

# ASSIGNMENT NO – 1

## ❖ To Install Cloudera QuickStart VM on VirtualBox:

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**
3. **Downloading the Cloudera QuickStart VM.**
4. **Cloudera QuickStart VM Installation on windows.**

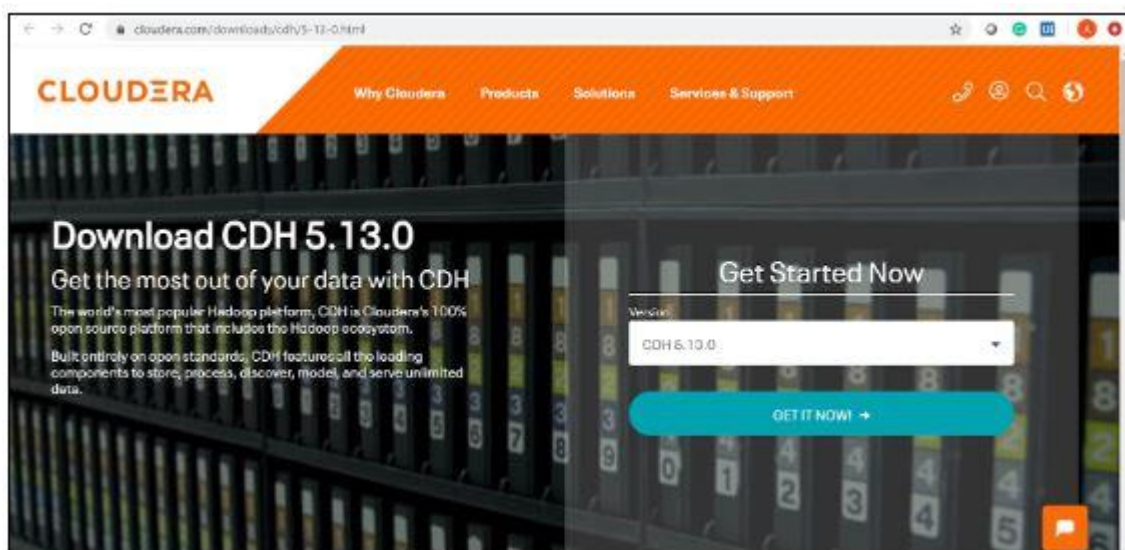
Cloudera QuickStart VM includes everything that you would need for using CDH, Impala, Cloudera Search, and Cloudera Manager. The Cloudera QuickStart VM uses a package-based install that allows you to work with or without the Cloudera Manager. It has a sample of Cloudera's platform for "Big Data."

### 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

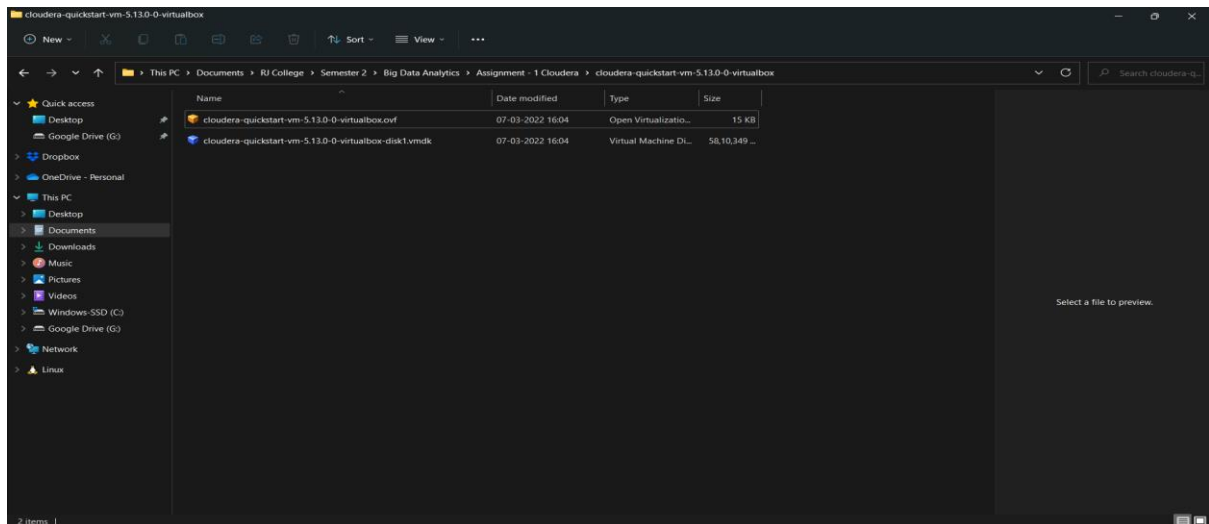
### Downloading the Cloudera QuickStart VM

