

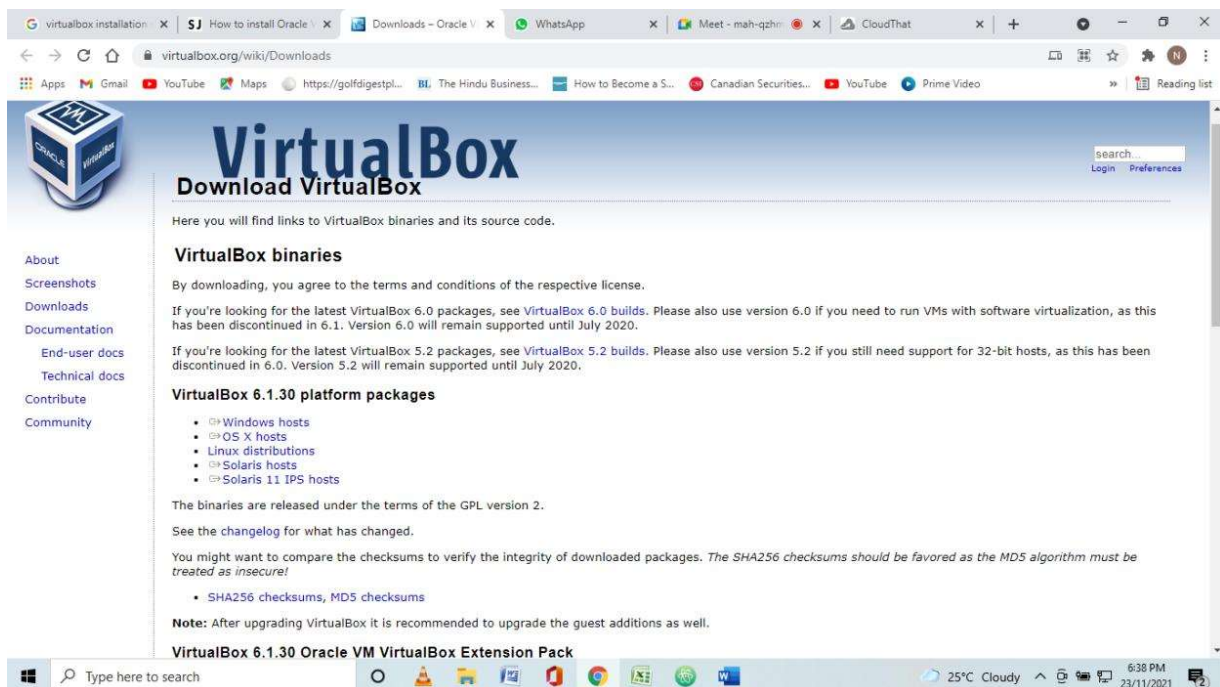
Nandani soni
18100BTCCE04031

EXPERIMENT-1

AIM-: Installation Process of Oracle VM VirtualBox.

Step 1-:

Download VirtualBox from the given link.
<https://www.virtualbox.org/wiki/Downloads>



Step 2-:

The file downloaded will have the file name format like VirtualBox-VersionNumber-BuildNumber-Win.exe. Something like this: VirtualBox-6.1.6-137129-Win.exe. Double click on the installer to launch the setup Wizard.

Nandani soni
18100BTCCE04031



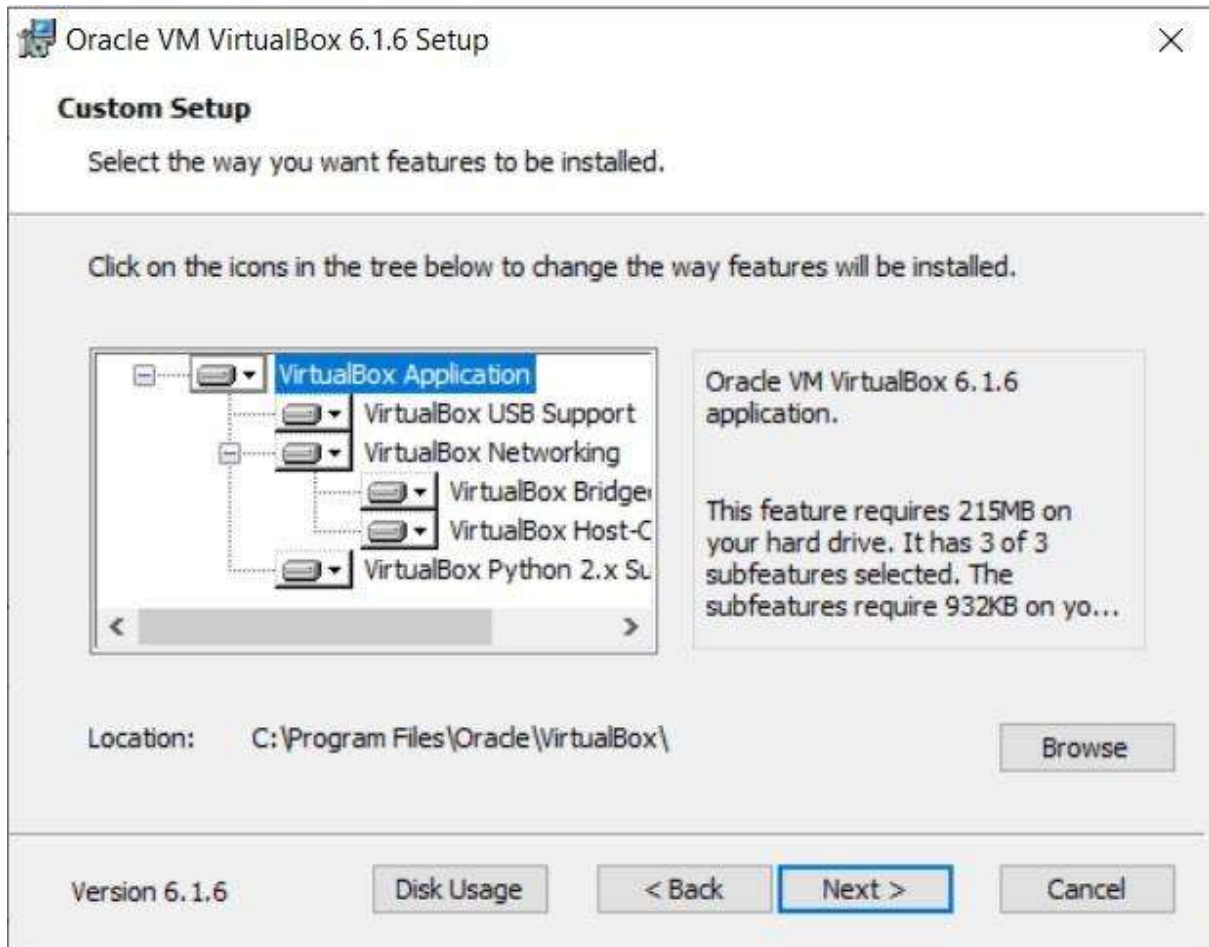
Click next to continue.

