NAME: Meith Navlakha
SAP: 60004190068
SEM-VI BDI Practical

EXPERIMENT 1

AIM: Installation of Hadoop on a single node cluster.

THEORY:

Hadoop:
Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model.

Cloudera:
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.

VM Player:
VMware Workstation Player is an ideal utility for running a single virtual machine on a Windows or Linux PC. Organizations use Workstation Player to deliver managed corporate desktops, while students and educators use it for learning and training.

Steps for installation:

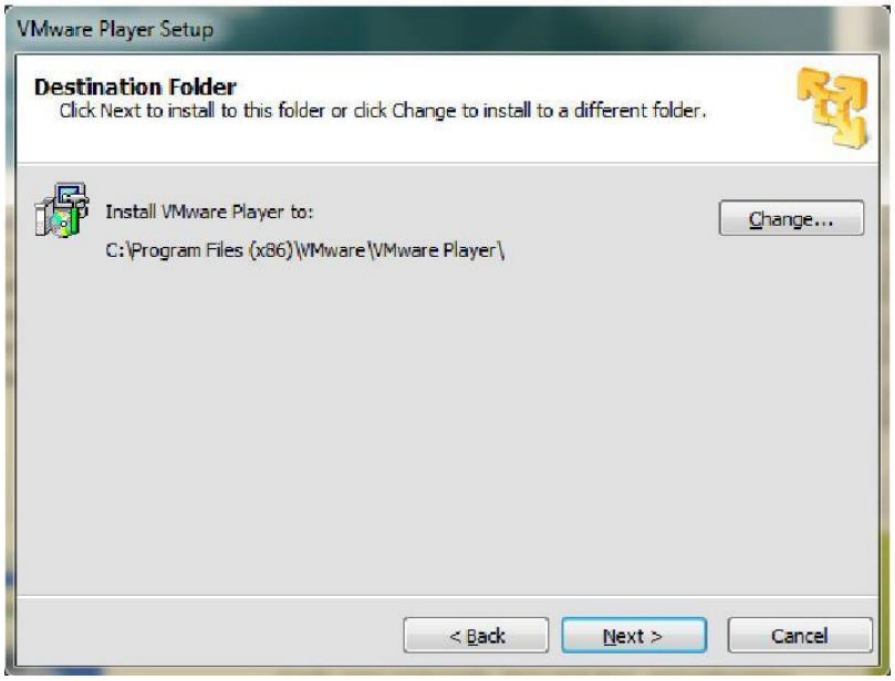
1. Install VM Player from https://www.vmware.com/in/products/workstation-player/workstation-player-evaluation.html
1. Then download the Cloudera quickstart virtual machine image from https://www.simplilearn.com/tutorials/big-data-tutorial/cloudera-quickstart-vm
1. Next, import Cloudera on VM Player by clicking “Open a Virtual Machine”. Then locate the cloudera image (vmdk file) you downloaded and click open.
2. This will now launch the virtual machine by Cloudera which will run on CentOS and have all big data tools like HDFS, Hive, etc installed.

NAME: Meith Navlakha
SAP: 60004190068
SEM-VI BDI Practical

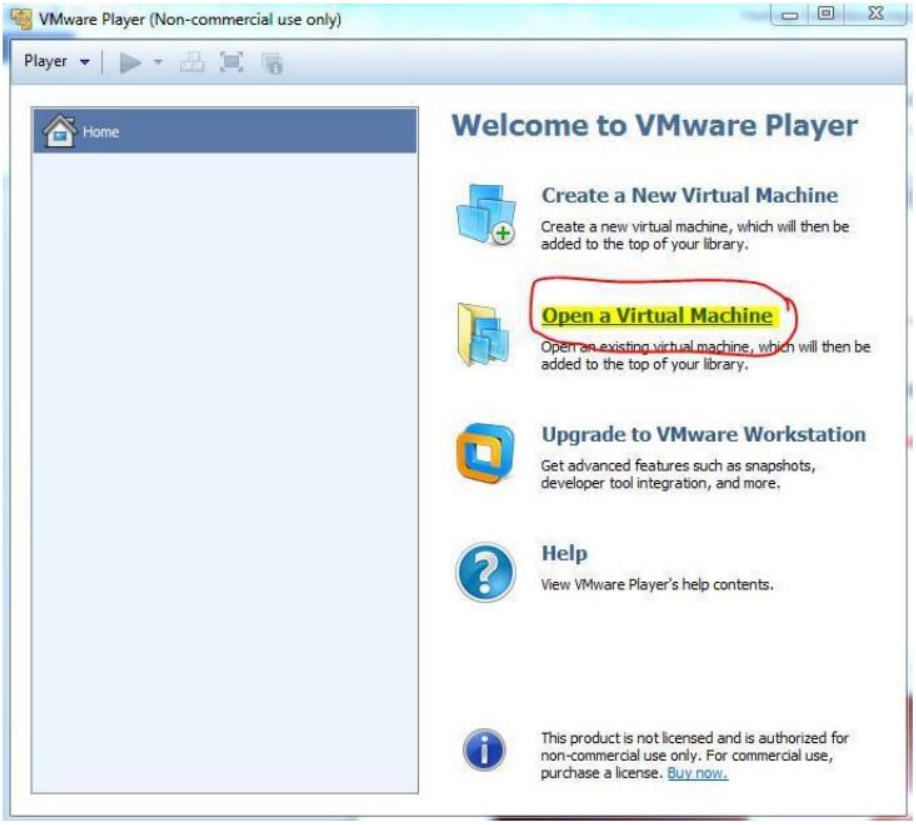
IMPLEMENTATION:

Step 1: We first download VMware Player and launch the setup wizard.

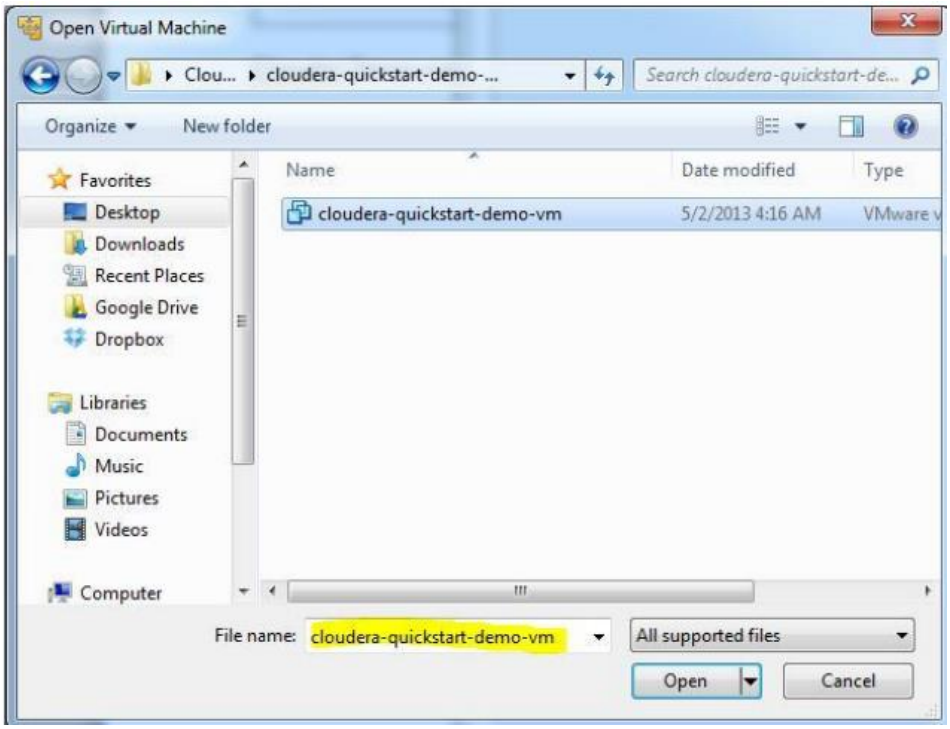
Step 2: Select the destination folder as shown below.



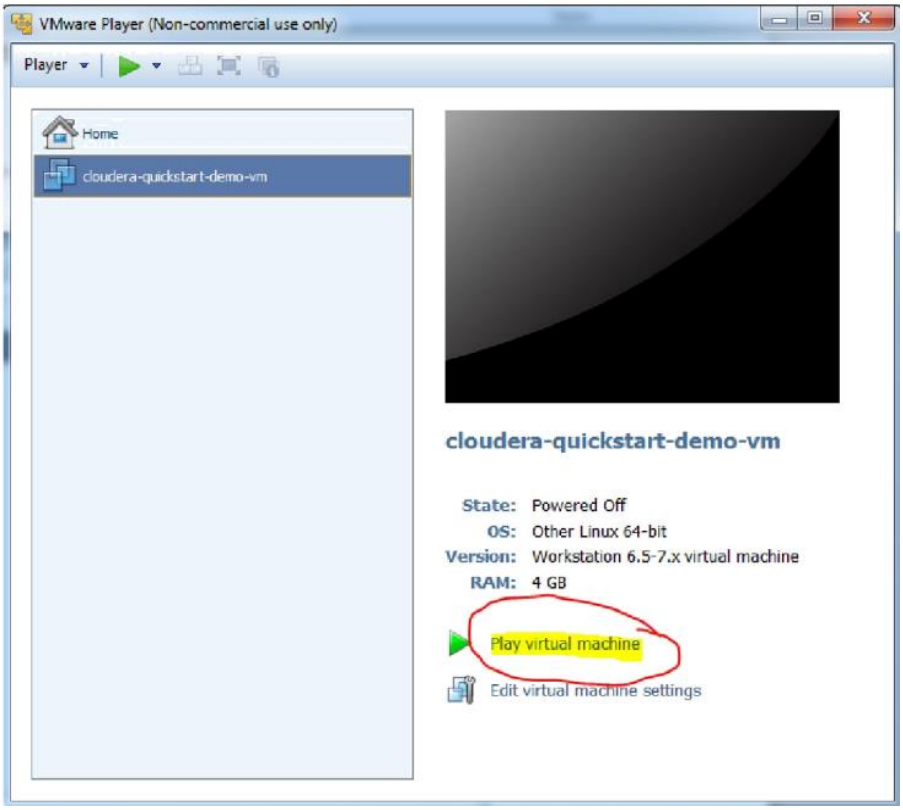
Step 3: Now we click “Open a Virtual Machine”



