: Big data analytics can provide insights that lead to better decisions and strategic moves.

It was the emergence of cloud computing which made it easier to provide the best of technology in the most cost-effective packages. Cloud computing has not only reduced costs, but also made a wide array of applications available to the companies of all sizes: small, mid-sized, big, and giant corporations.

2. Google Cloud Platform (GCP)

Google Cloud Platform (GCP) provides on-demand cloud computing platforms to individuals, companies and governments.

The technology allows subscribers to have a full-fledged virtual cluster of computers, available all the time, through the Internet. Google Cloud Platform (GCP) virtual machines (VM) have most of the attributes of a real computer including hardware (CPU(s) & GPU(s) for processing, local/RAM memory, hard-disk/SSD storage); a choice of operating systems; networking; and pre-loaded application software.

Each GCP virtual machine also virtualizes its console I/O (keyboard, display, and mouse), allowing the user to connect to their remote instances using a modern browser. The browser acts as a window into the virtual computer, letting subscribers log-in, configure and use their virtual systems just as they would a real physical computer.

IMPORTANT NOTES:
--) All the documents posted in the Canvas page GOOGLE CLOUD PLATFORM: GCP for Deep Learning should be used for HW 1.

## 3. PART I: Set Up Deep Learning Virtual Machine (VM) in GCP (40 Points)

TO-DO
--) Based on the lectures, set up a deep learning virtual machine (VM) in Google Cloud Platform (GCP).
--) Based on the lectures, install the GCLOUD SDK tool in the student's local computer.

SUBMISSION REQUIREMENT #1.1:
--) Write a brief report to summarize the major steps of setting up the remote server.
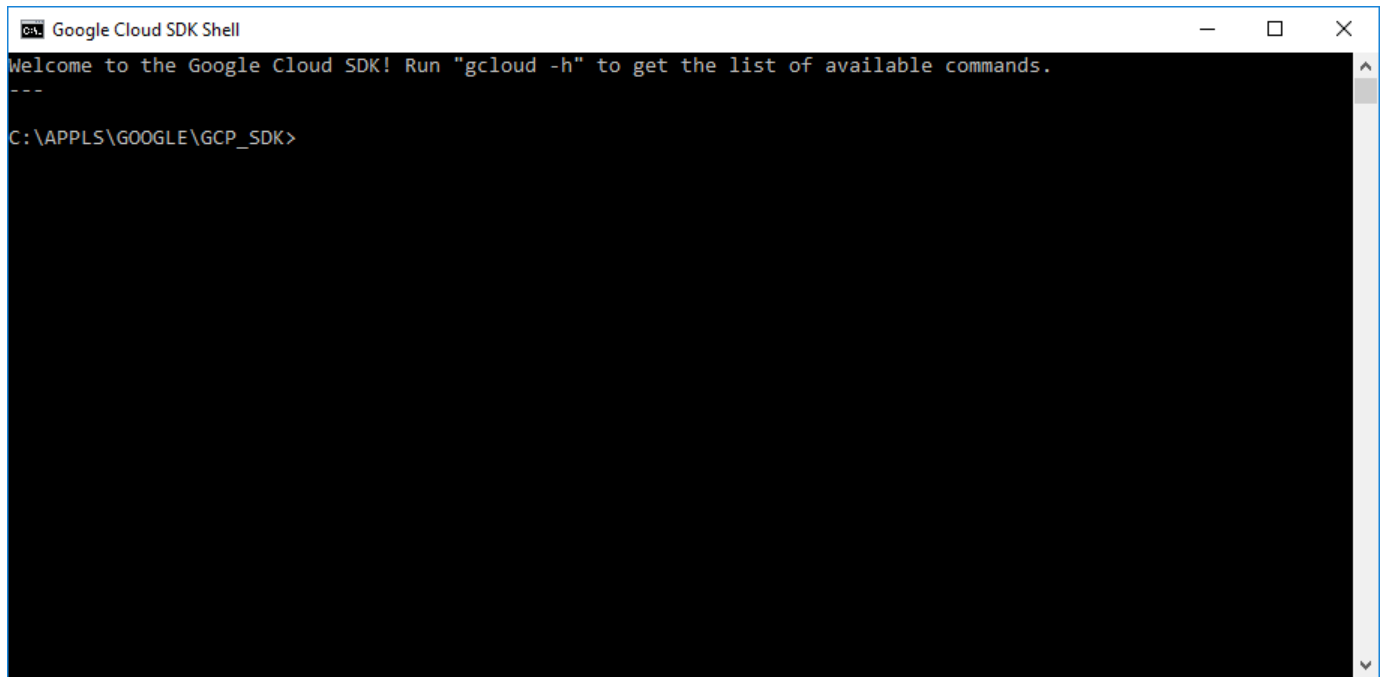--) Capture the screenshot that shows the critical information of the newly created remote instance.

