

## Practical No.: 1

**Aim: Install, configure and run Hadoop and HDFS and explore HDFS.**

### Step 1: Download and install VirtualBox

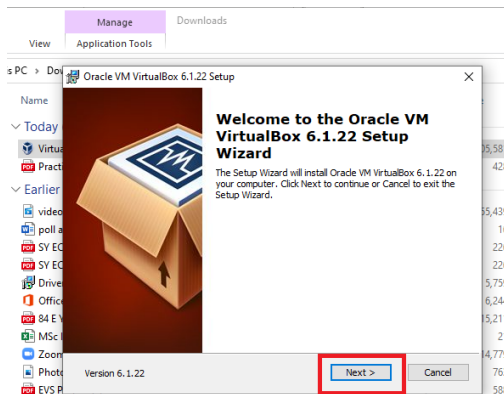
Go to the website of Oracle VirtualBox and get the latest stable version from the following site

<https://www.virtualbox.org/>

click on 'Download'

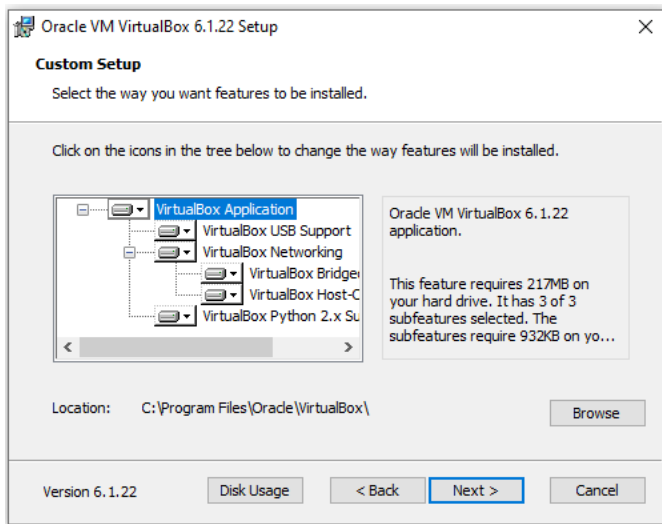


You will get VirtualBox-6.1.22-144080-Win.exe file downloaded. Double click and run it. Click on next.

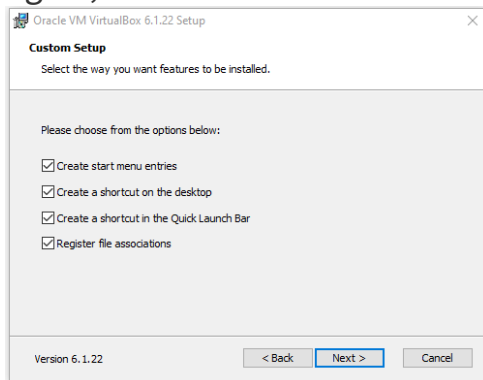


Click on 'next' without changing the default folder as shown below:

## BDA – Lab manual for practical 1



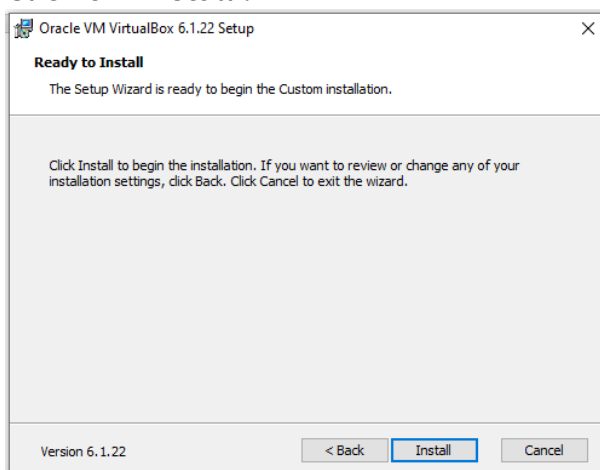
Again, click on next as shown below:



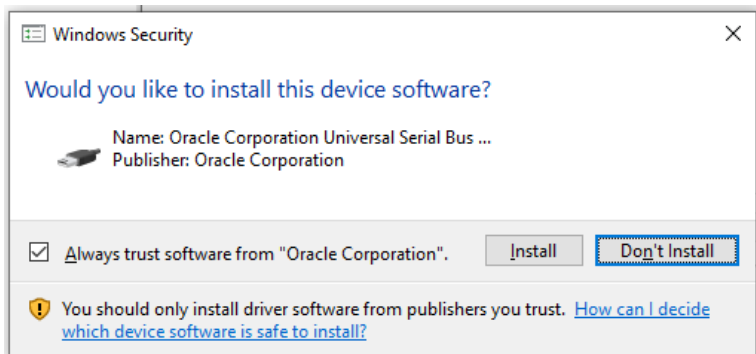
Finally, click on 'Yes'.



