PRACTICAL NO:1

AIM: To Install Cloudera Quickstart VM on VirtualBox

Theory:
Cloudera is a software that provides a platform for data analytics, data warehousing, and machine learning. Initially, Cloudera started as an open-source Apache Hadoop distribution project, commonly known as Cloudera Distribution for Hadoop or CDH. It contains Apache Hadoop and other related projects where all the components are 100% open-source under Apache License.

Cloudera provides virtual machine images of complete Apache Hadoop clusters, making it easy to get started with Cloudera CDH. The following topics will be covered in this assignment on Cloudera QuickStart VM Installation.
1. What is Cloudera QuickStart VM?
2. Cloudera QuickStart VM Installation - Prerequisites
3. Downloading the Cloudera QuickStart VM
4. Cloudera QuickStart VM Installation on windows

What Is Cloudera QuickStart VM?
Cloudera QuickStart VM includes everything that you would need for using CDH, Impala, Cloudera Search, and Cloudera Manager. The Cloudera QuickStart VM uses a package-based install that allows you to work with or without the Cloudera Manager. It has a sample of Cloudera's platform for "Big Data."

Cloudera QuickStart VM Installation - Prerequisites
A virtual machine such as Oracle Virtual Box or VMWare
RAM of 12+ GB. That is 4+ GB for the operating system and 8+ GB for Cloudera
80GB hard disk

Download Oracle Virtual Box from https://www.virtualbox.org/wiki/Downloads and install it in your system

Downloading the Cloudera QuickStart VM
- ✓ The Cloudera QuickStart VMs are openly available as Zip archives in VirtualBox, VMware and KVM formats. To download the VM, search for https://www.cloudera.com/downloads.html, and select the appropriate version of CDH that you require.
- ✓ Click on the 'GET IT NOW' button, and it will prompt you to fill in your details.
- ✓ Once the file is downloaded, go to the download folder and unzip these files. It can then be used to set up a single node Cloudera cluster.
- ✓ Shown below are the two virtual images of Cloudera QuickStart VM.
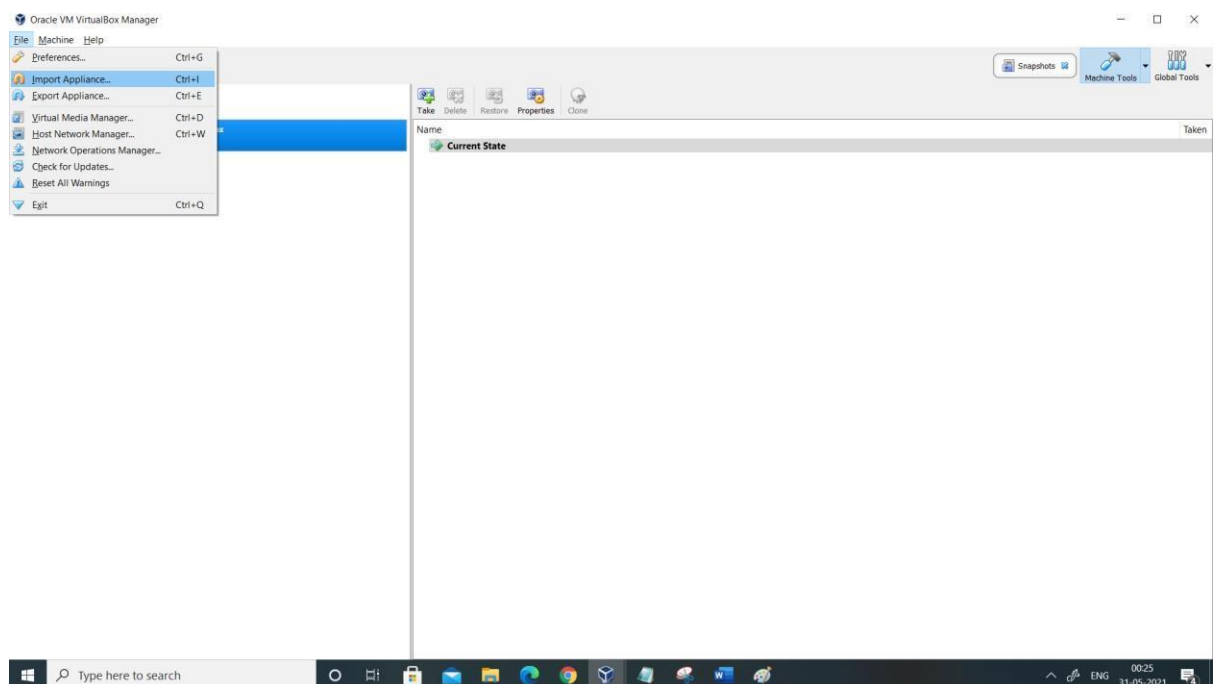
Acer (C:) > Python > SEM 2 Materials > BT > New folder > cloudera-quickstart-vm-5.13.0-0-virtualbox

