



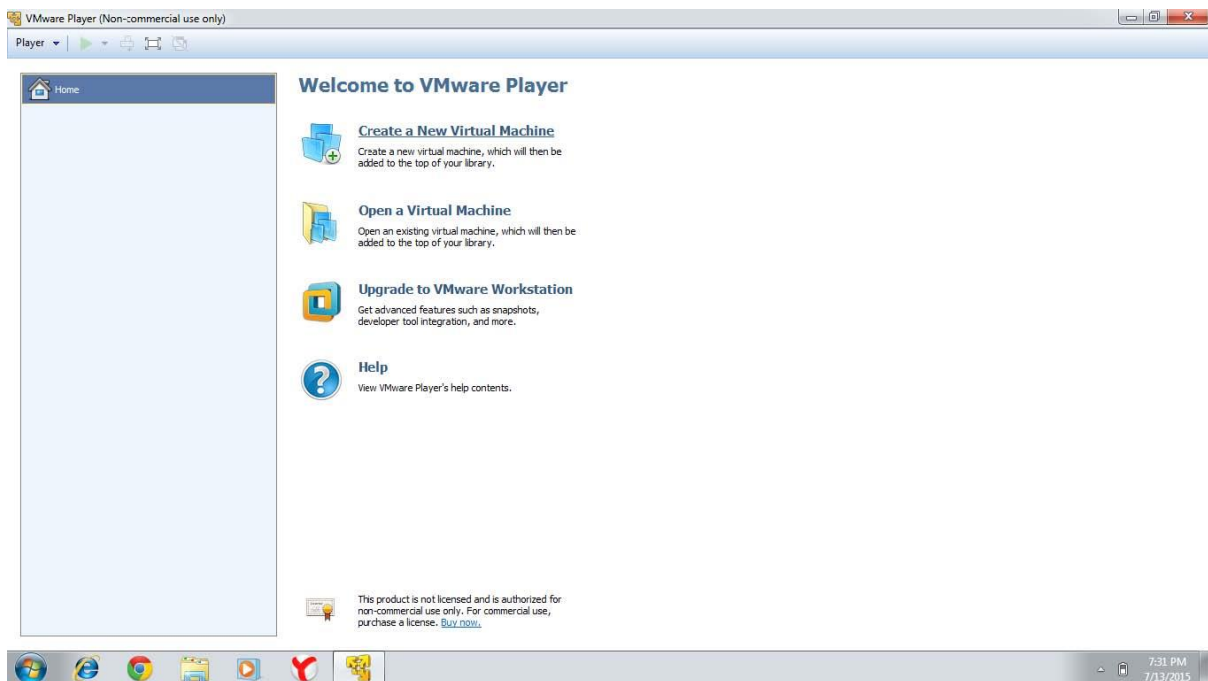
Create a Hadoop1.2.1 cluster in local mode on centos

# VMware & centos6.4 Installation

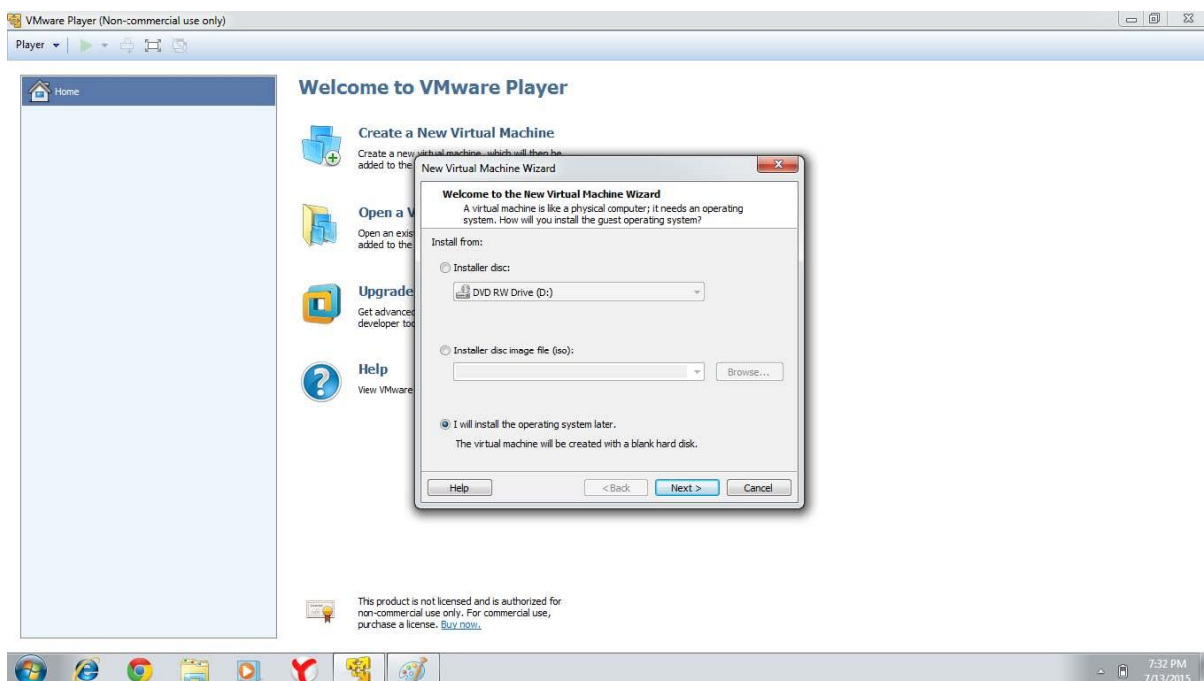
## Steps:

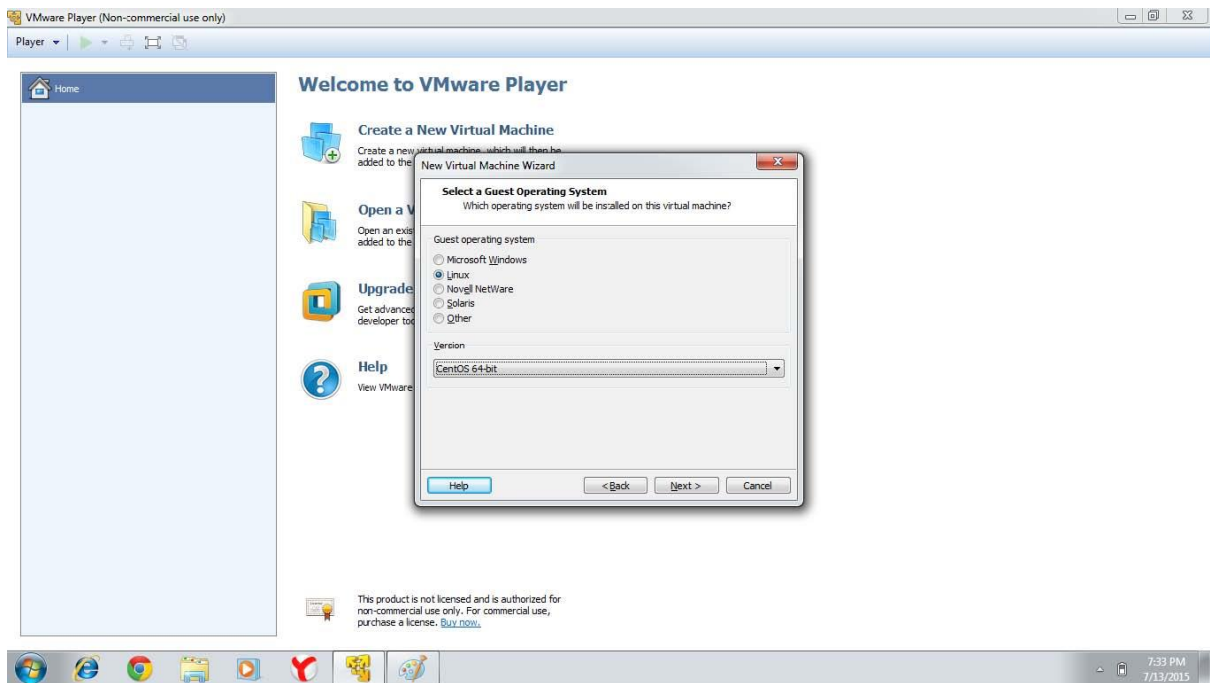
### 1. Set Up a Typical VMware Player Installation

