EXPERIMENT 1

INSTALLATION GUIDELINE FOR HADOOP

THEORY

What is Hadoop?

Apache Hadoop is an open-source framework that is used to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data. Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly.

Hadoop consists of four main modules:

- Hadoop Distributed File System (HDFS) – A distributed file system that runs on standard or low-end hardware. HDFS provides better data throughput than traditional file systems, in addition to high fault tolerance and native support of large datasets.
- Yet Another Resource Negotiator (YARN) – Manages and monitors cluster nodes and resource usage. It schedules jobs and tasks.
- MapReduce – A framework that helps programs do the parallel computation on data. The map task takes input data and converts it into a dataset that can be computed in key-value pairs. The output of the map task is consumed by reducing tasks to aggregate output and provide the desired result.
- Hadoop Common – Provides common Java libraries that can be used across all modules.

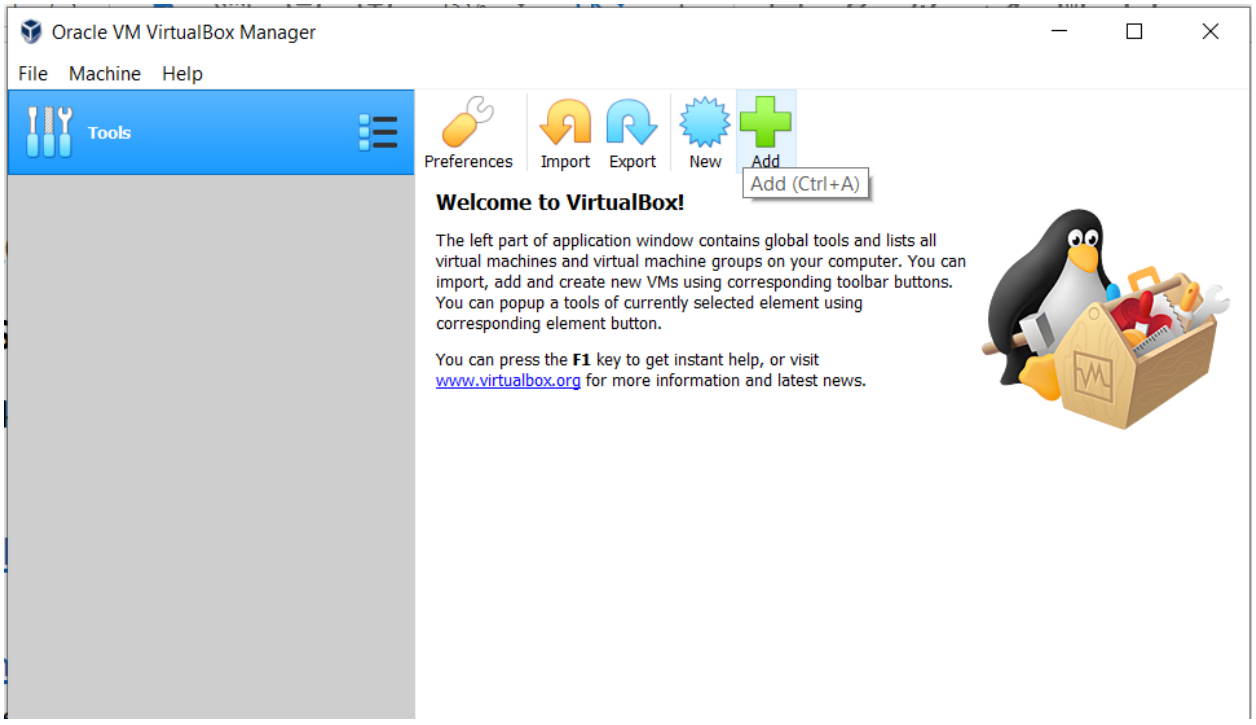
STEP 1

Use the following links to download above mentioned software successfully:
1. Link for Cloudera:
https://downloads.cloudera.com/demo_vm/virtualbox/cloderaquickstart-vm-5.12.0-0-virtualbox.zip