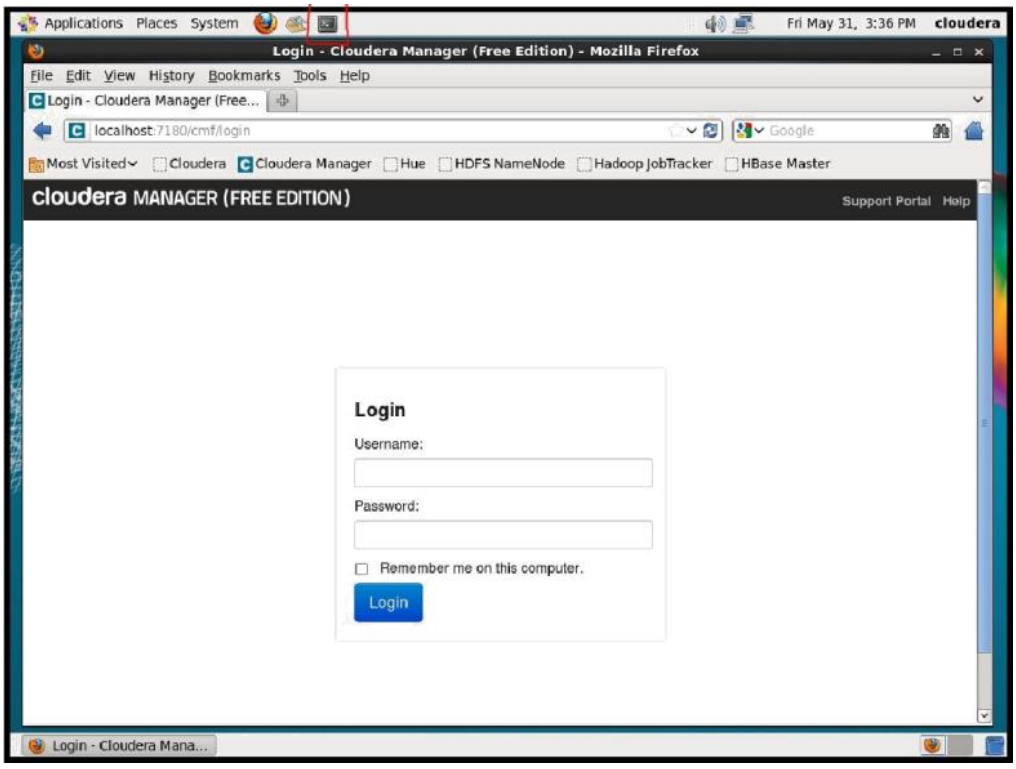
Step 4: We now locate the cloudera quickstart VM image and open it in VM Player.



