Hadoop Installation in Standalone Mode

1. Install VMware player

Hadoop software was developed using Java and was tested successfully on Linux environments. In order to be successful in setting up a single node cluster, we need to have Linux environment. Best way to run a Linux virtual machine would be using VWware player. Latest free version can be downloaded from below link. Any VMware player version above 5.0 would be good enough for this setup.

https://my.vmware.com/web/vmware/free#desktop_end_user_computing/vmware_player/6_0

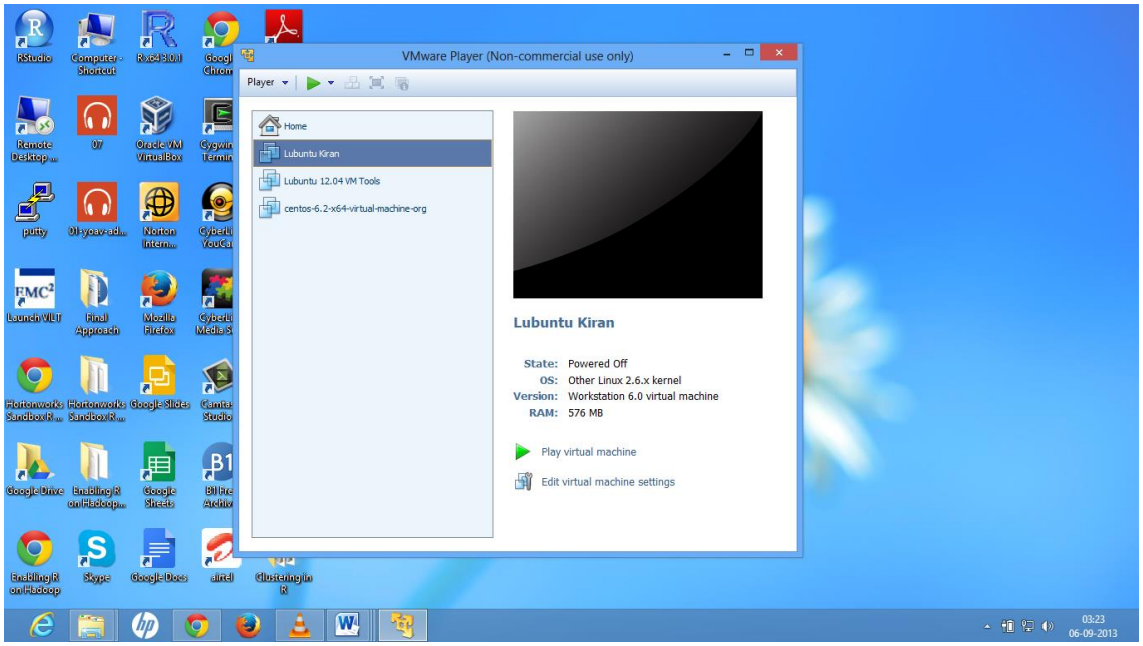
2. Download Linux image (Ubuntu 1204 version)

After installing the VMware player, you need to load a Linux virtual image for setting up hadoop cluster. For this, please download the Ubuntu Linux Image from below link. lubuntu1204t.zip file would be around 600 Mb in size and you have to extract the file after the download.