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.

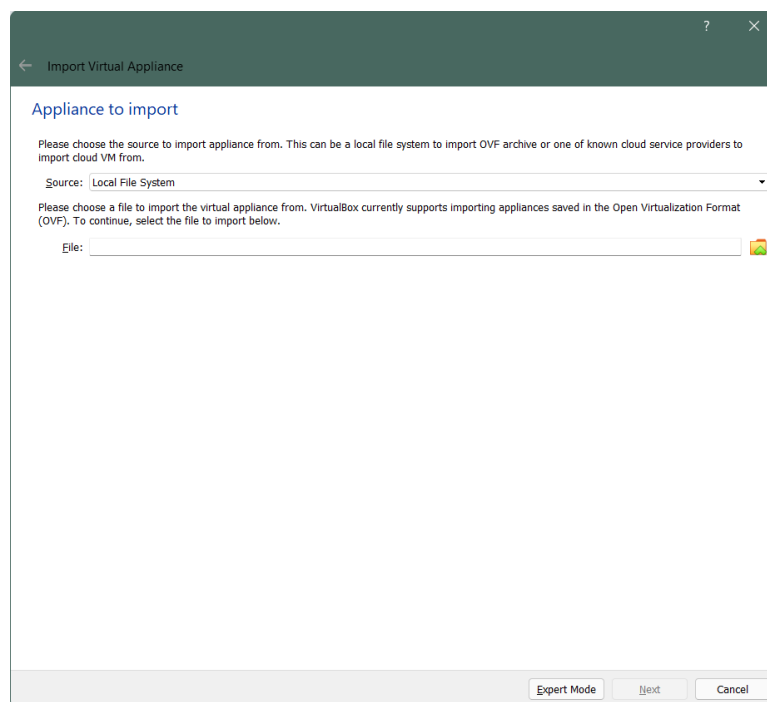
# ASSIGNMENT NO – 1



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

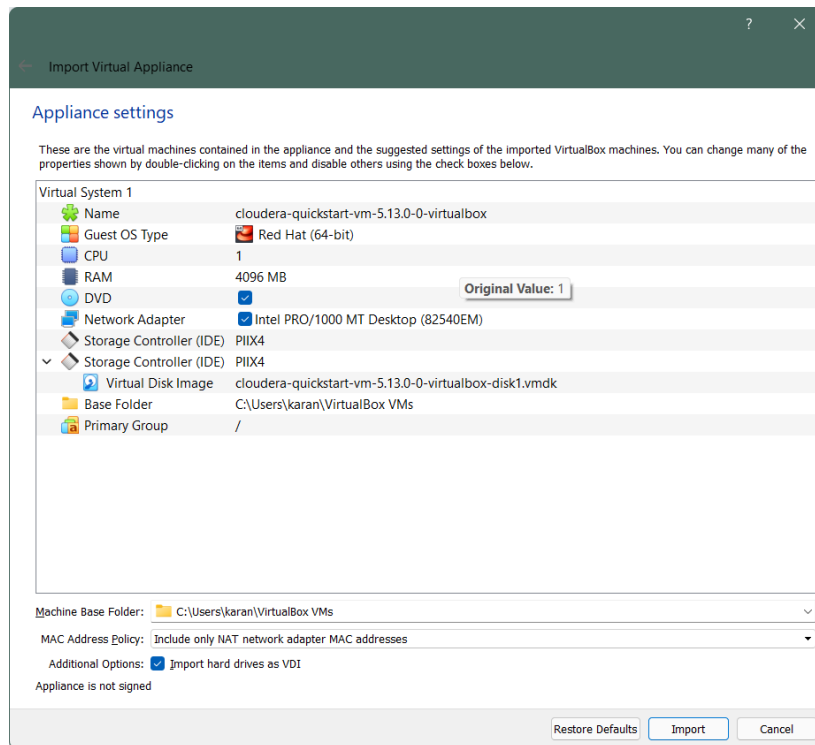
## Cloudera QuickStart VM Installation

- 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](#).
- To set up the Cloudera QuickStart VM in your Oracle VirtualBox Manager, click on 'File' and then select 'Import Appliance'.

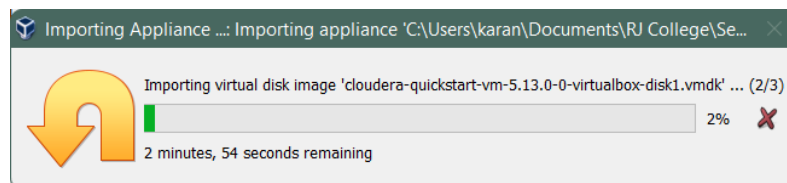