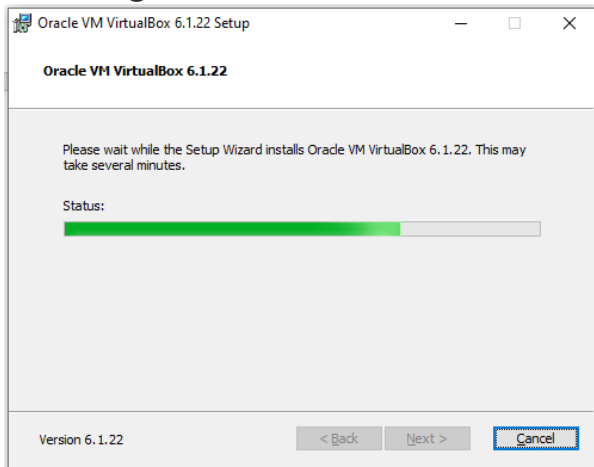
Click on 'Install'.



It may ask you for the permission to install, click 'yes' to allow.  
Select 'Install' as shown below:



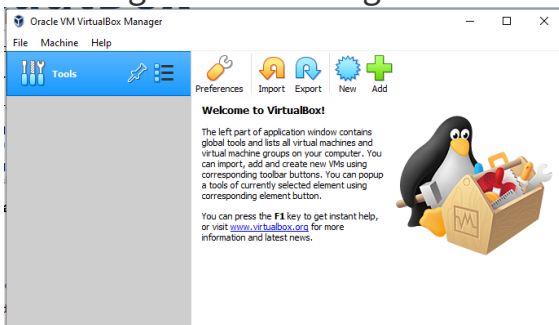
You will get the screen as shown below:



Click on 'Finish' to finish Installation of virtual box.



You will get the following screen:



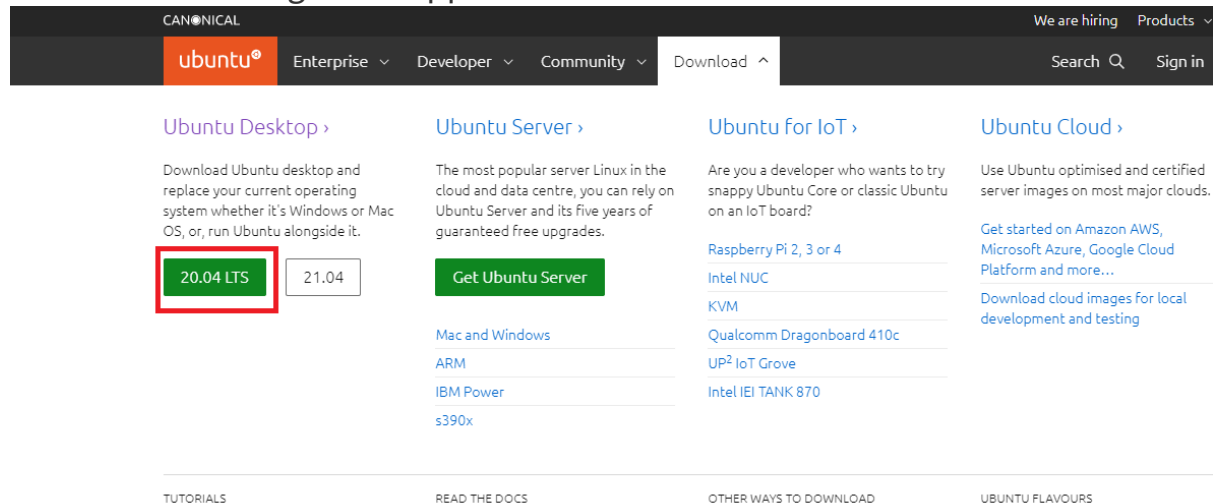
## Step 2: download Ubuntu

Download iso file ubuntu-20.04.2.0-desktop-amd64; which is required to install Ubuntu.

Browse ubuntu.com

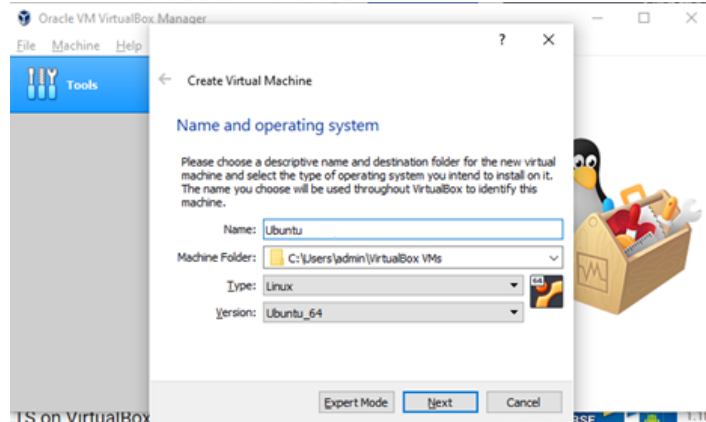
Click on download and 20.04 LTS as shown below:

LTS stands for Long term support

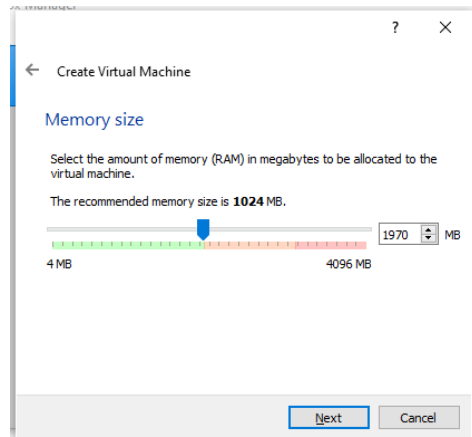


You will get file, which may take few minutes to download.