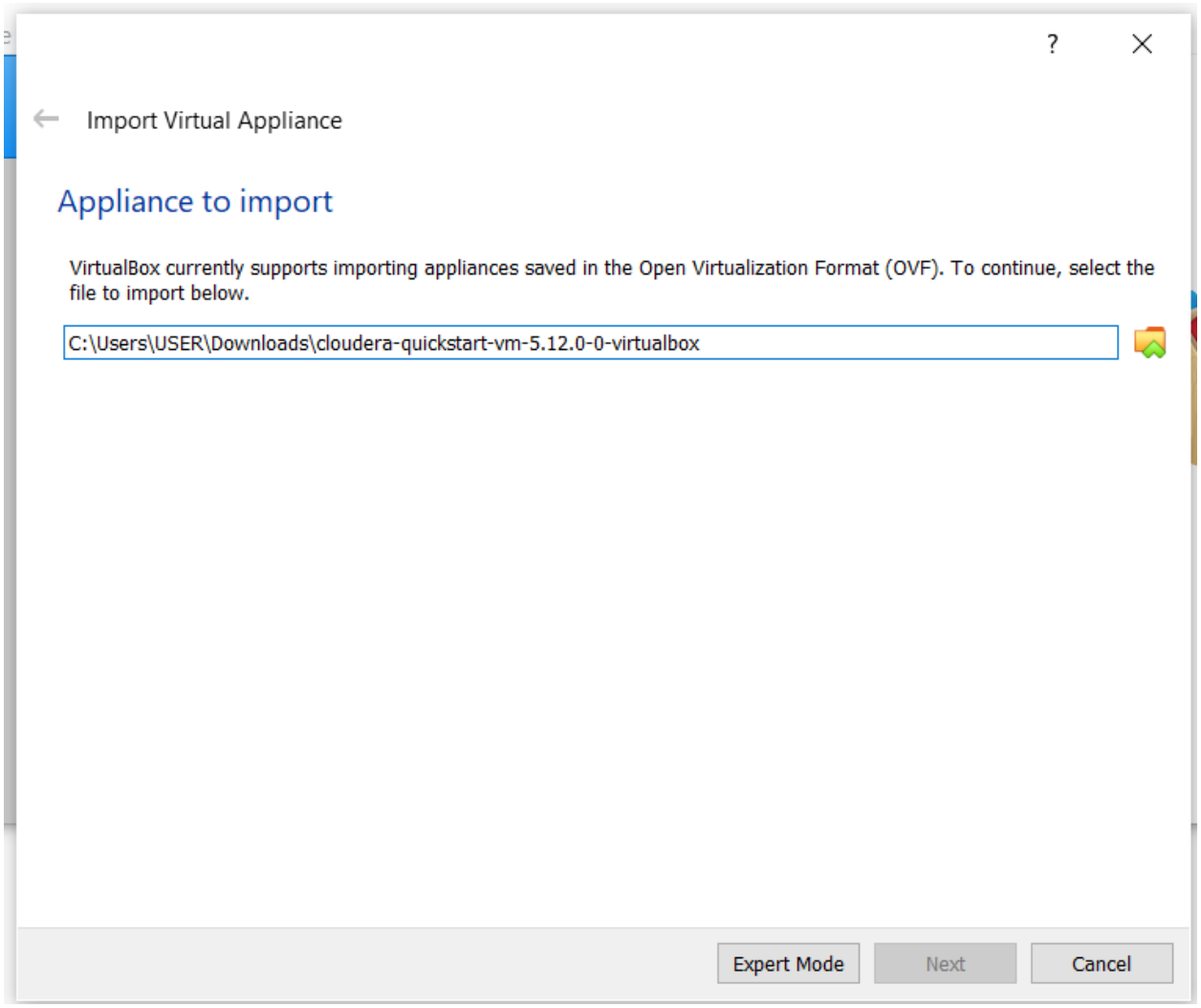
2. Link for VirtualBox: https://www.virtualbox.org/wiki/Download_Old_Builds_6_0 3
In case of any error, check the following link to enable Virtualization on your device (please look for the company whose machine(laptop/PC) you are using):
https://2nwiki.2n.cz/pages/viewpage.action?pageId=75202968