### 2. Click on crate a new virtual machine

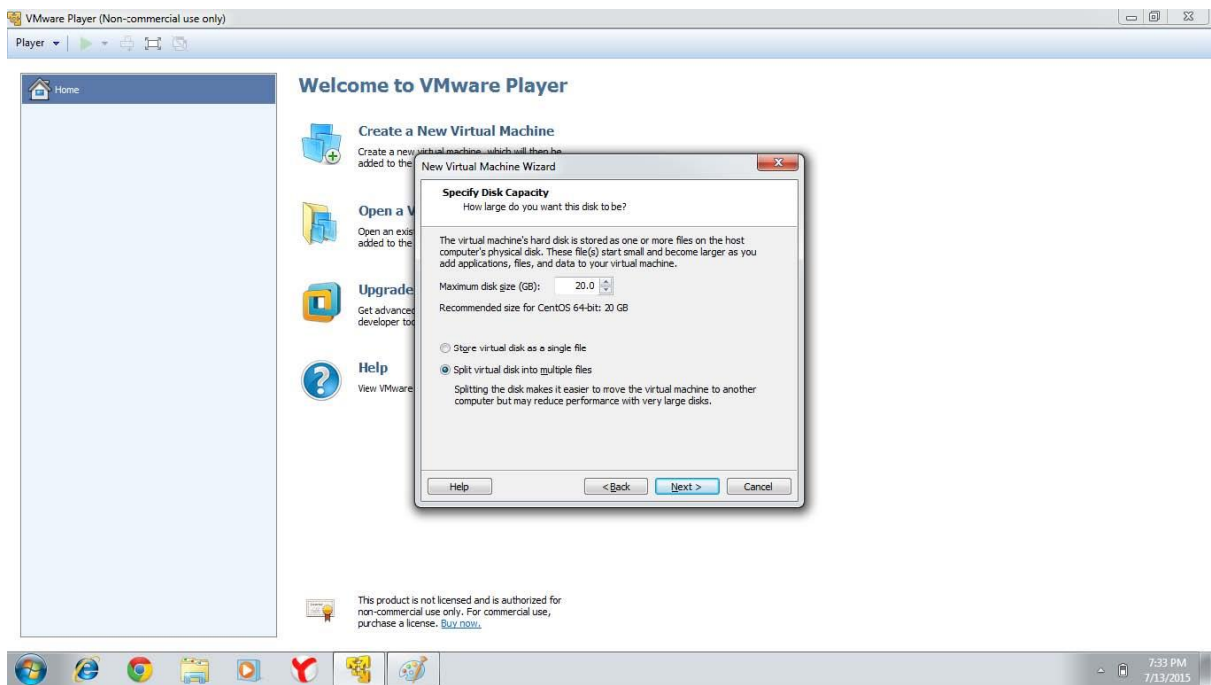


### 3. On the next screen choose I will install the OS later.



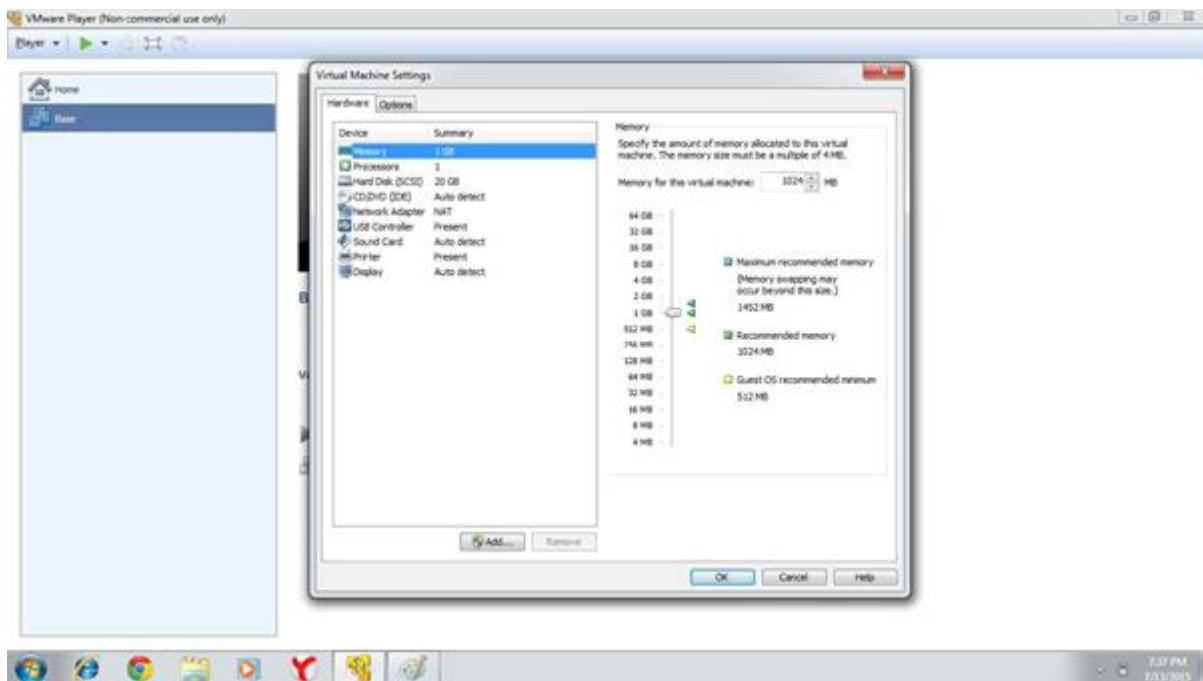
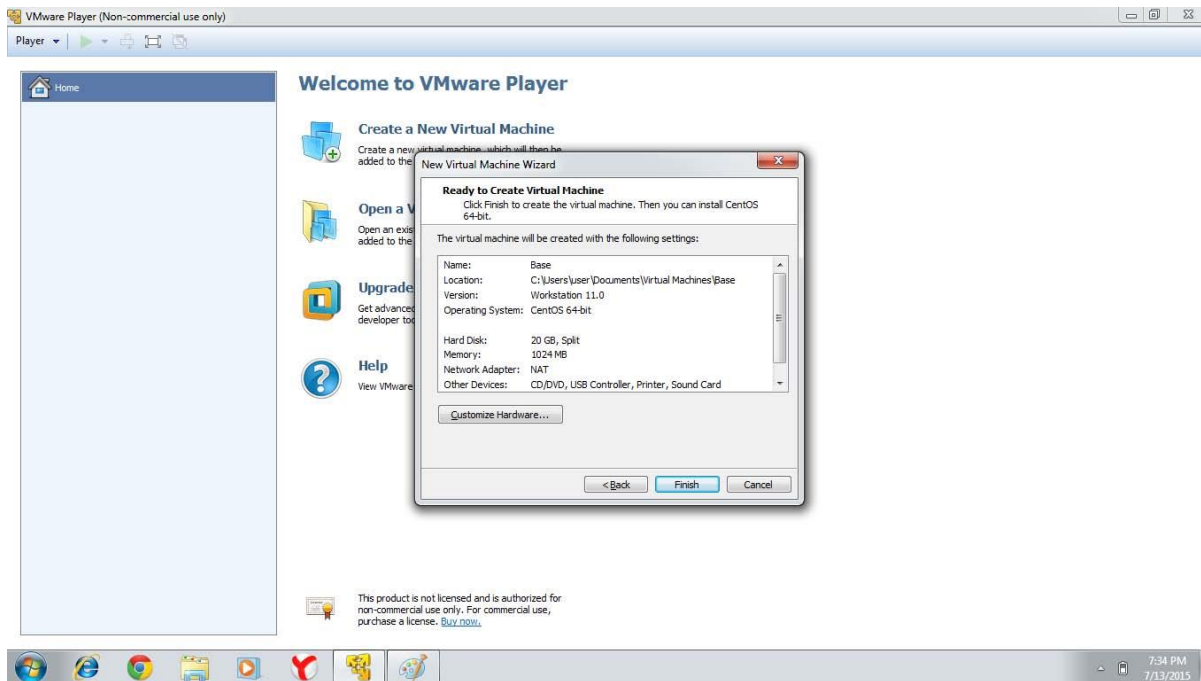


