Step for Hadoop 3.1.2 installation on Window 10

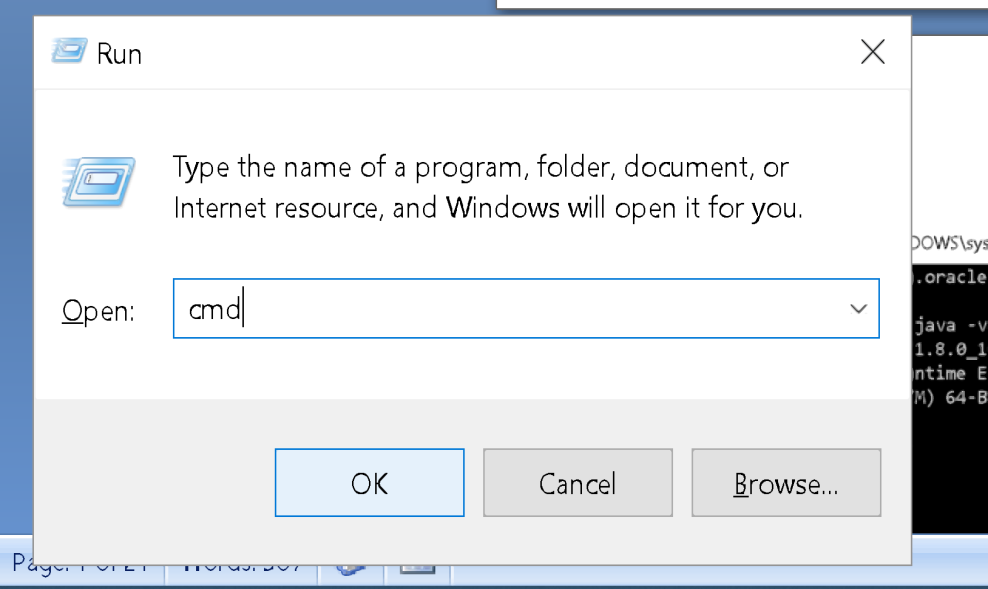
Key note:

* This was installed on windows 10 x64,before downloading please check your windows architecture and install.
* And it was installed with windows administrative account, Please don’t use any User or Guest account while installing.
* Please make sure your path location is correct while setting environment Variable and moving files from local to hdfs.

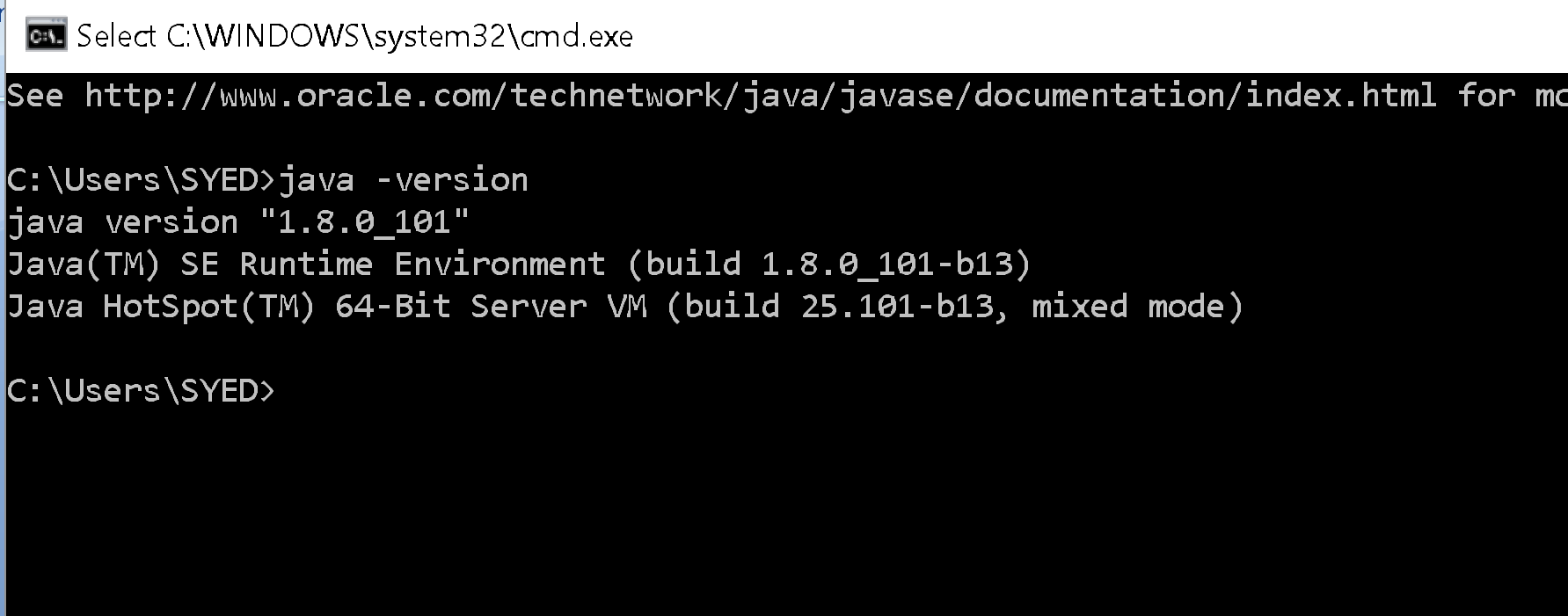
1. **Prepare:**

These softwares should be prepared to install Hadoop 3.1.2 on window 10 64bit.

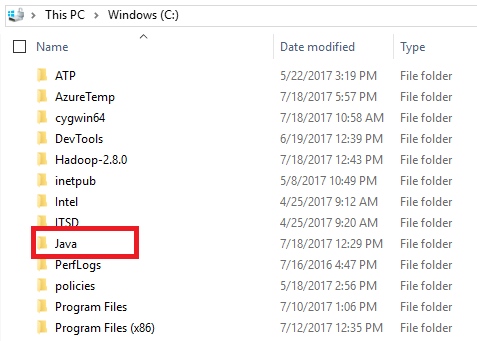
1. Download Hadoop 3.1.2 latest version <https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz>
2. Download **Development Kit 8u211**  (Link: <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>) latest version from their site. If you are already installed goto **STEP 2**
3. **Set up**
   1. Check either Java is already installed on your system or not, use **"Javac -version"** or **“ java –version”** to check. Press WINDOWS+R keys and type **cmd** and press enter to open command prompt.



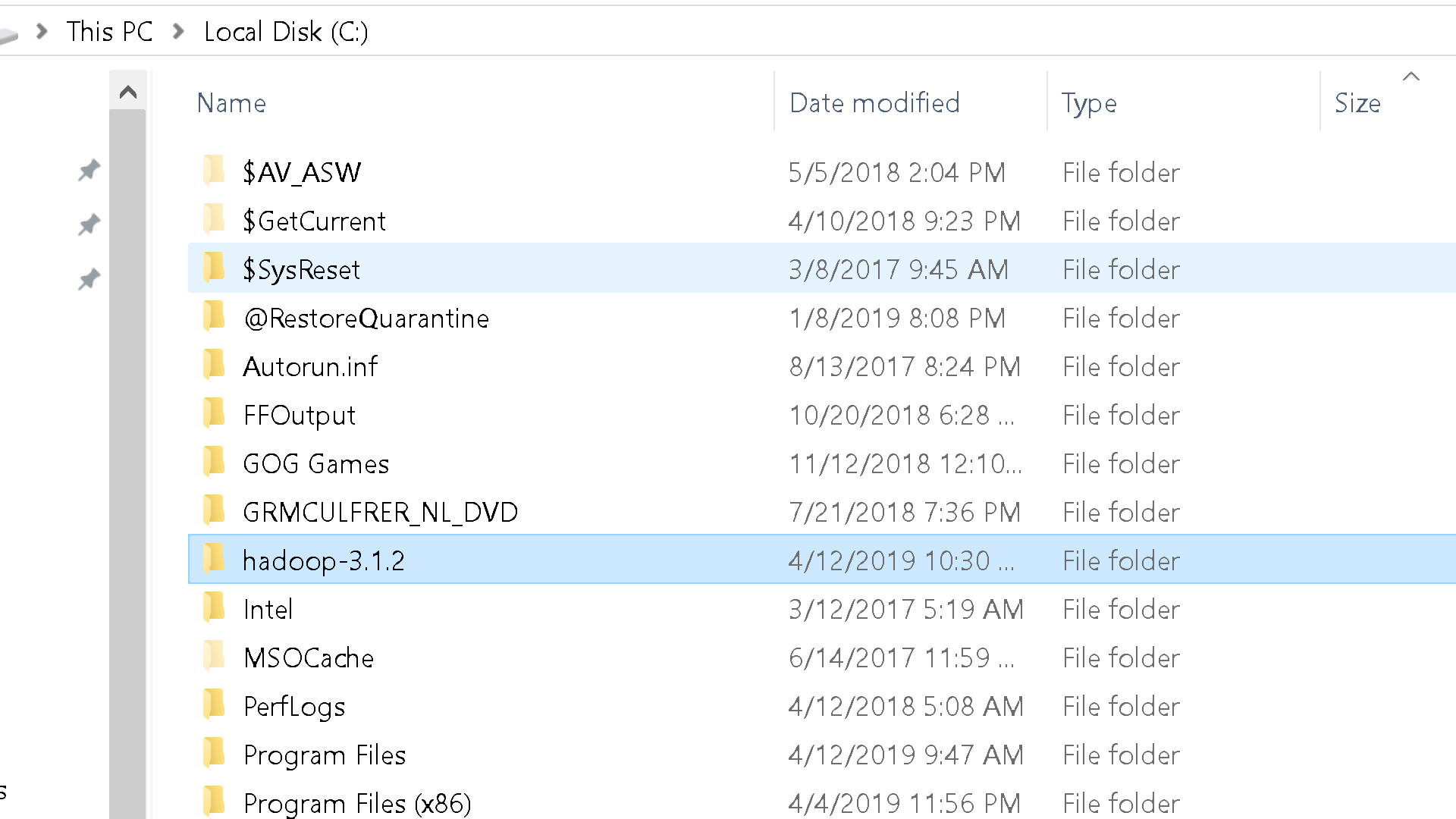
Then



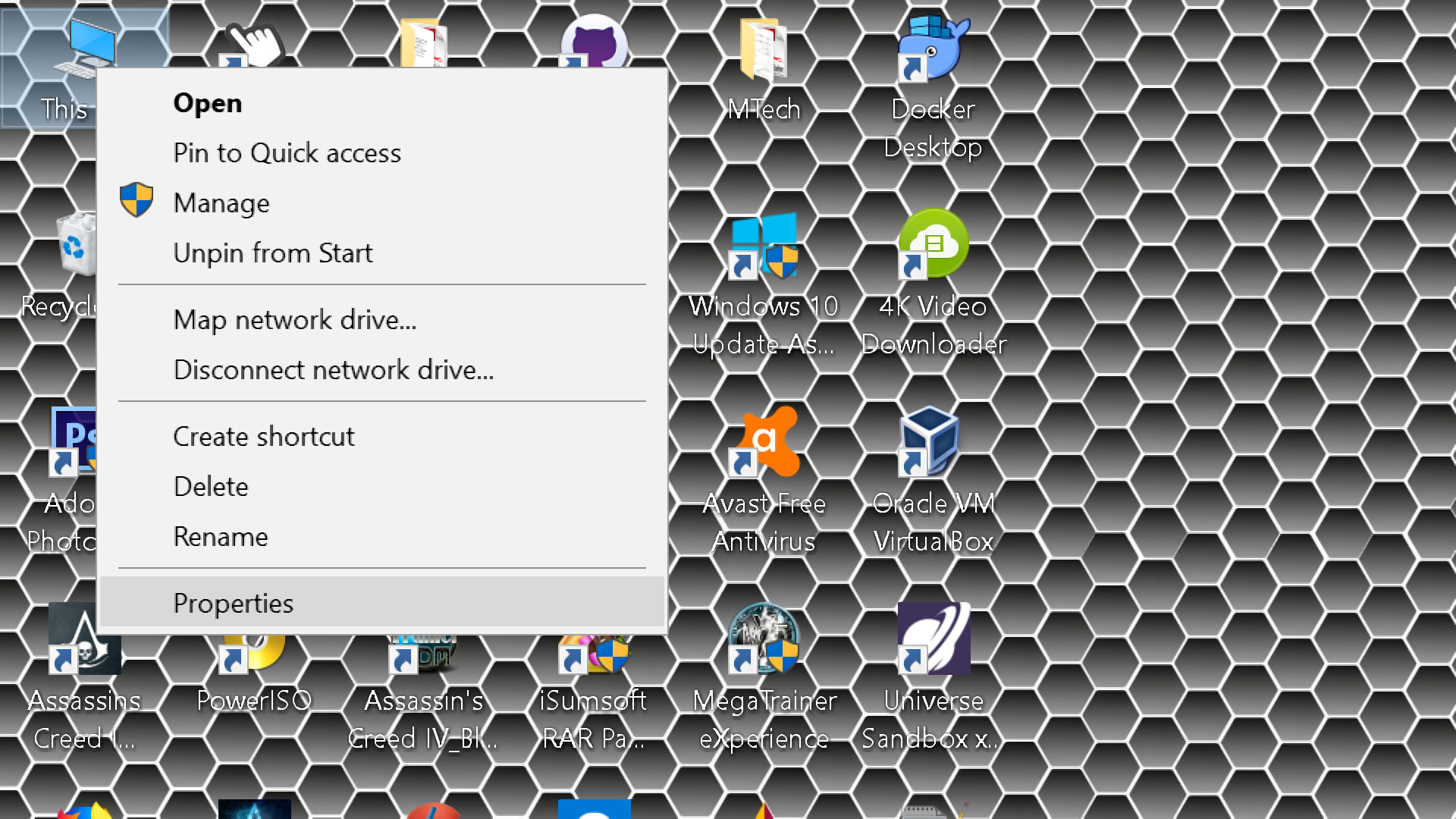
* 1. If Java is not installed on your system then first install java under **"C:\JAVA"** Or **“C:\Program Files\Java”** or justinstall , it will automatically allocate the path.

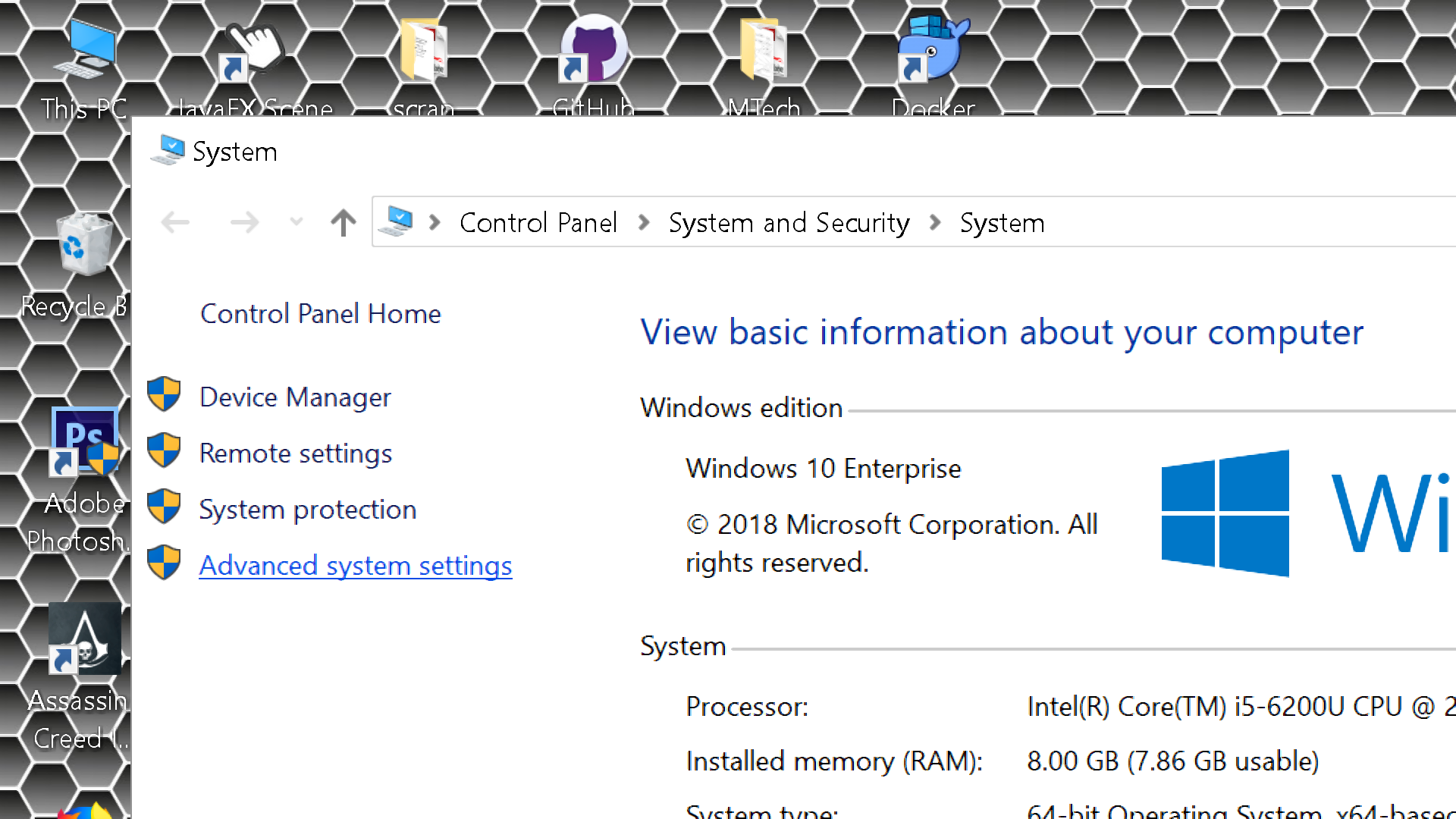


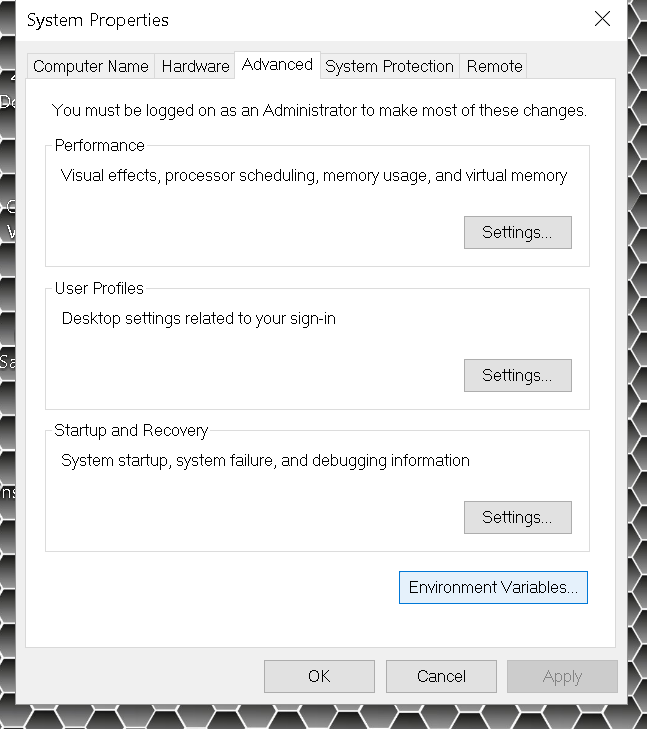
* 1. Extract the file Hadoop 3.1.2.tar.gz and place under C like **"C:\Hadoop-3.1.2"**.



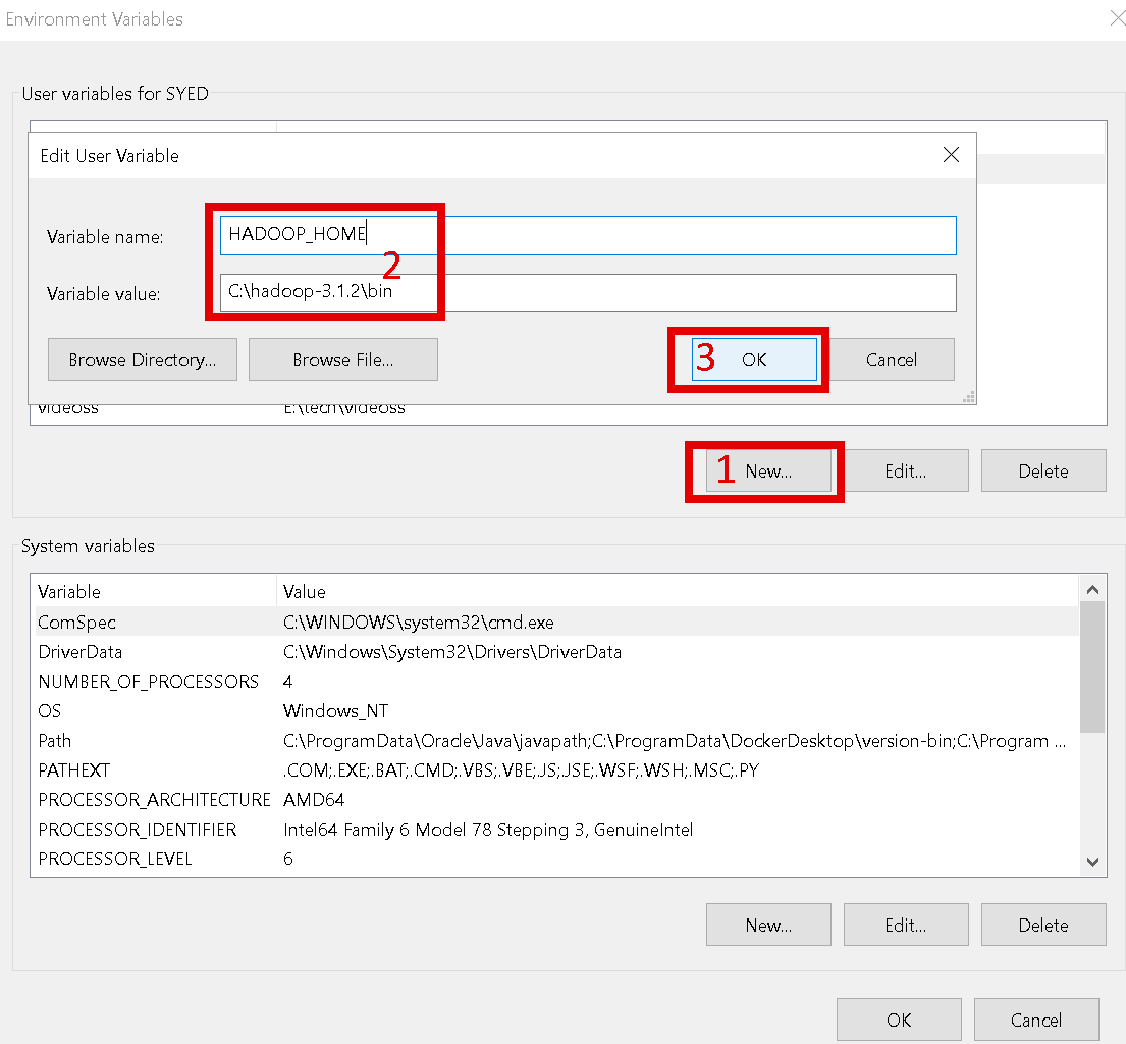
* 1. Set the path HADOOP\_HOME Environment variable on windows10 (Go to This PC right click on it -> go to properties ->Advance system Settings -> Environment variable ->New in the User variables->set variable name and value as given below)



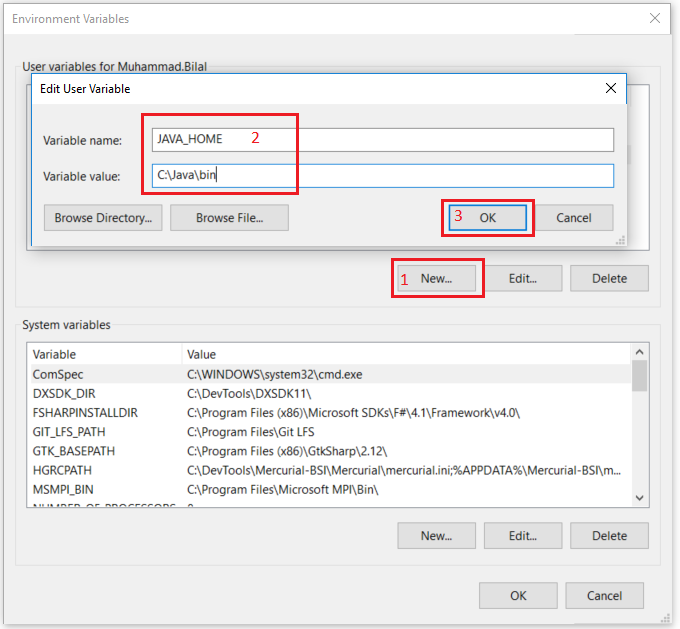




The **Variable name** should be as it is given below



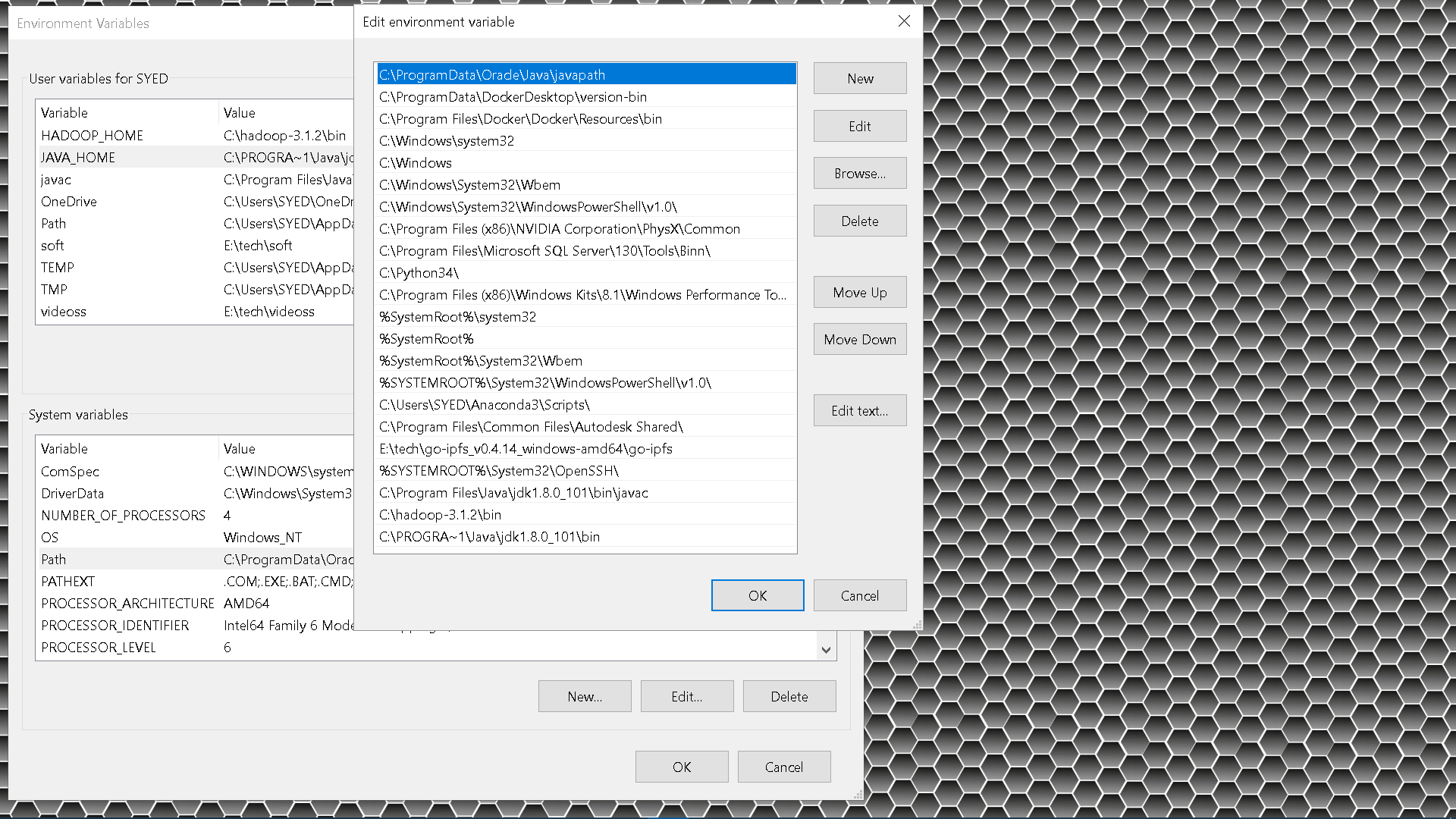
1. Set the path JAVA\_HOME Environment variable on windows 10 (check your java bin path)



OR Similarly do this by setting relative path name. (check your java\bin path)

Here I changed the path from **“C:\Program Files\Java “** to **“C:\PROGRA~1\Java”**  bcz in my first path it has space in the name . so replace **“Program Files “** to **“PROGRA~1”** wherever you find through the process. And **jdk** name also may change please verify from your directory location.

1. Next we set the Hadoop bin directory path and JAVA bin directory path.



8

**7**

1

6

4

**3**  **5**

**2**

1. **Configuration (** I’m using NPP+ for editing ,Don’t use notepad **)**
2. Edit file **C:/Hadoop-3.1.2/etc/hadoop/core-site.xml**, paste/append below xml paragraph and save this file.

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

1. Edit this file **C:/Hadoop-3.1.2/etc/hadoop/mapred-site.xml**, paste/append below xml paragraph and save this file. If you can’t find then rename "mapred-site.xml.template" to "mapred-site.xml" and edit it.

<configuration>

<property>

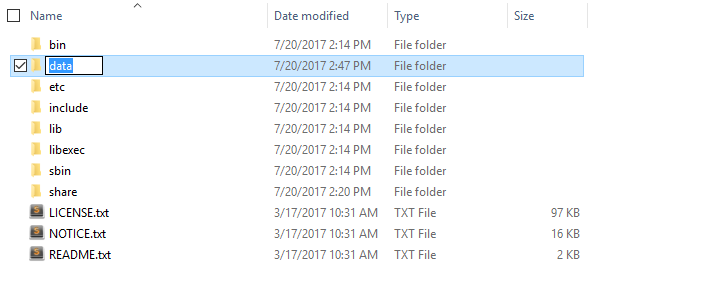
<name>mapreduce.framework.name</name>

<value>yarn</value>

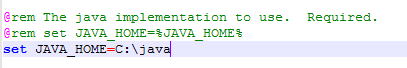
</property>

</configuration>

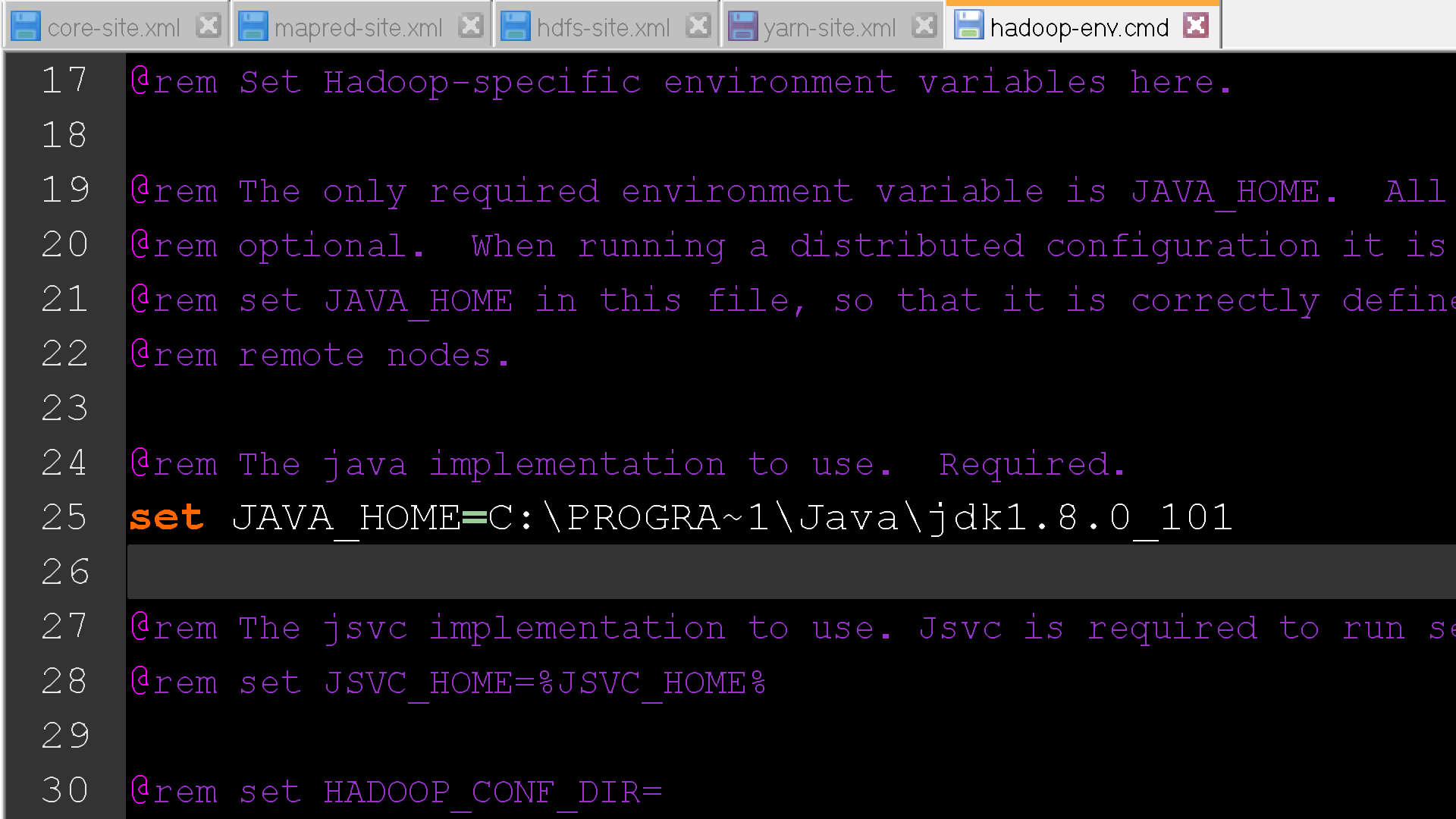
1. Create folder **"data"** under **"C:\Hadoop-3.1.2"**

* Create folder **"datanode"** under **"C:\Hadoop-3.1.2\data"**
* Create folder **"namenode"** under **"C:\Hadoop-3.1.2\data"**  

1. Edit file **C:/Hadoop-3.1.2/etc/hadoop/hdfs-site.xml**, paste/append below xml paragraph and save this file.
2. <property>
3. <name>dfs.replication</name>
4. <value>1</value>
5. </property>
6. <property>
7. <name>dfs.namenode.name.dir</name>
8. <value>file:/C:/hadoop-3.1.2/data/namenode</value>
9. </property>
10. <property>
11. <name>dfs.datanode.data.dir</name>
12. <value>file:/C:/hadoop-3.1.2/data/datanode</value>
13. </property>
14. Edit file **C:/Hadoop-3.1.2/etc/hadoop/yarn-site.xml**, paste/append below xml paragraph and save this file.
15. <property>
16. <name>yarn.nodemanager.aux-services</name>
17. <value>mapreduce\_shuffle</value>
18. </property>
19. <property>
20. <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
21. <value>org.apache.hadoop.mapred.ShuffleHandler</value>
22. </property>
23. <property>
24. <name>yarn.application.classpath</name>
25. <value>
26. %HADOOP\_HOME%\etc\hadoop,
27. %HADOOP\_HOME%\share\hadoop\common\\*,
28. %HADOOP\_HOME%\share\hadoop\common\lib\\*,
29. %HADOOP\_HOME%\share\hadoop\hdfs\\*,
30. %HADOOP\_HOME%\share\hadoop\hdfs\lib\\*,
31. %HADOOP\_HOME%\share\hadoop\mapreduce\\*,
32. %HADOOP\_HOME%\share\hadoop\mapreduce\lib\\*,
33. %HADOOP\_HOME%\share\hadoop\yarn\\*,
34. %HADOOP\_HOME%\share\hadoop\yarn\lib\\*
35. </value>
36. </property>
37. Edit file **C:/Hadoop-3.1.2/etc/hadoop/hadoop-env.cmd** by closing the command line**"JAVA\_HOME=%JAVA\_HOME%"** instead of set **"JAVA\_HOME=C:\Java"** (On C:\java this is path to file jdk.18.0) OR **"JAVA\_HOME=C:\PROGRA~1\Java\jdk1.\*"**

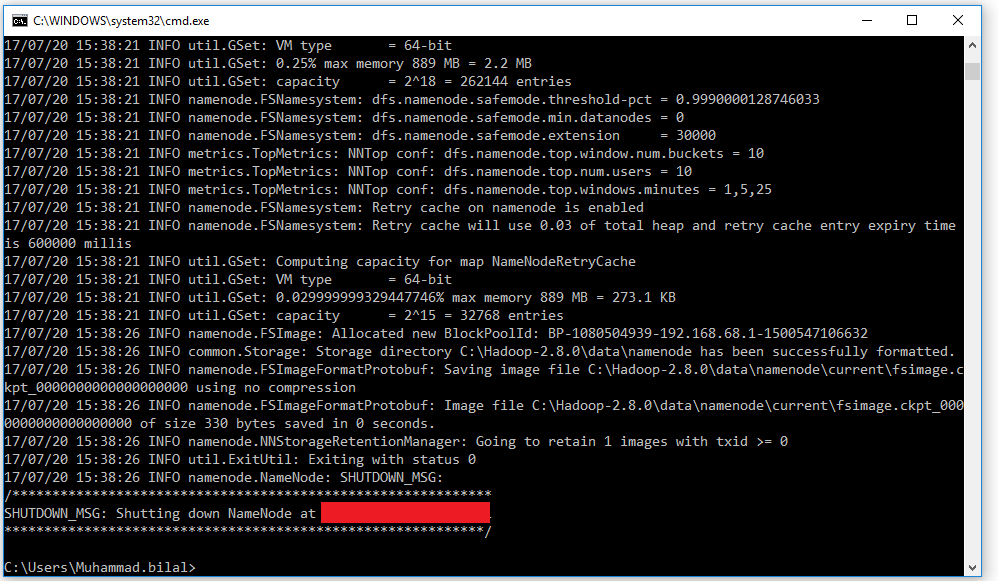


OR **( !!! check your path and jdk version and set)**

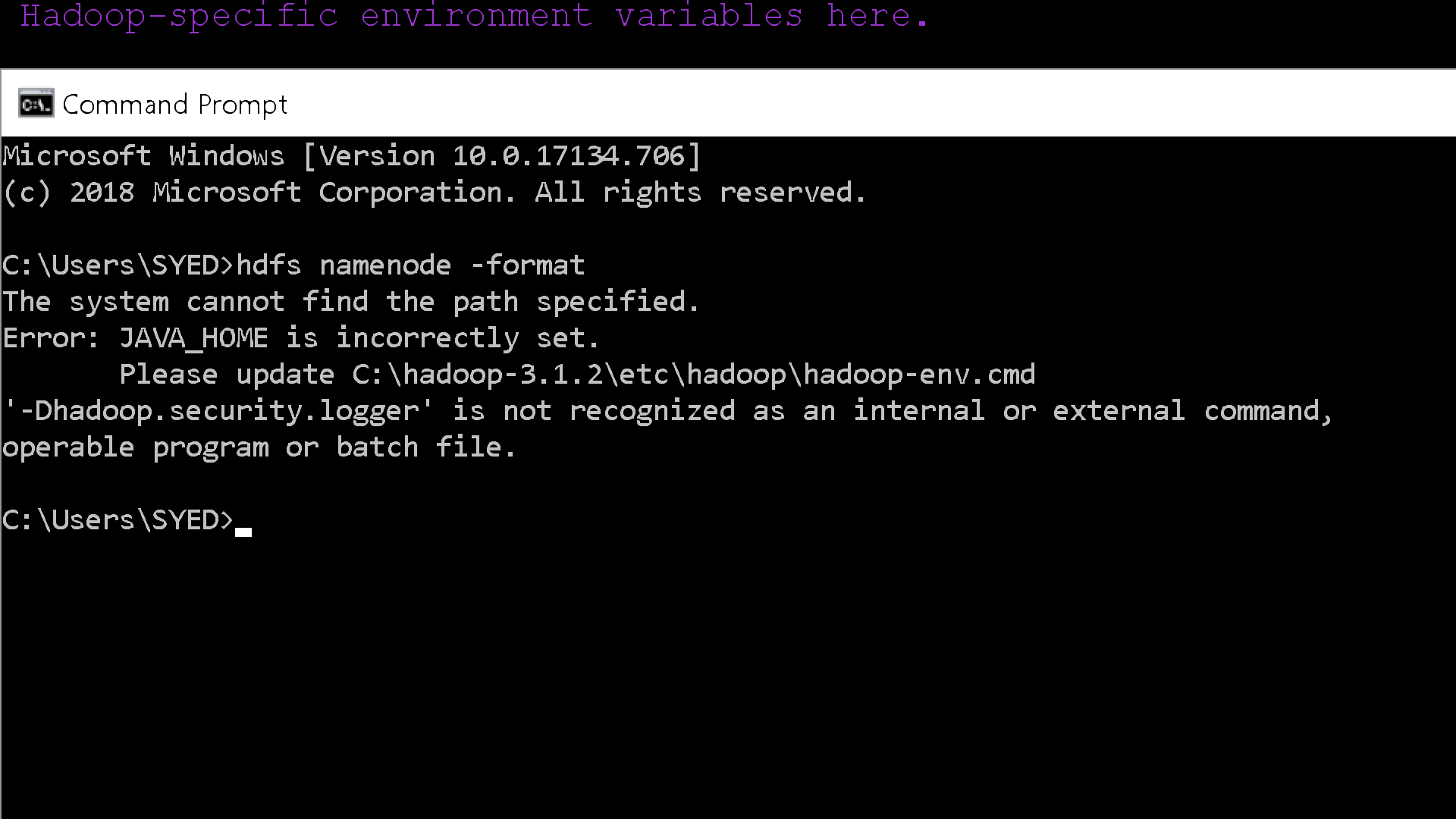


**Hadoop Configuration**

1. Hadoop needs windows OS specific files which does not come with default download of hadoop. To include those files, replace the bin folder in hadoop directory with the bin folder provided in this github link. <https://github.com/s911415/apache-hadoop-3.1.0-winutils>
2. Add the “bin” folder files you just downloaded in “C:\hadoop-3.1.2\bin” folder .Don’t replace this files if you are using different version of hadoop.
3. Open cmd prompt (in Administrative mode) and on typing command **"hdfs namenode –format"** you will see this.



* 1. If you get any error like this

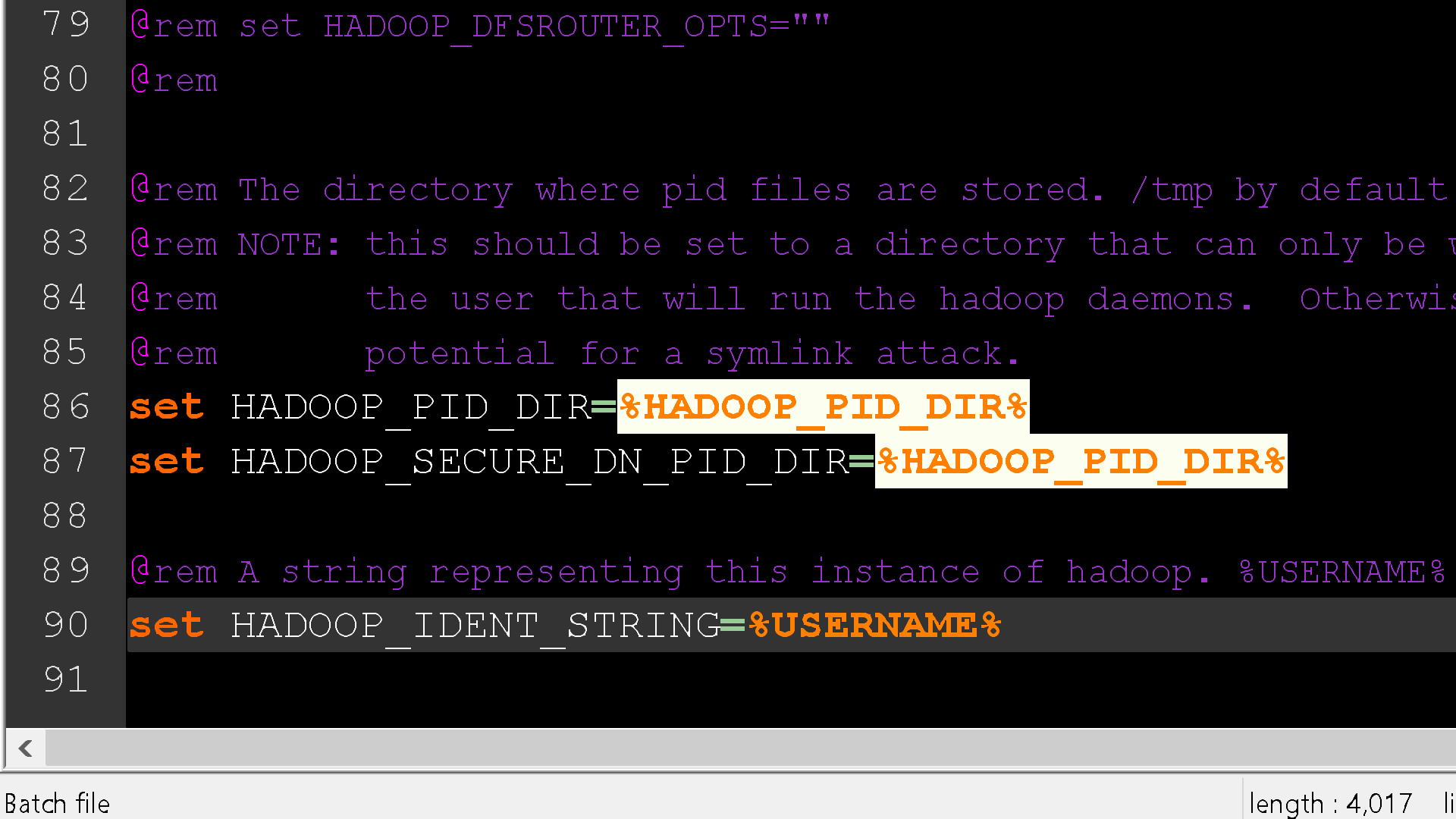


Please update your JAVA\_HOME path wherever you have used , like in creating an Environment Variable and Changing JAVA\_HOME in hadoop-env.cmd file .The path should **NOT** include Space in it like **“Program Files”.**

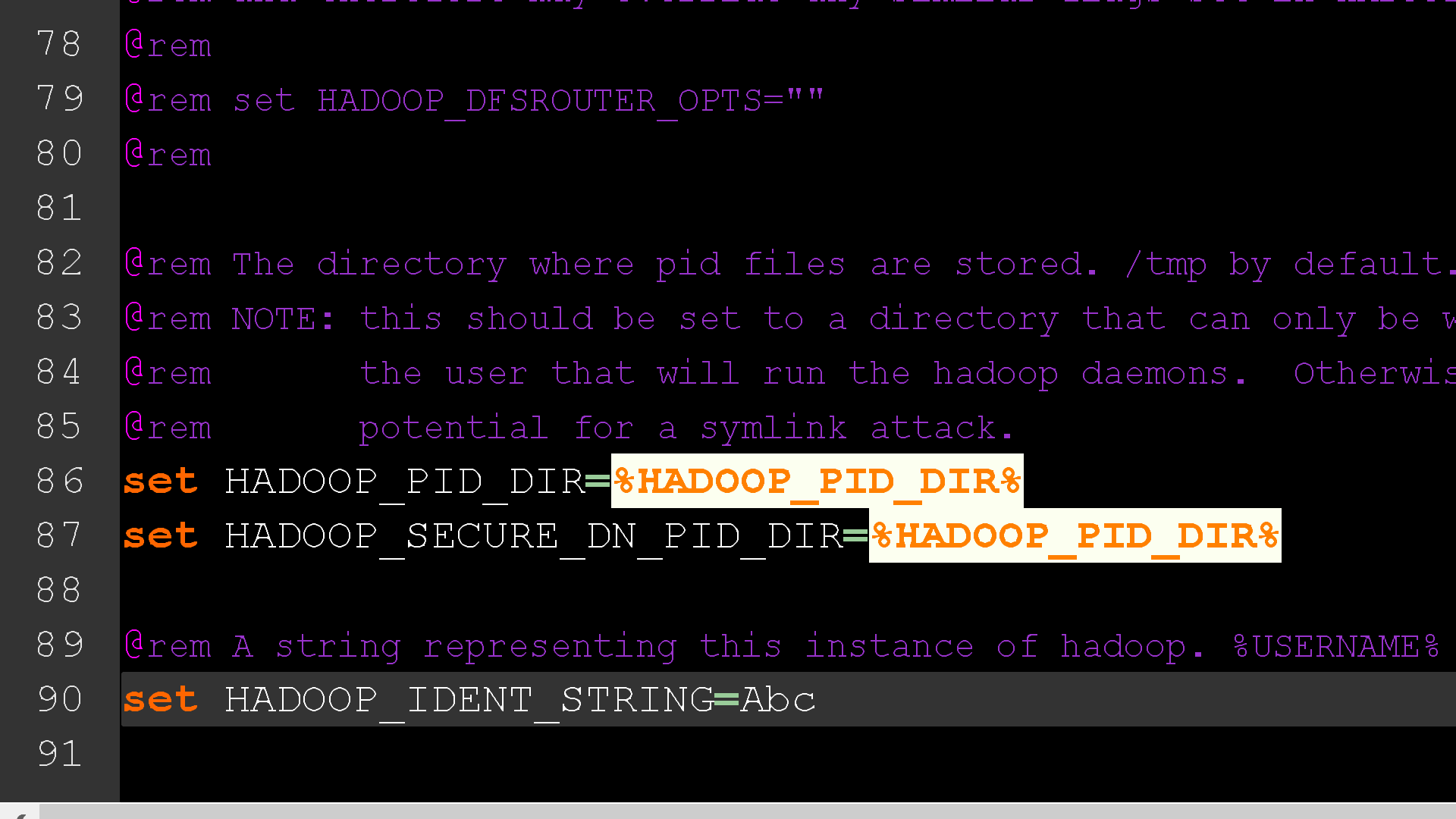
* 1. If your COMMAND PROMPT shows the path like this “**C:\Users\Abc Xyz>”** means if the path contain space in the tag or your name then follow this

Change the HADOOP\_IDENT\_STRING field in the hadoop-env.cmd file to your FIRST/any name which contain no spaces**.**

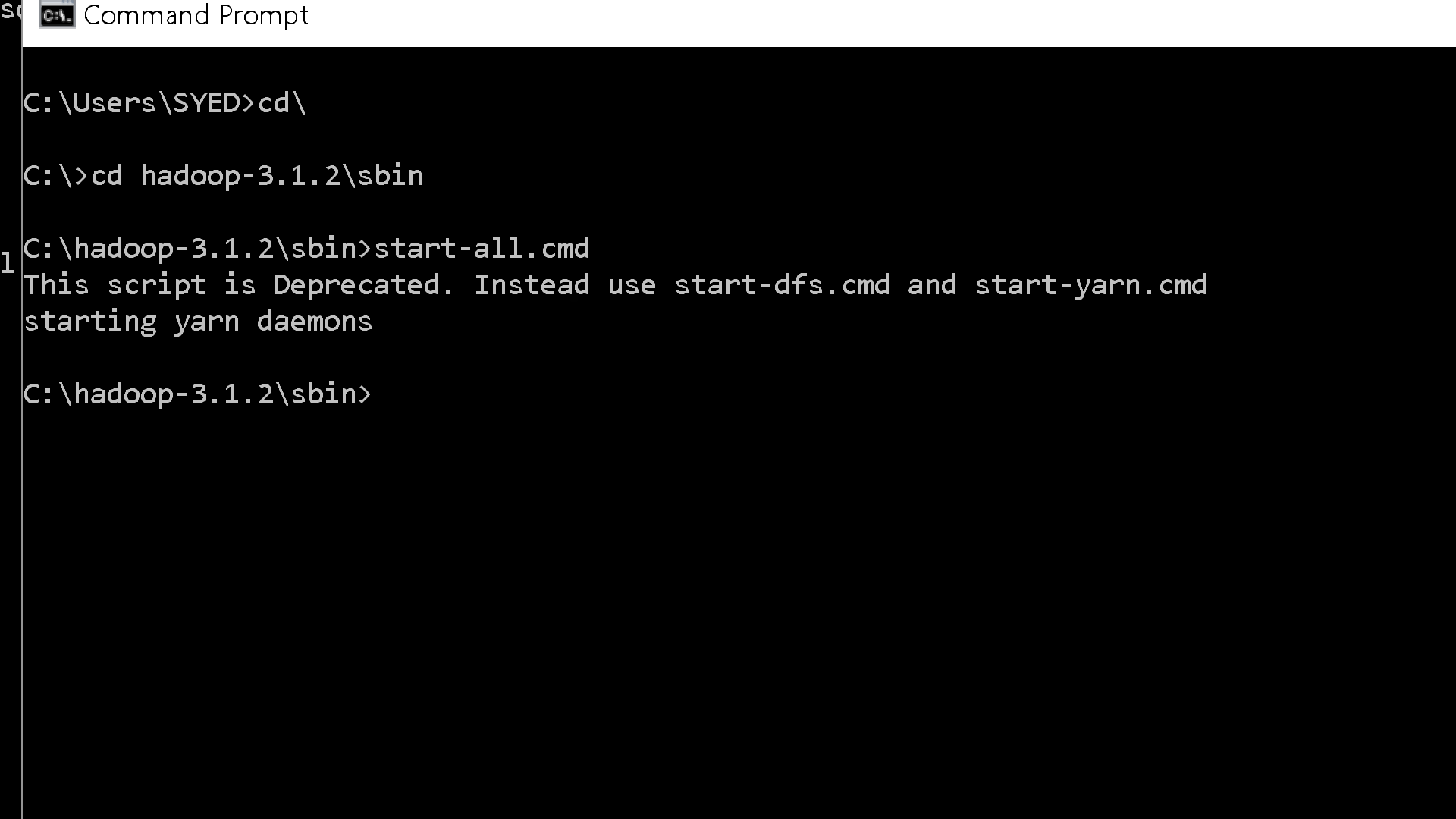
From this



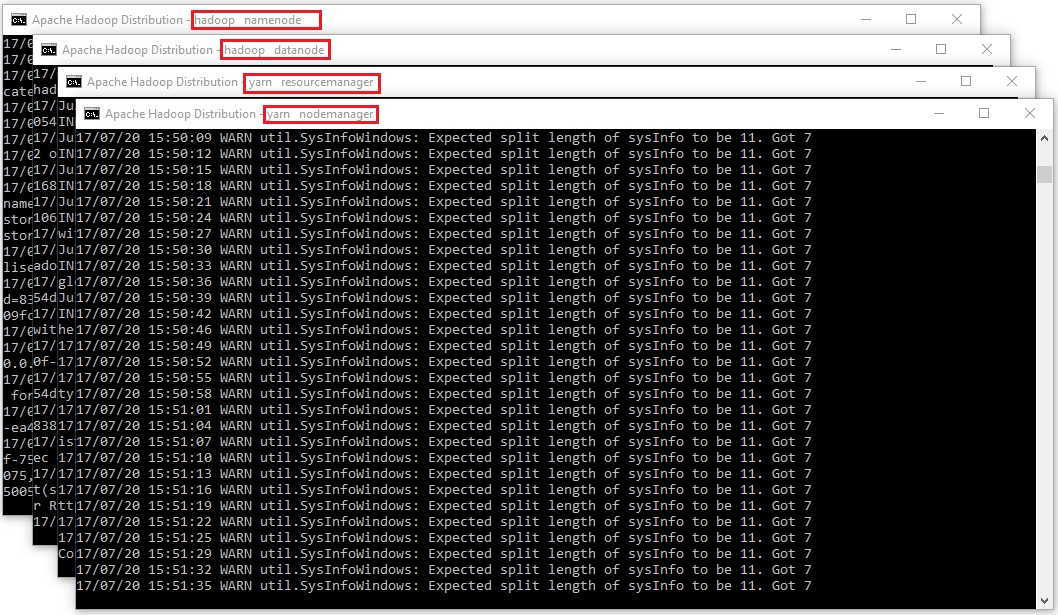
To



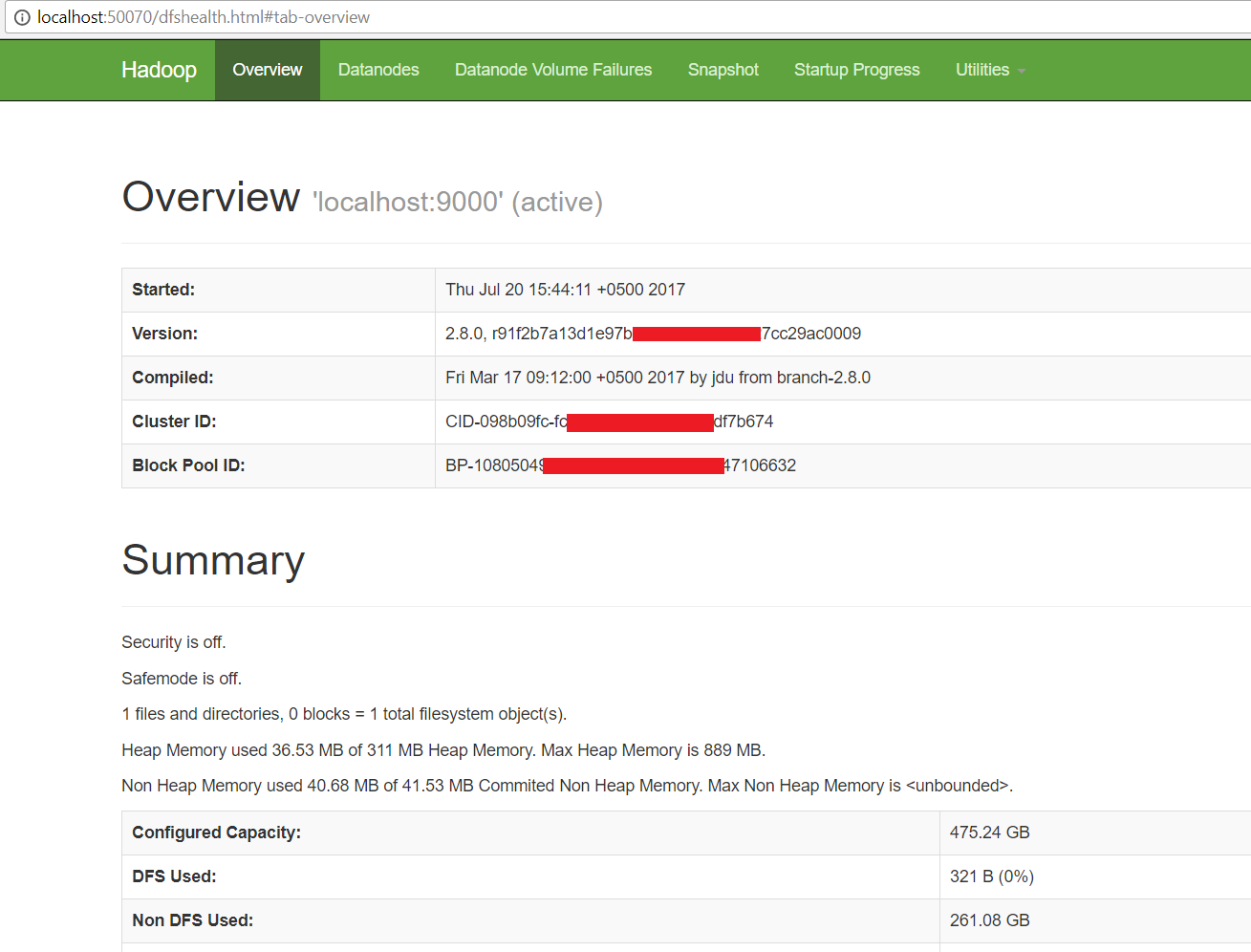
**Testing**

1. Open cmd and change directory to "C:\Hadoop-3.1.2\sbin" and type **"start-all.cmd"** to start apache. 
2. Make sure these apps are running

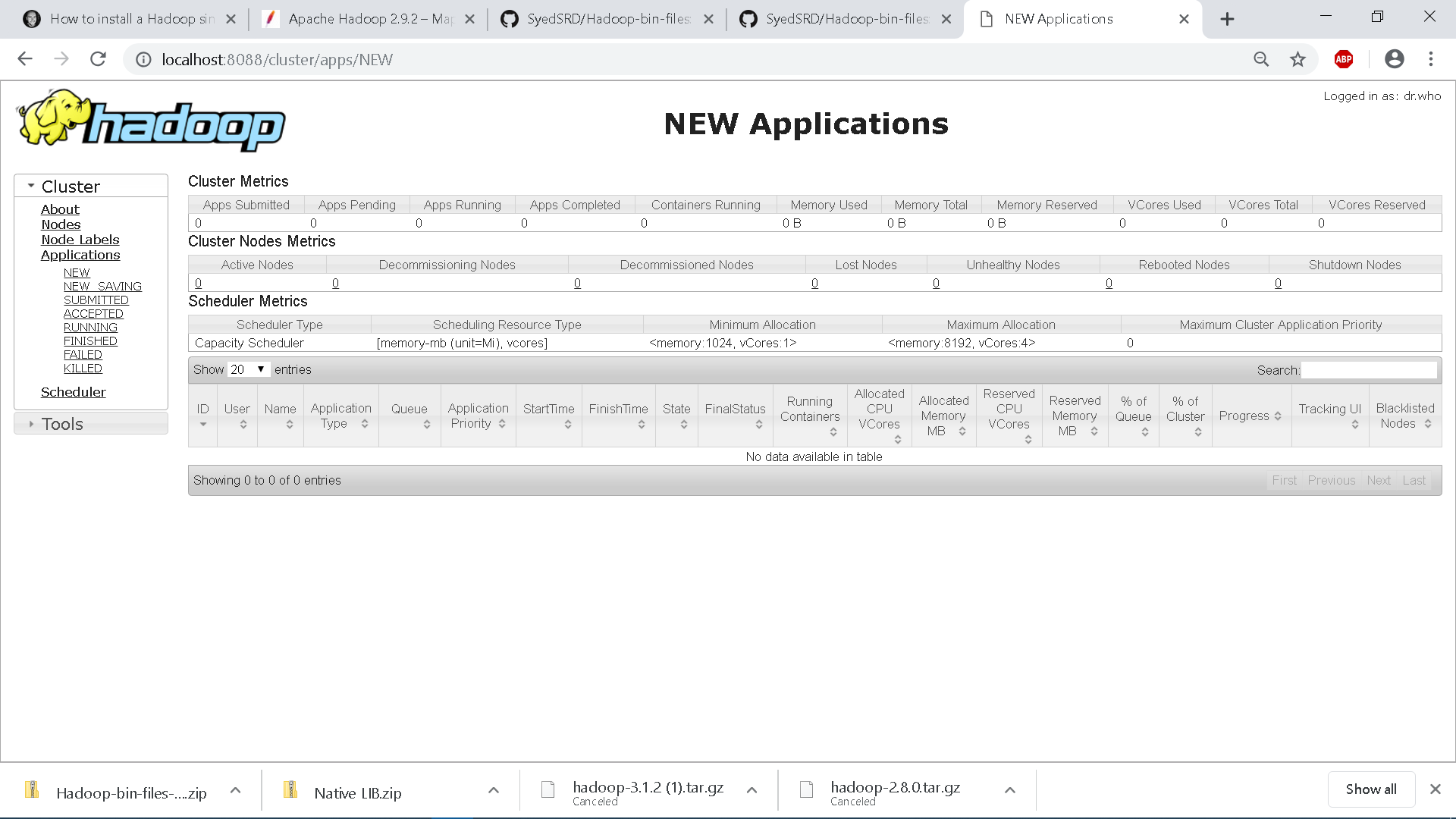
* Hadoop Namenode
* Hadoop datanode
* YARN Resource Manager
* YARN Node Manager



1. Open: [http://localhost:9870](http://localhost:9870088)

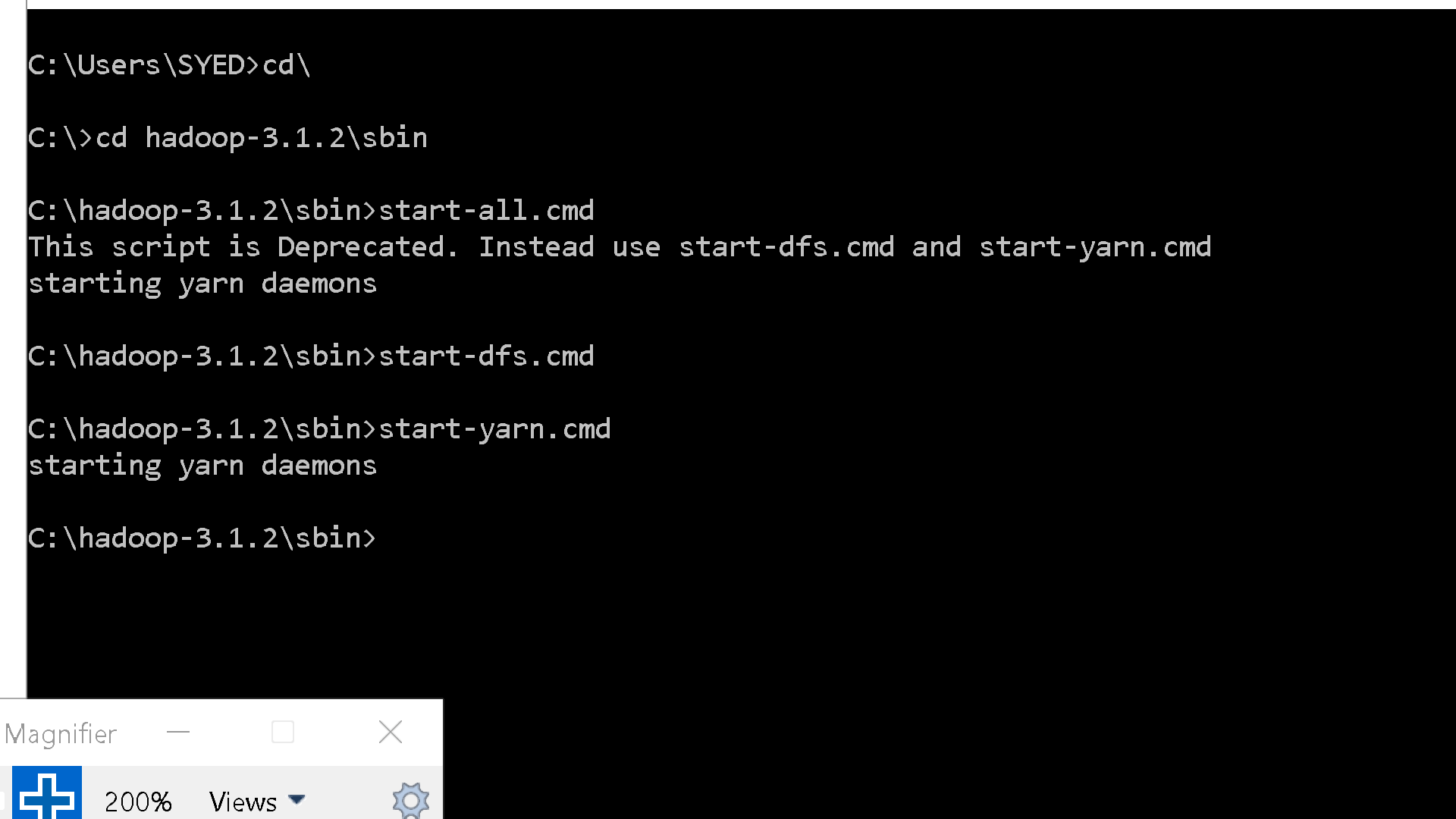


1. Open [http://localhost:8088](http://localhost:8088/)

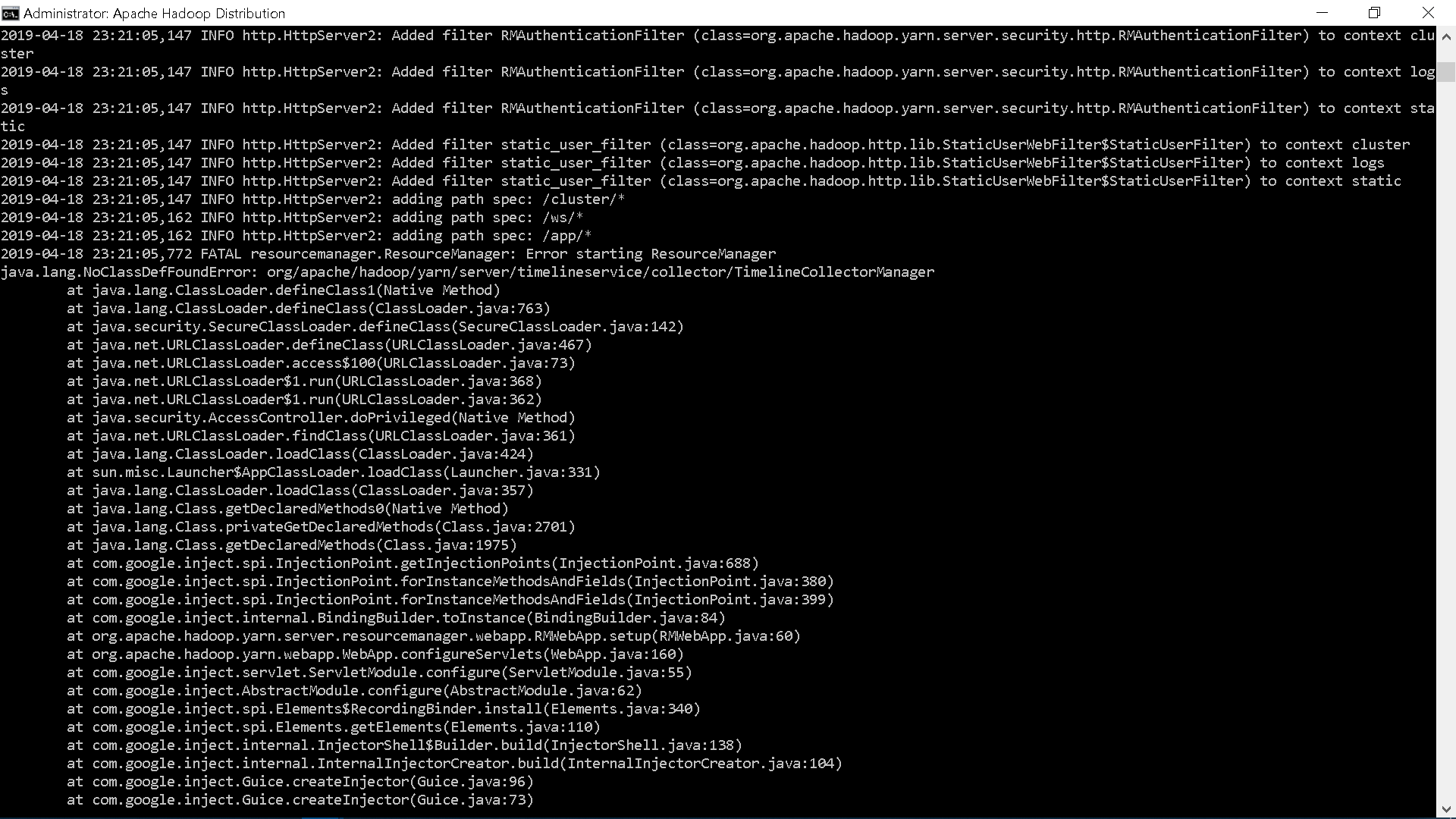


If the [http://localhost:9870](http://localhost:9870088) . or port 8088 doesn’t work then try typing these commands

1. Type **“start-dfs.cmd”** then on same console
2. Type **“start-yarn.cmd”**

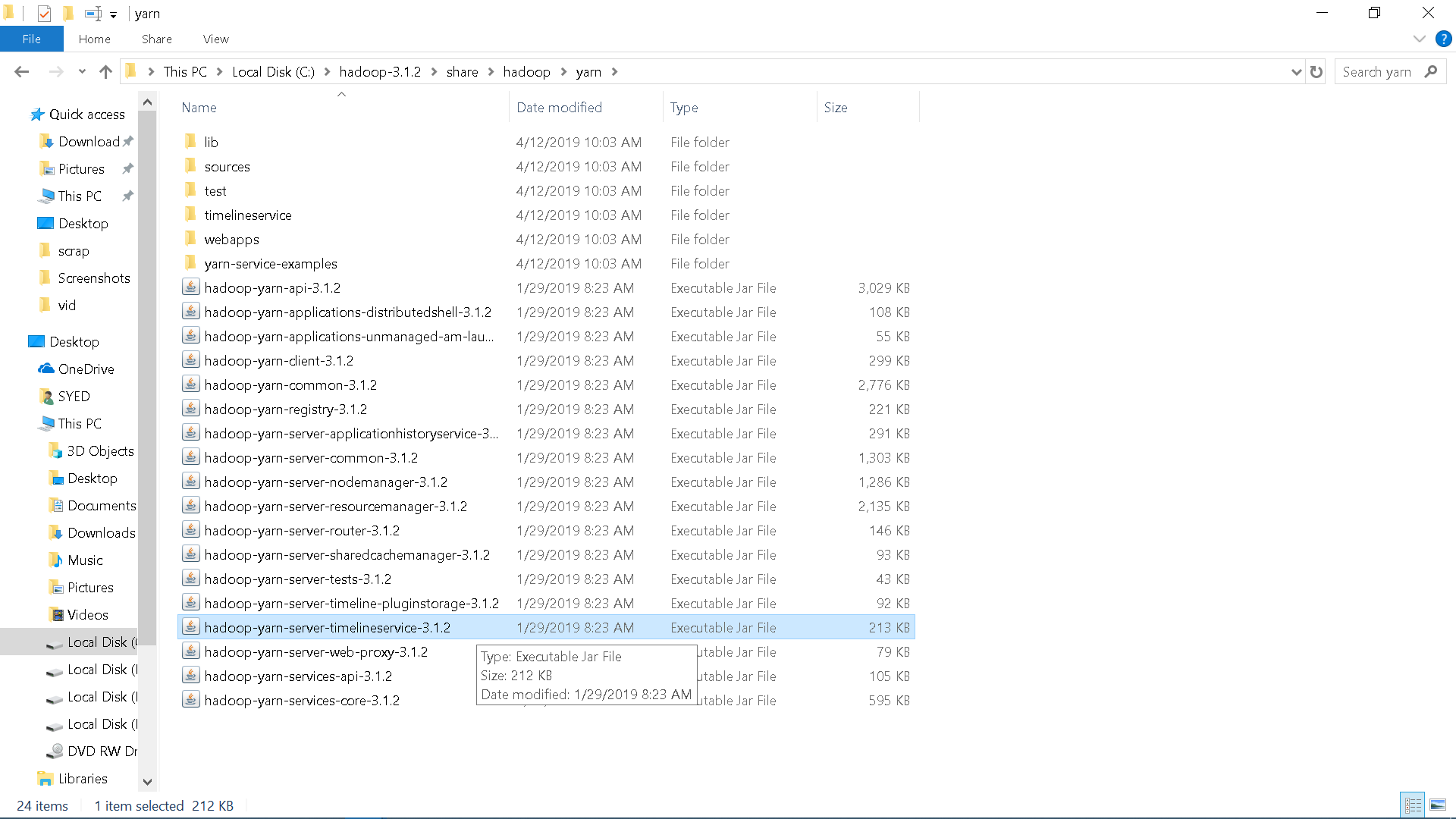


1. Then open [http://localhost:9870](http://localhost:9870088). AND Open [http://localhost:8088](http://localhost:8088/) ( if it still doesn’t open go to step 6)
2. If u get an error in administrative Apache hadoop Distribution or YARN Resource Manager Console like this

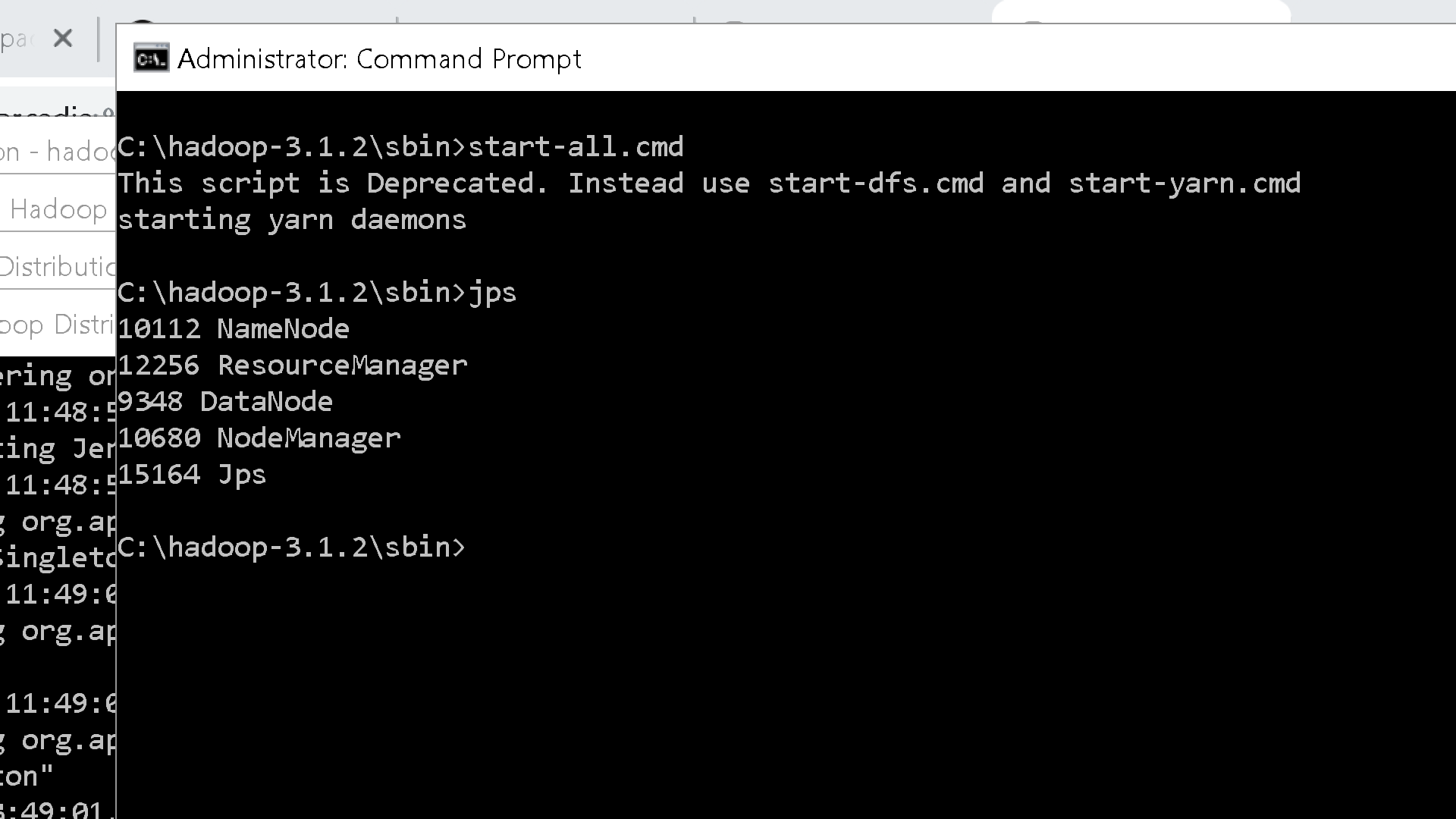


Then move the jar file “hadoop-yarn-server-timelineservice-3.1.2” **FROM** C:\hadoop-3.1.2\share\hadoop\yarn\timelineserviceor %HADOOP\_HOME% \share\hadoop\yarn\timelineservice

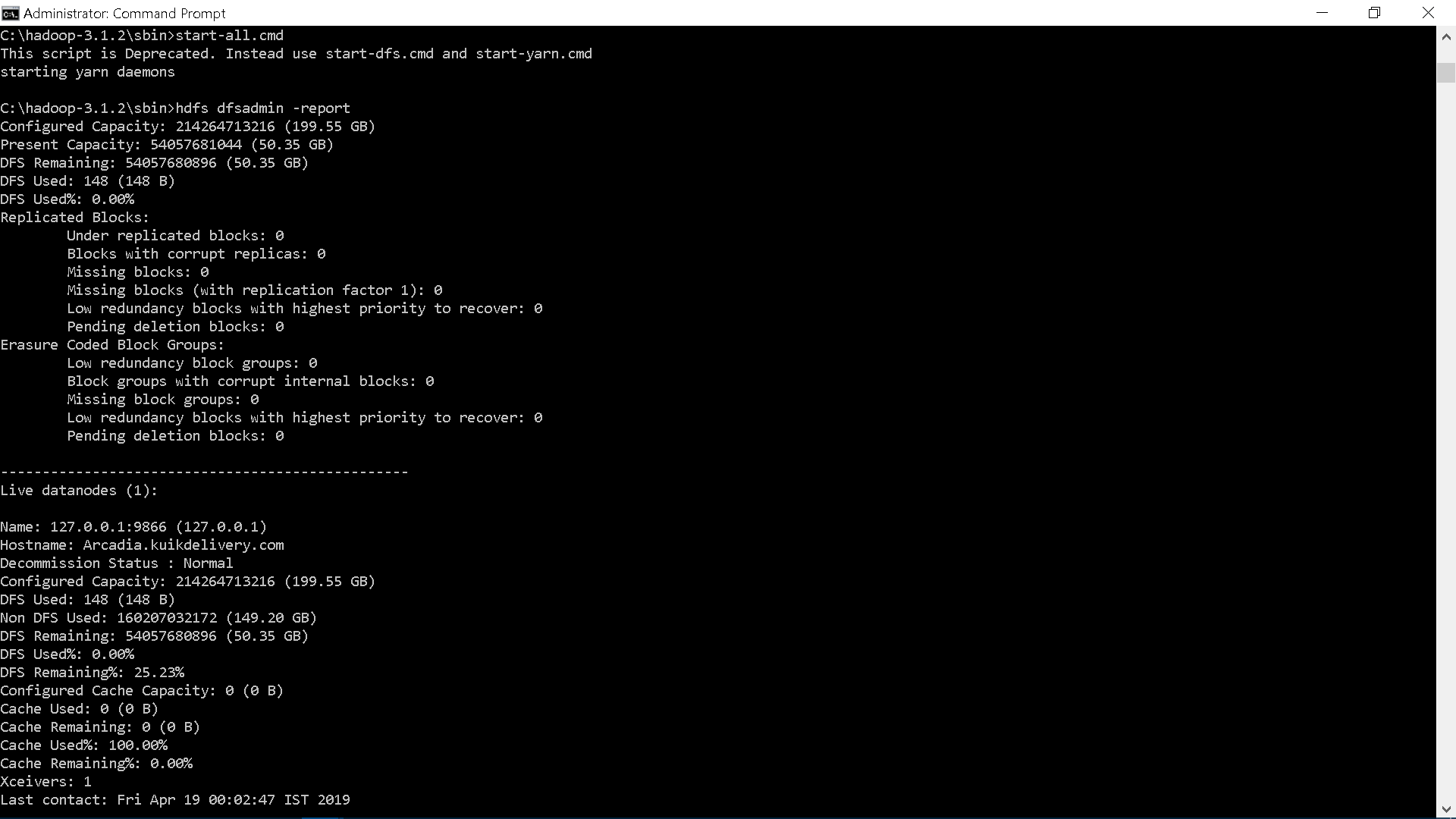
**TO** C:\hadoop-3.1.2\share\hadoop\yarn (to the yarn folder) this will fix the error , run the command once again (start-all.cmd) and check.



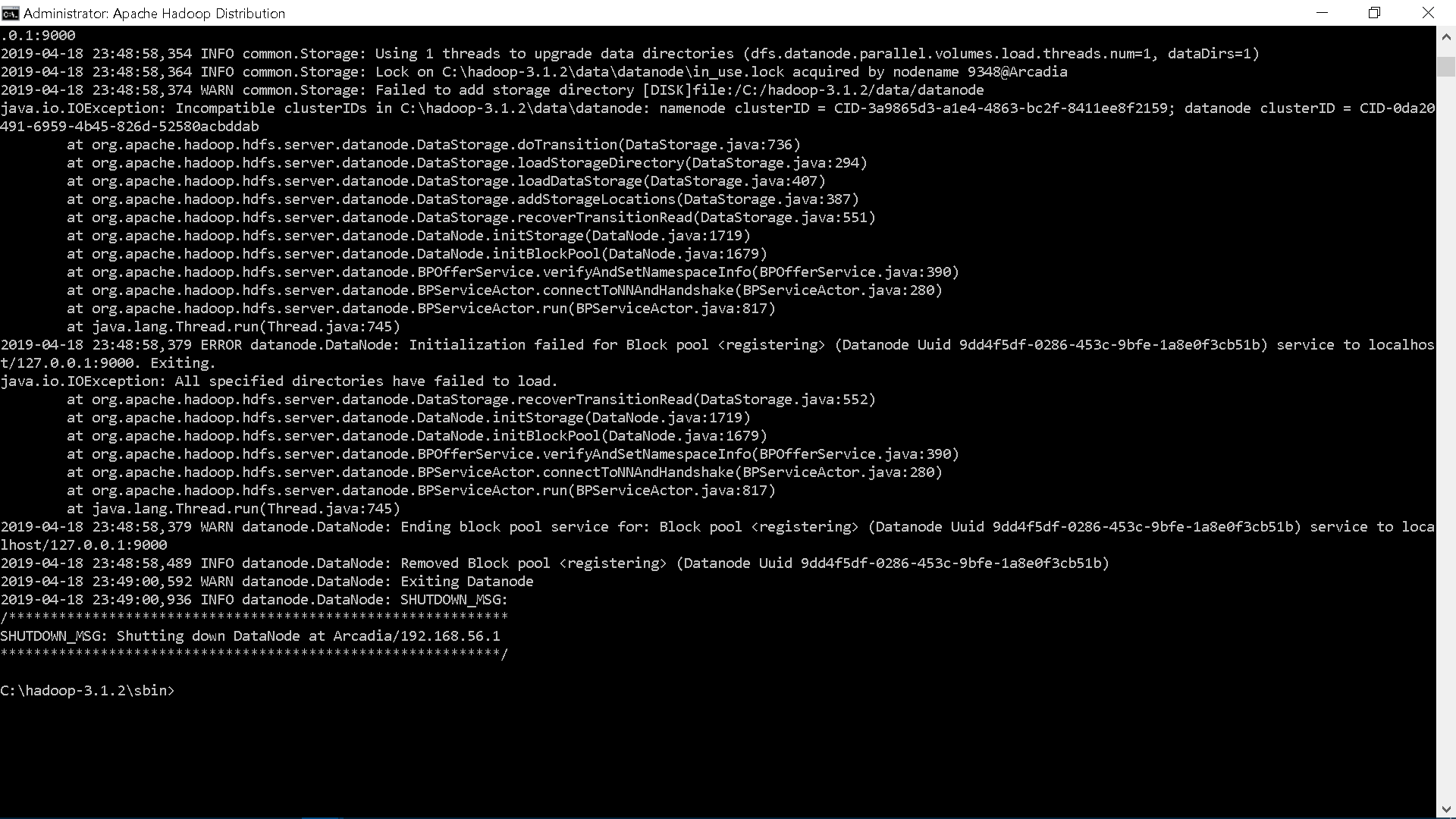
1. Testing whether all services running correctly or not , run this command **“jps”** after running **“start-all.cmd”**

****

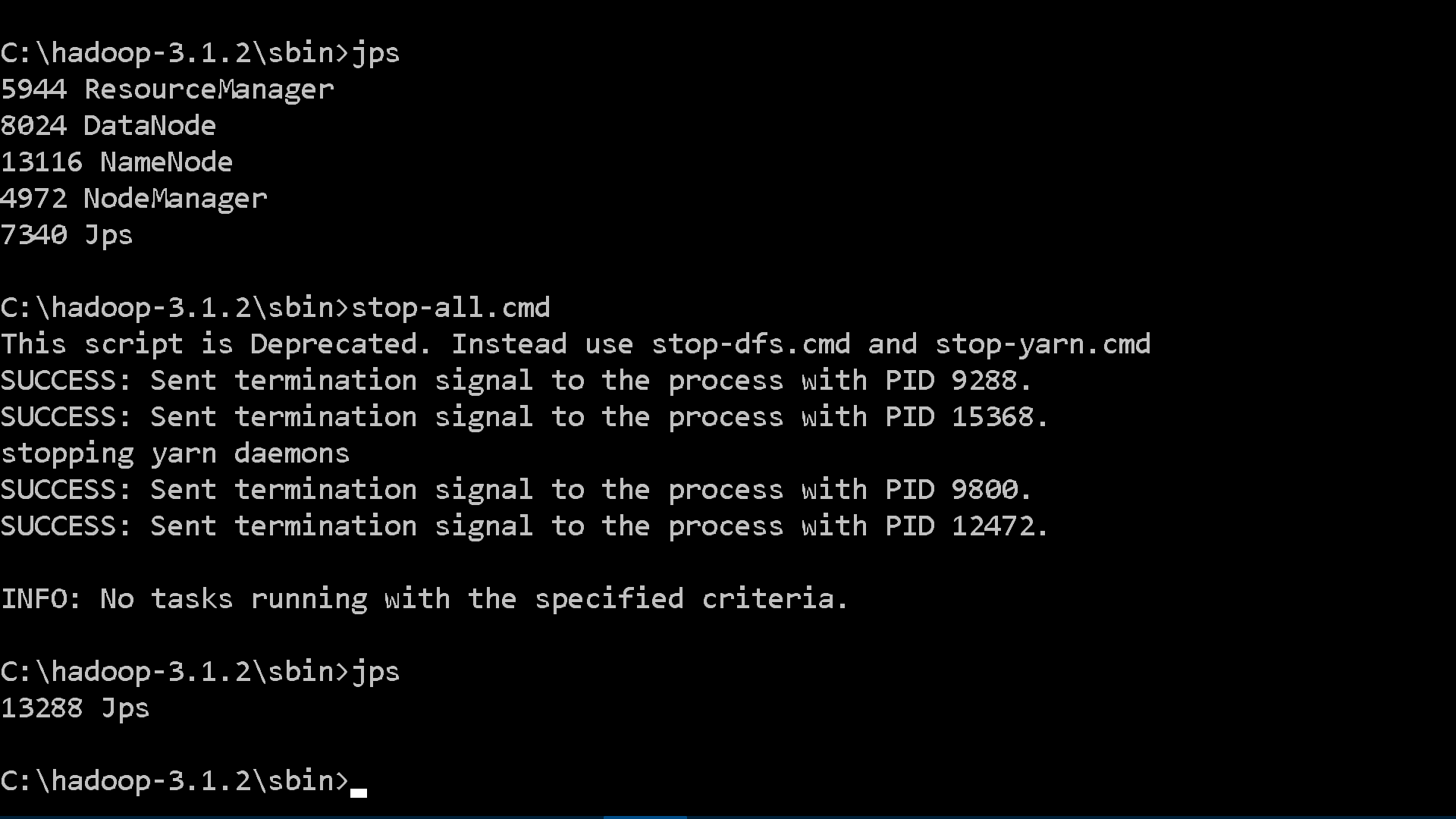
1. Run this command to view namenode correctly configured or not “**hdfs dfsadmin –report”** you should get the output like this



1. If u get output of yarn resource manager like this



Then type **“stop-all.cmd”** in the main console.



Then go to “**C:\hadoop-3.1.2\data\datanode”** and **clean(delete all files and folder in)** the directory or folder ,and also do the same for namenode **“C:\hadoop-3.1.2\data\namenode”** folder. Then type **“hdfs namenode –format” .**Then go head and type **“start-all.cmd”.** This time it will work.

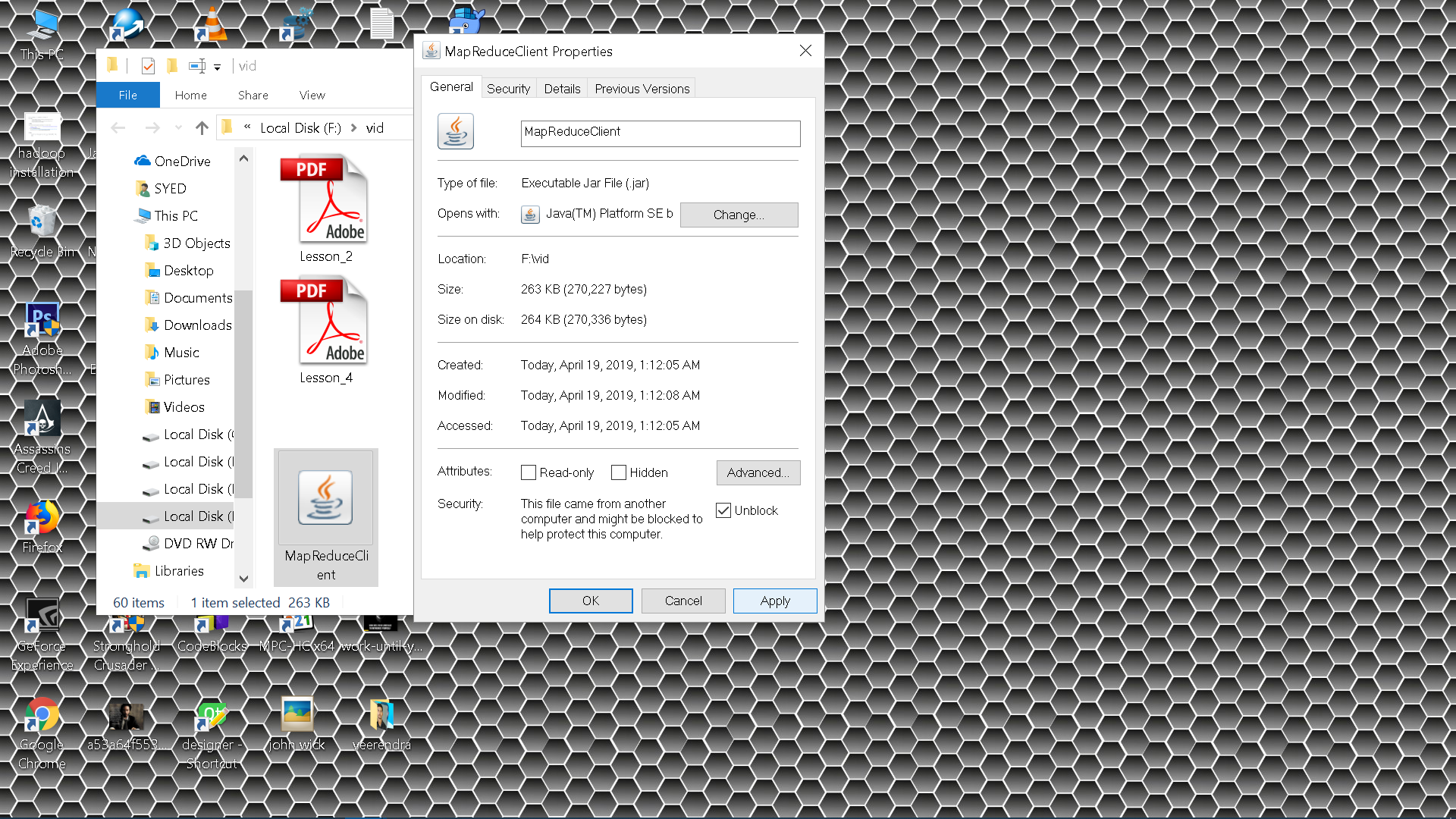
Congratulations !!!

Note:

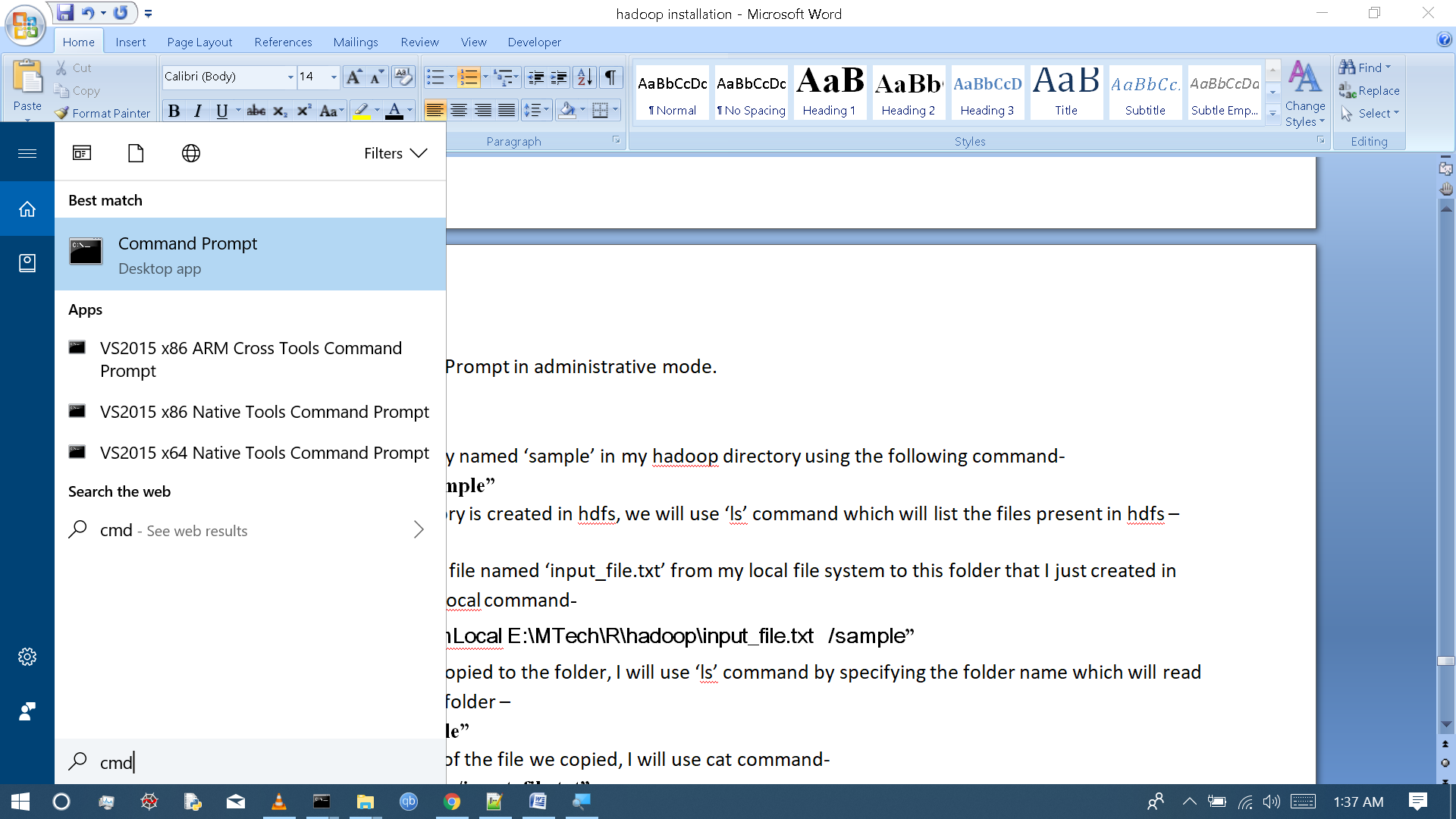
* If you are using other versions of Hadoop this may not work ,bcz the apache-hadoop-3.1.0-winutils-master.zip bin folder contains specific files of the hadoop 3.1.2 version’s bin files. It may work if you add only those files which are not present in the bin folder of your hadoop version. but there is no guarantee it work.so better to find those dependencies specific to your version.
* This is tested on windows 10 only. It may or may not work on other windows versions.
* If you are downloaded hadoop-2.8.0 this process may work (link: <https://github.com/MuhammadBilalYar/Hadoop-On-Window/wiki/Step-by-step-Hadoop-2.8.0-installation-on-Window-10> )

Running MapReduce:

1. Download these files and store it in local folder.
   1. Download MapReduceClient.jar (Link: <https://github.com/MuhammadBilalYar/HADOOP-INSTALLATION-ON-WINDOW-10/blob/master/MapReduceClient.jar>)
   2. Download Input\_file.txt (Link: <https://github.com/MuhammadBilalYar/HADOOP-INSTALLATION-ON-WINDOW-10/blob/master/input_file.txt>)
   3. Go to properties of these files and check the unblock box. And then click apply.



1. Now open Command Prompt in administrative mode.(press windows key and type cmd ,Right click on Command Prompt and select Run as Administrator),Now follow the hdfs commands.



* 1. I will create a directory named ‘sample’ in my hadoop directory using the following command-

**“hdfs dfs –mkdir /sample”**

1. To verify if the directory is created in hdfs, we will use ‘ls’ command which will list the files present in hdfs –

**“hdfs dfs –ls /”**

1. Then I will copy a text file named ‘input\_file.txt’ from my local file system to this folder that I just created in hdfs using copyFromLocal command**-**

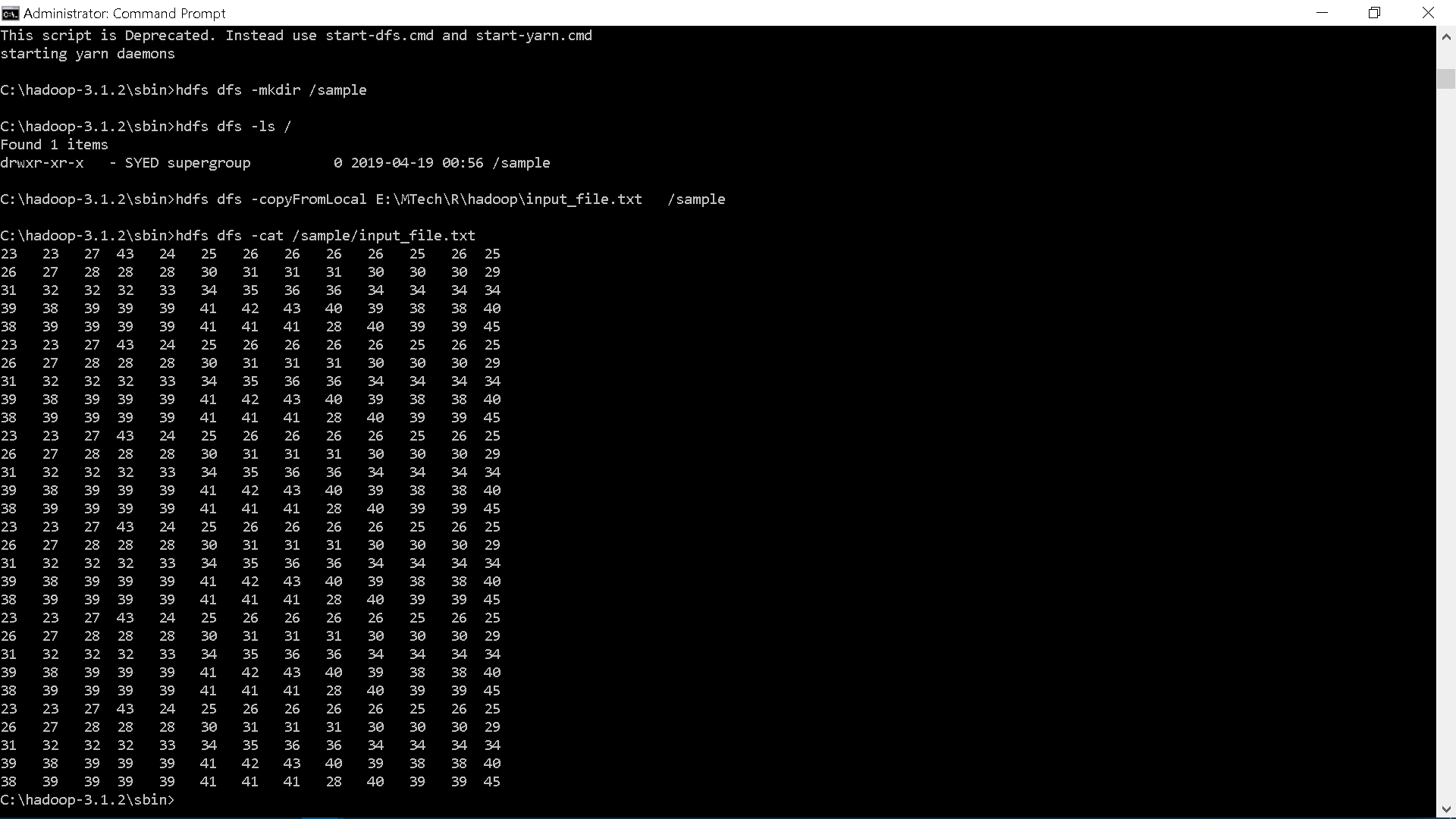
**“hdfs dfs -copyFromLocal E:\MTech\R\hadoop\input\_file.txt /sample”**

1. To verify if the file is copied to the folder, I will use ‘ls’ command by specifying the folder name which will read the list of files in that folder –

**“hdfs dfs –ls /sample”**

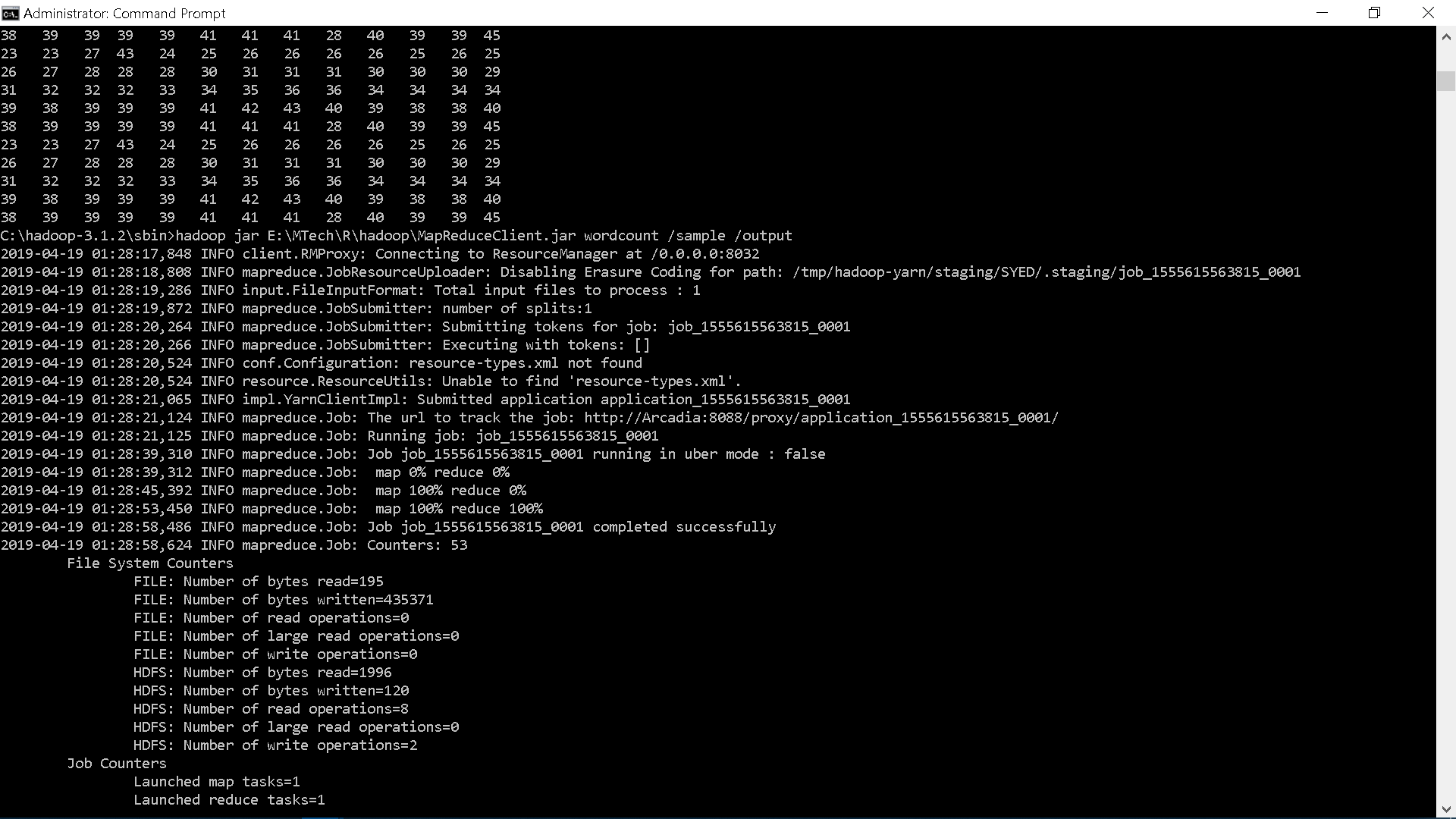
1. To view the contents of the file we copied, I will use cat command-

**“hdfs dfs –cat /sample/input\_file.txt”**

****

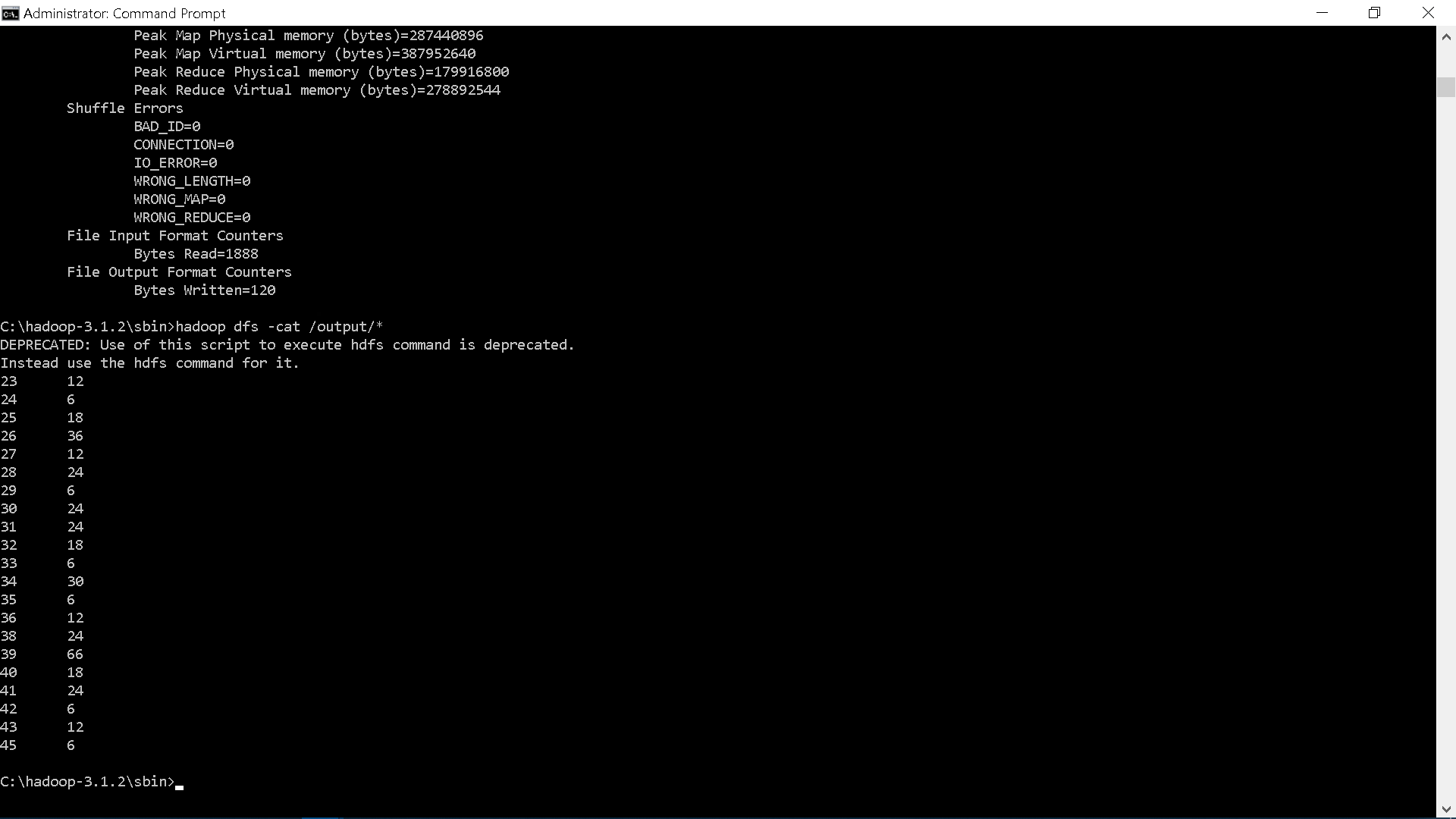
1. To run MapReduce I will use this command-

**“hadoop jar E:\MTech\R\hadoop\MapReduceClient.jar wordcount /sample /output “**

****

1. To view the output run this command –

**“hadoop dfs –cat /output/\*”**

****