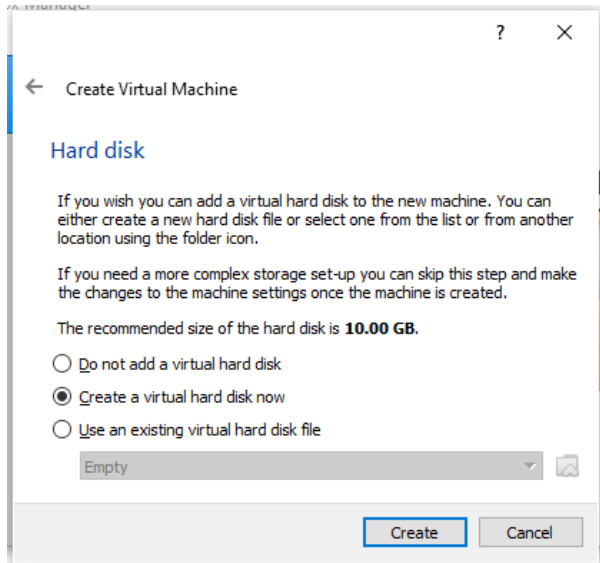
Now, click on 'New' to virtual box and write Name as 'Ubuntu' as shown below:



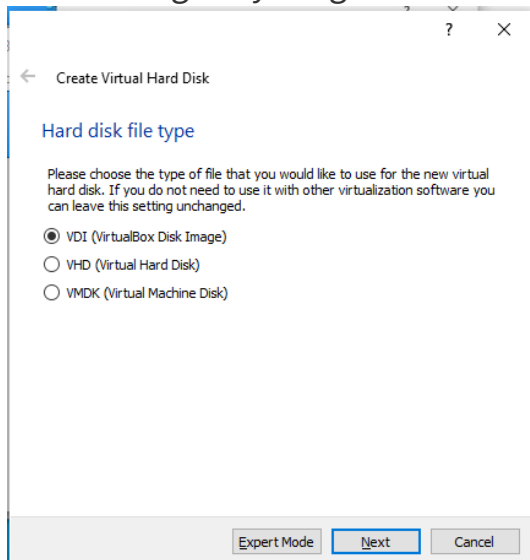
Click on 'Next'.



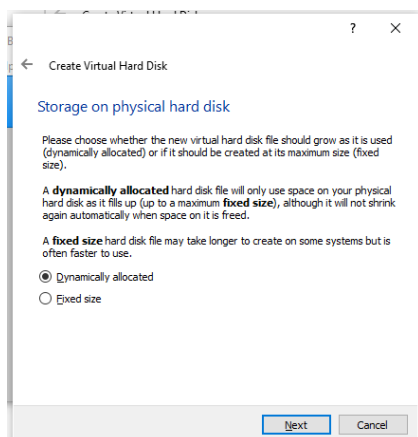
Here, you allow memory size up to green indicator (1970 MB). Click on 'Next'.



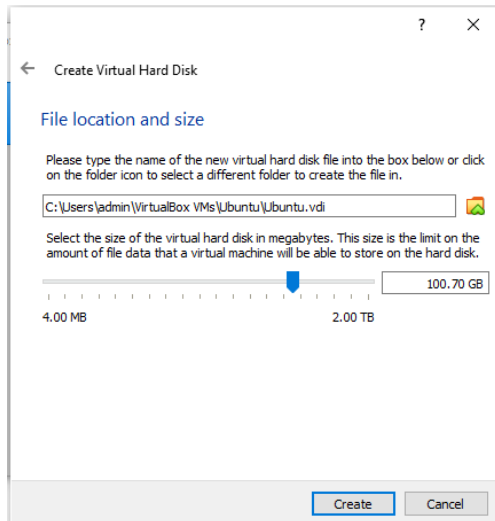
Don't change anything in this screen and click on 'Create'.



Click on 'Next', keeping the selection as it is (on VDI).'