On the next screen, select Linux and Cent OS 64 bit as version



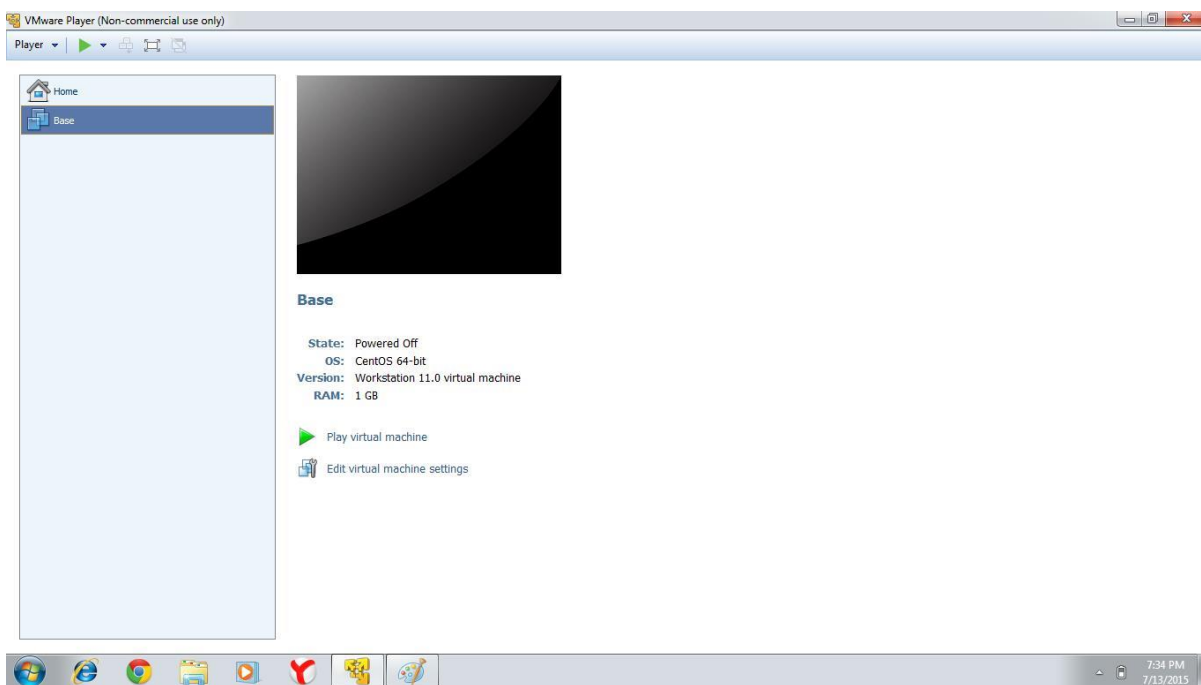
On the next screen, do not make any changes and click next

On the next screen, do not make any changes and click next.



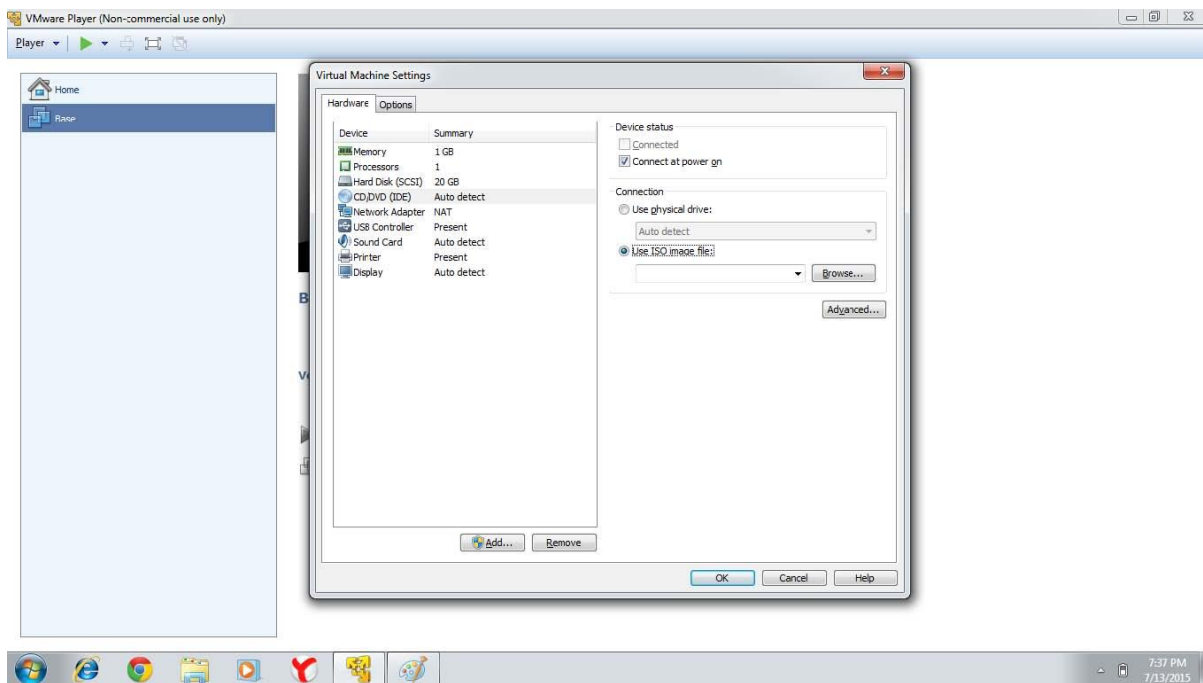
On the next screen give a name “Centos” to our virtual machine.