Step 3-:

You will see custom setup dialog box. There is not much to choose from. You can accept the default and click next.

If you wish to change the installation directory, you can change it by clicking on the browse button and selecting the new directory and clicking OK. Normally I leave it as the default as the whole installation process does not take much space on your hard drive.

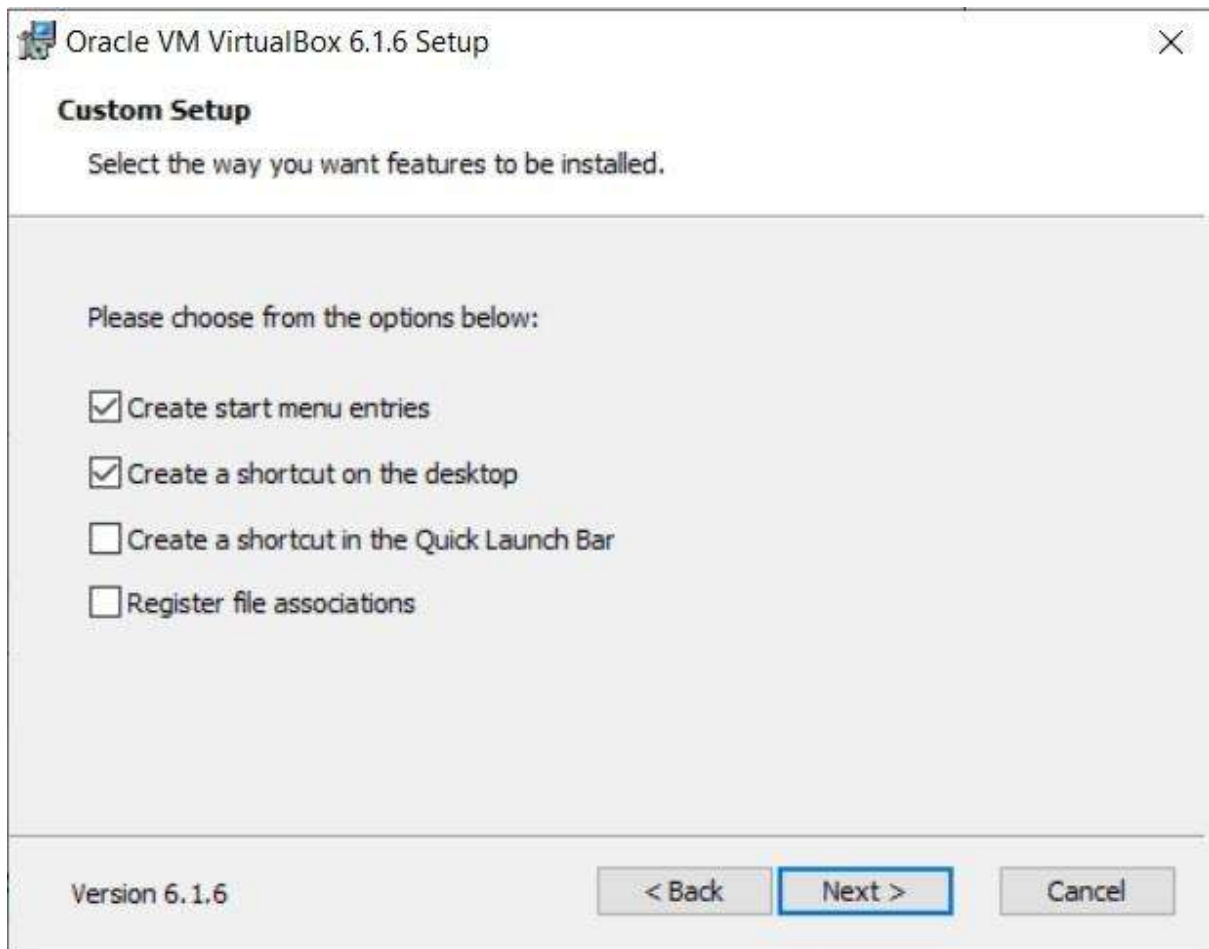
Nandani soni
18100BTCCE04031



Step 4-:

In this dialog box you can choose which features to install. As you can see, there is not much to choose. You can accept the default and click next. Normally I uncheck Create a shortcut in the quick launch bar and register file association.

Nandani soni
18100BTCCE04031



Step 5-:

This dialog box warns you about setting up a Network Interface. what this means that VirtualBox will install network interfaces that will interact with the installed virtual machines and the host operating system which in our case is windows. This will temporarily disconnect you from the internet but that OK, nothing to worry.