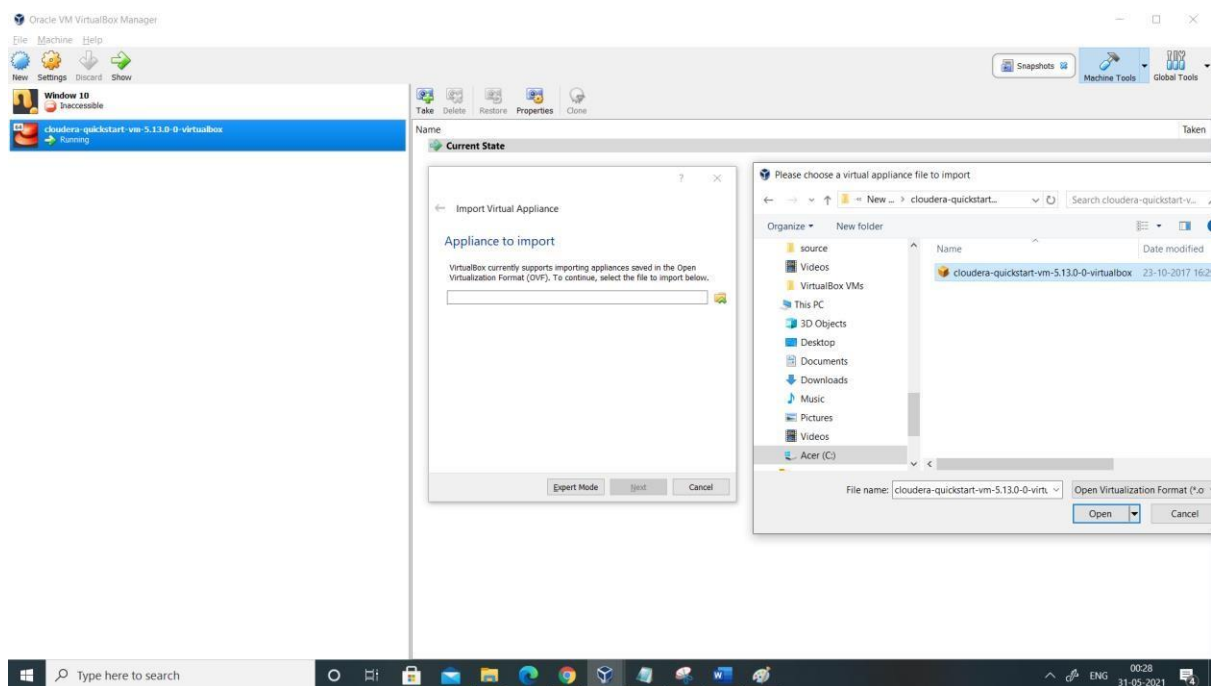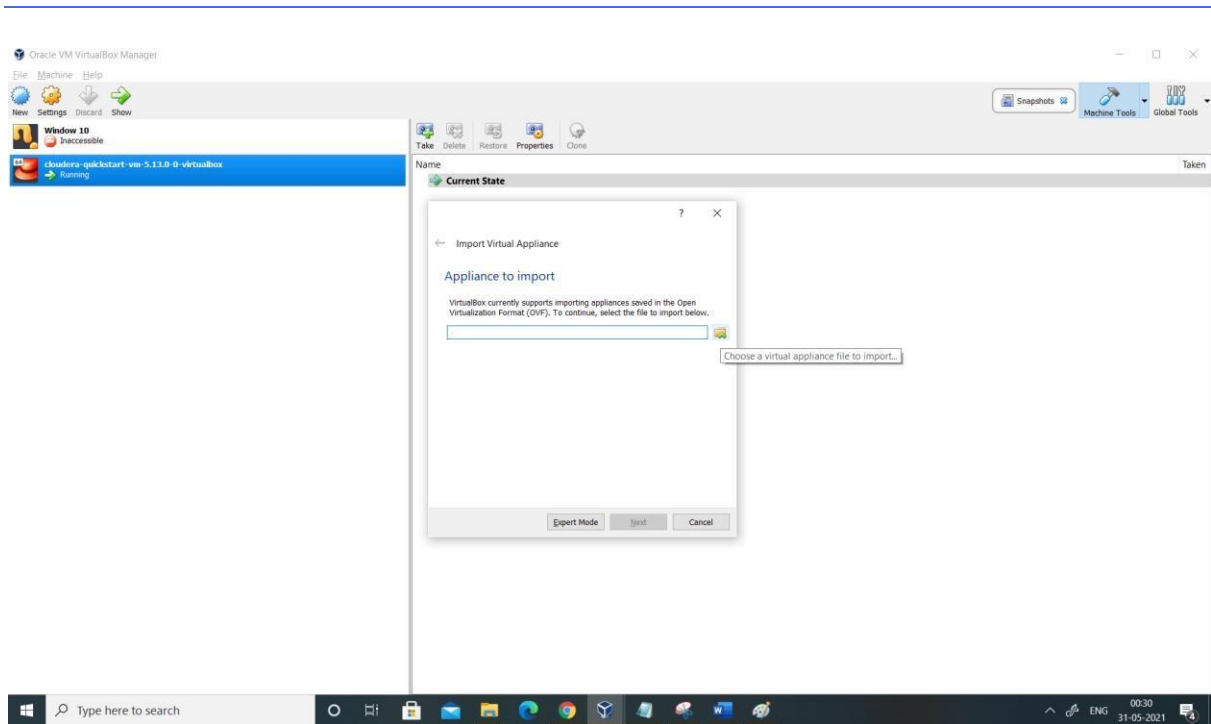| Name	Date modified	Type	Size
cloudera-quickstart-vm-5.13.0-0-virtualbox	23-10-2017 16:29	Open Virtualizatio...	15 KB
cloudera-quickstart-vm-5.13.0-0-virtualbox-disk1	23-10-2017 16:34	Virtual Machine Di...	58,10,349 ...

