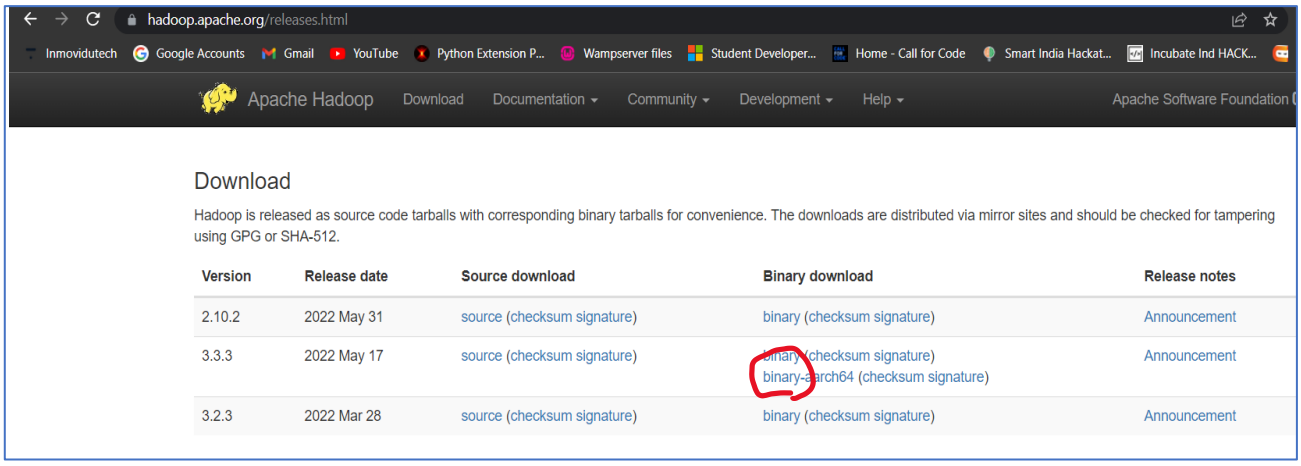
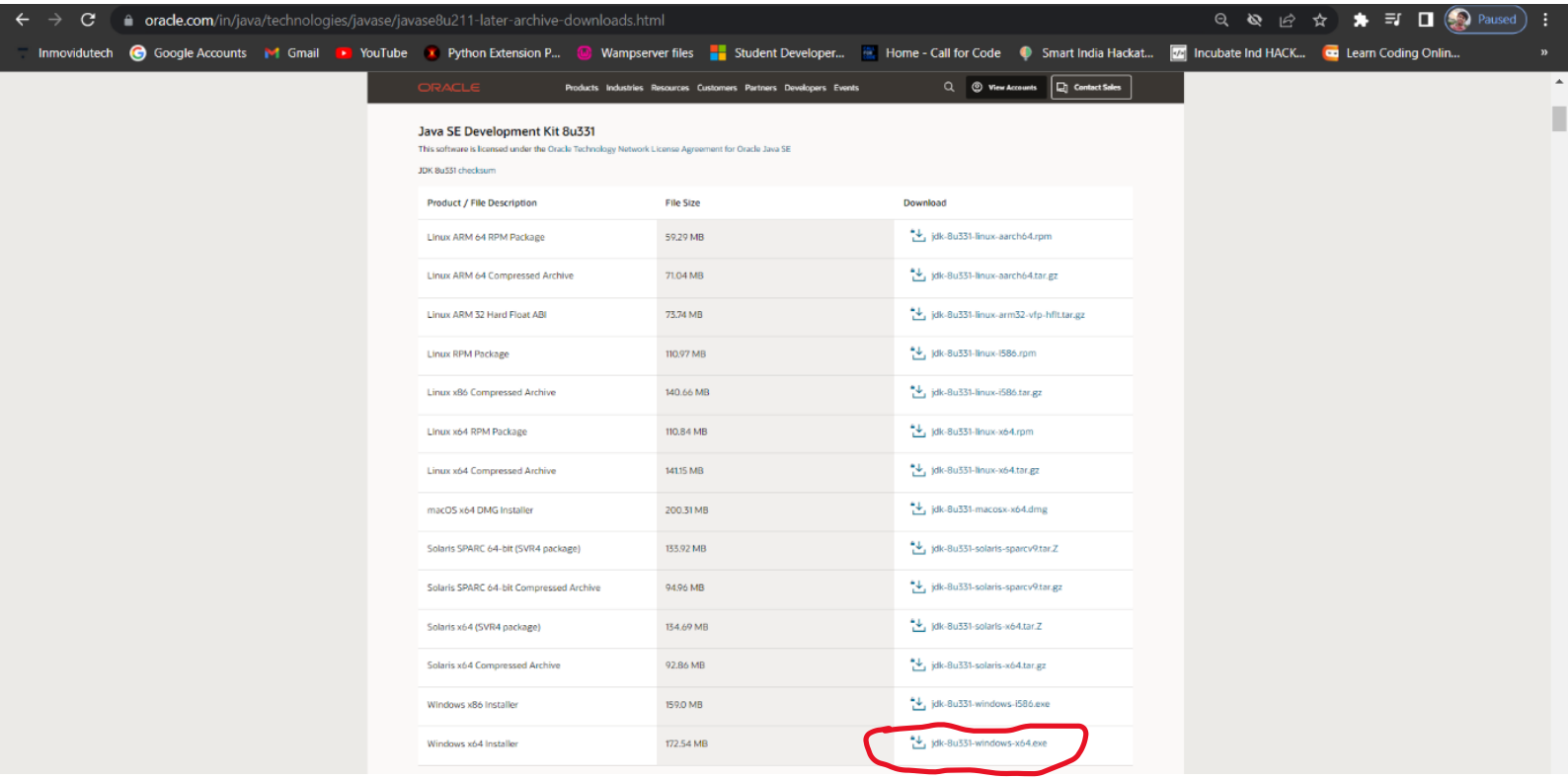
Steps to Install Hadoop on Windows

Step 1: Download Apache Hadoop from this website: https://hadoop.apache.org/releases.html by clicking the binary

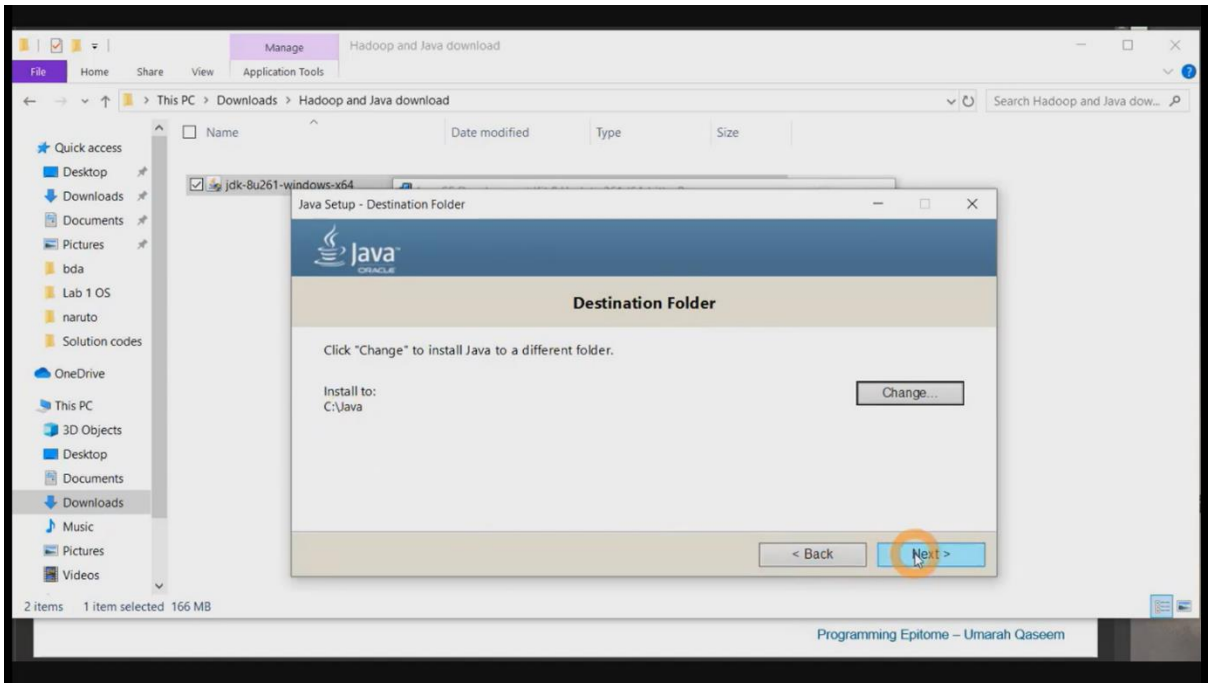


Step 2: Download JDK 8 from this website: https://www.oracle.com/in/java/technologies/javase/javase8u211-later-archive-downloads.html