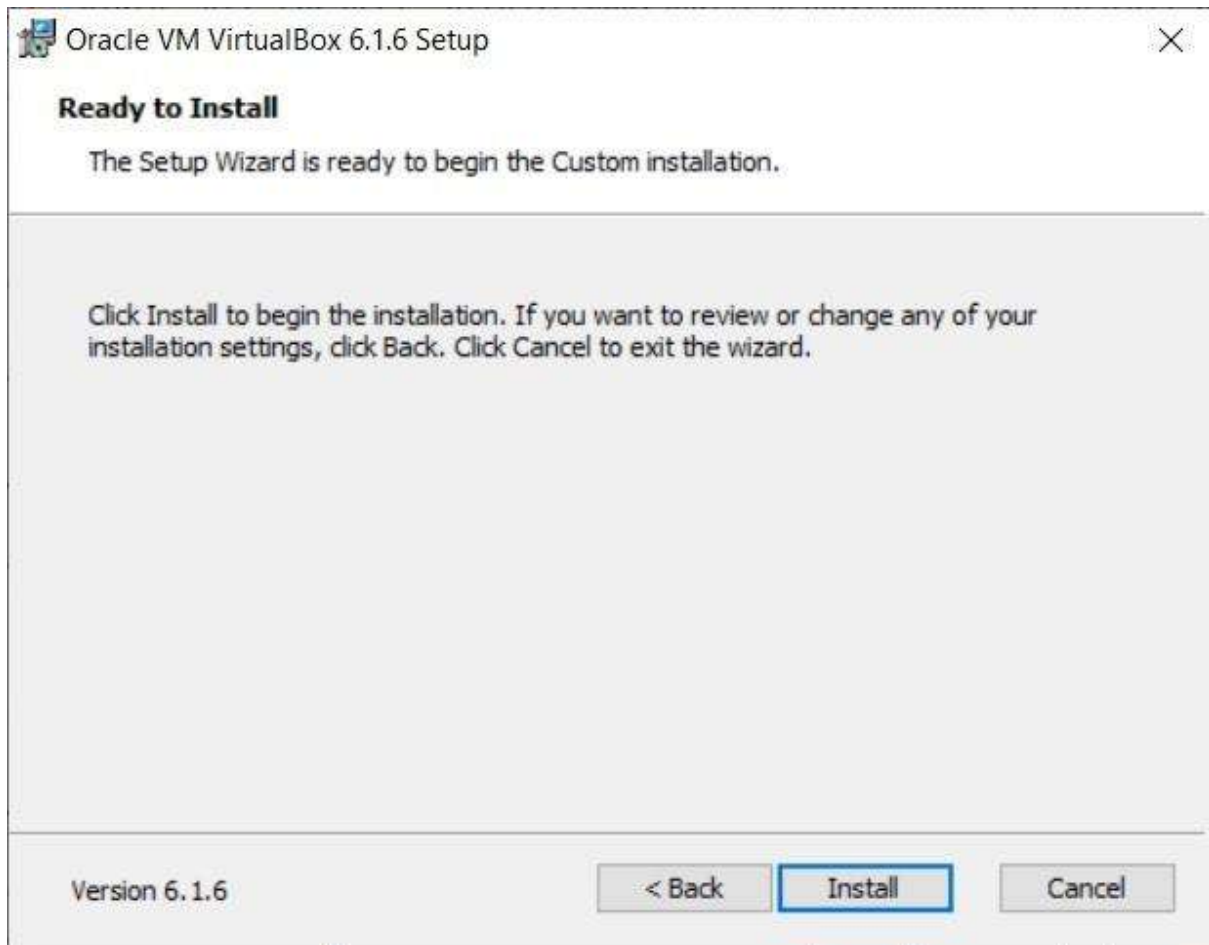
Nandani soni
18100BTCCE04031



Step 6-:

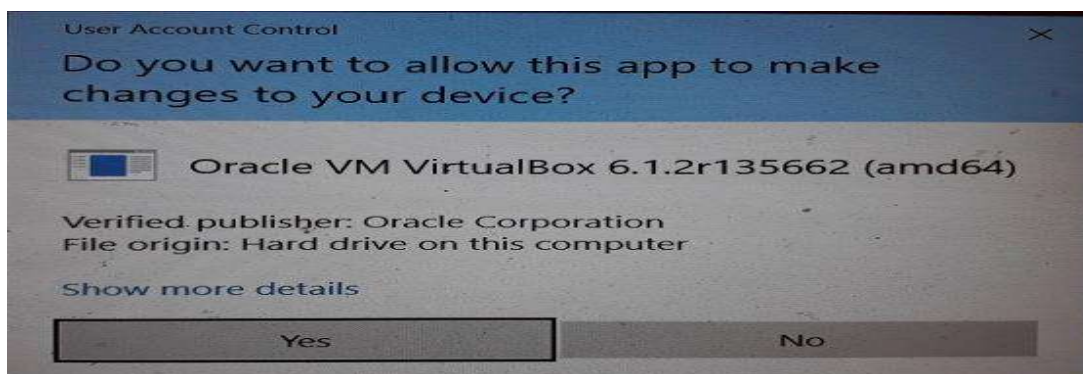
You will see ready to install dialog box.