- Choose the QuickStart VM image by looking into your downloads. Click on 'Open' and then 'Next'. Now you can see the specifications, then click on 'Import'. This will start importing the virtual disk image .vmdk file into your VM box.

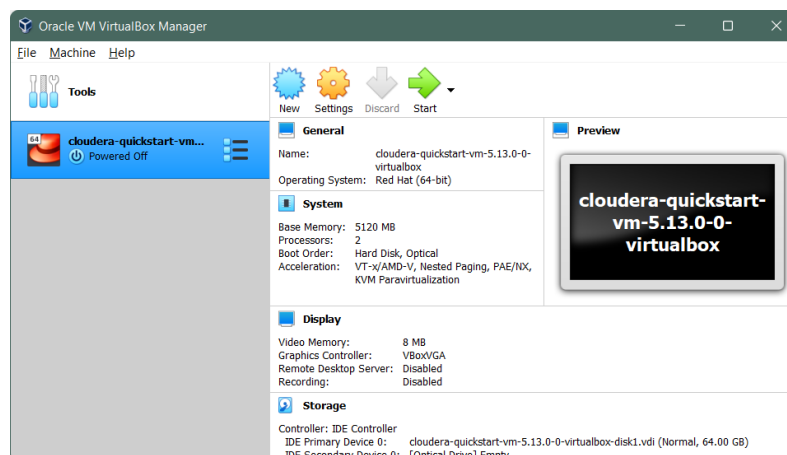
# ASSIGNMENT NO – 1



- Now, to give more RAM and CPU cores, click on 'Settings', followed by 'System', and increase the RAM to 5GB. Click on the processor and assign 2 CPU cores. Subsequently, select 'Network'. The Adapter 1 settings should be NAT by default. Click on 'OK' next.