http://www.traffictool.net/vmware/lubuntu1204t.html

3. Set up Linux image on VMware player

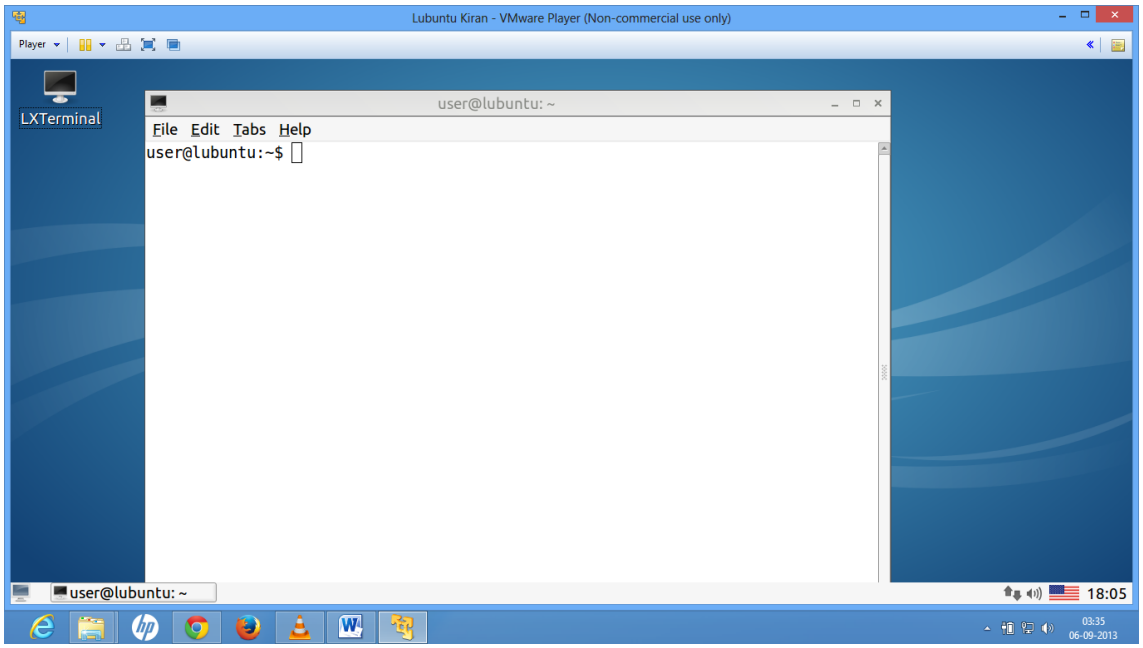
Click on the VMware player icon and once it is ready, click the “Open a virtual machine” tab. After this point the directory to the location Ubuntu 1204 extracted folder and in that, select the file named “Lubuntu” which is .vmx file (VMware virtual machine configuration). If you have selected the right file, then the image will look like the below screenshot. I renamed the version as “Lubuntu Kiran” and in your case, the name might differ.



After this, hit “Play virtual machine” tab and Linux virtual image would be loaded in few minutes.

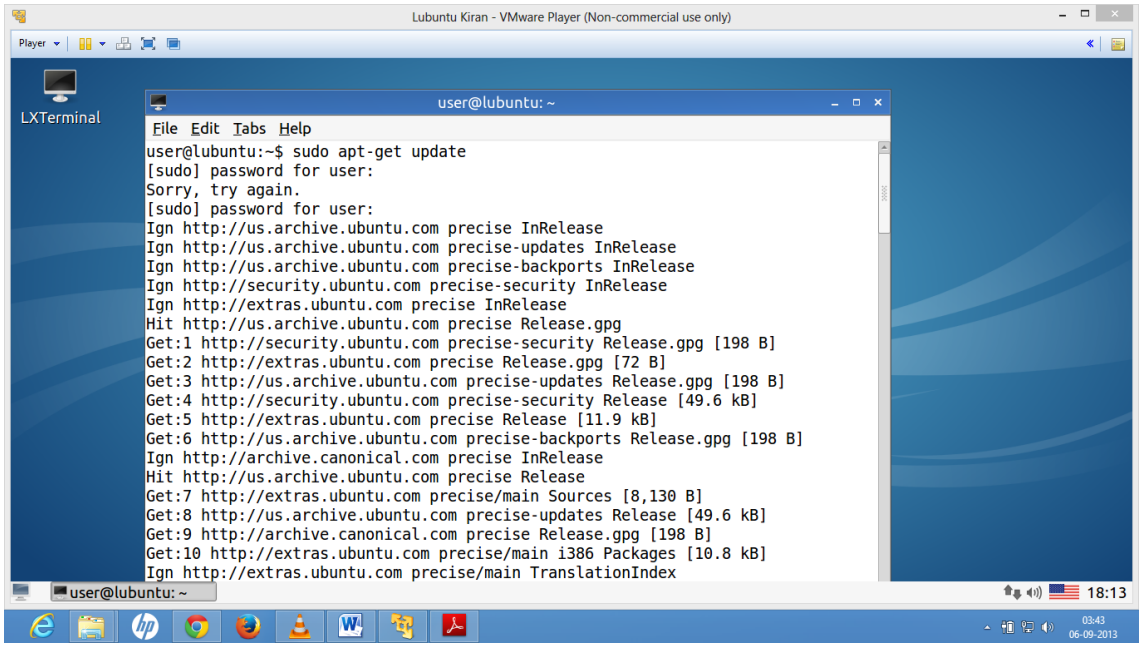
4. Install updates on the Linux virtual image

Once the Ubuntu virtual image is set, please install required updates. As soon as you see the desktop, update manager will prompt you for the required updates. In Linux, all the commands are executed in a terminal. You can access the same under Start (Computer Icon) -> Accessories -> LXTerminal.