Nandani soni
18100BTCCE04031

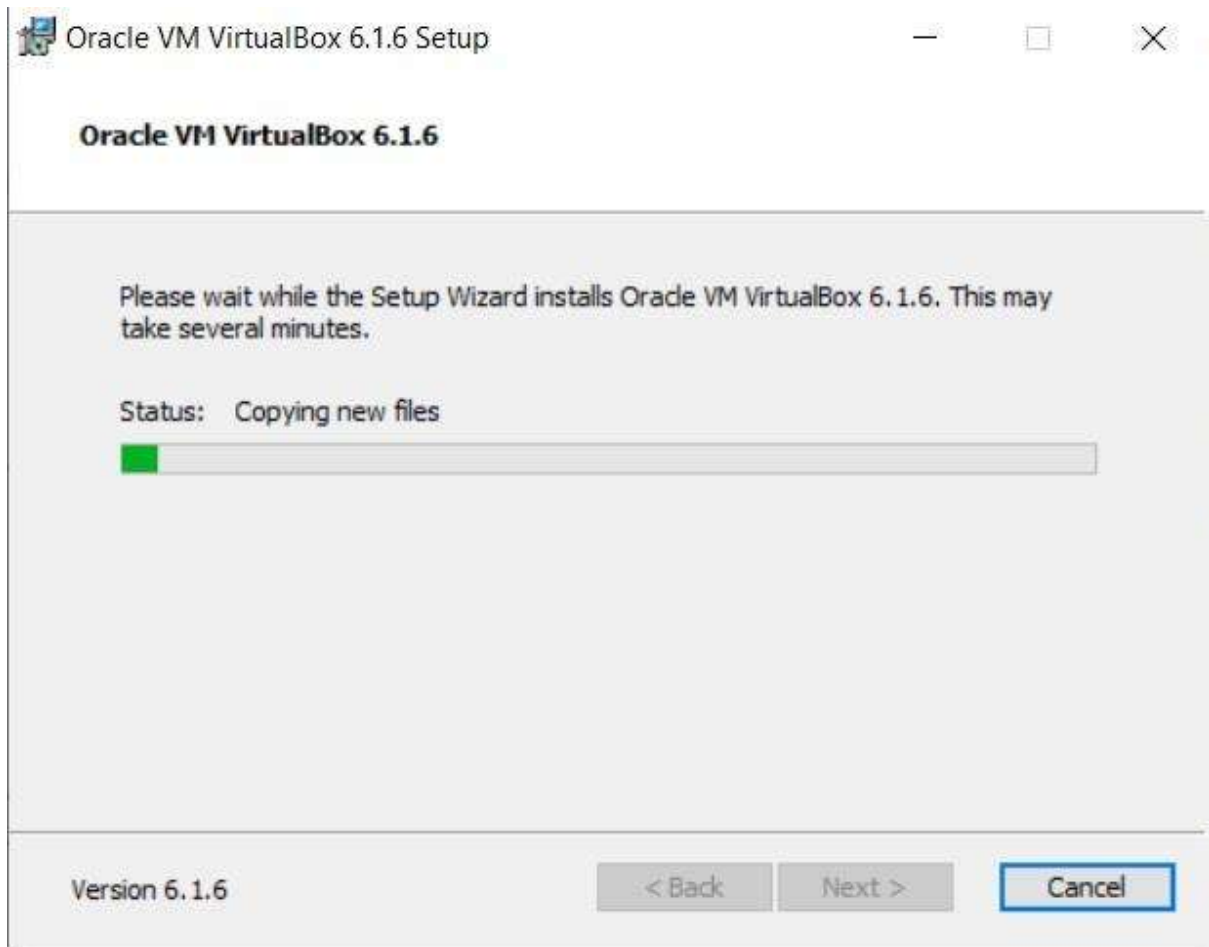


Step 7-:

After clicking install, you will mostly probably see User access control confirmation dialog box from Windows OS. This is a security feature in Windows that wants to confirm if the application should be allowed to proceed with the installation process. Click Yes to continue and you will see that the installation process will begin. Wait for the installation to complete.

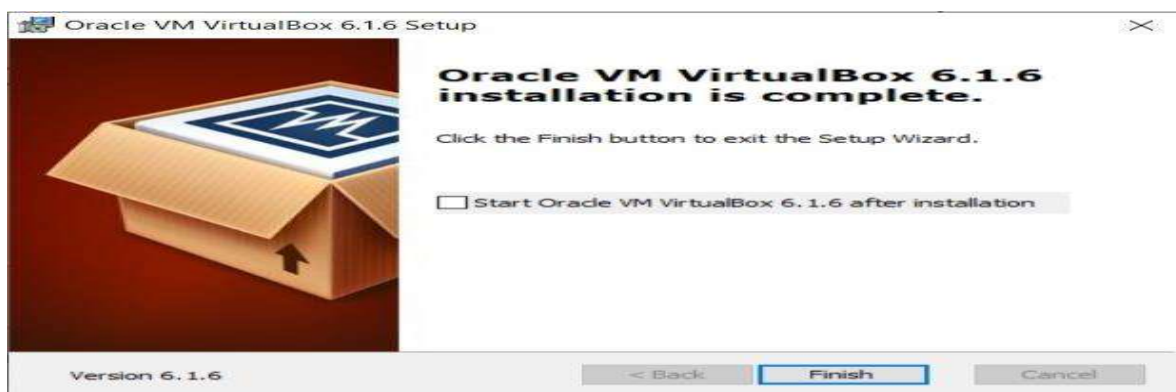


Nandani soni
18100BTCCE04031



Step 8:-

After the installation completes, you will see installation completion dialog box. Click finish. If you leave Start Oracle VM VirtualBox after installation checked, VirtualBox will launch automatically. If you uncheck it, you will have to launch VirtualBox manually. It's no big deal anyways. Click finish to exit installation.



EXPERIMENT-2

AIM-: To Install Hadoop.

These software should be prepared to install Hadoop 2.8.0 on window 10 64bit

1. Download Hadoop 2.8.0

(Link: <http://www-eu.apache.org/dist/hadoop/common/hadoop-2.8.0/hadoop-2.8.0.tar.gz>)

OR

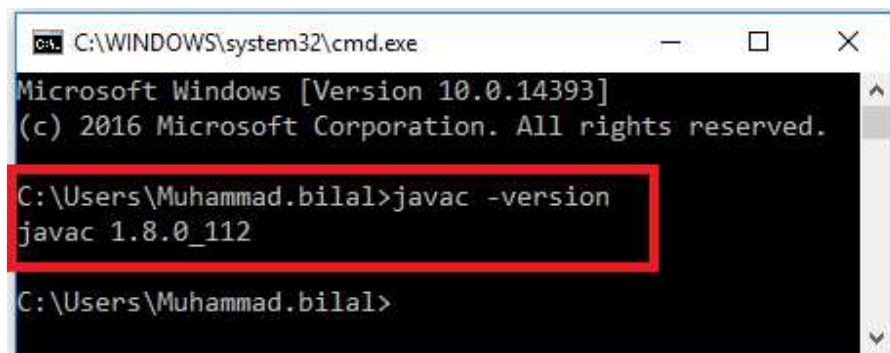
<http://archive.apache.org/dist/hadoop/core/hadoop-2.8.0/hadoop-2.8.0.tar.gz>)

2. Java JDK 1.8.0.zip

(Link: <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>)

Set up

- 1) Check either Java 1.8.0 is already installed on your system or not, use "Javac -version" to check.



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

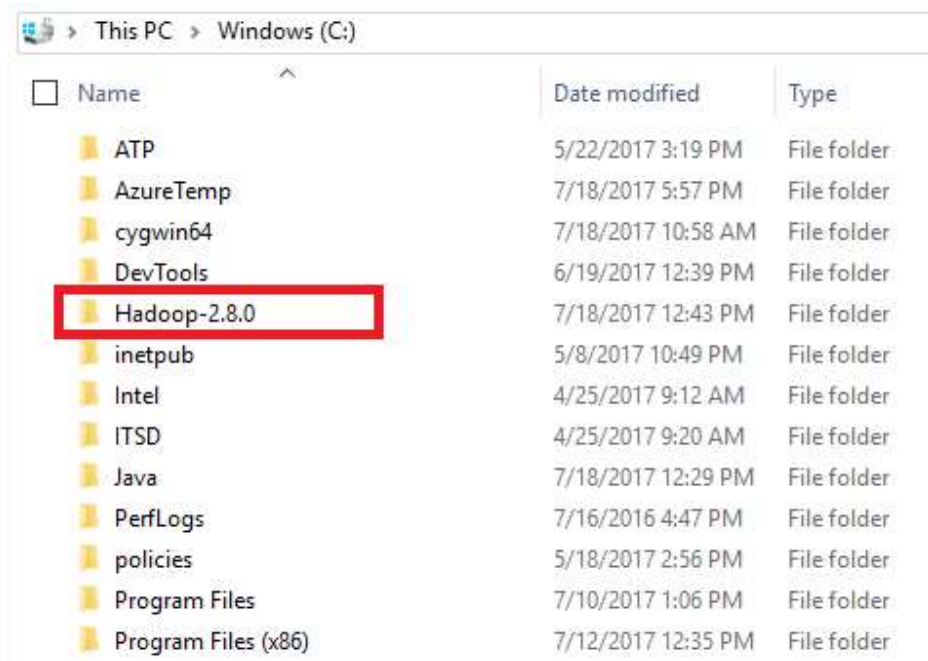
C:\Users\Muhammad.bilal>javac -version
javac 1.8.0_112

