Setting up your Hadoop System

There are three main options to set up your Hadoop system:

Option 1: Deploying Hadoop on your computer

Building Hadoop from the ground up by yourself, you will learn a lot. If you want to go further in HPC or distributed system computing areas, you should choose this option. You can either use Dr. Zhou's supervised student *Ahmed Mostafa's* [step-by-step installation manual](#) or follow the directions at https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html#Installing_Software.

Option 2: Using Virtual Machine

You can also leverage virtual machine where a straightforward, pre-configured, learning environment has been included so that you don't have to configure the Hadoop system by yourself. All you need to do is install a virtual machine such as VirtualBox or VMware then apply an existing virtual machine image of Hadoop system into it.

Setup Hadoop by *Cloudera QuickStart VM* (it is no longer supported by *Cloudera,* however, for practice purposes, it is enough at this point of time**) and VirtualBox as below:**

You can also employ CDH (A complete, tested, and popular distribution of Apache Hadoop and other related open-source projects, including Apache Impala and Cloudera Search) as your Bigdata (Hadoop) platform to work on some practical exercises. To quickly setup your environment, you can also install a CDH VM into your virtual box rather than deploying and configuring your environment manually. *QuickStart virtual machines (VMs)* provides everything you need to try CDH, Cloudera Manager, Impala, and Cloudera Search.

Steps of *Virtual Machine Option C* above – Set up Hadoop by *Cloudera QuickStart VM and VirtualBox:*

1. Install download VirtualBox:

    https://www.virtualbox.org/wiki/Downloads
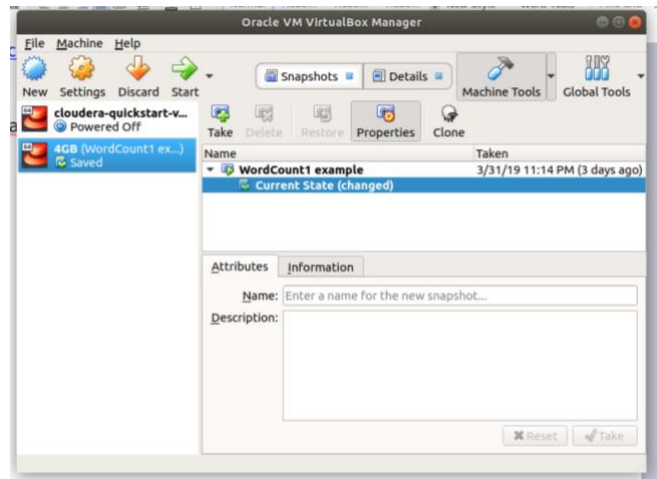
2. Download a Cloudera QuickStart VM and install it into your VirtualBox:

    Go to our CougarVIEW course, in the Big Data Processing module.   Download the VM at the Cloudera QuickStart VM.website for the VM images since the *Cloudera QuickStart VM* now may not be maintained by Cloudera.

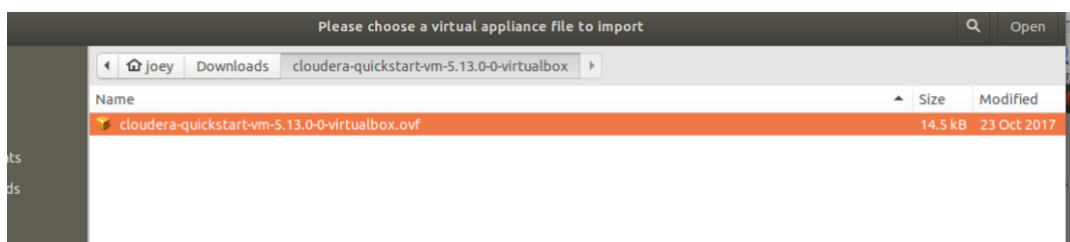3. How to install ***Cloudera QuickStart VM*** in your VirtualBox:

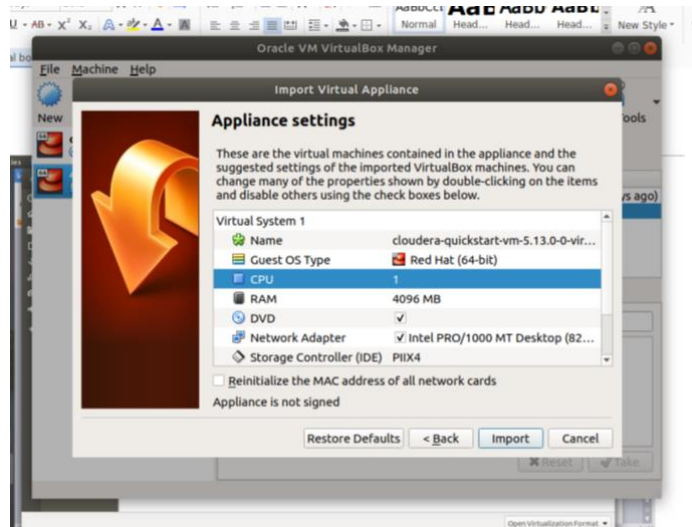   1) Launch your VirtualBox and click File->Import Appliance

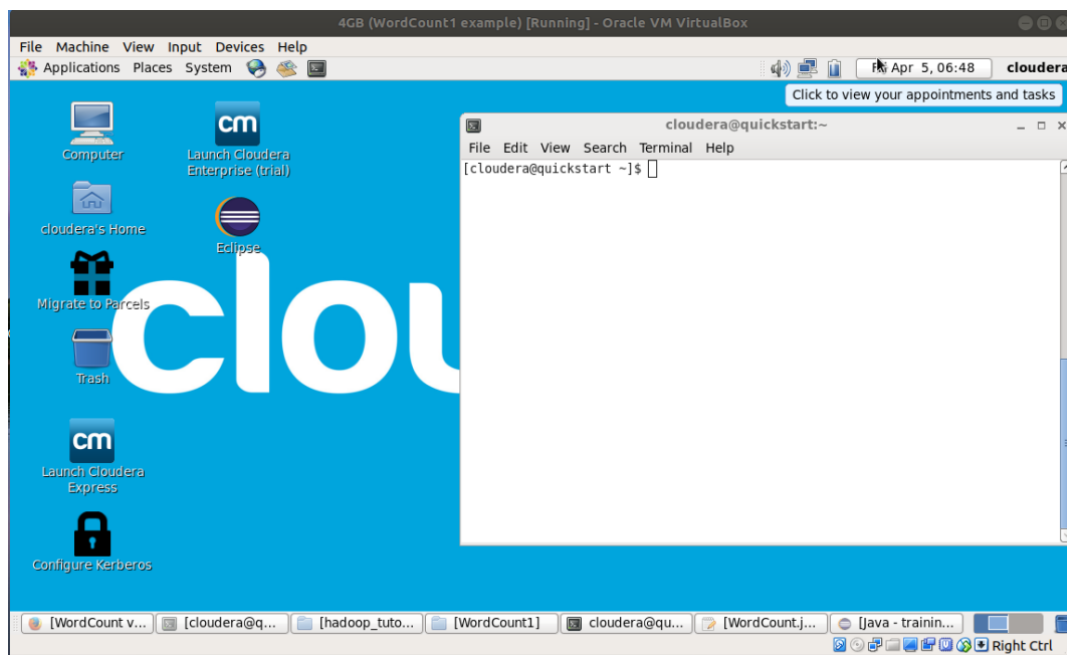2) Brows to the folder where you downloaded and extracted the VM image

3) Select .ovf file to import into your VirtualBox

4) Configure your virtual machine by setting number of CPUS, size of RAM according to your needs

5) Launch your VM and you will have a similar environment shown like below, you are then ready to program and test.

6) First MapReduce example: word count

https://docs.cloudera.com/documentation/other/tutorial/CDH5/topics/ht_wordcount1.html

Option 3: Run Hadoop in the Cloud

Create a cloud resource that can run Hadoop.   For example, https://techfortalk.co.uk/2025/02/24/set-up-a-hadoop-cluster-on-aws-emr-a-step-by-step-guide/.