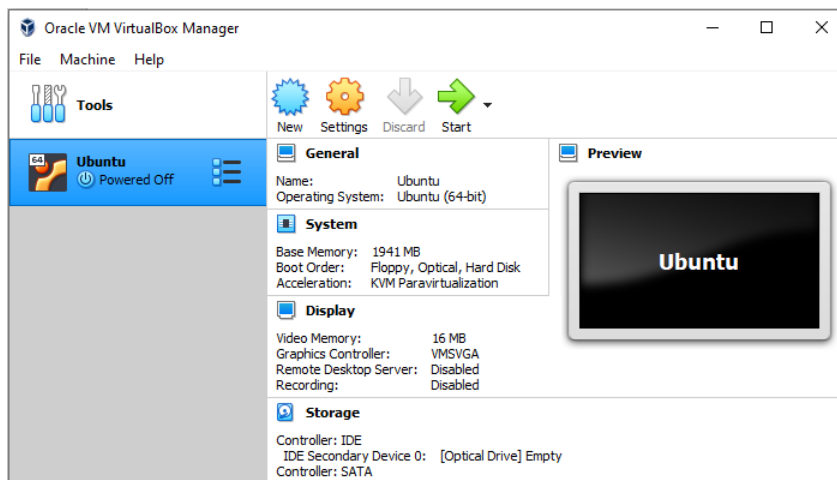


Keep this screen also as it is and click on ‘Next’.



Keep the file location as it is but preferably keep size 100 GB and click on ‘Create’.

You may see the following screen having Ubuntu on Virtual Machine.



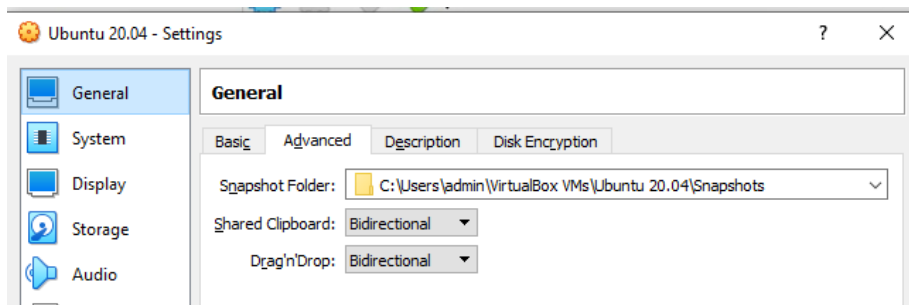
Select ‘General’ -> ‘Basic’ as shown below:

You may change the name from Ubuntu to Ubuntu 20.04

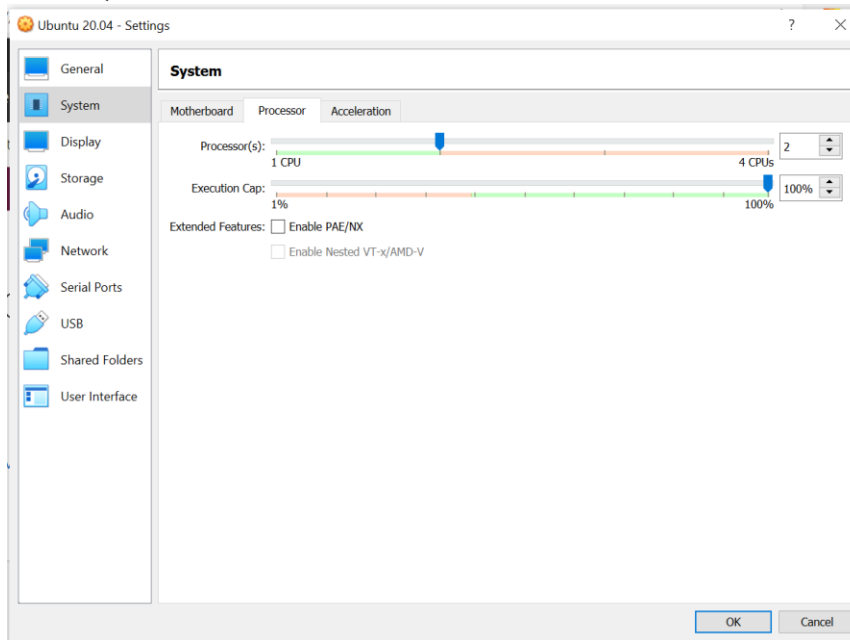
Click on Ubuntu and then click on settings option as shown below:

Select bidirectional in ‘General’ -> ‘Advanced’ as shown below:

## BDA – Lab manual for practical 1



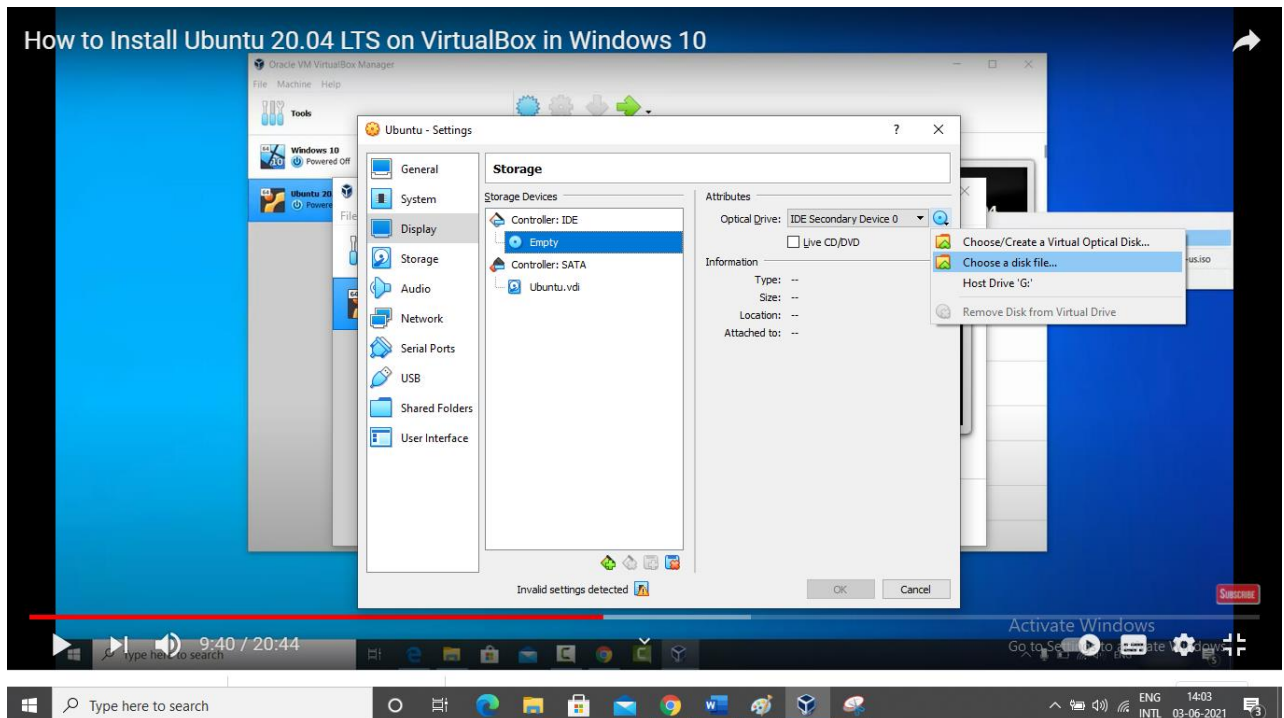
Go to 'System' option and change the processor up to green bar, usually 4.(if it allows)



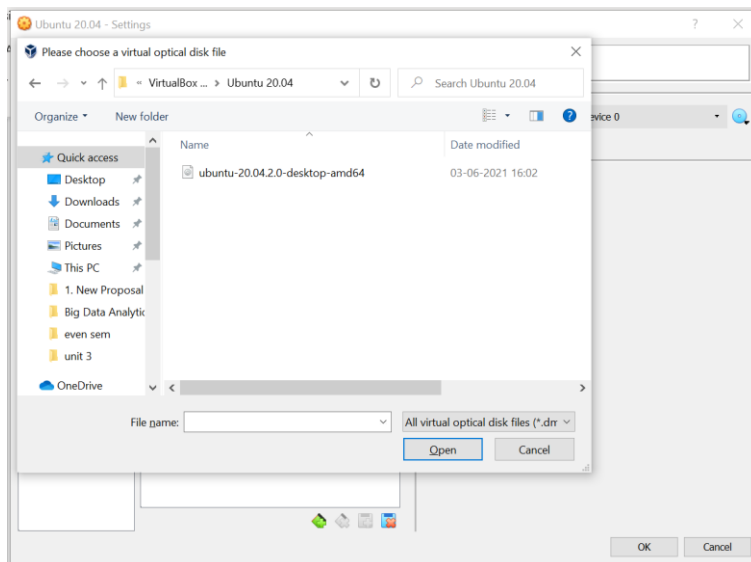
Cut and paste your ubuntu .iso file from current folder to C:\Users\ADMIN\VirtualBox VMs\Ubuntu 20.04 folder.

Click on 'Storage' and click on 'Empty' followed by 'Choose a disk file' as shown below:

## BDA – Lab manual for practical 1



Browse the folder where you have selected ubuntu iso file.

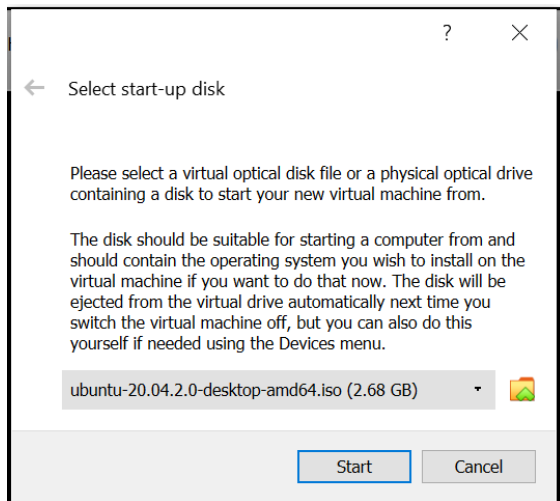


Click on Ubuntu....iso file and click on open and then click on ok.