C:\Users\Muhammad.bilal>
```

Figure 2.1. Checking Java Version

- 2) If Java is not installed on your system then first install java under "C:\JAVA"
- 3) Extract file Hadoop 2.8.0.tar.gz or Hadoop-2.8.0.zip and place under "C:\Hadoop-2.8.0".

Nandani soni
18100BTCCE04031



The image shows a Windows File Explorer window titled 'This PC > Windows (C:)'. It displays a list of folders in the C: drive. The folder 'Hadoop-2.8.0' is highlighted with a red rectangular box. The table below represents the data shown in the File Explorer window.

<input type="checkbox"/> Name	Date modified	Type
ATP	5/22/2017 3:19 PM	File folder
AzureTemp	7/18/2017 5:57 PM	File folder
cygwin64	7/18/2017 10:58 AM	File folder
DevTools	6/19/2017 12:39 PM	File folder
Hadoop-2.8.0	7/18/2017 12:43 PM	File folder
inetpub	5/8/2017 10:49 PM	File folder
Intel	4/25/2017 9:12 AM	File folder
ITSD	4/25/2017 9:20 AM	File folder
Java	7/18/2017 12:29 PM	File folder
PerfLogs	7/16/2016 4:47 PM	File folder
policies	5/18/2017 2:56 PM	File folder
Program Files	7/10/2017 1:06 PM	File folder
Program Files (x86)	7/12/2017 12:35 PM	File folder

Figure 2.2. Installation of Hadoop

4) Set the path HADOOP_HOME Environment variable on windows 10(see Step 1,2,3 and 4 below).

Nandani soni
18100BTCCE04031

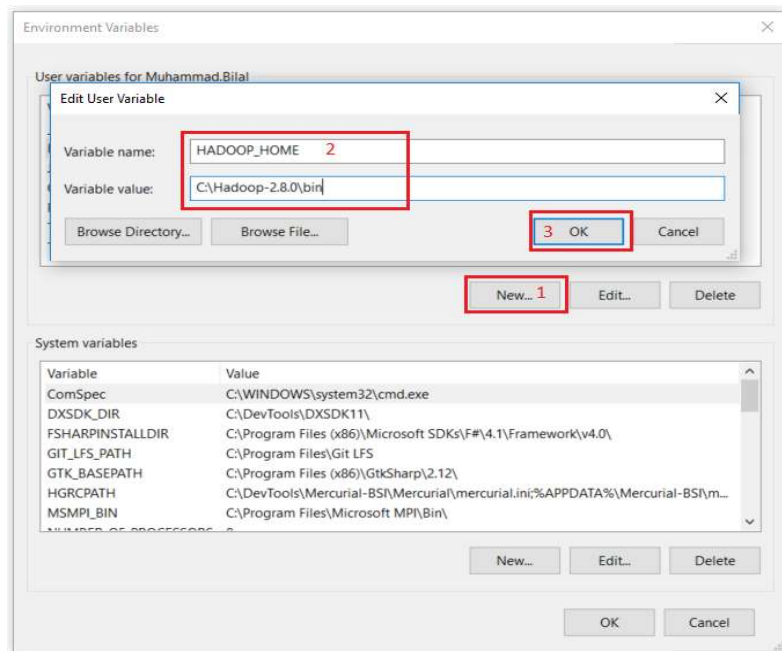


Figure 2.3. Setting Hadoop Environment Path

5) Set the path JAVA_HOME Environment variable on windows 10(see Step 1,2,3 and 4 below).

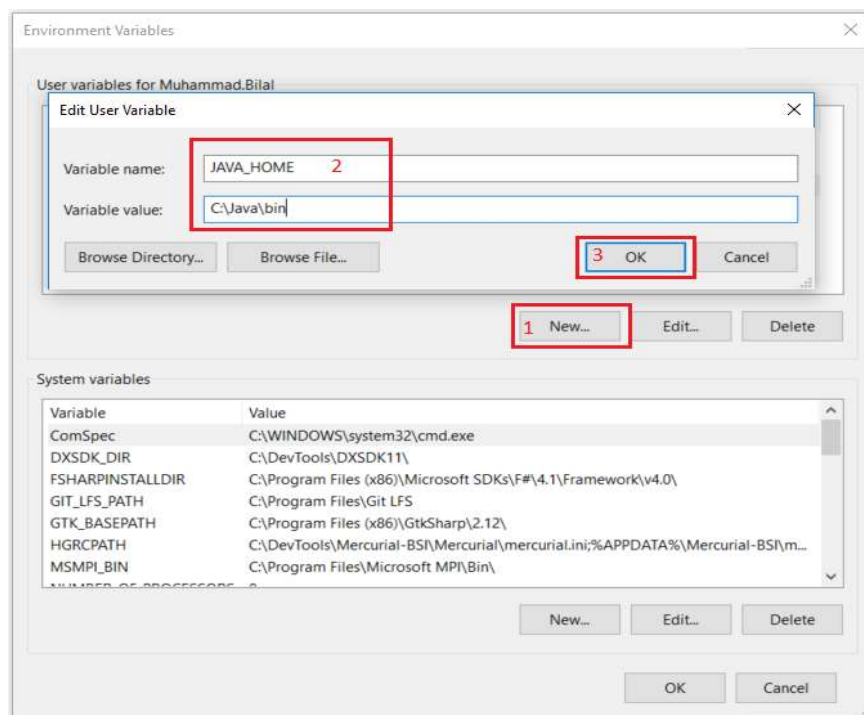


Figure 2.4. Setting Java Environment Path

Nandani soni
18100BTCCE04031

6) Next we set the Hadoop bin directory path and JAVA bin directory path.

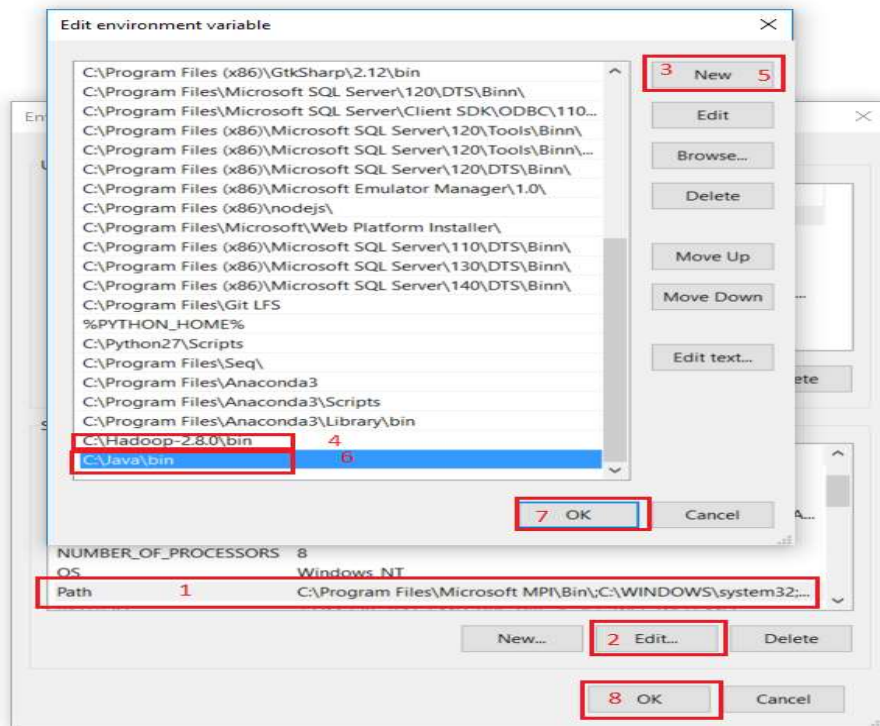


Figure 2.5. Setting Hadoop bin Directory Path

2.3 Configuration

1) Edit file C:/Hadoop-2.8.0/etc/hadoop/core-site.xml, paste below xml paragraph and save this file.

