

DOCUMENTATION FOR INSTALLATION OF HADOOP:

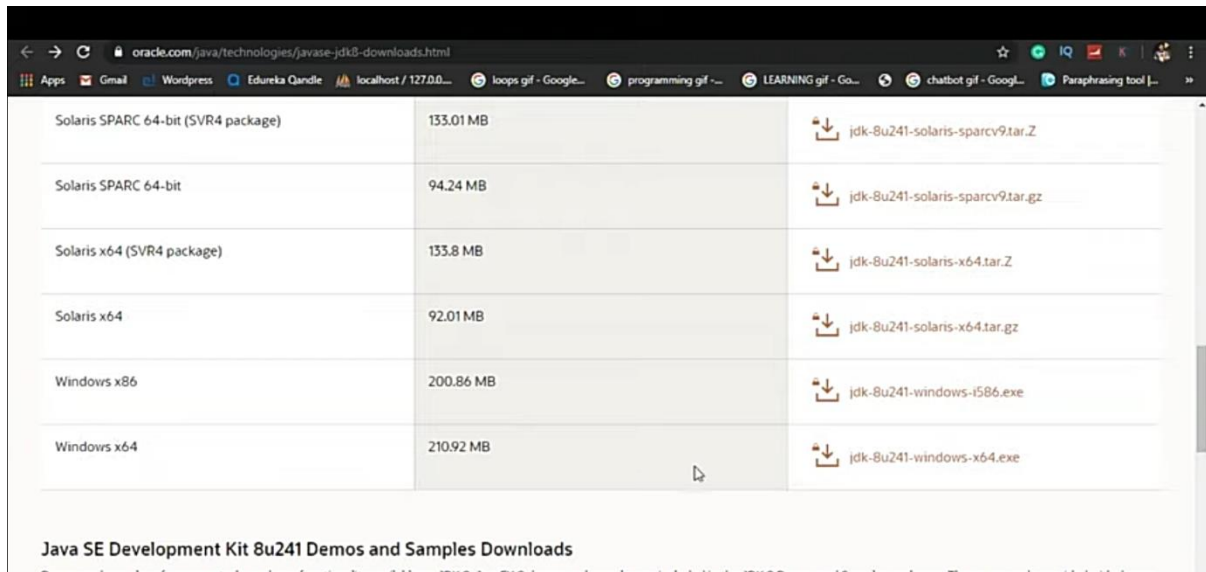
Group G1

Step 1:

Prerequisites to install Hadoop: Java v8 must be installed

Download jdk-8 for your system as your System Compatibility

<https://www.oracle.com/in/java/technologies/javase/javase-jdk8-downloads.html>



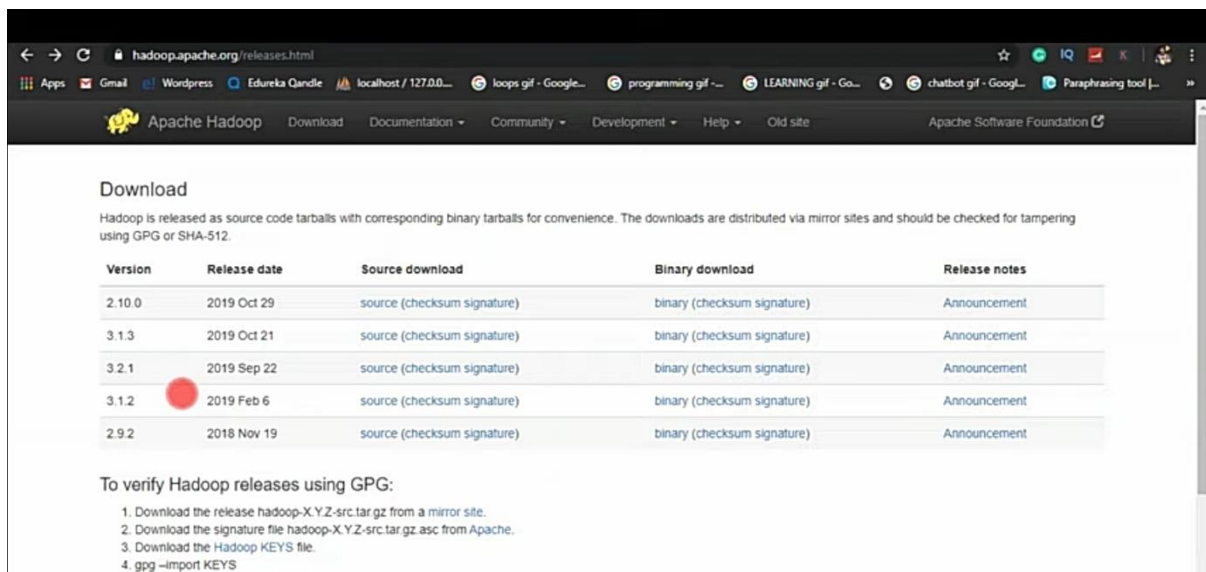
Oracle will ask to create an account for downloading, So create it.