✓   Now that the downloading process is done with, let's move forward with this Cloudera QuickStart VM Installation guide and see the actual process.

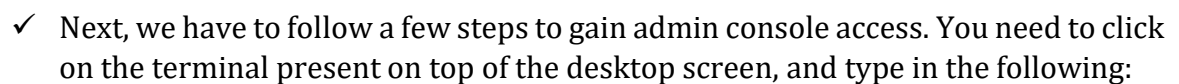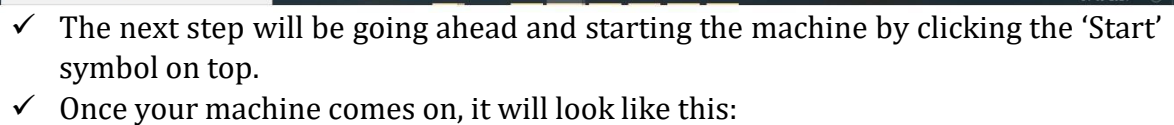Cloudera QuickStart VM Installation

- ✓ Before setting up the Cloudera Virtual Machine, you would need to have a virtual machine such as VMware or Oracle VirtualBox on your system.
- ✓ In this case, we are using Oracle VirtualBox to set up the Cloudera QuickStart VM.
- ✓ In order to download and install the Oracle VirtualBox on your operating system, click on the following link: Oracle VirtualBox(https://www.virtualbox.org/wiki/Downloads).
- ✓ To set up the Cloudera QuickStart VM in your Oracle VirtualBox Manager, click on 'File' and then select 'Import Appliance'.