If you did well, you will have the Base virtual machine on the start-up page, as shown below

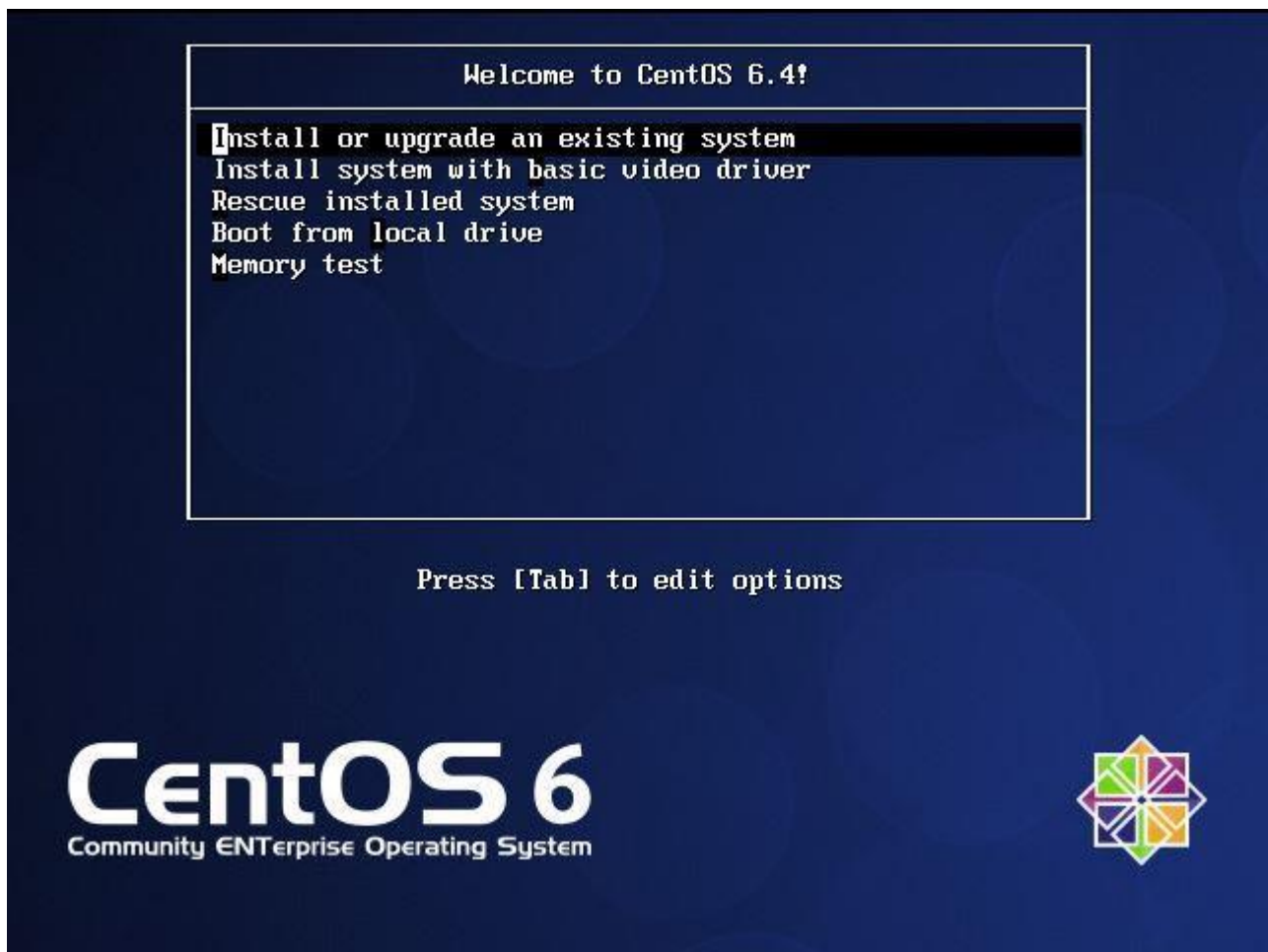


Now, click on edit virtual machine settings

Click on CD/DVD and select use ISO image. Browse and give the Cent OS ISO file that we have just downloaded.



Click OK. Now we are ready to install Cent OS. Click on Play this virtual machine. The VM will boot from the ISO. The following screen will be displayed.