Step 2:

Download Apache Hadoop

<https://hadoop.apache.org/releases.html>

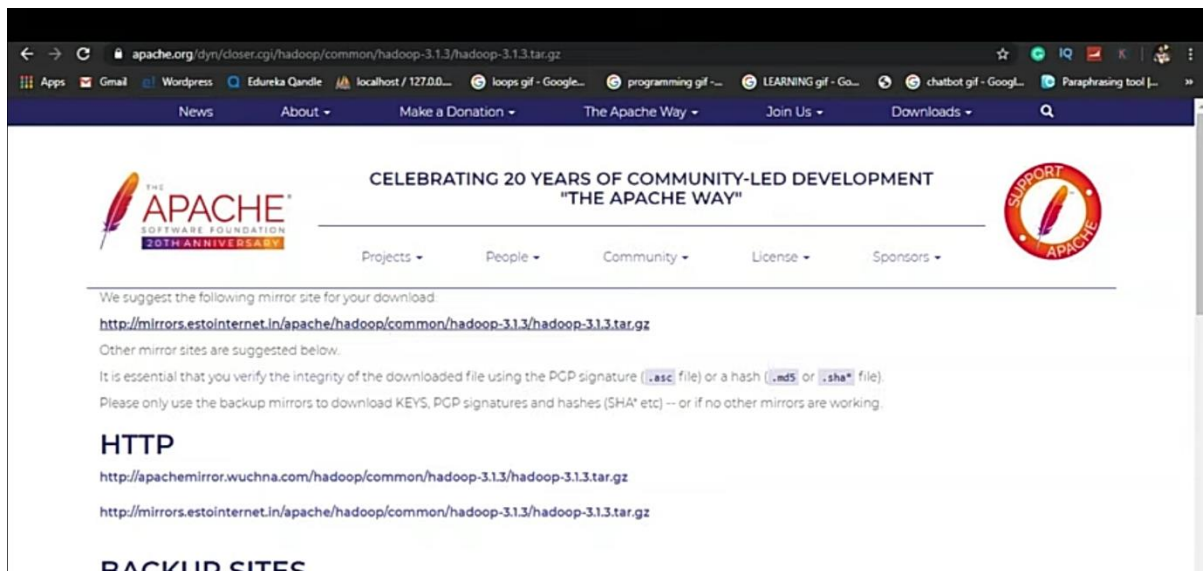
You can download the latest version, but recommended version is 3.2.1