Here is one example of the screenshot:
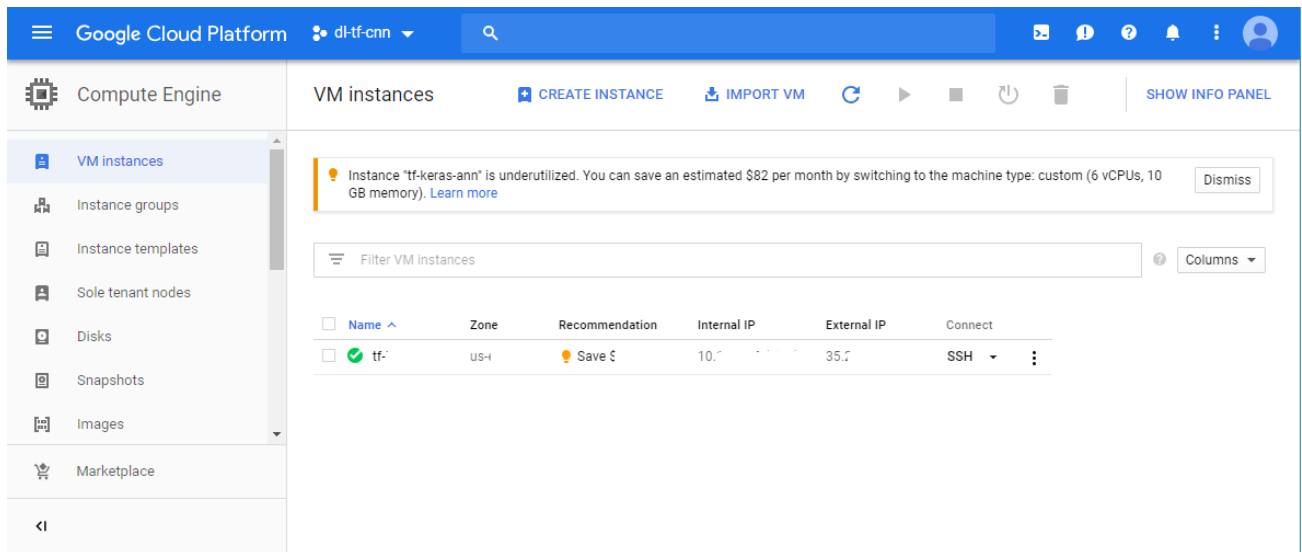
SUBMISSION REQUIREMENT #1.2:
--) Write a brief report to summarize the major steps of installing the GCLOUD SDK.
--) Capture the screenshot that shows the tool has been successfully installed.

Here is one example of the screenshot:

```
Google Cloud SDK Shell                                          —    □    ×
Welcome to the Google Cloud SDK! Run "gcloud -h" to get the list of available commands.
---

C:\APPLS\GOOGLE\GCP_SDK>
```

4. PART II: Connect Explore Remote VM Using SSH (20 Points)

TO-DO
--) Question 2.1:
Based on the lectures, open an SSH connection from the local computer to the remote VM.

--) Question 2.2
Using the basic Linux command lines to explore the contents of the home directory

--) Question 2.3
Create a new sub-folder named "JPTR_NTBK" under the home directory

--) Question 2.4
Change the current directory to the newly created folder

SUBMISSION REQUIREMENT #2:
--) Write a brief report to describe the major activities the student has finished in PART III
--) For each question, capture the screenshot that shows what he/she has done

IMPORTANT NOTES:
--) The answer should be clearly labeled with which question that the student is working on.

5. PART III: Start and Connect to Jupyter Notebook in Remote VM (20 Points)

TO-DO
--) Question 3.1:
Based on the lectures, start the Jupyter Notebook server in the remote virtual machine

--) Question 3.2:
Connect to the Jupyter Notebook server in the remote virtual machine

--) Question 3.3:
Run Jupyter Notebook in a local browser

SUBMISSION REQUIREMENT #3:
--) Write a brief report to explain the steps the student has done to start, connect, and use Jupyter Notebook that runs in the remote virtual machine.
--) For each question, capture the screenshot that shows what he/she has done

IMPORTANT NOTES:
--) The answer should be clearly labeled with which question that the student is working on.

6. PART IV: Write Simple Python Code in Jupyter Notebook in Remote VM (20 Points)

TO-DO
--) Write Python code to provide the solutions to the following simple problems **using Python Numpy library**:

1. Create a vector (1D array) of size 20. All the elements are initialized with 0 (zero) except for the 8th element that is set with the value 8.

2. Create a vector of size 16 with random values ranging from 0 to 63, print the vector, then sort it and print the vector again.

3. Create a 5x5 matrix with values ranging from 0 to 24.

4. Create an 8x8 array with random values, then find the min and max values stored in this matrix.

5. Create a vector of size 32 that is initialized with random values inside the range (0, 99) and then find the mean of all the initial values.

IMPORTANT NOTES:
--) Write the code of each problem in a separate Jupyter Notebook cell.

SUBMISSION REQUIREMENT #4:
--) Run the code of each problem
--) For each question, capture the screenshots that show the results of running the code for each problem.

7. HOWTO Submit

The student is required to submit all the sections, i.e., submission requirements, in a Microsoft Words document that is sent to the instructor (Thuan.Nguyen@unt.edu) as an attachment to a UNT email.

The subject of the email must be: "ADTA 5550: Assignment 1 – Submission."

Due date & time: 11:00 PM – Wednesday 06/10/2020