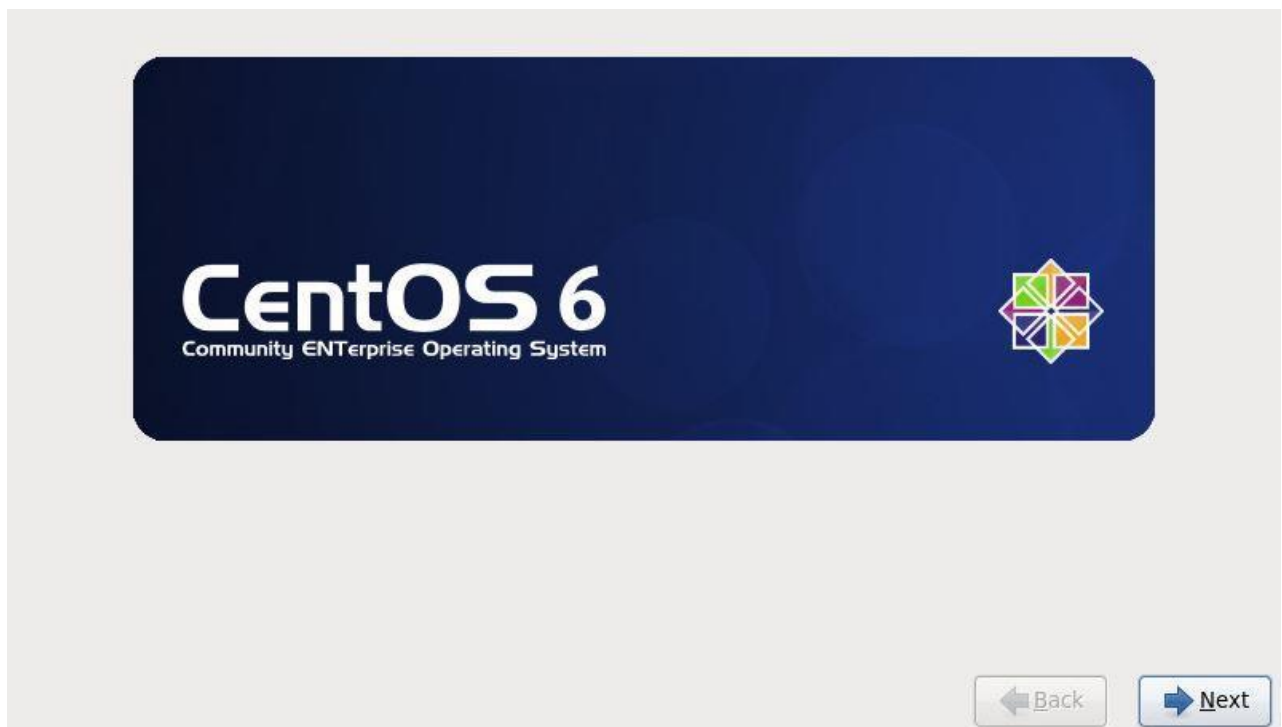


2. Choose **Skip** to cancel the installation media check. But if you choose ok, this will check your installation



media for any sort of missing installation files of the operating system.

3. Now you can see the installation welcome screen. Here click **next** button.



4. Now choose the language you prefer to install and click **next** button.





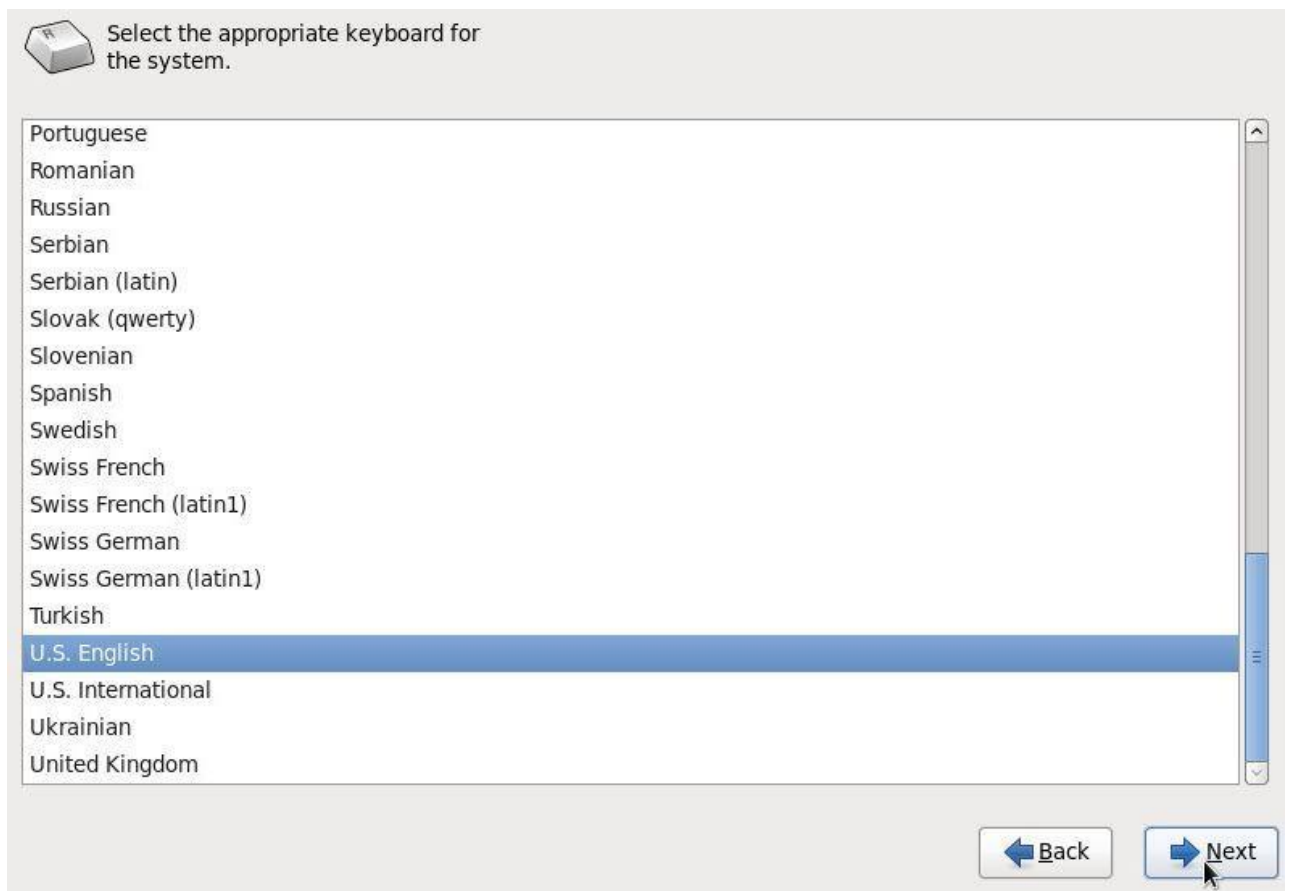
What language would you like to use during the installation process?

Bulgarian (Български)  
Catalan (Català)  
Chinese(Simplified) (中文 (简体))  
Chinese(Traditional) (中文 (正體))  
Croatian (Hrvatski)  
Czech (Čeština)  
Danish (Dansk)  
Dutch (Nederlands)  
English (English)  
Estonian (eesti keel)  
Finnish (suomi)  
French (Français)  
German (Deutsch)  
Greek (Ελληνικά)  
Gujarati (ગુજરાતી)  
Hebrew (עברית)  
Hindi (हिन्दी)

← Back

Next →

5. Choose the appropriate keyboard layout from the option and click **next** button.




6. Here choose **Basic Storage Device** and click **next** button.



7. Click Yes, discard any data to continue option in storage device warning.




8. Type unique host name for this system and click Next button.

 Please name this computer. The hostname identifies the computer on a network.