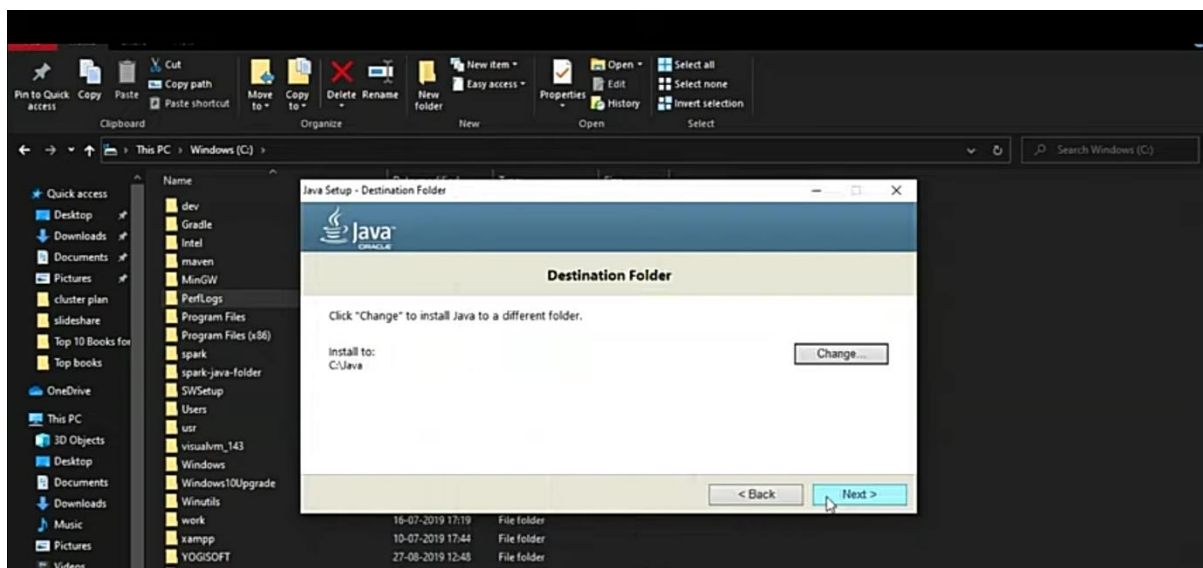
Step 3:

Select binary file in front of your desired version of Hadoop. You will be redirected to a Page. Select the mirror link on that web page. Downloading of tar file will be downloaded.