```
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

2) Rename "mapred-site.xml.template" to "mapred-site.xml" and edit this file C:/Hadoop-2.8.0/etc/hadoop/mapred-site.xml, paste below xml paragraph and save this file.

```
<configuration>
<property>
<name>mapreduce.framework.name</name>
```

Nandani soni
18100BTCCE04031

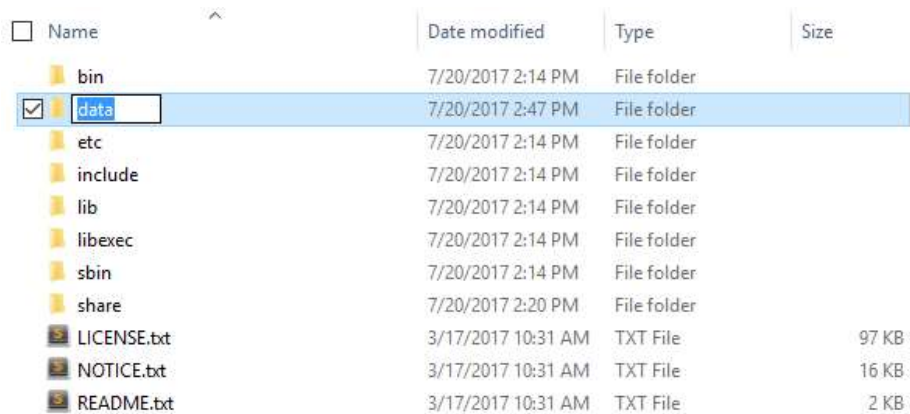
```
<value>yarn</value>
```

```
</property>
```

```
</configuration>
```

3) Create folder "data" under "C:\Hadoop-2.8.0"

- Create folder "datanode" under "C:\Hadoop-2.8.0\data"
- Create folder "namenode" under "C:\Hadoop-2.8.0\data"



<input type="checkbox"/> Name	Date modified	Type	Size
bin	7/20/2017 2:14 PM	File folder	
<input checked="" type="checkbox"/> data	7/20/2017 2:47 PM	File folder	
etc	7/20/2017 2:14 PM	File folder	
include	7/20/2017 2:14 PM	File folder	
lib	7/20/2017 2:14 PM	File folder	
libexec	7/20/2017 2:14 PM	File folder	
sbin	7/20/2017 2:14 PM	File folder	
share	7/20/2017 2:20 PM	File folder	
LICENSE.txt	3/17/2017 10:31 AM	TXT File	97 KB
NOTICE.txt	3/17/2017 10:31 AM	TXT File	16 KB
README.txt	3/17/2017 10:31 AM	TXT File	2 KB

Figure 2.6. Creating folders in Hadoop

4) Edit file C:\Hadoop-2.8.0/etc/hadoop/hdfs-site.xml, paste below xml paragraph and save this file.

```
<configuration>
```

```
<property>
```

```
<name>dfs.replication</name>
```

```
<value>1</value>
```

```
</property>
```

```
<property>
```

```
<name>dfs.namenode.name.dir</name>
```

```
<value>C:\hadoop-2.8.0\data\namenode</value>
```

```
</property>
```

```
<property>
```

```
<name>dfs.datanode.data.dir</name>
```

```
<value>C:\hadoop-2.8.0\data\datanode</value>
```