In the terminal session, you are logged in as user with super user privileges. In this case, user would be one such created user name and password would be password. Refer to the link provided for Ubunut 1204 download for more information on these details. Firstly, run the below command for updating the linux repository.

$ sudo apt-get update



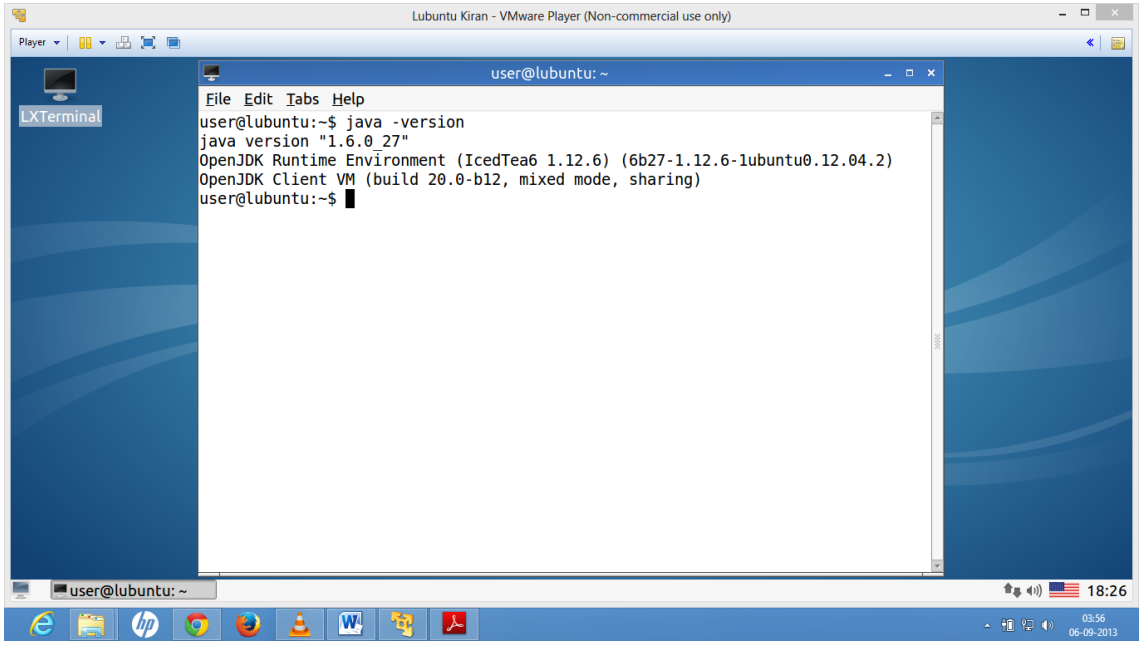
5. Install JVM

Once the update is done, install java package on the virtual machine. This package is a pre-requisite for Hadoop.