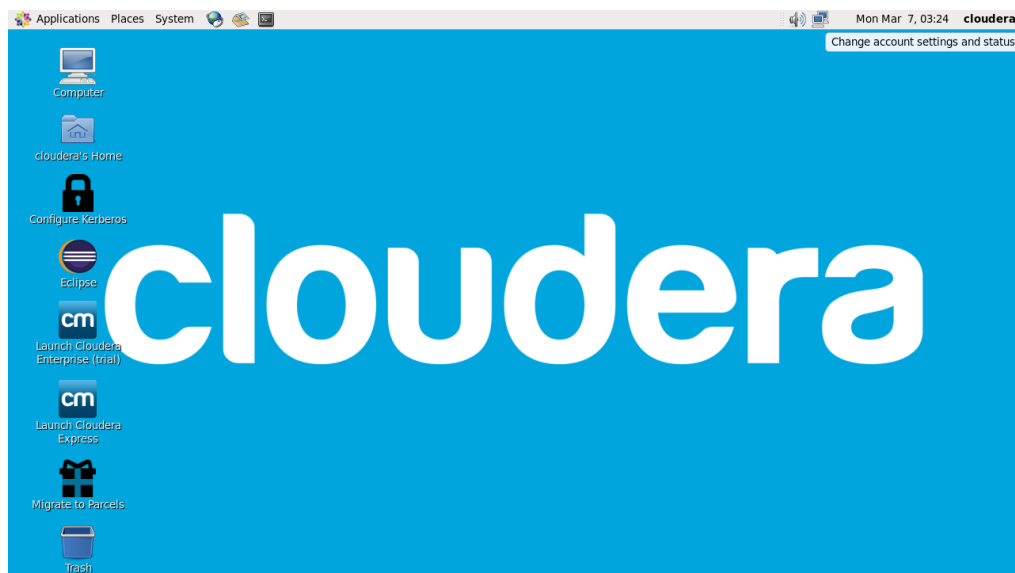


- Now you are required to start the machine, so that it uses 2 CPU cores, 5GB RAM, and brings up the Cloudera QuickStart VM.
- The next step will be going ahead and starting the machine by clicking the 'Start' symbol on top.



# ASSIGNMENT NO – 1

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



- Next, we must 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:
  - hostname** # This shows the hostname which will be quickstart.cloudera
  - hdfs dfs -ls /** # Checks if you have access and if your cluster is working. It displays what exists on your HDFS location by default
  - service cloudera-scm-server status** # Tells what command you must type to use cloudera express free
  - su -** # Login as root
  - service cloudera-scm-server status** # The password for root is cloudera

```
cloudera@quickstart:/home/cloudera
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hostname
quickstart.cloudera
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-03-07 03:15 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-03-07 03:15 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ service cloudera-scm-server status
By default, the Cloudera QuickStart VM run Cloudera's Distribution including
Apache Hadoop (CDH) under Linux's service and configuration management. If you
wish to migrate to Cloudera Manager, you must run one of the following
commands.

To use Cloudera Express (free), run:

    sudo /home/cloudera/cloudera-manager --express

This requires at least 8 GB of RAM and at least 2 virtual CPUs.

To begin a 60-day trial of Cloudera Enterprise with advanced management
features, run:

    sudo /home/cloudera/cloudera-manager --enterprise

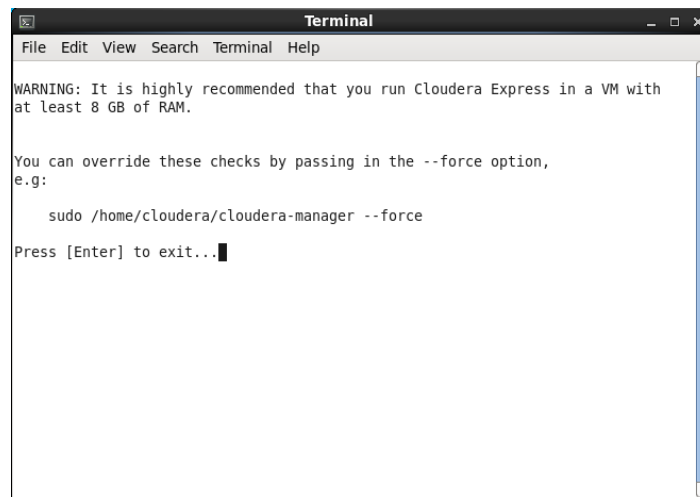
This requires at least 10 GB of RAM and at least 2 virtual CPUs.

Be aware that after rebooting, it may take several minutes before Cloudera
Manager has started all of the services it manages and is ready to accept
connections from clients.
[cloudera@quickstart ~]$ su
Password:
[root@quickstart cloudera]#
```

- Once you see that your HDFS access is working fine, you can close the terminal. Then, you must click on the following icon that says 'Launch Cloudera Express'.