Nandani soni
18100BTCCE04031

```
</property>  
</configuration>
```

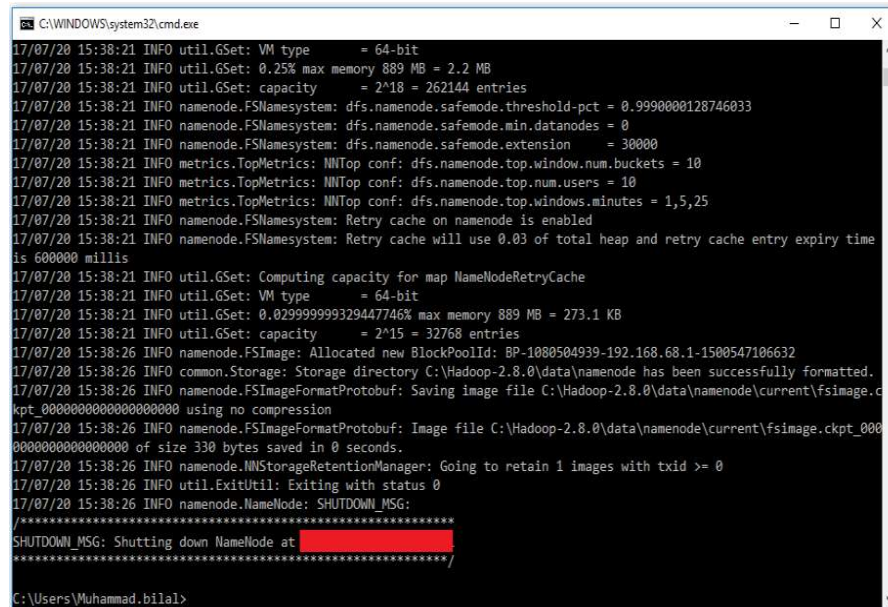
5) Edit file C:/Hadoop-2.8.0/etc/hadoop/yarn-site.xml, paste below xml paragraph and save this file.

```
<configuration>  
  <property>  
    <name>yarn.nodemanager.aux-services</name>  
    <value>mapreduce_shuffle</value>  
  </property>  
  <property>  
    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>  
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>  
  </property>  
</configuration>
```

Hadoop Configuration

1. Download file Hadoop Configuration.zip
(Link: <https://github.com/MuhammadBilalYar/HADOOP-INSTALLATION-ON-WINDOW-10/blob/master/Hadoop%20Configuration.zip>)
2. Delete file bin on C:\Hadoop-2.8.0\bin, replaced by file bin on file just download (from Hadoop Configuration.zip).
3. Open cmd and typing command "hdfs namenode –format" . You will see

Nandani soni
18100BTCCE04031

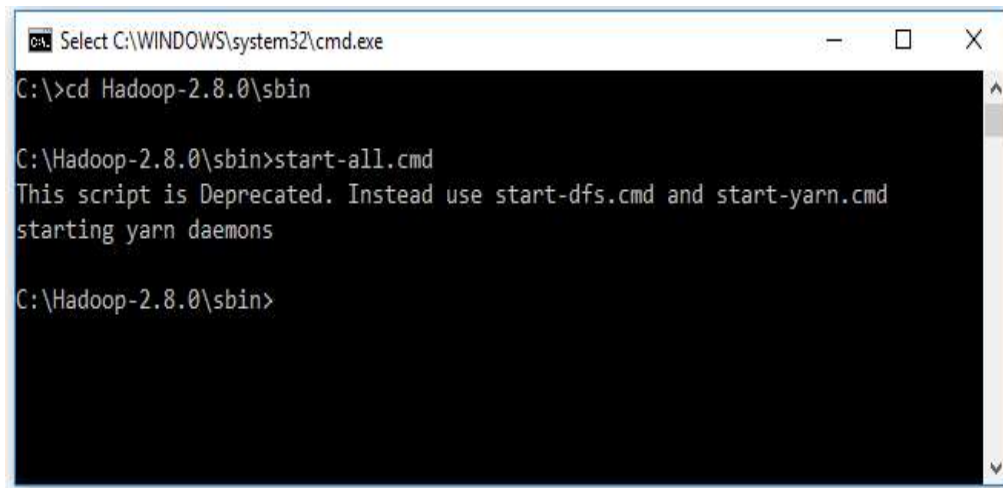


```
C:\WINDOWS\system32\cmd.exe
17/07/20 15:38:21 INFO util.GSet: VM type = 64-bit
17/07/20 15:38:21 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
17/07/20 15:38:21 INFO util.GSet: capacity = 2^18 = 262144 entries
17/07/20 15:38:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
17/07/20 15:38:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
17/07/20 15:38:21 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
17/07/20 15:38:21 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
17/07/20 15:38:21 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
17/07/20 15:38:21 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
17/07/20 15:38:21 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
17/07/20 15:38:21 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
17/07/20 15:38:21 INFO util.GSet: Computing capacity for map NameNodeRetryCache
17/07/20 15:38:21 INFO util.GSet: VM type = 64-bit
17/07/20 15:38:21 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
17/07/20 15:38:21 INFO util.GSet: capacity = 2^15 = 32768 entries
17/07/20 15:38:26 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1080504939-192.168.68.1-1500547106632
17/07/20 15:38:26 INFO common.Storage: Storage directory C:\Hadoop-2.8.0\data\namenode has been successfully formatted.
17/07/20 15:38:26 INFO namenode.FSImageFormatProtobuf: Saving image file C:\Hadoop-2.8.0\data\namenode\current\fsimage.ckpt_00000000000000000000 using no compression
17/07/20 15:38:26 INFO namenode.FSImageFormatProtobuf: Image file C:\Hadoop-2.8.0\data\namenode\current\fsimage.ckpt_00000000000000000000 of size 330 bytes saved in 0 seconds.
17/07/20 15:38:26 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
17/07/20 15:38:26 INFO util.ExitUtil: Exiting with status 0
17/07/20 15:38:26 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at [REDACTED]
*****/
C:\Users\Muhammad.bilal>
```

Figure 2.7. Hadoop Configuration

Testing

1. Open cmd and change directory to "C:\Hadoop-2.8.0\sbin" and type "start-all.cmd" to start apache.



```
Select C:\WINDOWS\system32\cmd.exe
C:\>cd Hadoop-2.8.0\sbin

C:\Hadoop-2.8.0\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

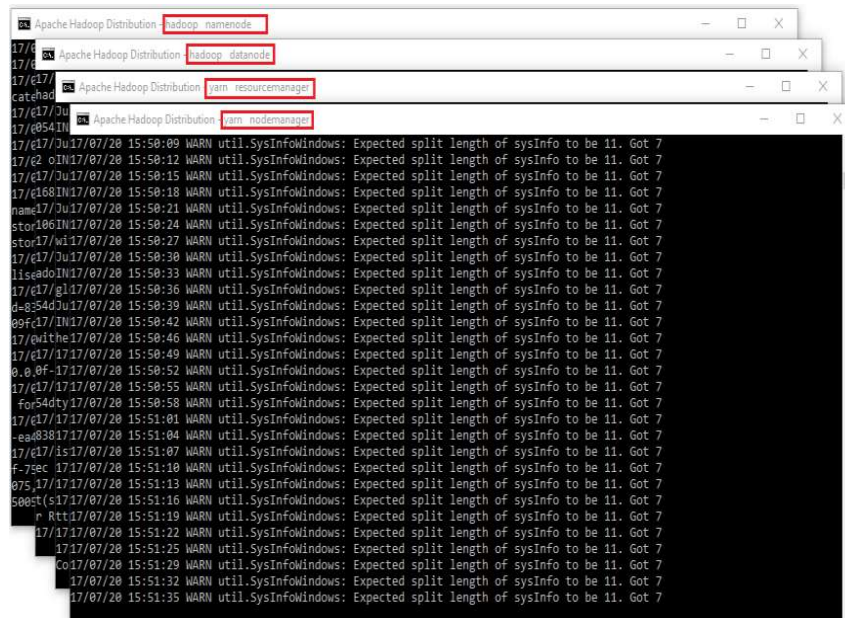
C:\Hadoop-2.8.0\sbin>
```