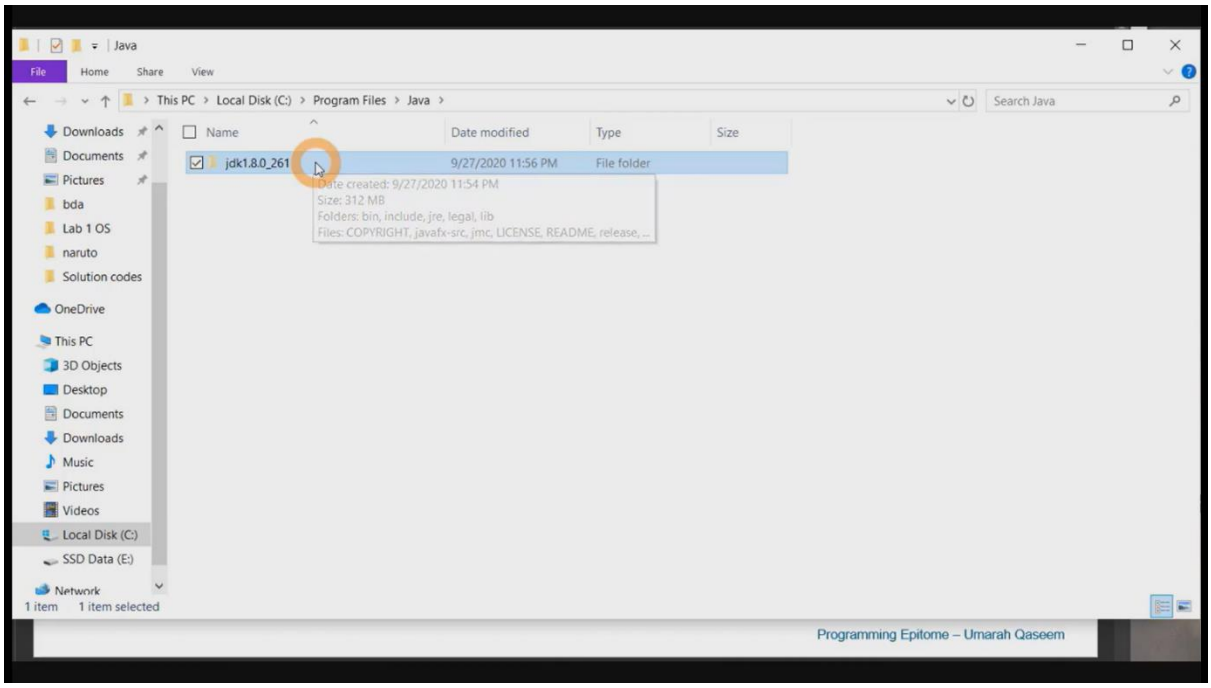


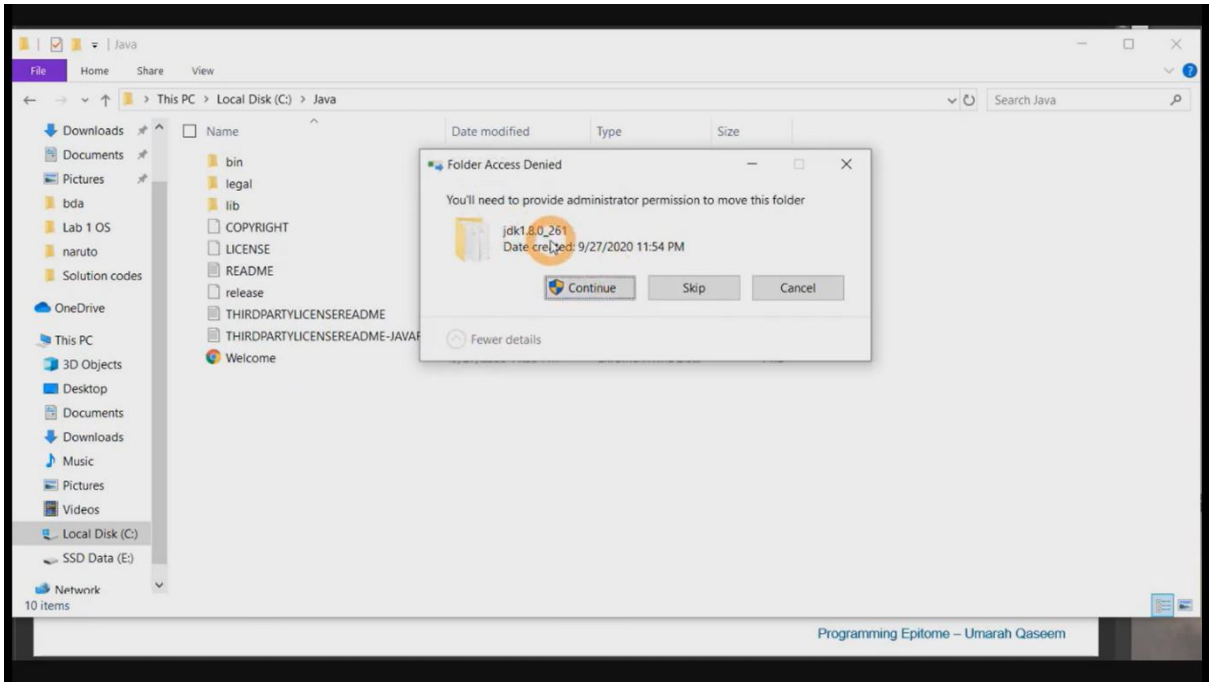
JDK Installation:

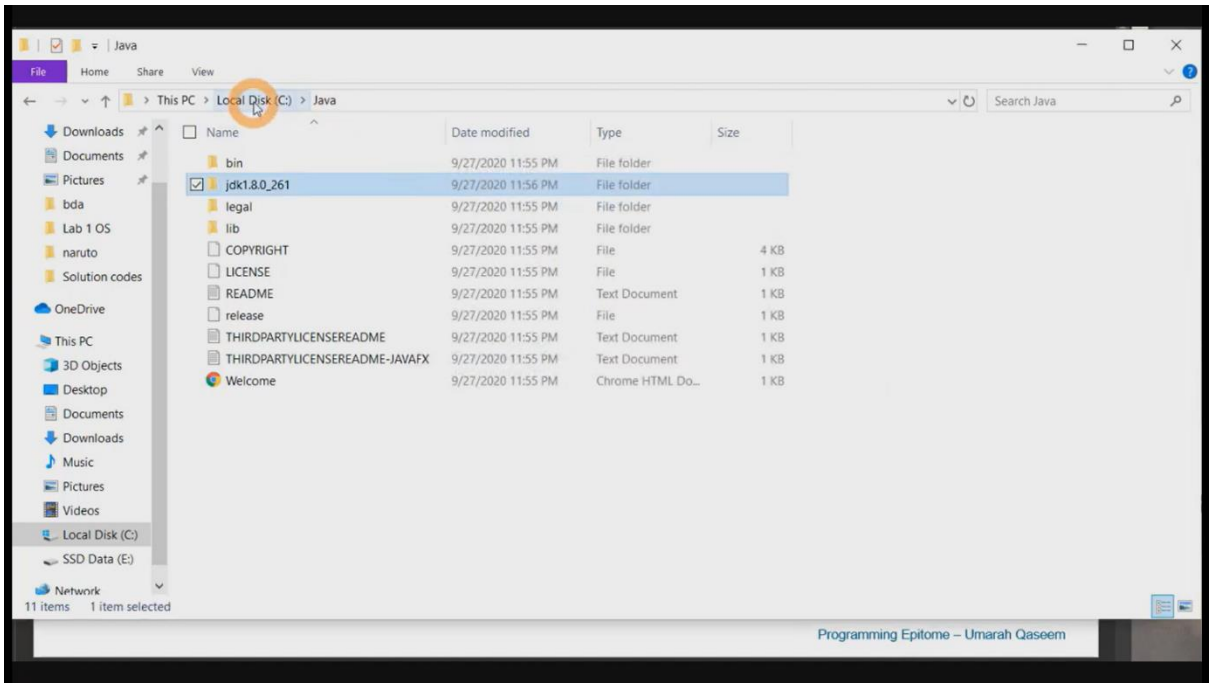
Step 3: Create a new folder named “Java” in your C drive. Now double click the downloaded Jdk file. While installing the JDK set the Destination folder as “C:\Java”.



Step 4: Now go to your C drive, then Program Files>java then move the file in that java folder to the “Java” folder which we created in the C drive.



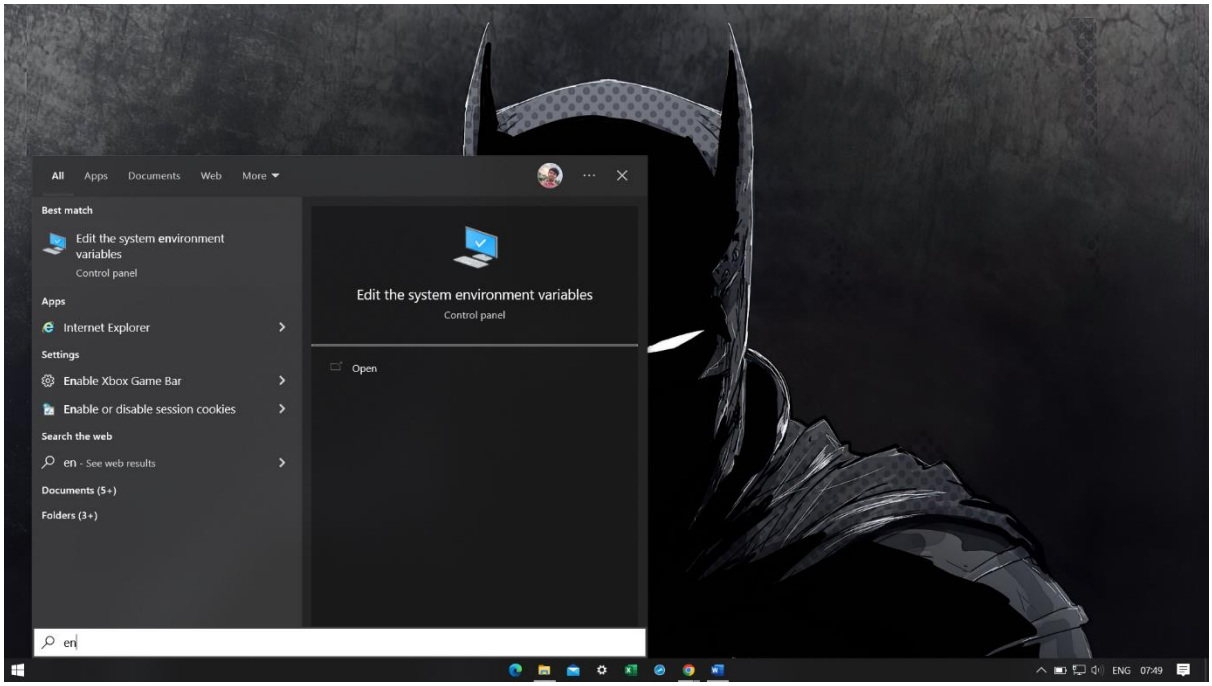




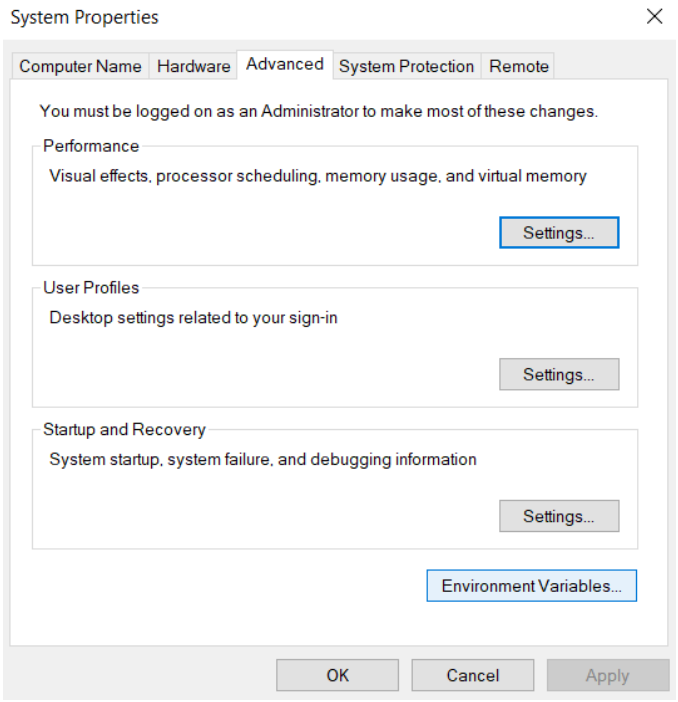
Step 5: Delete the java folder in Program Files folder

Step 6: Now set environment variables and path by:

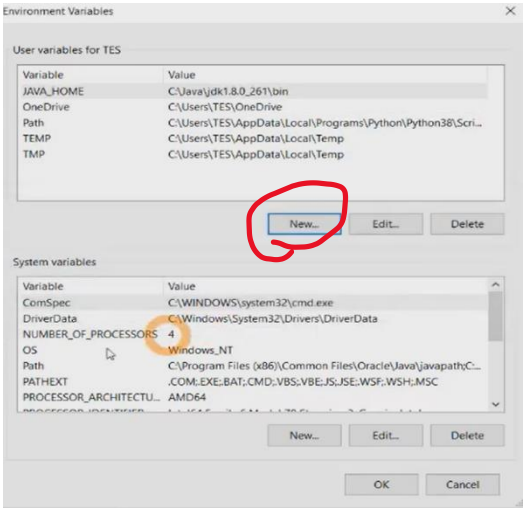
6.1. Search for Edit the system environment variables in control panel