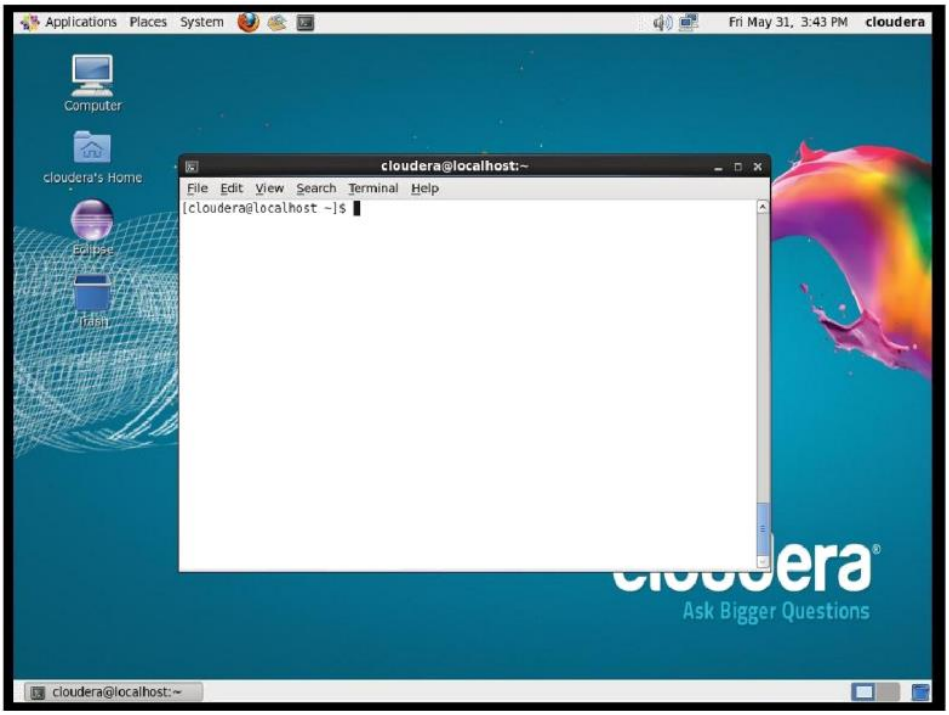
Step 5: Play the newly opened virtual machine in VM Player.



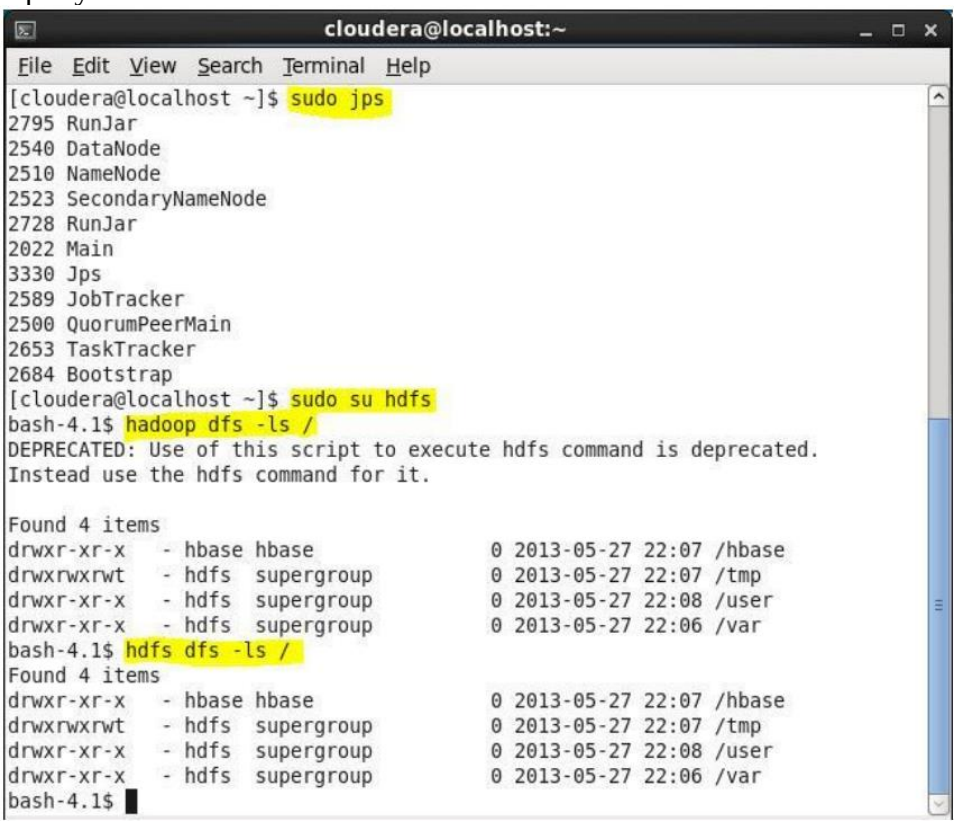
Step 6: The login page has opened by default. Click the black box shape icon highlighted in the image below to open the terminal in CentOS.



You now have the terminal open where you can run the Hadoop commands.



Step 7: Checking the installation using jps. The jps tool lists the instrumented HotSpot Java Virtual Machines (JVMs) on the target system. We use it to see if the hadoop processes are running properly or not.



CONCLUSION:

In this experiment, I first installed VM Player, a virtual machine, then installed Cloudera software package which contained the Hadoop ecosystem. Loaded the cloudera .vmdk file onto my VM Player and ran the virtual machine. This virtual machine has the Hadoop Ecosystem. Thus, I have successfully installed Hadoop and completed the experiment.