

# ADTA 5550.400: 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 that 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 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: Confirm UNT Student Email Address (10 Points)

Before answering the following questions, the student is required to read the document [VERIFY\\_UNT\\_email\\_address.pdf](#) posted in the Canvas module: [.../WEEK\\_1](#).

The instructor sent a Welcome message using the UNT email to all the students in the class.

The student is **required** to [access](#) his/her UNT student email ([\\*.my.unt.edu](#)) and take a screenshot or print out the email to PDF and upload it in the "Welcome email upload link" in week one assignment section.

#### **IMPORTANT NOTES:**

*If a student is an employee of either UNT, UNT Systems, or an organization belonging to UNT Systems, and he/she prefers to use his/her work email address – instead of ...@my.unt.edu, the student can email the instructor and let him know about that.*

## 4. PART II: Select an Operating System (10 Points)

During the course, the student uses cloud technologies in classwork. The student needs to use his/her local computing devices to access the remote deep learning server set up in cloud. For this purpose, it is required that the student should know how the operating system installed in his/her local computers work and can use it effectively.

### TO-DO

--> **Select an operating system (OS), either Windows, MAC, or Linux, that the student knows it well and can use it effectively.**

--> For the **selected operating system**, it is **required** that:

- The student should have fundamental knowledge about the operating system and its major, important components such as **drives, directories and folders, files, the ownership status of these components, the administrative privileges**, to name a few.
  - Fundamental knowledge includes, but not limited to:
    - How to **find where** a file or media contents are saved in the computing device
    - How to **get the real, physical path** of a file or a directory created or saved in the local storage of the computing device?
    - How to **access real, physical media contents** saved in the computing device
    - ...
- The student should be able to run basic operations successfully on any computing devices in which the selected OS is installed.
  - Basic operations include, but not limited to:
    - **Creating new directory and folders**
    - **Accessing real, physical directories and folders and their contents**
    - Creating and saving new contents in the folders and can **access their real, physical files**, not only the alias names or links as in the MAC OS.
    - **Download** software applications and **save** them in a specified folder
    - **Install** software applications in a specified **real, physical** directory or folder
    - **Open terminals in the windows environment of the OS**
    - **Run command lines** in terminals **in the windows environment of the OS**
    - ...
- The student should be **able to resolve issues** related to the **OS and the host computing device** if they happen while he/she uses the selected OS and its host computing device.
  - The student should **know how to do research**, e.g., using Google search, and **get technical support** such as contacting the vendors of the devices and the maker of the OS, and other supporting facilities like public technical forums, ...

### SUBMISSION REQUIREMENT #1:

--> Provide a **short paragraph** to **specify which operating system has been selected** and to **state that the student can use it effectively**.

## 5. PART III: Set Up Deep Learning Virtual Machine (VM) in GCP (20 Points)

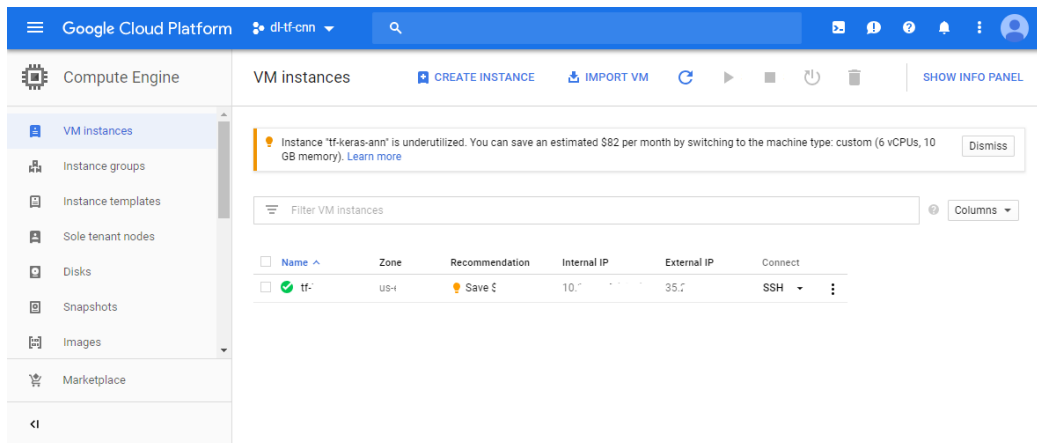
### TO-DO

- > 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 #2.1:

- > 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.