$ sudo apt-get install openjdk-6-jdk

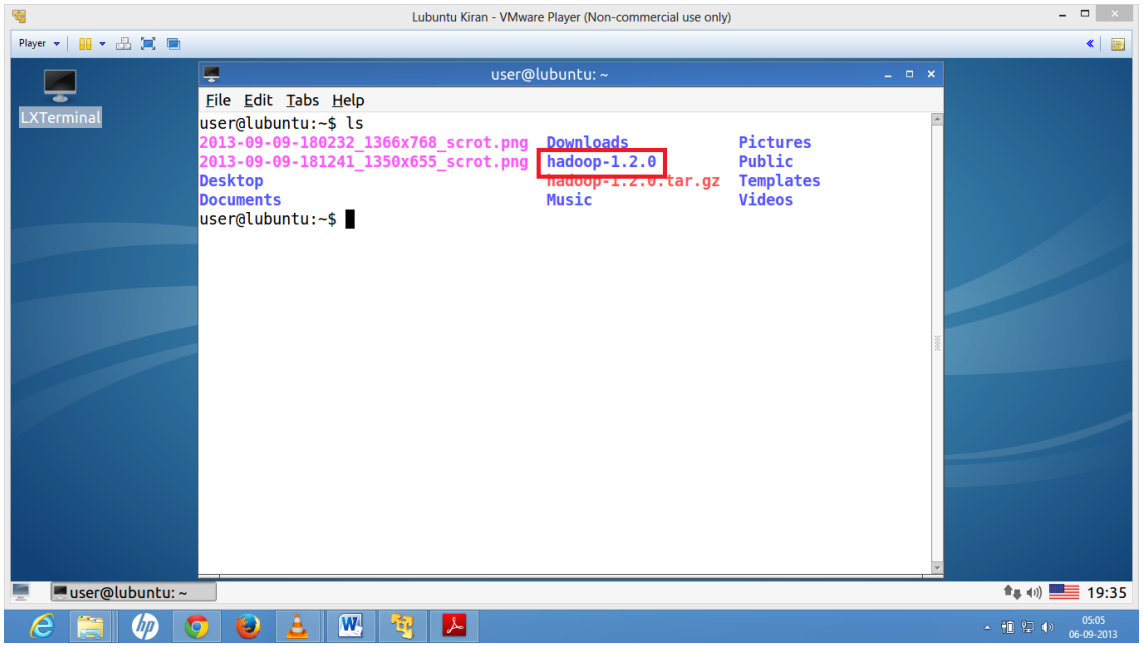
After Java is installed, enter the below command to check whether java is present in the machine.

$ java –version



6. Download and extract hadoop package from official site

Now we will have to download the stable release of Hadoop which is 1.2.0 version.



$ wget http://archive.apache.org/dist/hadoop/core/hadoop-1.2.0/hadoop-1.2.0.tar.gz

$ tar -xvf hadoop-1.2.0.tar.gz

$ ls

7. Configure hadoop environment file

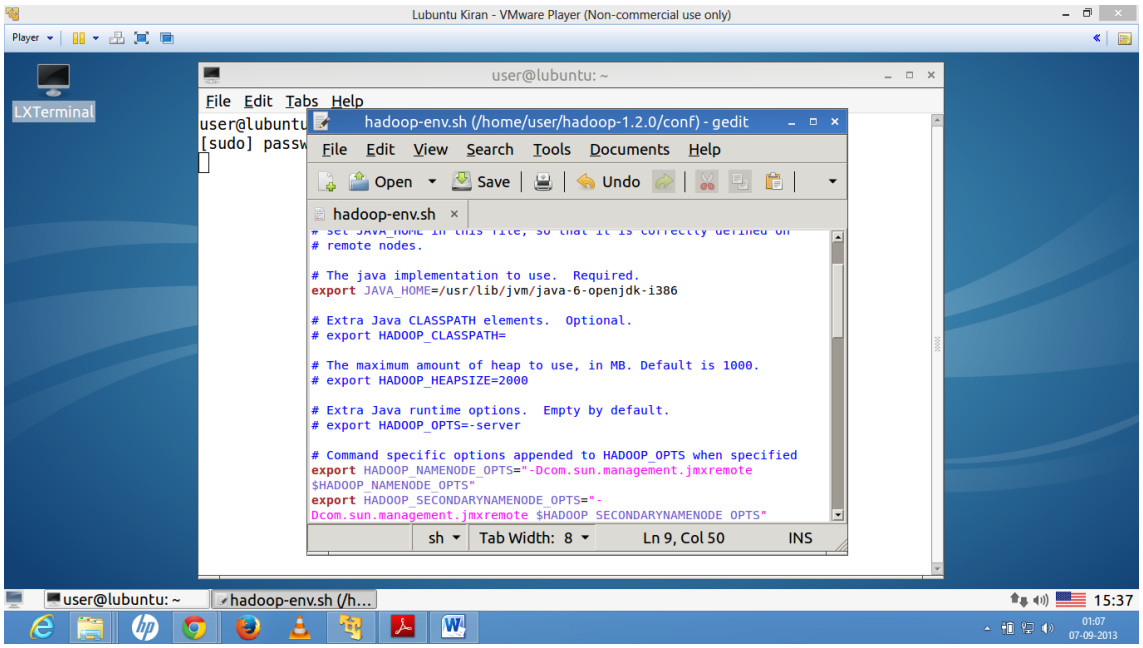
Normally gedit command should be used which helps in making changes to text files. If your machine do not have gedit installed already, use the below command.