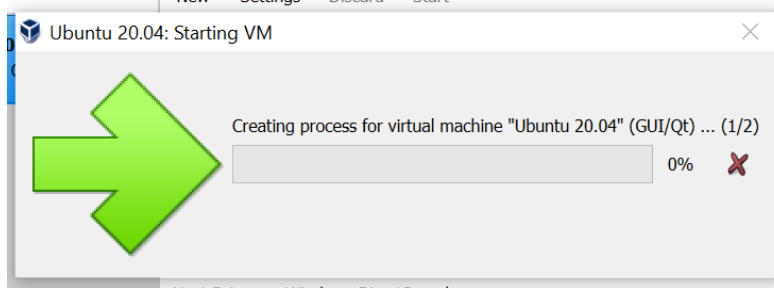
Click on Ubuntu -> start button.



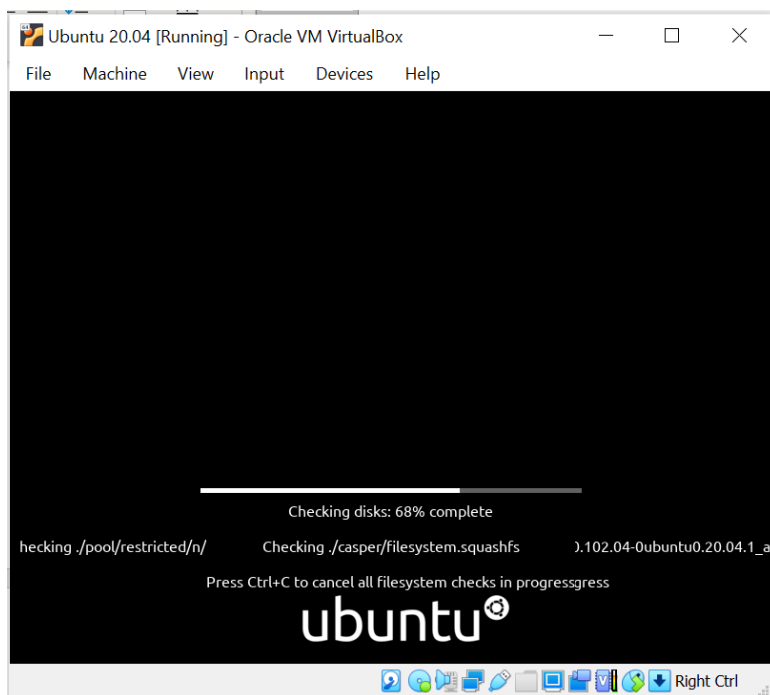
## BDA – Lab manual for practical 1



Again, click on 'Start' button. It will show you the following screen.

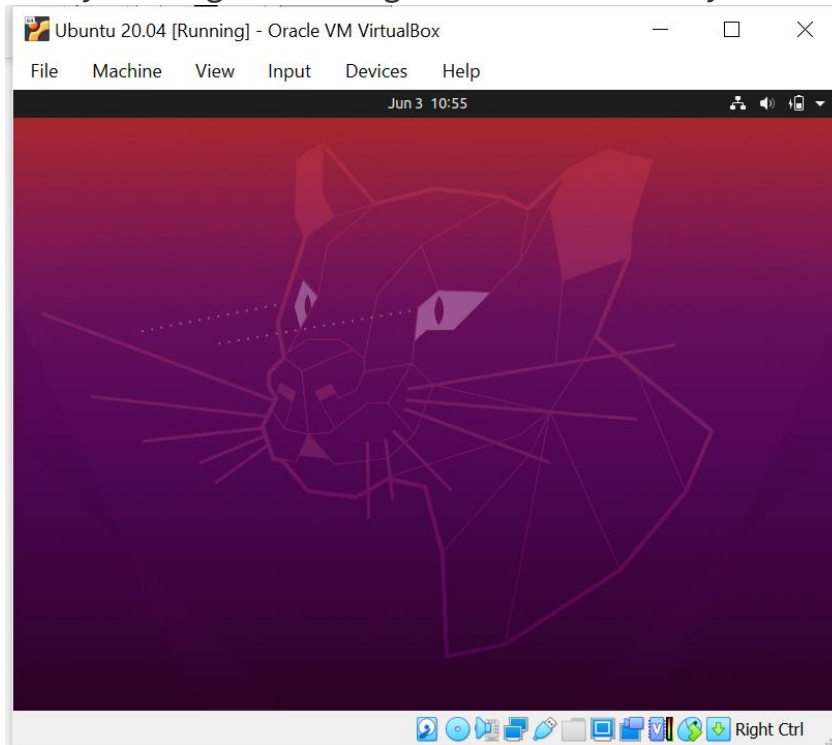


And simultaneously one more screen as follows:



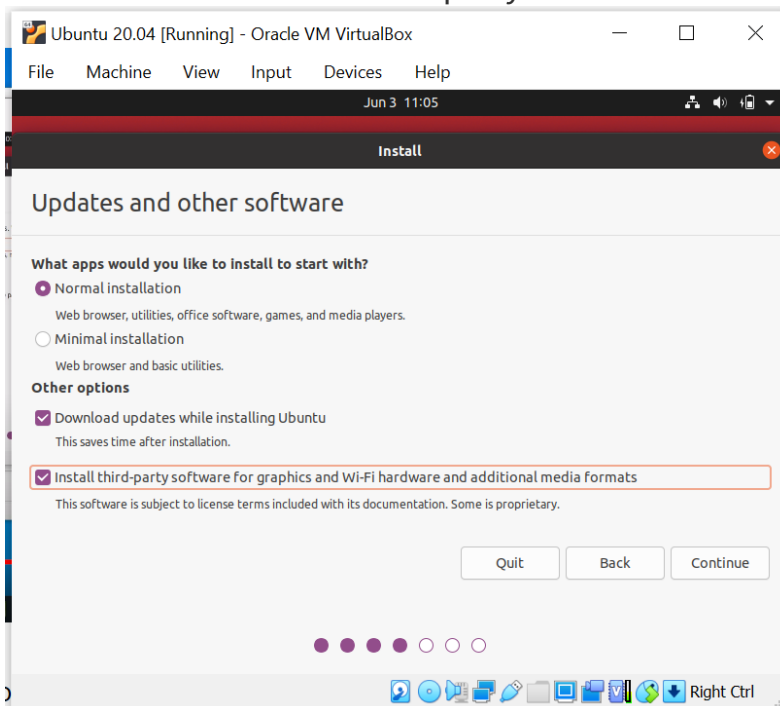
Keep on closing all warnings.

Next you will get following screen automatically.



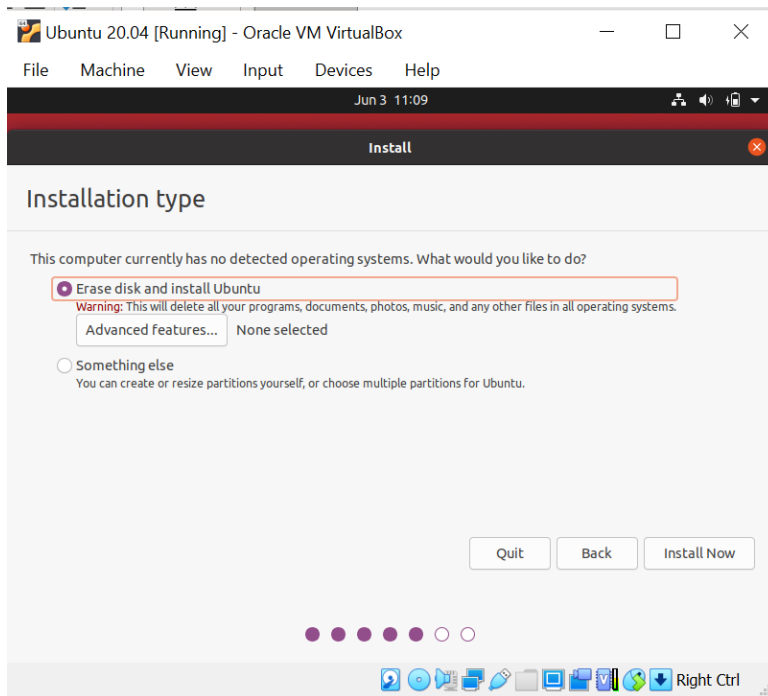
Select language -> English and click on 'Install Ubuntu'. In 'Keyboard Layout' screen, select 'English UK'. Click on 'Continue'.

Select the checkbox for third party software as shown below:

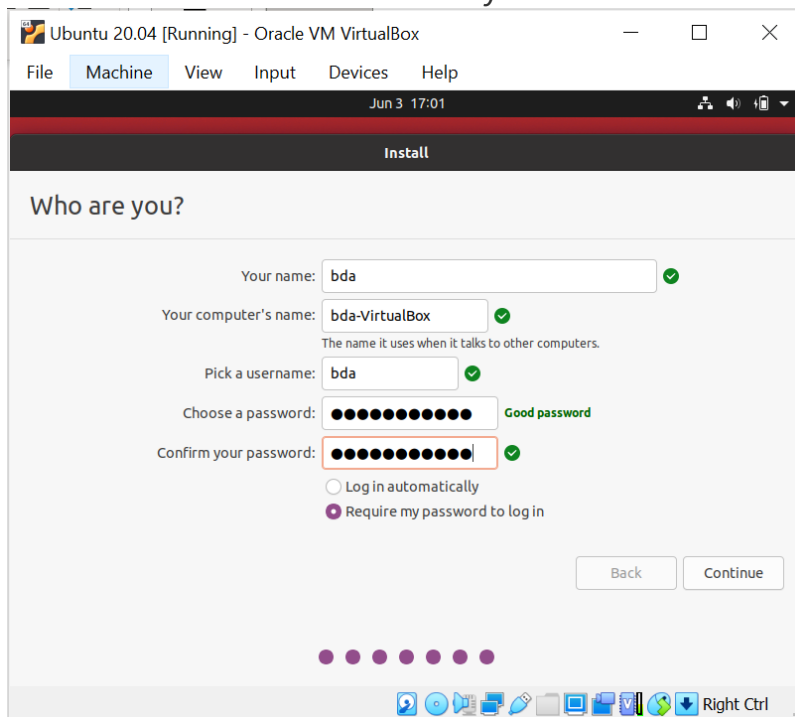


Click on 'continue'.