<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz>

**Step 4:****If java is already installed Skip to Step 8**

Install jdk in the C: Drive as shown

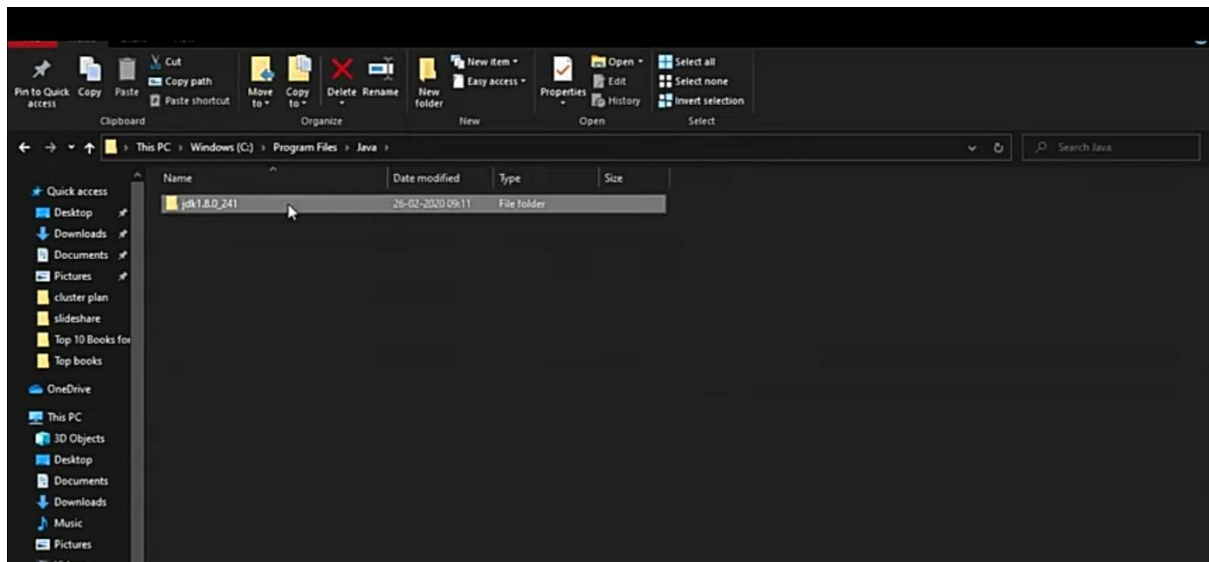


Step 5:

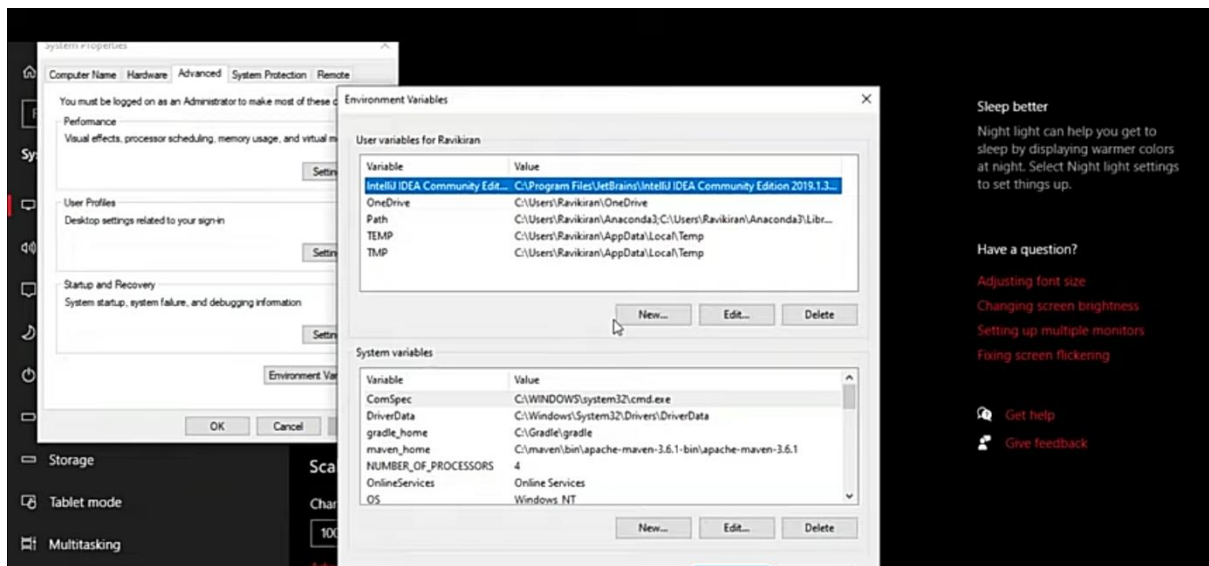
After installation there will be a **jdk** folder in “C:\Program Files\Java\jdk”

Cut Paste this **jdk** folder in our installed path “C:\Java”