STEP 2

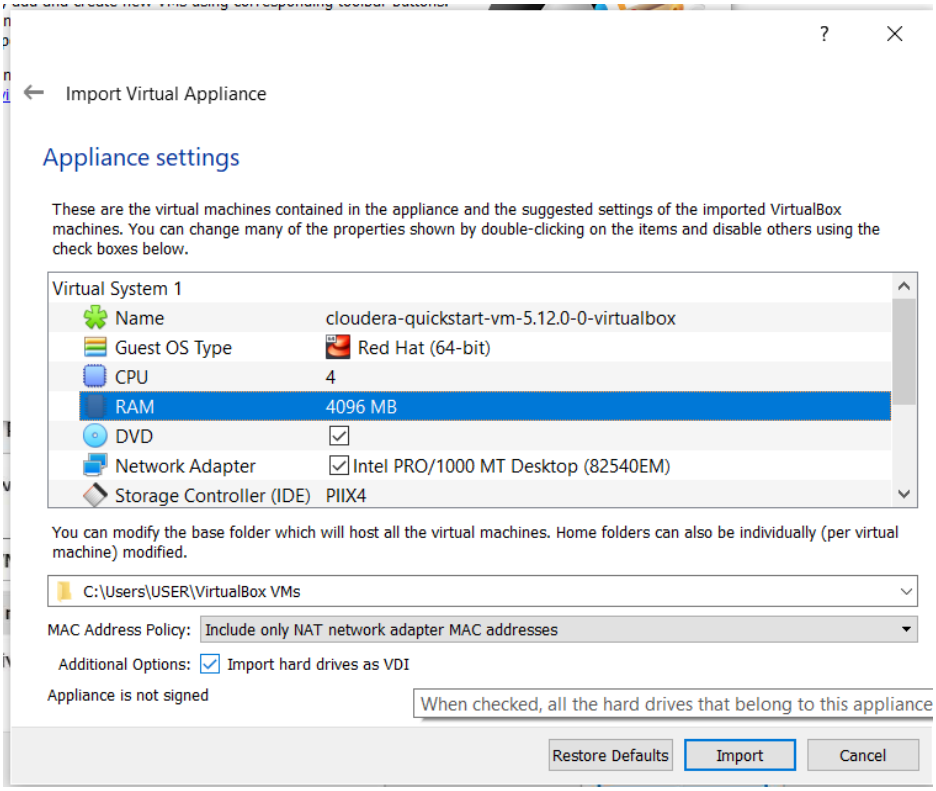
After downloading Cloudera, unzip it using a zip extractor and extract the files. Upon completion, open the virtual box software and select the Import option.

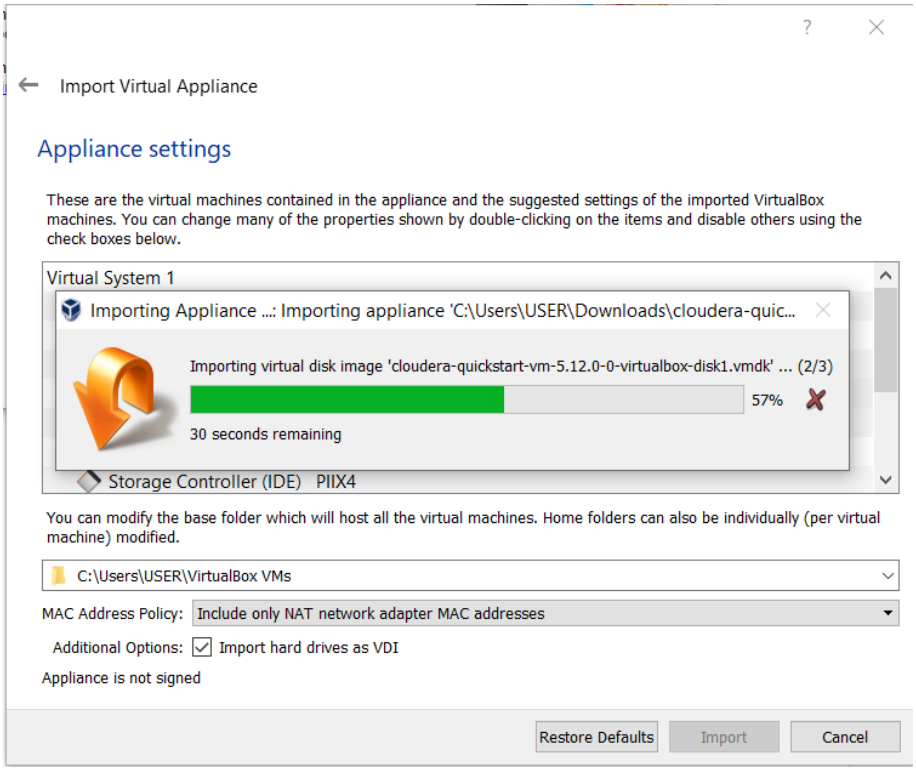


After selecting import, include the path of the previously downloaded Cloudera software.