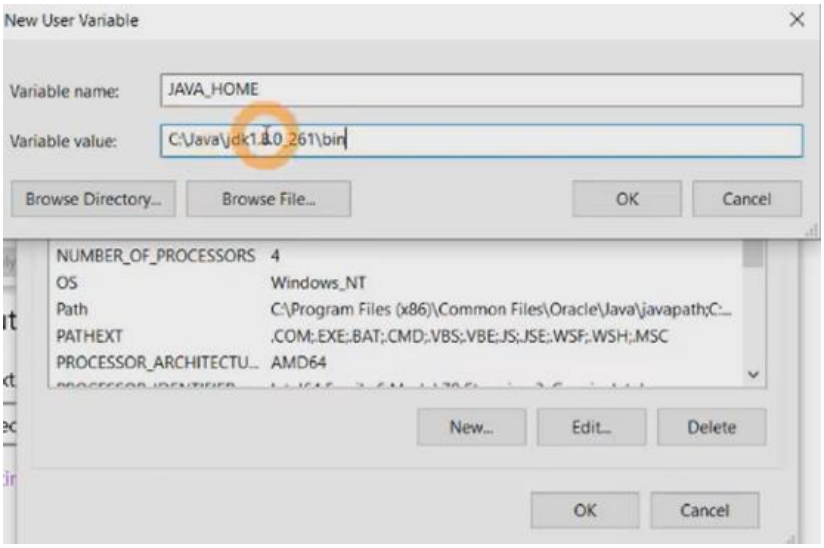


6.2. Now click on Environment variables.

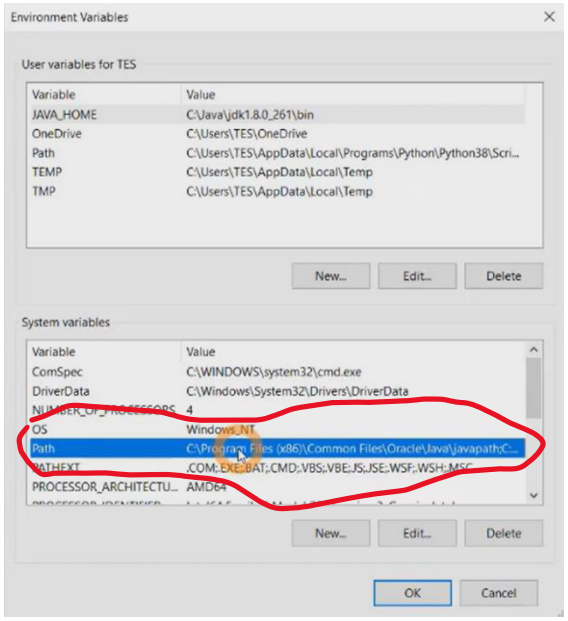


6.3. Now click new and give the variable name as “JAVA_HOME” and variable value as the path where the java is installed (Eg: C:\Java\jdk1.8.0_261\bin) and then click OK.





6.4. Now create the path by selecting the path and click edit. In that paste the path of the java (Eg: C:\Java\jdk1.8.0_261\bin) and click OK.



6.5. **Now Close all the Tabs.**

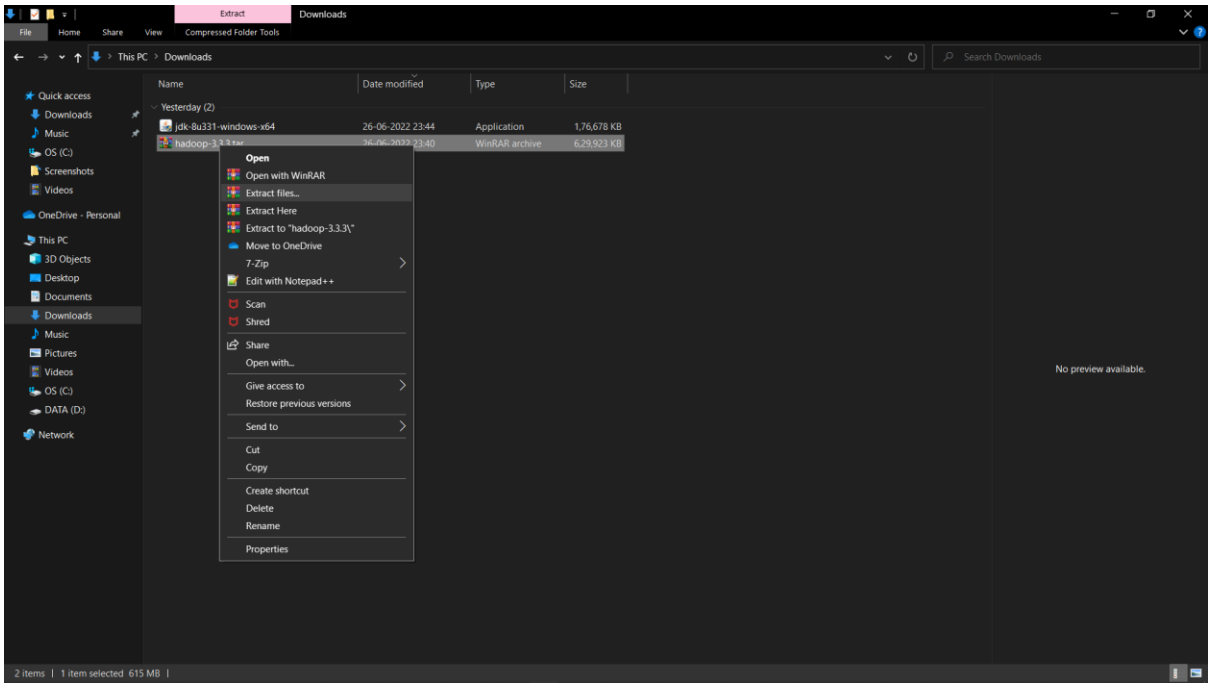
6.6. **You have Java package installed successfully.**

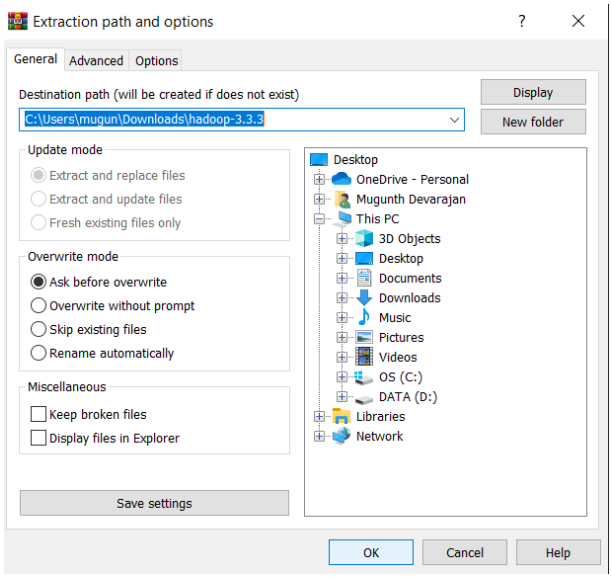
Testing Java:

1. **Open cmd prompt in admin mode.**
2. **Type “javac”, which should display various options available in java package.**
3. **Type “java –version” to check the version of the java package.**

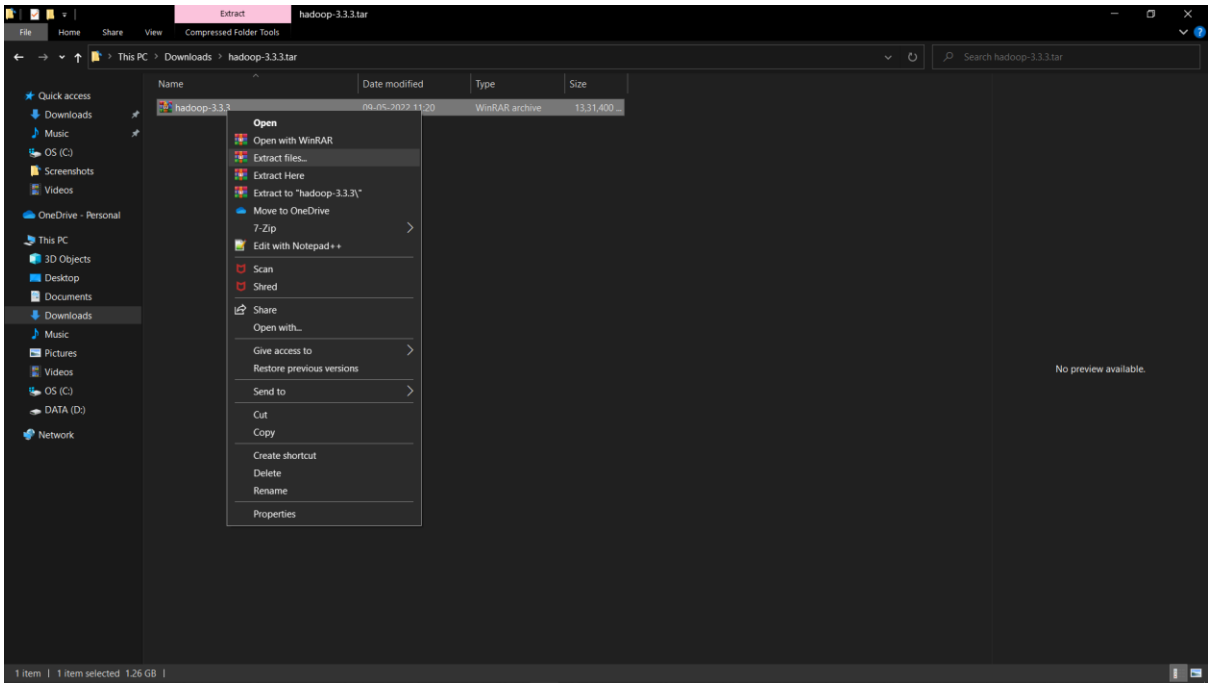
Hadoop Installation:

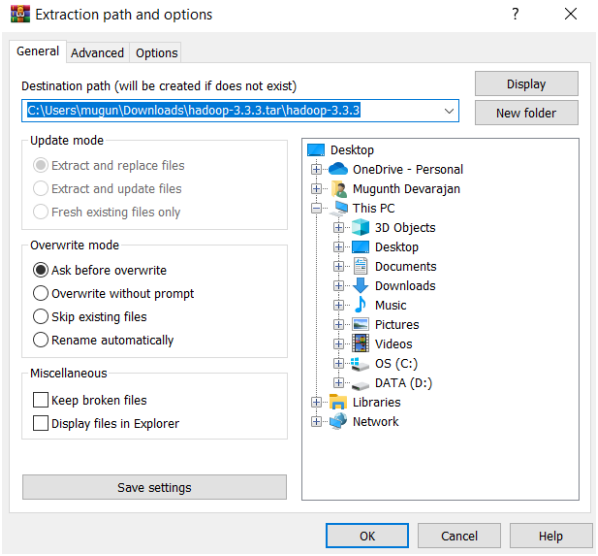
Step 7: Unzip the downloaded Hadoop file.



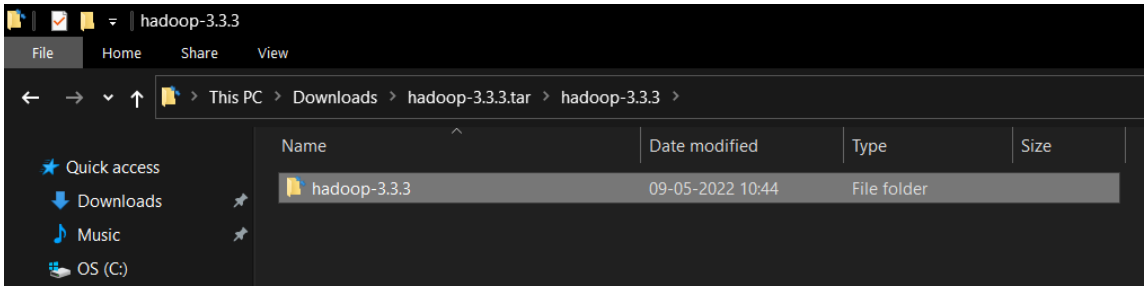


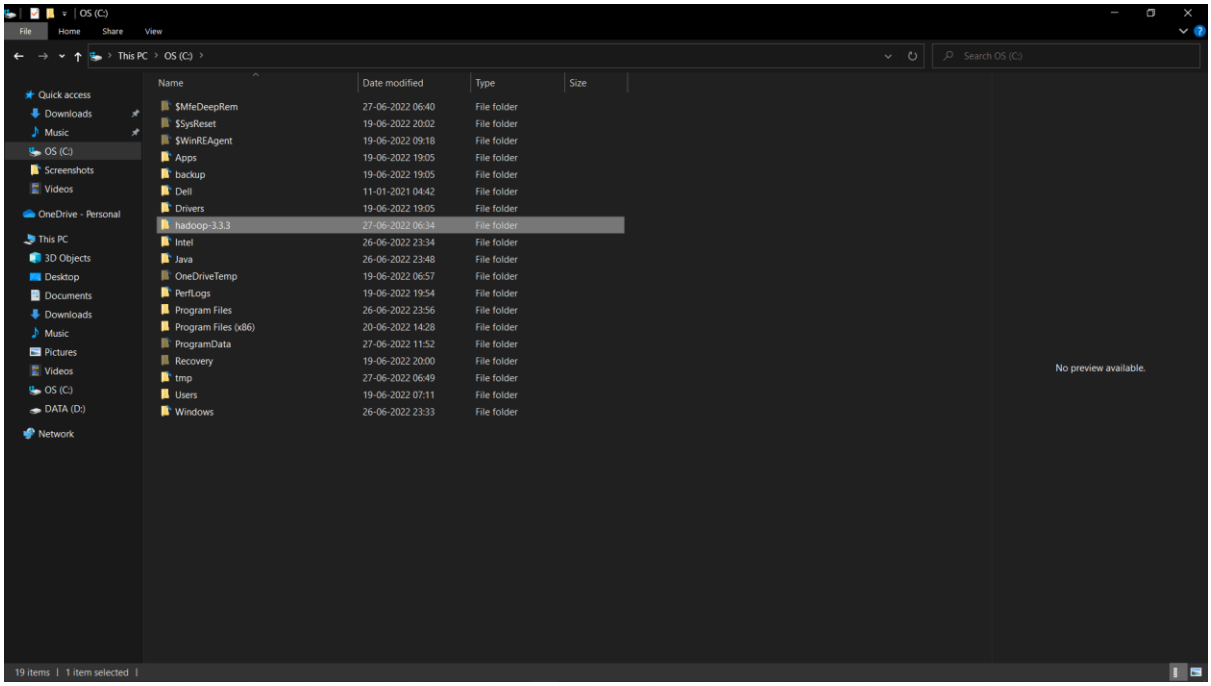
Step 8: Now open the unzipped file (hadoop-3.3.3.tar) which contains another zip file (hadoop-3.3.3), now unzip it.





Step 9: Now open the unzipped file (hadoop-3.3.3) and move the hadoop-3.3.3 file to C drive.





Configuring Hadoop:

Step 10: Now go to "C:\hadoop-3.3.3\etc\hadoop".

Note: Before Step 11 install Notepad ++.

Step 11: Open the core-site.xml, mapred-site.xml, hdfs-site.xml, yarn-site.xml, hadoop-env in Notepad ++.

Step 12: Paste the following in each file

a) File C:/Hadoop-3.3.3/etc/hadoop/core-site.xml, paste below xml paragraph and

save this file.

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

Now give ctrl+s to save the file

b) C:/Hadoop-3.3.3/etc/hadoop/mapred-site.xml, paste below xml paragraph and save

this file.

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

Now give ctrl+s to save the file

c) Create folder "data" under "C:\Hadoop-3.3.3"

1) Create folder "datanode" under "C:\Hadoop-3.3.3\data"

2) Create folder "namenode" under "C:\Hadoop-3.3.3\data"

d) Edit file C:\Hadoop-3.3.3/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>file:///C:/hadoop-3.3.3/data/namenode</value>

</property>

<property>

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

<value>file:///C:/hadoop-3.3.3/data/datanode</value>

</property>

</configuration>

Now give ctrl+s to save the file

e) Edit file C:/Hadoop-3.3.3/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>

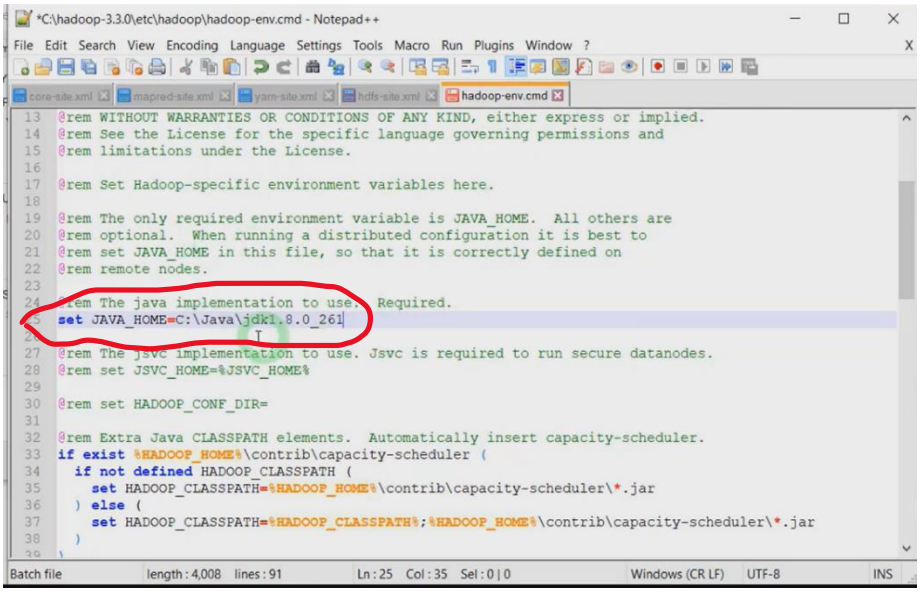
</property>

</configuration>

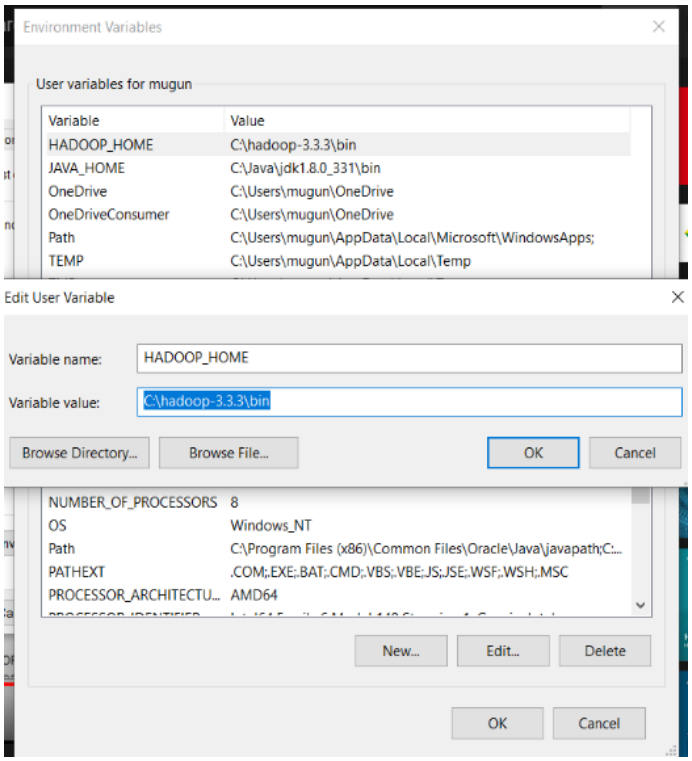
Now give ctrl+s to save the file

f) Edit C:/Hadoop-3.3.3/etc/hadoop/hadoop-env.cmd,

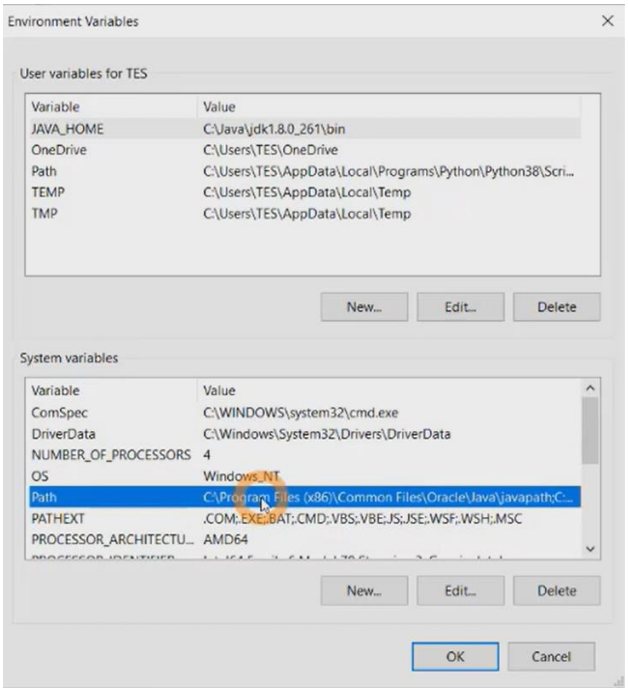
type the variable value given in Step 6.3 (Eg: C:\Java\jdk1.8.0_261) (without bin) to "set JAVA_HOME" and give ctrl+s to save the file

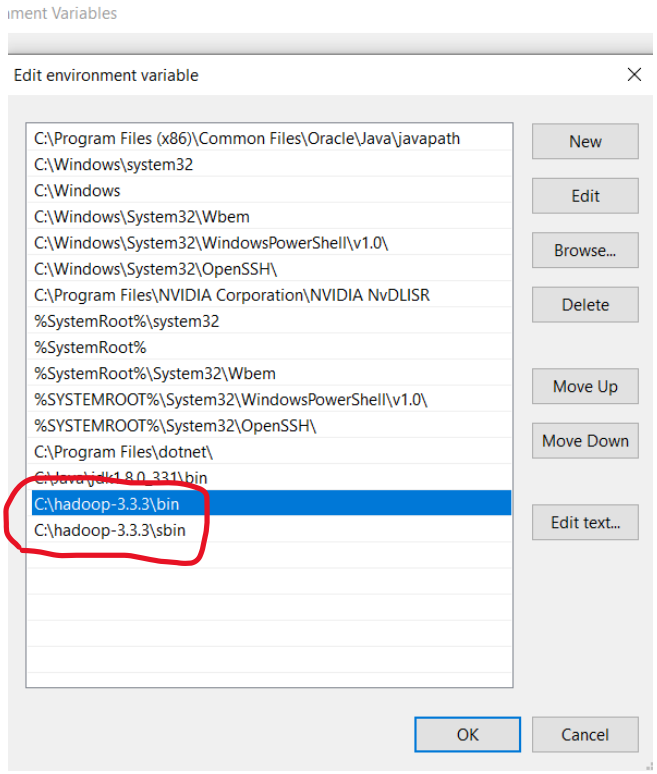


Step 13: Open the "Edit the system Environment Variable" in control Panel. Now click new and give the variable name as "HADOOP_HOME" and variable value as the path where the hadoop is installed (Eg: C:\hadoop-3.3.3\bin) and then click OK.



Step 14: Now create the path by selecting the path and click edit. In that click new and type "C:\hadoop-3.3.3\bin" and "C:\hadoop-3.3.3\sbin" (separately) and click OK.





Step 15: In hadoop bin folder some necessary files are missing so now click on this link to download those files:
https://drive.google.com/file/d/1kVhX9snOZ3oLUxDjh3AVI8fcRnEWAAE4/view

Step 16: Now unzip the bin file downloaded from the above link and replace it with the bin file in the hadoop folder (C:\hadoop-3.3.3).

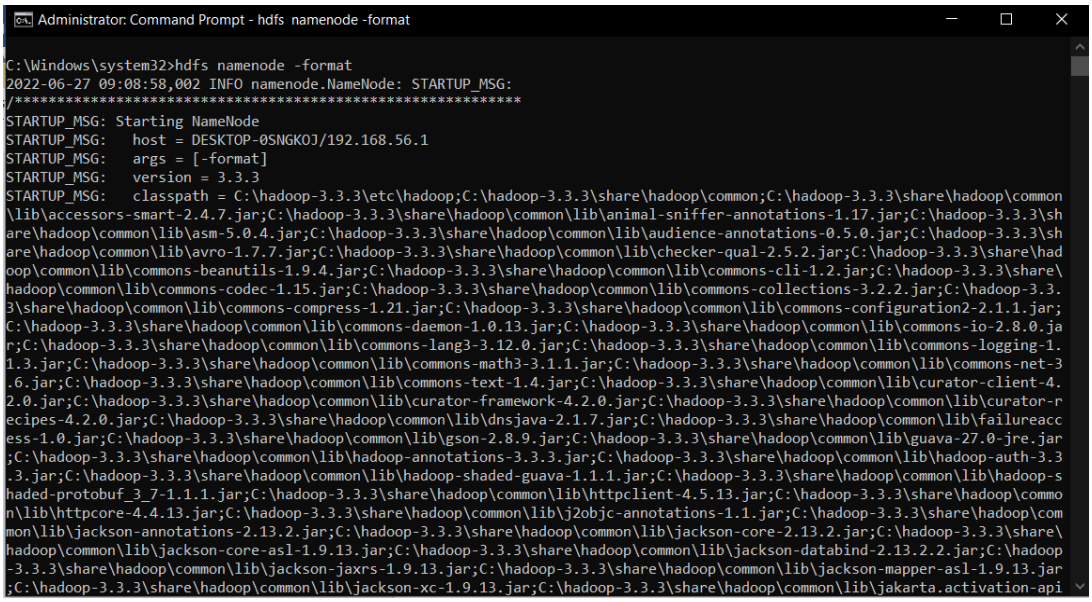
Hadoop Installation and Configuration completed

Verifying whether the installation is successful or not:

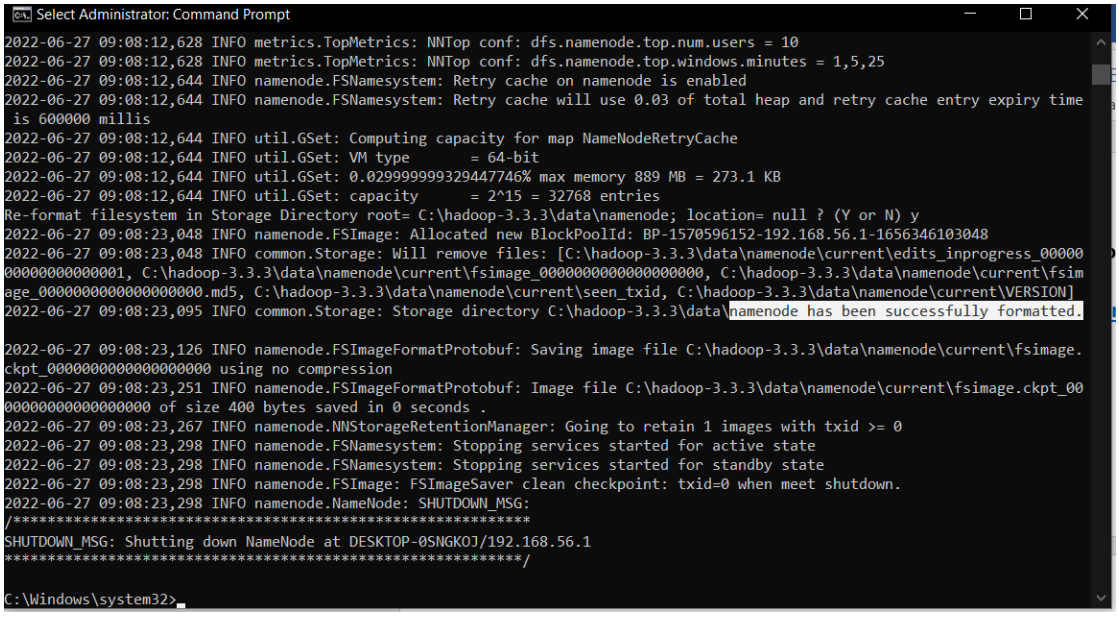
Step 17: Open cmd in admin mode.

Step 18: Now type the following commands:

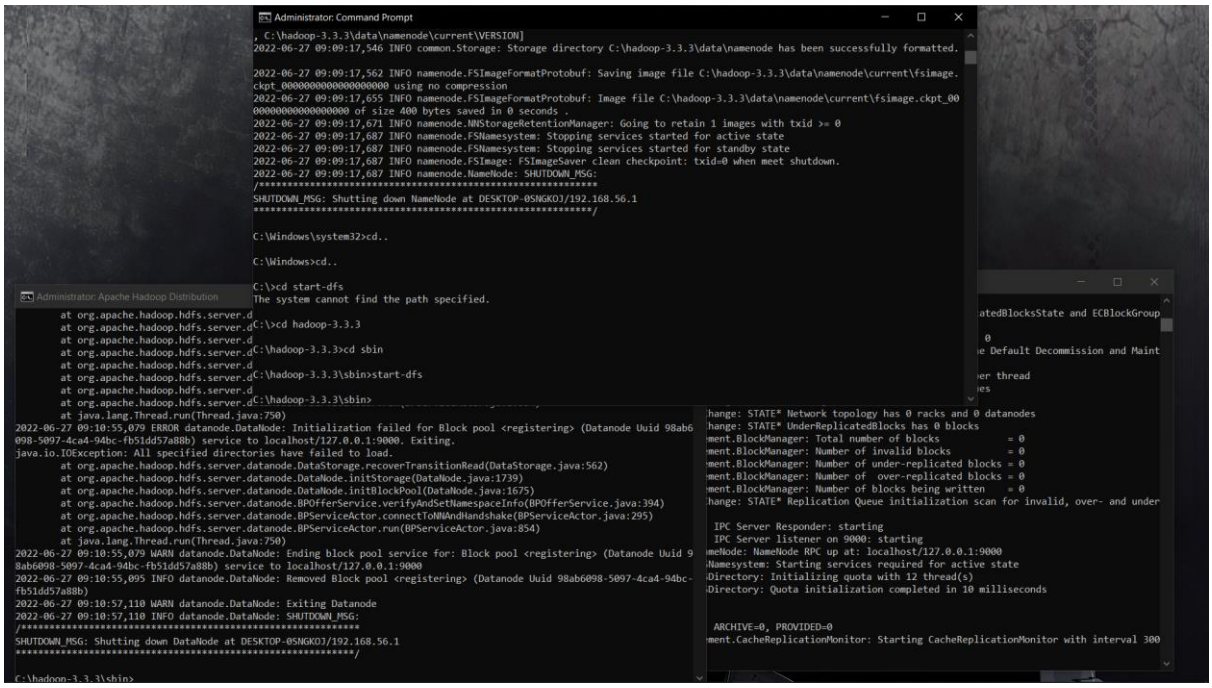
1. **hdfs namenode -format**



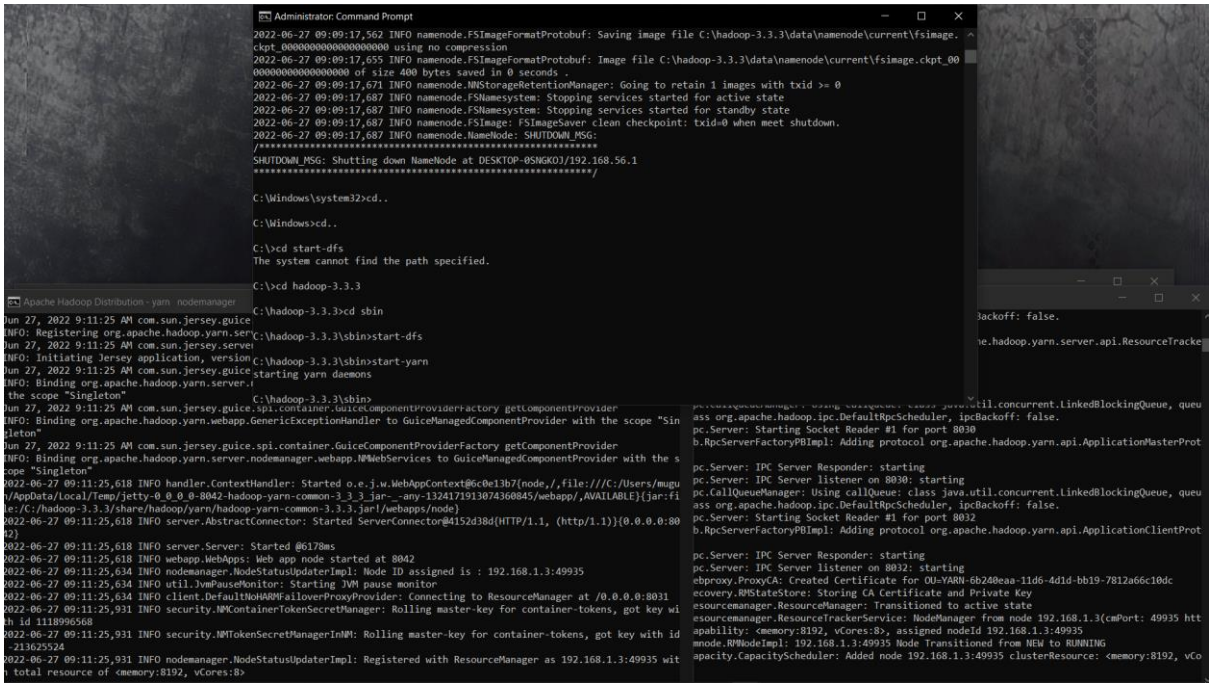
(after executing the command wait for few secs, it should display "namenode has been successfully formatted")



2. **Type “cd..” 2 times and then type:**
   2.1. **cd hadoop-3.3.3**
   2.2. **cd sbin**
3. **now type “start-dfs” to start hadoop.**

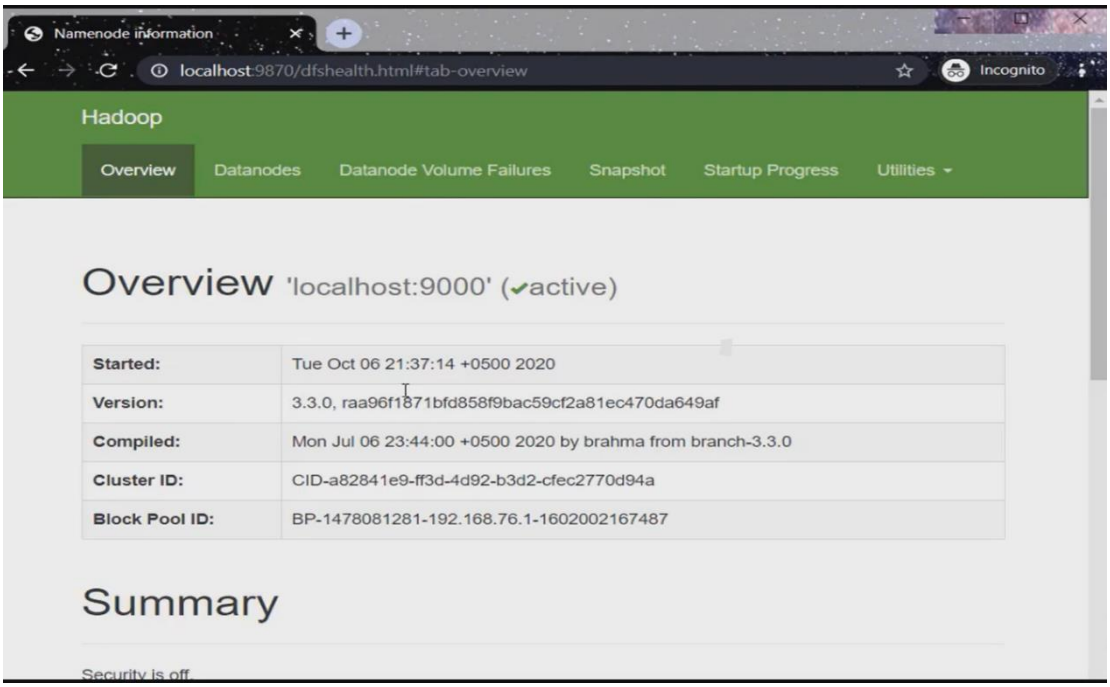


4. **then type “start-yarn” to start yarn.**

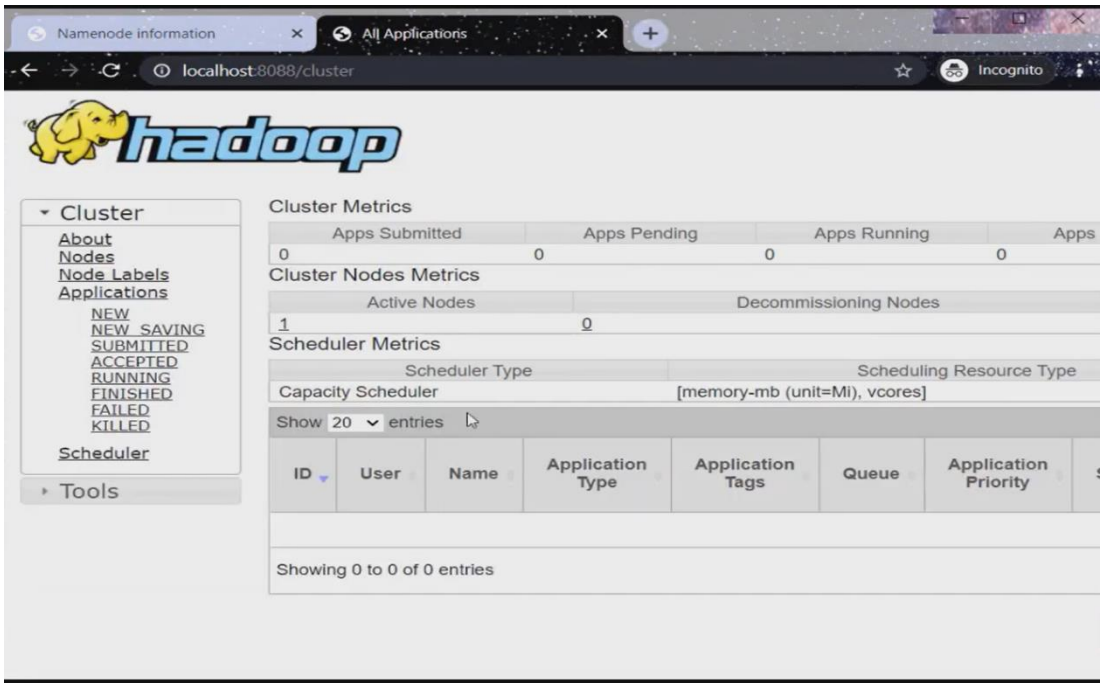


Note: the cmd prompts should run without getting terminated immediately

5. **In browser type “localhost:9870” to view/ check the working of hadoop.**



6. **In browser type “localhost:8088” to view/ check the working of yarn.**



7. **Type “tjs”.**
8. **Type “stop-yarn” to stop yarn.**
9. **Type “stop-dfs” to terminate hadoop.**