ADTA 5900/5750.501: Applied Natural Language Processing

Thuan L Nguyen, PhD

Assignment 1

1. Overview

The rise of cloud computing has been a facilitator of 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): Service Account

Various enterprise services of Google Cloud Platform (GCP) will be used in classwork. When using cloud services to run applications, cyber-security is one of the top-priority concerns. In order to use the services to create and run Natural Language Processing (NLP) applications in GCP, the user is required to set up a service account that can be used in the GCP's sophisticated authentication system.

IMPORTANT NOTES:
*--) All the documents posted on the Canvas page **GOOGLE CLOUD PLATFORM: GCP for Deep Learning – TF2** should be used for HW 1.*

*--) All the documents posted on the Canvas page **GOOGLE CLOUD PLATFORM: GCP for Natural Language Processing (NLP)** should be used for HW 1.*

3. PART I: Select an Operating System (5 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 the cloud. For this purpose, it is required that the student should know how the operating system installed in his/her local computers works 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**, **and administrative privileges**, to name a few.
    - Fundamental knowledge including 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 including but not limited to:
        - **Creating new directories and folders**
        - **Accessing real, physical directories and folders and their contents**
        - Creating and saving new contents in the folders and **accessing 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 PART I #1:

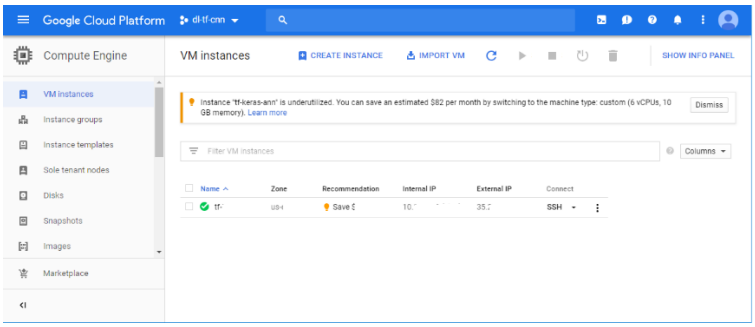
--) Provide a **short paragraph** to **specify** which operating system has been selected and to **state** that the student can use it effectively.

4. PART II: 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 on the student's local computer.

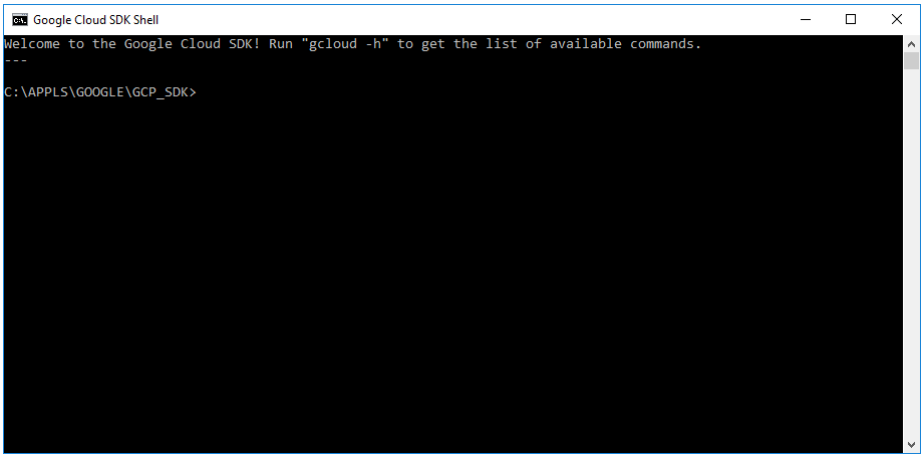
SUBMISSION REQUIREMENT PART II #1:
--) Write a brief report to summarize the major steps of setting up the remote server.
--) Capture the screenshot that shows the critical information of the newly created remote instance.

Here is one example of the screenshot:



SUBMISSION REQUIREMENT PART II #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:



IMPORTANT NOTES:
*--) **BE SURE**: Remote deep learning server has the VM image with **TensorFlow Enterprise 2.7** or **later**..*

5. PART III: Connect and Explore Remote VM Using SSH (5 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 PART III #1:
--) 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 the student is working on.

6. PART IV: Start and Connect to Jupyter Notebook in Remote VM (20 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 PART IV #1:
--) Write a brief report to explain the steps the student has taken 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

7. PART V: Create Service Account for GCP Project (20 Points)

TO-DO

- Create a service account for the project used in classwork, following the steps discussed in the documents posted on the Canvas page (***GOOGLE CLOUD PLATFORM: GCP for Natural Language Processing (NLP)***)

SUBMISSION REQUIREMENT PART V #1:
--) Write a brief report to explain the steps the student has taken to create the service account.
--) For each step, capture the screenshot that shows what he/she has done

IMPORTANT NOTES:
--) The screenshots must include one that shows the service has been created successfully.

8. PART VI: Set Up Natural Language API for GCP Project (20 Points)

TO-DO

- Set up the Natural Language API for the project, following the steps discussed in the documents posted on the Canvas page (***GOOGLE CLOUD PLATFORM: GCP for Natural Language Processing (NLP)***)

SUBMISSION REQUIREMENT PART VI #1:
--) Write a brief report to explain the steps the student has taken to set up the service account.
--) For each step, capture the screenshot that shows what he/she has done

IMPORTANT NOTES:
--) The screenshots must include one that shows the API has been set up successfully.

9. PART VII: Join a Group (10 Points)

TO-DO

- Discuss with classmates to join a group.

SUBMISSION REQUIREMENT PART VII #1:
--) Provide the list of group members of the group that you have joined to work on class assignments.

10. HOWTO Submit

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

The subject of the email must be:

- "ADTA 5750: Assignment 1 – Submission."

Due date & time: 11:00 PM – Wednesday 09/11/2024