Hostname:

9. Select time/zone and click **next** button.

Please select the nearest city in your time zone:

A world map showing various cities marked with yellow dots. The city of Kathmandu is highlighted with a red star and labeled. The map includes zoom in/out buttons and a scroll bar.

Selected city: Kathmandu, Asia


☒ System clock uses UTC




The root account is used for administering the system. Enter a password for the root user.

Root Password:

Confirm:






 Back

 Next


10. Enter strong root password and click **next** button.

11. Choose the type of installation. Here I will choose **replace existing Linux system(s)**. You can choose use all space.

Which type of installation would you like?

- ☐  **Use All Space**  
Removes all partitions on the selected device(s). This includes partitions created by other operating systems.  
**Tip:** This option will remove data from the selected device(s). Make sure you have backups.
- ☒  **Replace Existing Linux System(s)**  
Removes only Linux partitions (created from a previous Linux installation). This does not remove other partitions you may have on your storage device(s) (such as VFAT or FAT32).  
**Tip:** This option will remove data from the selected device(s). Make sure you have backups.
- ☐  **Shrink Current System**  
Shrinks existing partitions to create free space for the default layout.
- ☐  **Use Free Space**  
Retains your current data and partitions and uses only the unpartitioned space on the selected device(s), assuming you have enough free space available.
- ☐  **Create Custom Layout**  
Manually create your own custom layout on the selected device(s) using our partitioning tool.

- ☐ Encrypt system
- ☐ Review and modify partitioning layout

 Back

 Next

12. Click write changes to disk.


### Writing storage configuration to disk



The partitioning options you have selected will now be written to disk. Any data on deleted or reformatted partitions will be lost.

Go back

Write changes to disk

 Back

 Next

13. Choose basic server and click next.

The default installation of CentOS is a minimum install. You can optionally select a different set of software now.

- ☐ Desktop
- ☐ Minimal Desktop
- ☐ Minimal
- ☒ Basic Server
- ☐ Database Server
- ☐ Web Server
- ☐ Virtual Host
- ☐ Software Development Workstation

Please select any additional repositories that you want to use for software installation.

- ☒ CentOS

 Add additional software repositories

 Modify repository

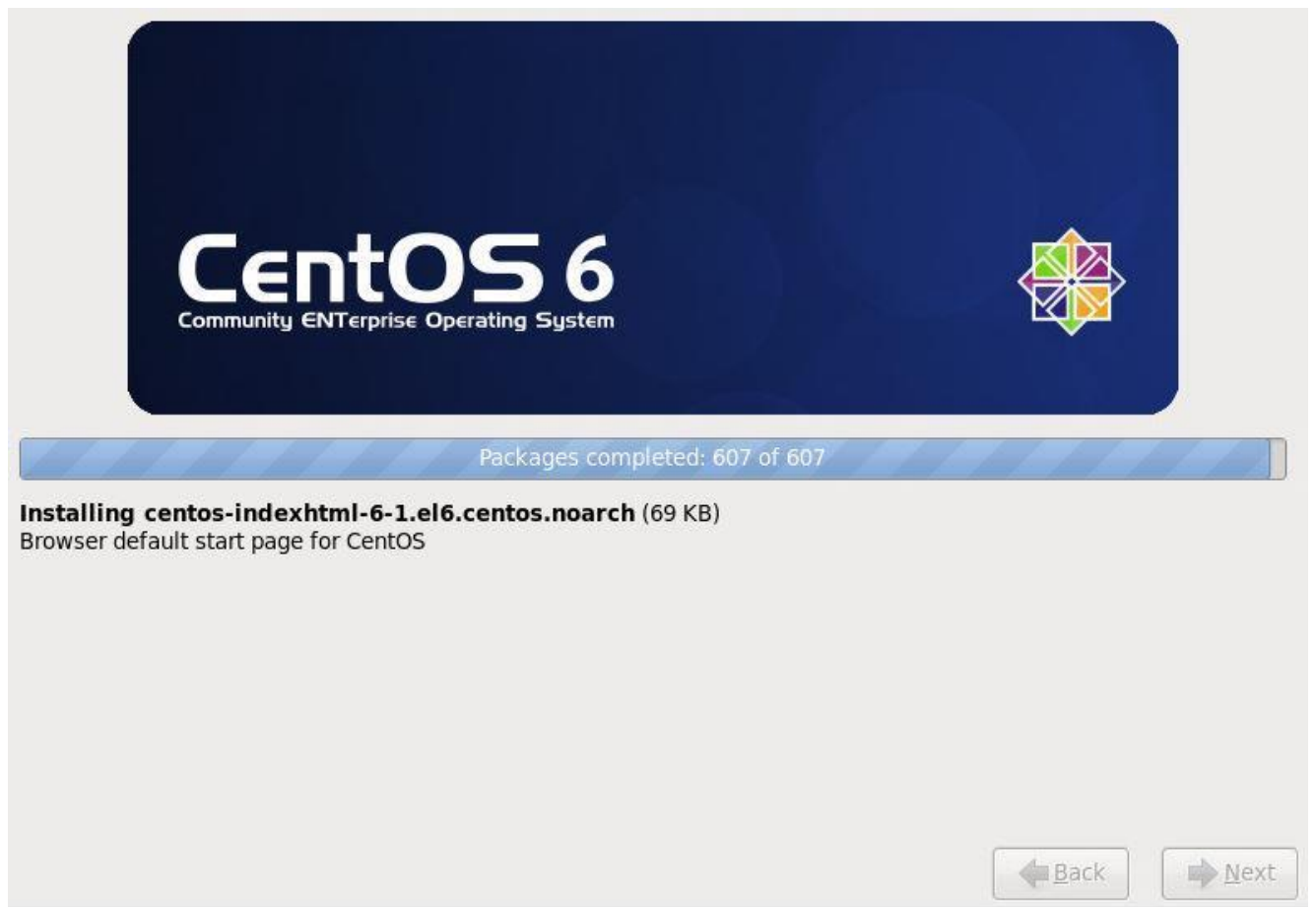
You can further customize the software selection now, or after install via the software management application.