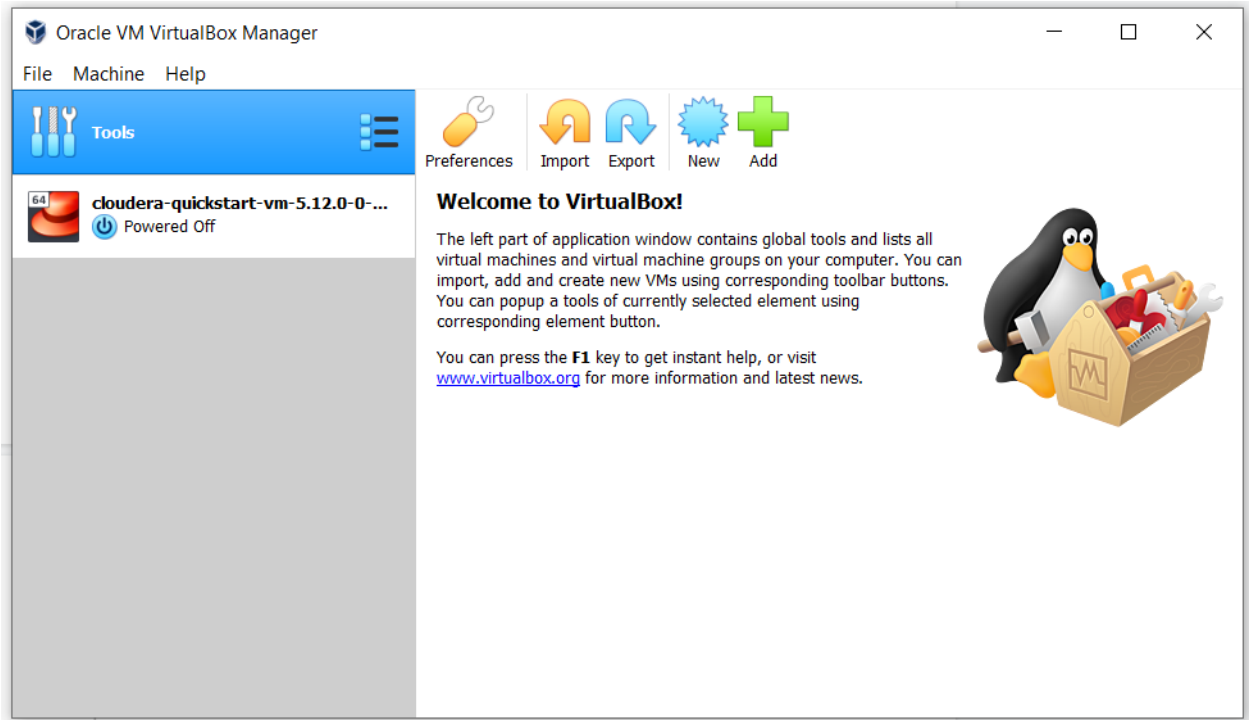


STEP 3 In the appliance settings, change the CPU section value from ‘1’ to ‘4’.