$ sudo apt-get install gedit

In the hadoop environment file, we need to add JAVA_HOME path variable which is mentioned as below.

export JAVA_HOME=/usr/lib/jvm/java-6-openjdk-i386

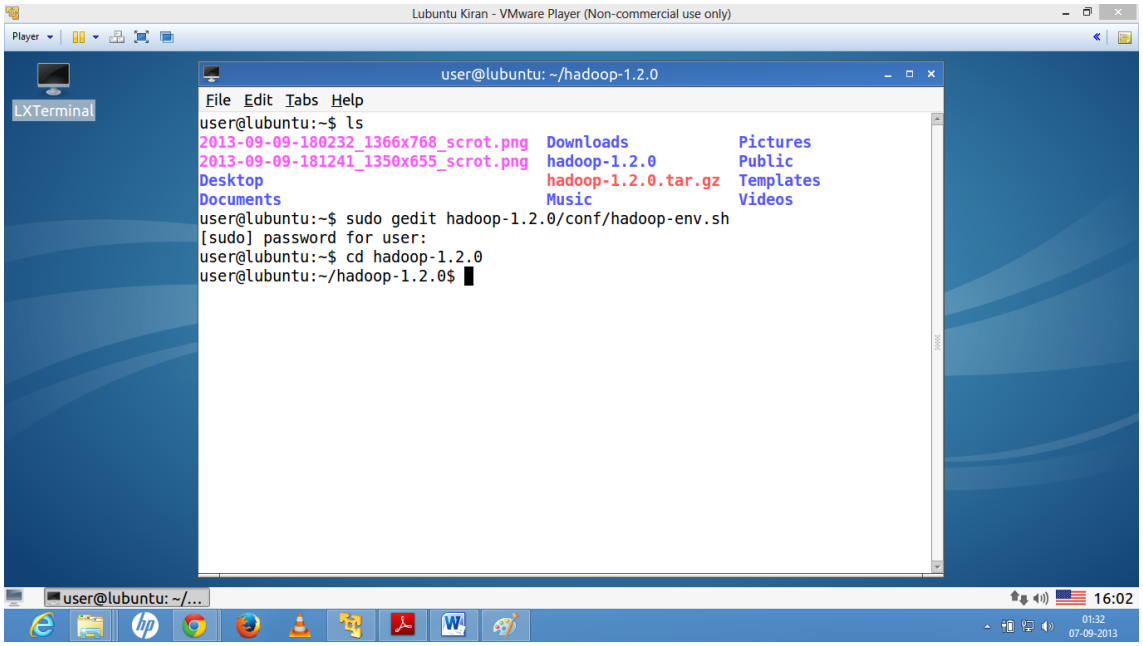
$ sudo gedit hadoop-1.2.0/conf/hadoop-env.sh



8. Run hadoop commands

With the above steps, hadoop is ready to be run in standalone mode on your machine. Firstly, the directory of hadoop folder should be made as local using the following command.

$ cd hadoop-1.2.0



Once we are in hadoop folder, other commands should always be prefixed by “bin/”.

Command for checking version of hadoop

$ bin/hadoop version

Formatting the name node of hadoop cluster. This command needs to be executed when you set up the cluster for the first time.

$ bin/hadoop namenode –format

More commands can be found by running the below command

$ bin/hadoop

9. Test hadoop installation

This is simple code which will test whether hadoop installation is successful and mapreduce programs would run smoothly.

At the linux terminal,

$ cd hadoop-1.2.0

Then once you are on hadoop home folder, run

$ bin/hadoop jar hadoop-examples-*.jar pi 10 100

If the value of pi would be printed as an output at the end of the query, then it is a success else you need to debug what went wrong.