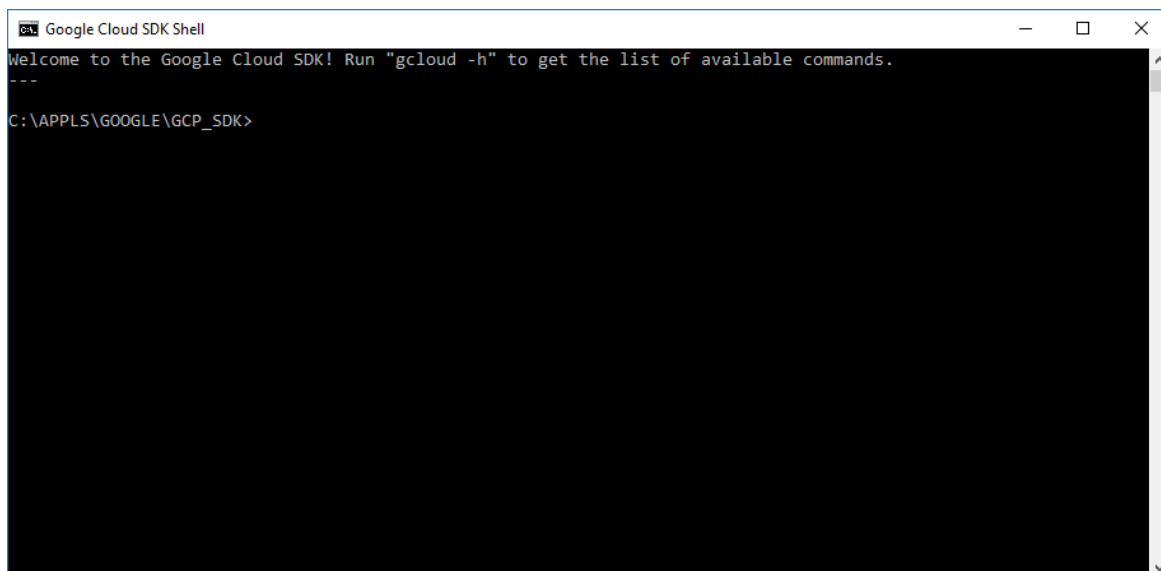
Here is one example of the screenshot:



### SUBMISSION REQUIREMENT #2.2:

- > 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:



## 6. PART IV: Connect and Explore Remote VM Using SSH (10 Points)

### IMPORTANT NOTES:

--> It is expected that the student should learn the steps discussed in the lecture document at the link [“Connect to Remote Virtual Machine in Google Cloud Platform \(GCP\) Using SSH”](#) before working on the following questions.

### TO-DO

--> Question 3.1:

Based on the lectures, open an SSH connection from the local computer to the remote VM.

--> Question 3.2

Using the basic Linux command lines to explore the contents of the home directory.

--> Question 3.3

Create a new sub-folder named “JPTR\_NTBK” under the home directory

--> Question 3.4

Change the current directory to the newly created folder

### SUBMISSION REQUIREMENT #3:

--> Write a brief report to describe the major activities the student has finished in PART IV

--> 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.

## 7. PART V: Start and Connect to Jupyter Notebook in Remote VM (30 Points)

### TO-DO

--> Question 4.1:

Based on the lectures, **start** the Jupyter Notebook server [in the remote virtual machine](#).

--> Question 4.2:

[Connect](#) to the [Jupyter Notebook](#) server in the [remote virtual machine](#) (by connecting a Local Computer Port, i.e., 8000, to the Remote Server Port, i.e., 8888)

--> Question 4.3:

[Use](#) Jupyter Notebook that is [currently running in the Remote Server](#) (in a browser on the local computer)

### SUBMISSION REQUIREMENT #4:

--> 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.*  
--) ***Question 5.3:** The question does not ask to **start Jupyter Notebook on the local computer**. For example, if the student has installed an Anaconda package on his/her local computer. He/she can start Jupyter Notebook on his/her local computer. The question does not ask to do that.*

## **8. PART VI: 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 the 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 #5:**

- ) Run the code of each problem  
--) For each question, capture the screenshots that show the results of running the code for each problem.

## **9. HOWTO Submit**

The student is required to submit all the sections, i.e., submission requirements, in a Microsoft Word document and upload it to Canvas. No emailed homework, assignment, or exams will be accepted

**Due date & time: 11:59 PM – Thursday 06/13/2024**