- ✓ Choose the QuickStart VM image by looking into your downloads. Click on 'Open' and then 'Next'. Now you can see the specifications, then click on 'Import'. This will start importing the virtual disk image .vmdk file into your VM box.

- ✓ Click on "Open" and Wait for a while, as the importing finishes.
- ✓ The next step is to go ahead and set up a Cloudera QuickStart VM for practice. Once the importing is complete, you can see the Cloudera QuickStart VM on the left side panel.
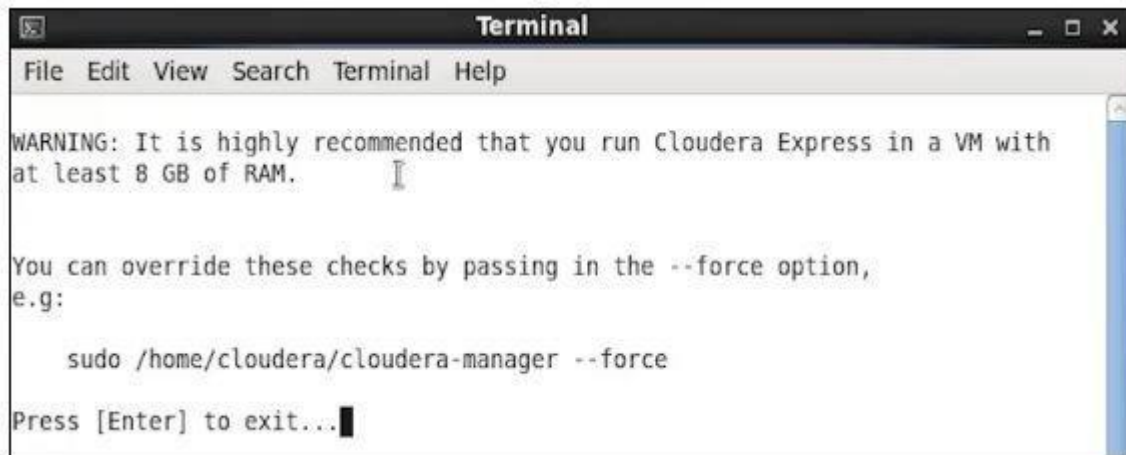
✓ The next step will be going ahead and starting the machine by clicking the 'Start' symbol on top.

✓ Once your machine comes on, it will look like this:

✓ Next, we have to follow a few steps to gain admin console access. You need to click on the terminal present on top of the desktop screen, and type in the following:

1. hostname # This shows the hostname which will be quickstart.cloudera
2. hdfs dfs -ls / # Checks if you have access and if your cluster is working. It displays what exists on   your HDFS location by default
3. service cloudera-scm-server status # Tells what command you have to type to use cloudera express free
4. su - #Login as root
5. service cloudera-scm-server status # The password for root is cloudera
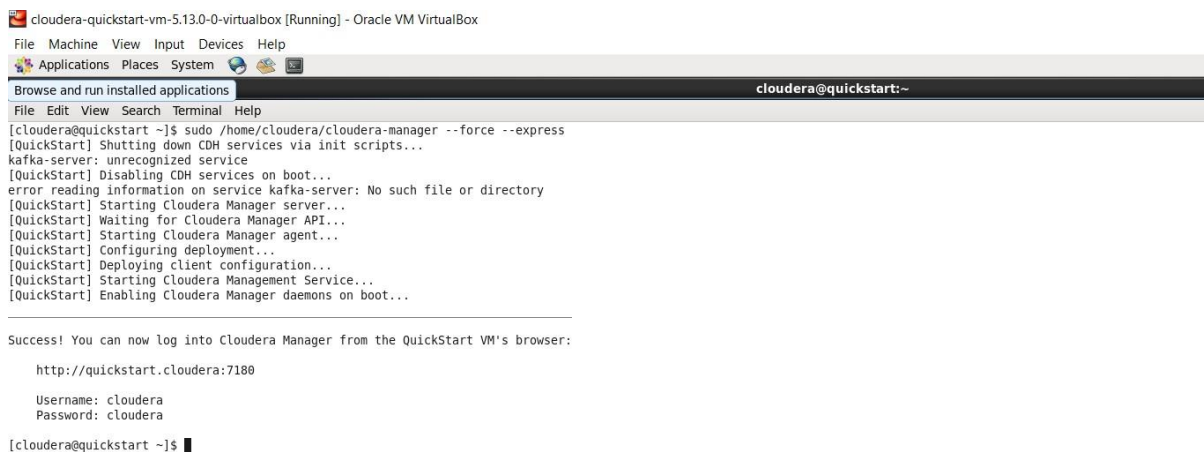
✓   Once you see that your HDFS access is working fine, you can close the terminal. Then, you have to click on the following icon that says 'Launch Cloudera Express'.
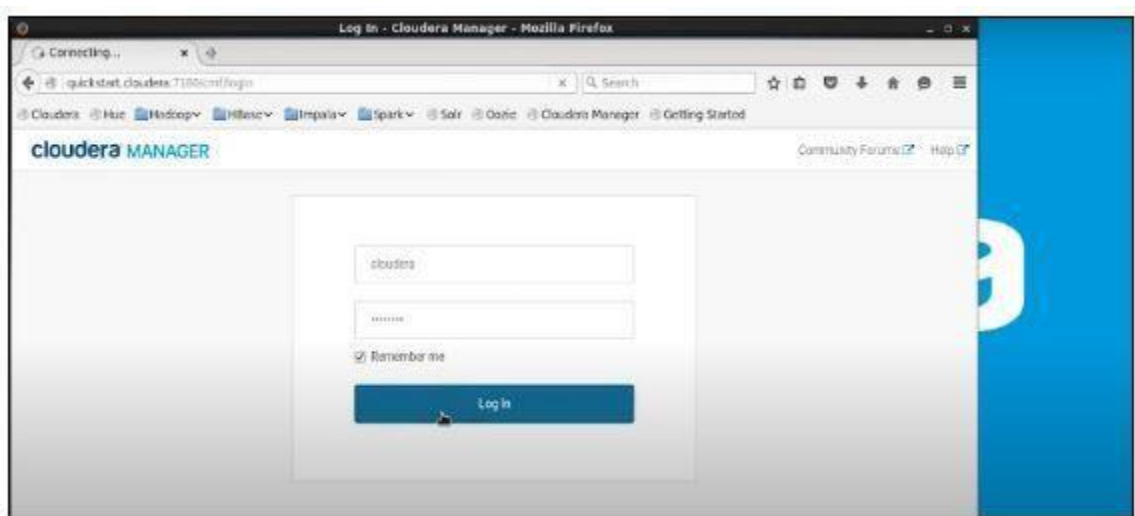

✓ Once you click on the express icon, a screen will appear with the following command:

✓ You are required to copy the command, and run it on a separate terminal. Hence, open a new terminal, and use the below command to close the Cloudera based services. It will restart the services, after which you can access your admin console.

✓ Now that our deployment has been configured, client configurations have also been deployed. Additionally, it has restarted the Cloudera Management Service, which gives access to the Cloudera QuickStart admin console with the help of a username and password.

✓ Go on and open up the browser and change the port number to 7180.
✓ You can log in to the Cloudera Manager by providing your username and password.

✓ You can go ahead and restart the services now. It will ensure that the cluster becomes accessible either by Hue as a web interface or Cloudera QuickStart Terminal, where you can write your commands.