Figure 2.8. Testing Hadoop

2. Make sure these apps are running
 - Hadoop Namenode

Nandani soni
18100BTCCE04031

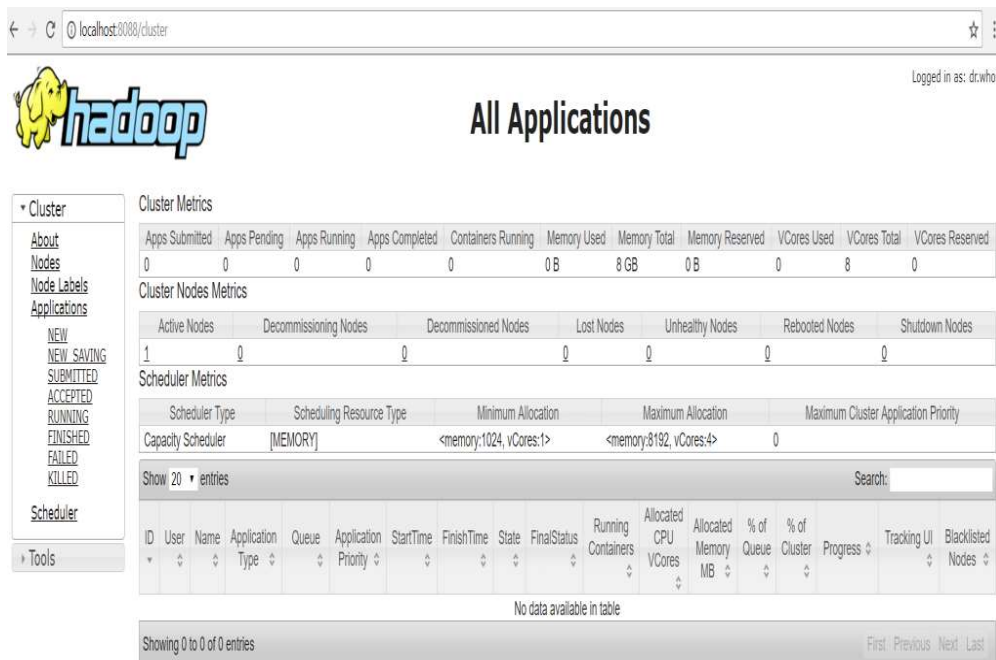
- Hadoop datanode
- YARN Resourc Manager
- YARN Node Manager



The image shows four overlapping terminal windows from the 'Apache Hadoop Distribution'. The windows are titled 'hadoop_namenode', 'hadoop_datanode', 'yarn_resourcemanager', and 'yarn_nodemanager'. Each window displays a series of log messages. The logs for the namenode, datanode, and resourcemanager show 'WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7' repeated multiple times. The nodemanager window shows a similar warning message.

Figure 2.9. Checking files running or not

3. Open: <http://localhost:8088>



The screenshot shows the Hadoop 'All Applications' page. The page has a header with the Hadoop logo and the title 'All Applications'. Below the header, there is a sidebar with navigation links: 'Cluster', 'About', 'Nodes', 'Node Labels', 'Applications', 'NEW', 'NEW SAVING', 'SUBMITTED', 'ACCEPTED', 'RUNNING', 'FINISHED', 'FAILED', 'KILLED', and 'Scheduler'. The main content area displays various metrics and a table of applications.

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Applications Table

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU V-Cores	Allocated Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																	

Showing 0 to 0 of 0 entries

Nandani soni
18100BTCCE04031

Figure 2.10. Localhost:8080

4. Open: <http://localhost:50070>

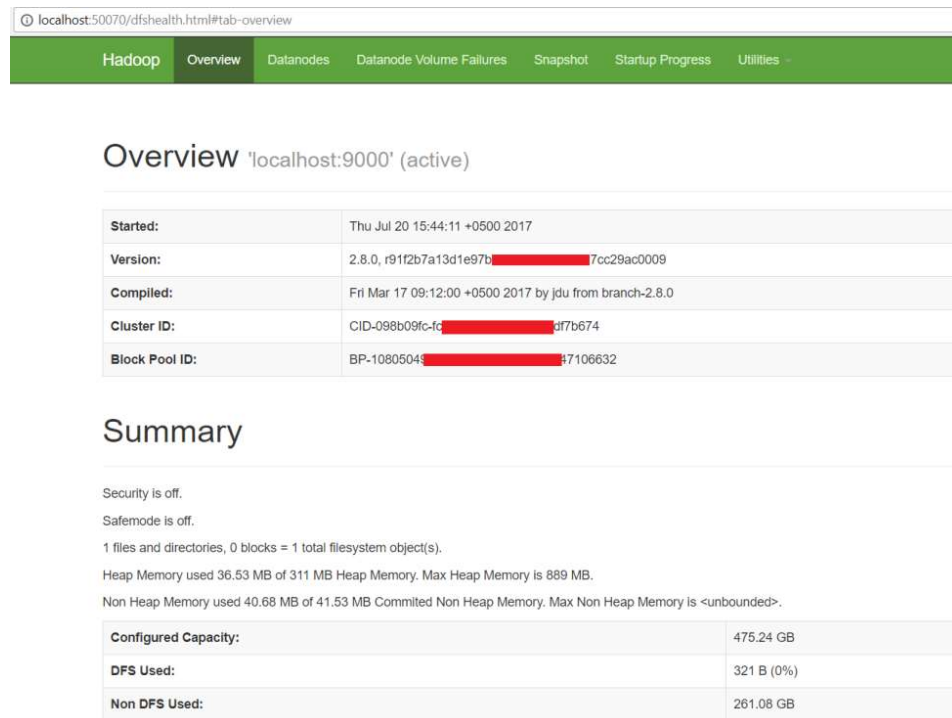


Figure 2.11. Localhost:50070

So, now you have successfully installed Hadoop