# ASSIGNMENT NO – 1

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



```
Terminal
File Edit View Search Terminal Help

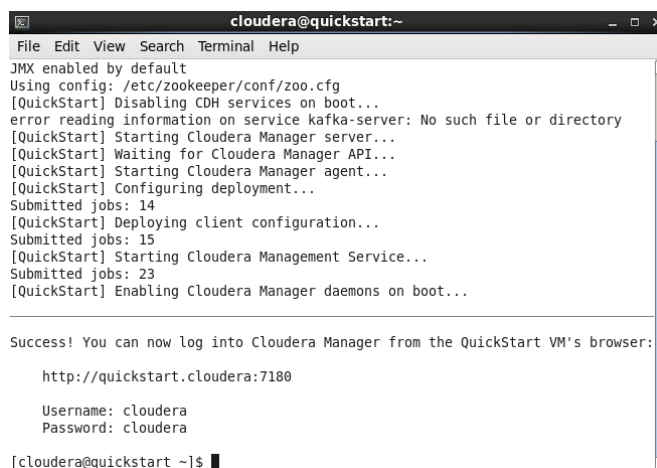
WARNING: It is highly recommended that you run Cloudera Express in a VM with
at least 8 GB of RAM.

You can override these checks by passing in the --force option,
e.g:

    sudo /home/cloudera/cloudera-manager --force

Press [Enter] to exit..
```

- 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.



```
cloudera@quickstart:~
File Edit View Search Terminal Help

JMX enabled by default
Using config: /etc/zookeeper/conf/zoo.cfg
[QuickStart] Disabling CDH services on boot...
error reading information on service kafka-server: No such file or directory
[QuickStart] Starting Cloudera Manager server...
[QuickStart] Waiting for Cloudera Manager API...
[QuickStart] Starting Cloudera Manager agent...
[QuickStart] Configuring deployment...
Submitted jobs: 14
[QuickStart] Deploying client configuration...
Submitted jobs: 15
[QuickStart] Starting Cloudera Management Service...
Submitted jobs: 23
[QuickStart] Enabling Cloudera Manager daemons on boot...

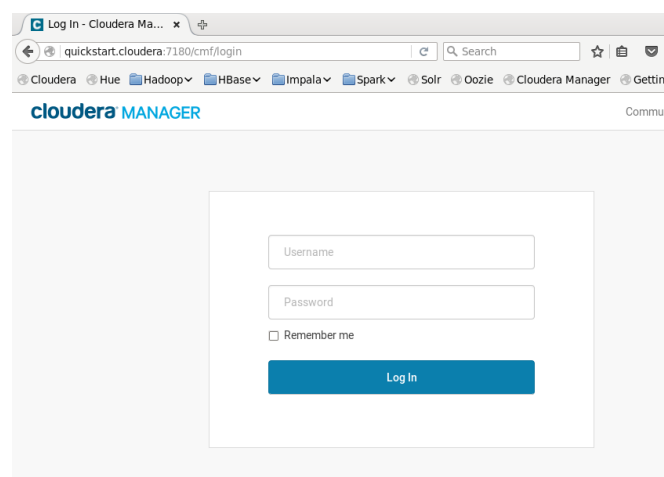
Success! You can now log into Cloudera Manager from the QuickStart VM's browser:

    http://quickstart.cloudera:7180

    Username: cloudera
    Password: cloudera

[cloudera@quickstart ~]$
```

- 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.



# ASSIGNMENT NO – 1

- 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.