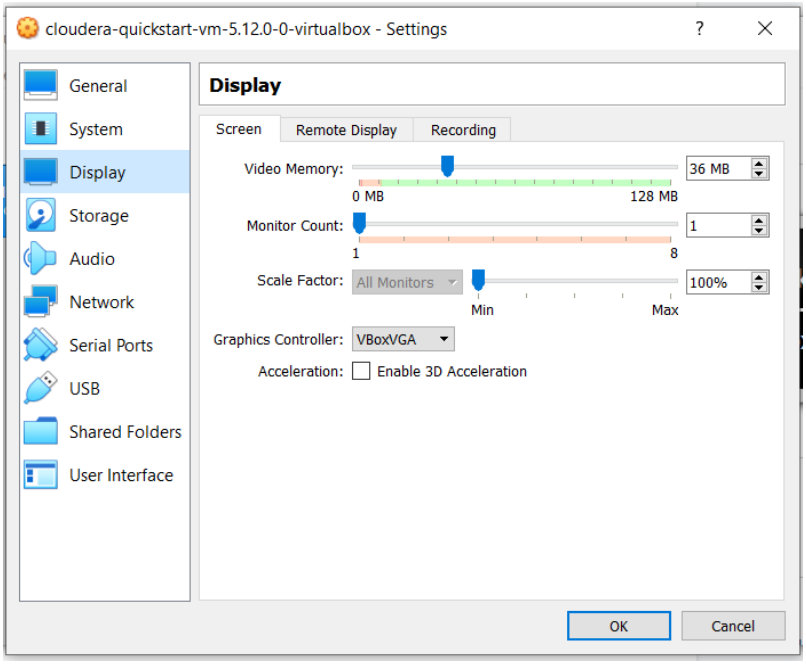




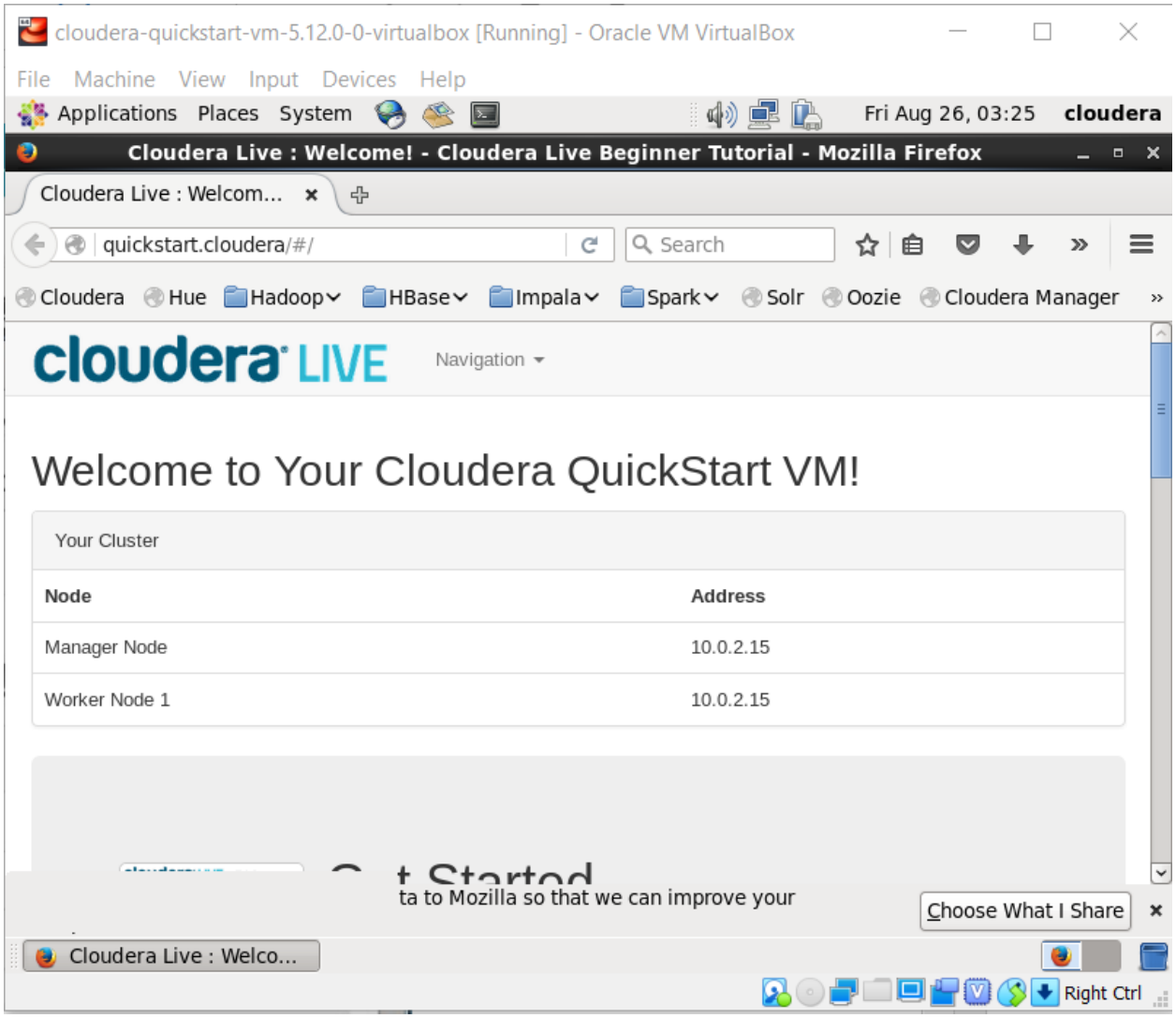
STEP 4 Proceed further if your VirtualBox homepage looks like this



Now click on the cloudera-quickstart-vm file which was initially showing powered off. Once you click on it, change the display settings and keep the video memory value between 0-40MB.



STEP 5 Now click on the Start button and wait for a few minutes. Initially, your window will look like this.



Once loading is completed, this window will appear.