☒ Customize later    ☐ Customize now

 Back

 Next

14. The installation will start now.



15. Click restart once finished.





Once done, it will also ask you to create a local user account. Create a local user named Hadoop. We are going to have lot of fun! That's it so far.

## Create a Hadoop1.2.1 cluster in local mode

### **Environment**

Centos 6 or above

JDK 6 or above

Hadoop-1.2.1 (Any stable release)

Follow these steps for installing and configuring Hadoop on a single node:

### **Step-1. Install Java**

In this tutorial, we will use Java 1.6 therefore describing the installation of Java 1.6 in detail.

Use the below command to begin the installation of Java

```
1$ sudo apt-get install openjdk-6-jdk
```

or

```
1$ sudo apt-get install sun-java6-jdk
```

This will install the full JDK under /usr/lib/jvm/java-6-sun directory.

## Step-2. Verify Java installation

You can verify java installation using the following command

```
1$ java -version
```

On executing this command, you should see output similar to the following:

```
java version "1.6.0_27"
```

```
Java(TM) SE Runtime Environment (build 1.6.0_45-b06)
```

```
Java HotSpot(TM) 64-Bit Server VM (build 20.45-b01, mixed mode)
```

## Step-3. SSH configuration

- Install SSH using the command.  

```
1sudo apt-get install ssh
```
- Generate ssh key  

```
ssh -keygen -t rsa -P ""
```

 (press enter when asked for a file name; this will generate a passwordless ssh file)
- Now copy the public key (id\_rsa.pub) of current machine to authorized\_keys. Below command copies the generated public key in the .ssh/authorized\_keys file:  

```
1cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```
- Verify ssh configuration using the command  

```
1ssh localhost
```

Pressing yes will add localhost to known hosts

## Step-4. Download Hadoop

Download the latest stable release of Apache Hadoop from <http://hadoop.apache.org/releases.html>.

Unpack the release tar – 

```
zxvf hadoop-1.2.1.tar.gz
```

Save the extracted folder to an appropriate location, HADOOP\_HOME will be pointing to this directory.

## Step-5. Verify Hadoop

Check if the following directories exist under HADOOP\_HOME: bin, conf, lib, bin

Use the following command to create an environment variable that points to the Hadoop installation directory (HADOOP\_HOME)

```
1export HADOOP_HOME=/home/user/hadoop
```

Now place the Hadoop binary directory on your command-line path by executing the command

```
1export PATH=$PATH:$HADOOP_HOME/bin
```

Use this command to verify your Hadoop installation:

```
hadoop version
```

The o/p should be similar to below one

```
Hadoop 1.2.1
```

### **Step-6. Configure JAVA\_HOME**

Hadoop requires Java installation path to work on, for this we will be setting JAVA\_HOME environment variable and this will point to our Java installation dir.

Java\_Home can be configured in ~/.bash\_profile or ~/.bashrc file. Alternatively you can also let hadoop know this by setting Java\_Home in hadoop **conf/hadoop-env.sh** file.

Use the below command to set JAVA\_HOME on Ubuntu

```
1 export JAVA_HOME=/usr/lib/jvm/java-6-sun
```

JAVA\_HOME can be verified by command

```
1 echo $JAVA_HOME
```

### **Step-7. Create Data Directory for Hadoop**

An advantage of using Hadoop is that with just a limited number of directories you can set it up to work correctly. Let us create a directory with the name hdfs and three sub-directories name, data and tmp.

Since a Hadoop user would require to read-write to these directories you would need to change the permissions of above directories to 755 or 777 for Hadoop user.

### **Step-8. Configure Hadoop XML files**

Next, we will configure Hadoop XML file. Hadoop configuration files are in the HADOOP\_HOME/conf dir.

#### **conf/core-site.xml**

```
1 <!--?xml version="1.0"-->
2 <!--?xml -stylesheet type="text/xsl" href="configuration.xsl"?-->
3 <!-- Putting site-specific property overrides the file. -->
4
5 fs.default.name
6 hdfs://localhost:9000
7
8 hadoop.temp.dir
9 /home/girish/hdfs/temp<span style="font-family: Georgia, 'Times New Roman',
10 'Bitstream Charter', Times, serif; font-size: 13px; line-height: 19px;"> </span>
```

#### **conf/hdfs-site.xml**

```
1 <!-- Putting site specific property overrides in the file. -->
2
3 dfs.name.dir
4 /home/girish/hdfs/name
5
6 dfs.data.dir
```

```
6 /home/girish/hdfs/data
7
8 dfs.replication
9 1
10
11<strong style="font-family: Georgia, 'Times New Roman', 'Bitstream Charter',
1Times, serif; font-size: 13px; line-height: 19px;">conf/mapred-site.xml</strong>
1<! -- Putting site-specific property overrides this file. -->
2
3mapred.job.tracker
4localhost:9001
```

### **conf/masters**

Not required in single node cluster.

### **conf/slaves**

Not required in single node cluster.

## **Step-9. Format Hadoop Name Node-**

Execute the below command from hadoop home directory

```
1$ ~/hadoop/bin/hadoop namenode -format
```

## **Step-10. Start Hadoop daemons**

```
1$ ~/hadoop/bin/start-all.sh
```

## **Step-11. Verify the daemons are running**

```
1$ jps (if jps is not in path, try /usr/java/latest/bin/jps)
```

output will look similar to this

9316 SecondaryNameNode

9203 DataNode

9521 TaskTracker

9403 JobTracker

9089 NameNode

Now we have all the daemons running: