

Practical – 1

Aim: Set up and Configuration Hadoop Using Cloudera. Creating a HDFS System with minimum 1 Name Node and 1 Data Nodes HDFS Commands.

Theory:

Hadoop is an open-source framework designed for processing and storing large volumes of data in a distributed computing environment. It was originally created by Doug Cutting and Mike Cafarella in 2005 and is now maintained by the Apache Software Foundation. Hadoop is used for big data processing and is particularly well-suited for tasks that involve processing vast amounts of unstructured or semi-structured data.

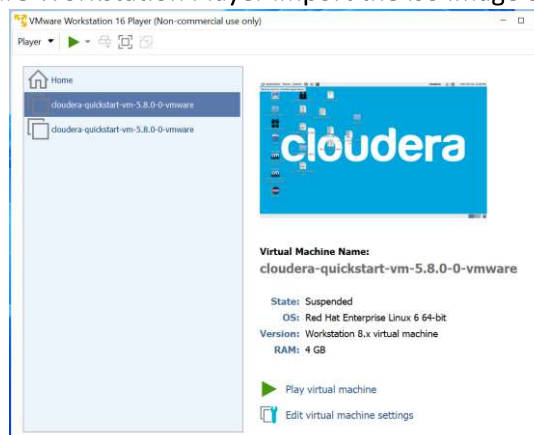
HDFS, which stands for Hadoop Distributed File System, is the primary storage component of the Hadoop ecosystem. It is a distributed, faulttolerant file system designed to store and manage large volumes of data across a cluster of commodity hardware. HDFS is a key component of the Hadoop framework and is responsible for the storage layer of big data processing.

Steps for installing Cloudera on VMWare Workstation Player 17 (Non Commercial Version):

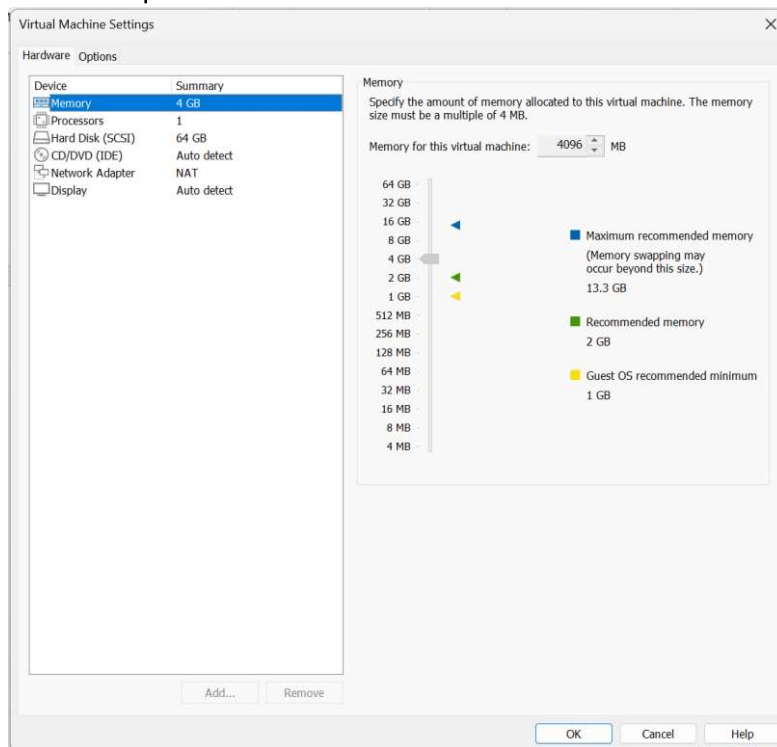
- a) Go to <https://www.vmware.com/in/products/workstation-player/workstation-player-evaluation.html> and download the VMWare Workstation Player.



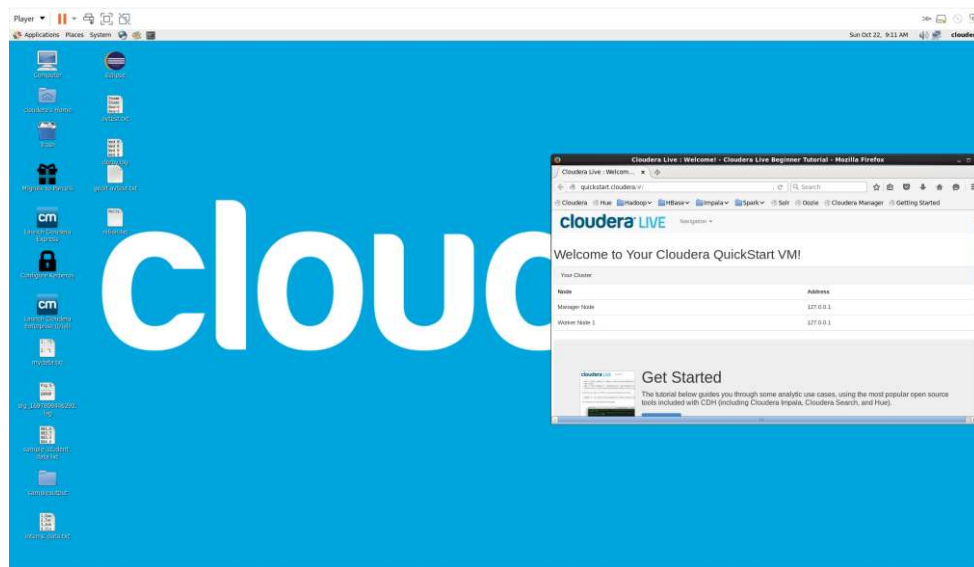
- b) After installing VMWare Workstation Player import the iso image as virtual machine.



c) Configure the VM as per need.

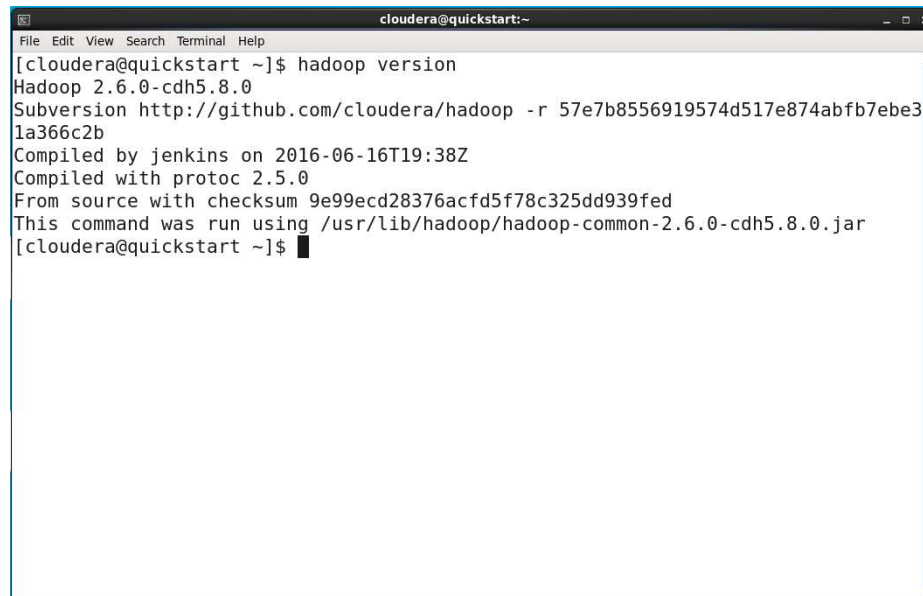


d) Launch the VM



e) Hadoop Version:

The hadoop version command is used to display information about the installed version of the Hadoop framework on your system. When you run this command, it provides you with details such as the Hadoop version number, the build date, and the Hadoop Git revision.

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the command '[cloudera@quickstart ~]\$ hadoop version' and its output: 'Hadoop 2.6.0-cdh5.8.0', 'Subversion http://github.com/cloudera/hadoop -r 57e7b8556919574d517e874abfb7ebe31a366c2b', 'Compiled by jenkins on 2016-06-16T19:38Z', 'Compiled with protoc 2.5.0', 'From source with checksum 9e99ecd28376acfd5f78c325dd939fed', and 'This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.8.0.jar'. The prompt '[cloudera@quickstart ~]\$' is followed by a cursor.

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop version  
Hadoop 2.6.0-cdh5.8.0  
Subversion http://github.com/cloudera/hadoop -r 57e7b8556919574d517e874abfb7ebe31a366c2b  
Compiled by jenkins on 2016-06-16T19:38Z  
Compiled with protoc 2.5.0  
From source with checksum 9e99ecd28376acfd5f78c325dd939fed  
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.8.0.jar  
[cloudera@quickstart ~]$
```

Conclusion: Thus, we have successfully installed, set up and Configured Hadoop Using Cloudera on VMWare Workstation Player