## BDA – Lab manual for practical 1

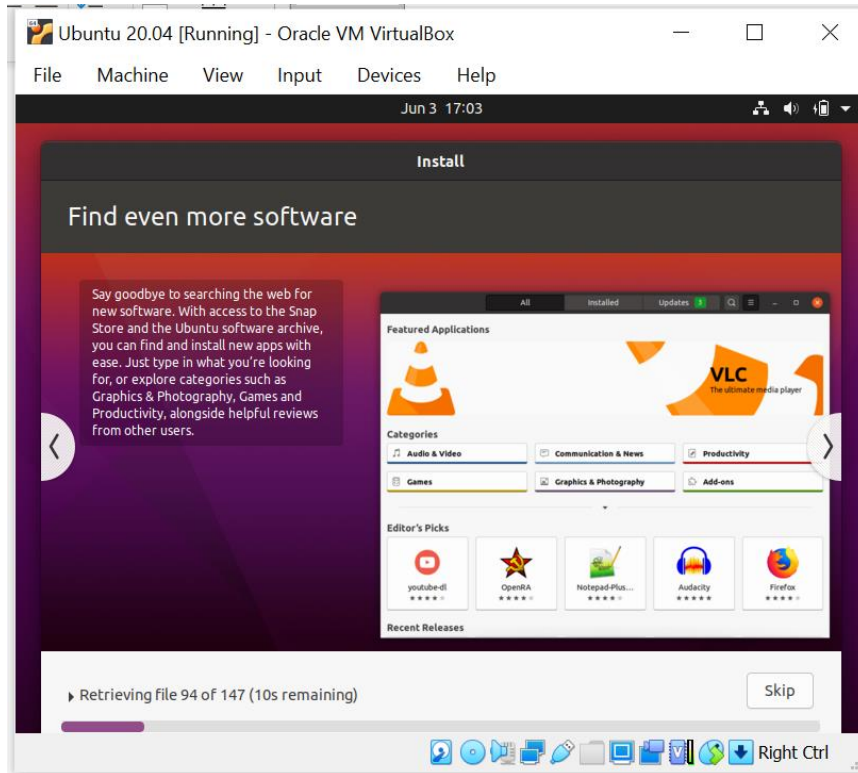


Select Erase disk and Install Ubuntu and click on **‘Install Now’**.  
Click on **‘Continue’** on the next screen.  
Select “Kolkata” for “where are you?” and click on **‘Continue’**.



Click on continue after entering name, company name, username, password and confirm your password.

## BDA – Lab manual for practical 1



Installation of Ubuntu started. Click on finish once installation done.  
Click on restart and press Enter key.

### Step 3 Install Hadoop

Login to ubuntu

Some keys may change like you try to type @ and it types “.

\*\* please refer to note - Some Keys for Ubuntu under UK keyboard layout – at the end.

Search for Ubuntu terminal on search bar, after login done.

Apply following commands from ubuntu terminal

```
$ sudo apt update
$ sudo apt install default-jdk
$ java -version
$ wget https://hadoop.apache.org/release/3.2.2.html/hadoop-3.2.2.tar.gz
$ tar xzvf hadoop-3.2.2.tar.gz
$ sudo mv hadoop-3.2.2 /usr/local/hadoop
$ readlink -f /usr/bin/java | sed "s:bin/java::"
```

# : Configuring Hadoop's Java Home; To begin, open hadoop-env.sh

```
$ sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

File will be opened. Add the following line at the end of .sh file

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64/
```

to save the changes in the file, press ctrl and x together.

then press Y

then press Enter key

then apply following commands:

```
$ /usr/lib/jvm/java-11-openjdk-amd64/
```

#### **Step 4: Running Hadoop**

```
$ /usr/local/hadoop/bin/hadoop
```

```
$ mkdir ~/input
```

```
$ cp /usr/local/hadoop/etc/hadoop/*.xml ~/input
```

We can use the following command to run the MapReduce hadoop mapreduce-examples program, a Java archive with several options. We'll invoke its grep program, one of the many examples included in hadoop-mapreduce-examples, followed by the input directory, input and the output directory grep\_example. The MapReduce grep program will count the matches of a literal word or regular expression. Finally, we'll supply the regular expression allowed[.]\* to find occurrences of the word allowed within or at the end of a declarative sentence. The expression is case-sensitive, so we wouldn't find the word if it were capitalized at the beginning of a sentence:

```
$ /usr/local/hadoop/bin/hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar grep ~/input ~/grep_example 'allowed[.]*'
```

```
$ cat ~/grep_example/*
```

#### **\*\*Note:**

##### **Some Keys for Ubuntu under UK keyboard layout**

“ -> @

@ -> “

pipe -> take from this file or on google search for pipe in linux

~ -> pipe