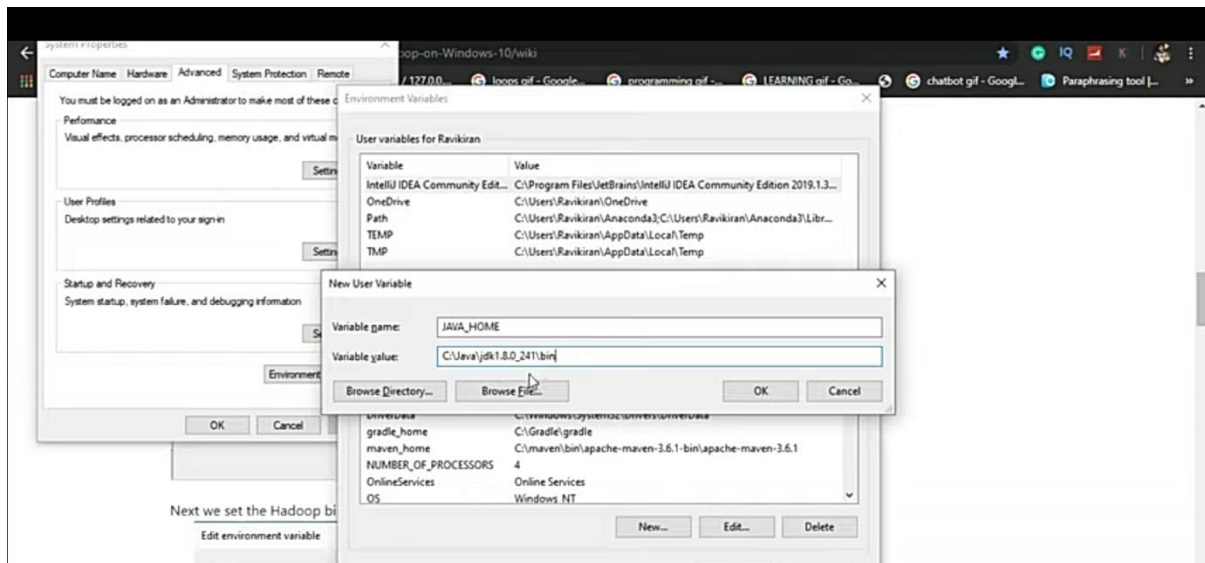
Delete the previous Java folder in Program files

**Step 6:**

Set the System Environment Variables for Java as shown in below screenshot (**Set it in User Variable Section**)

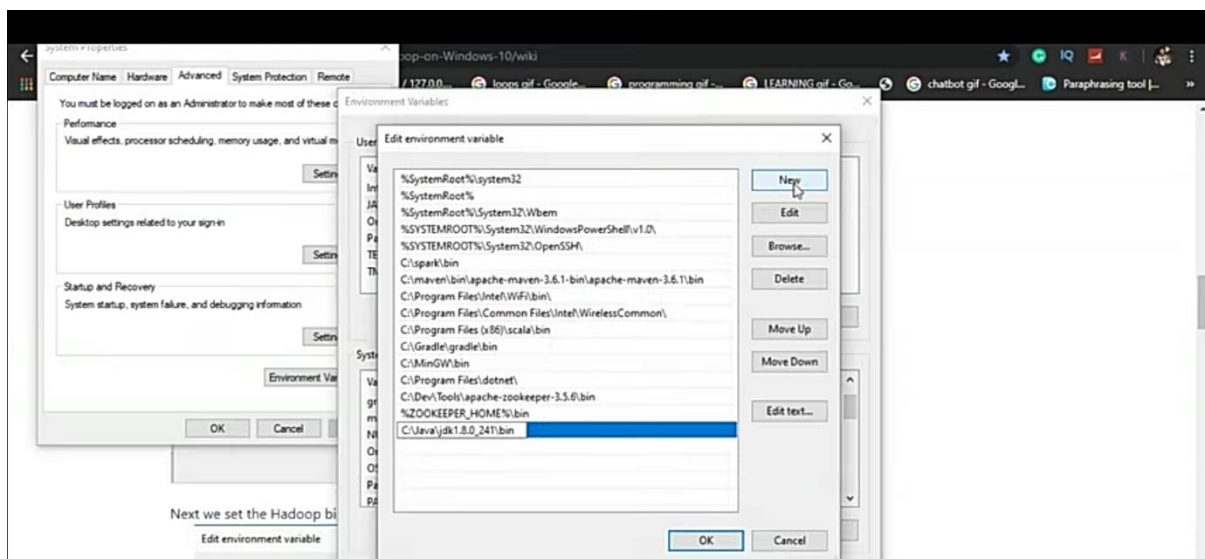


Name the variable “JAVA_HOME”

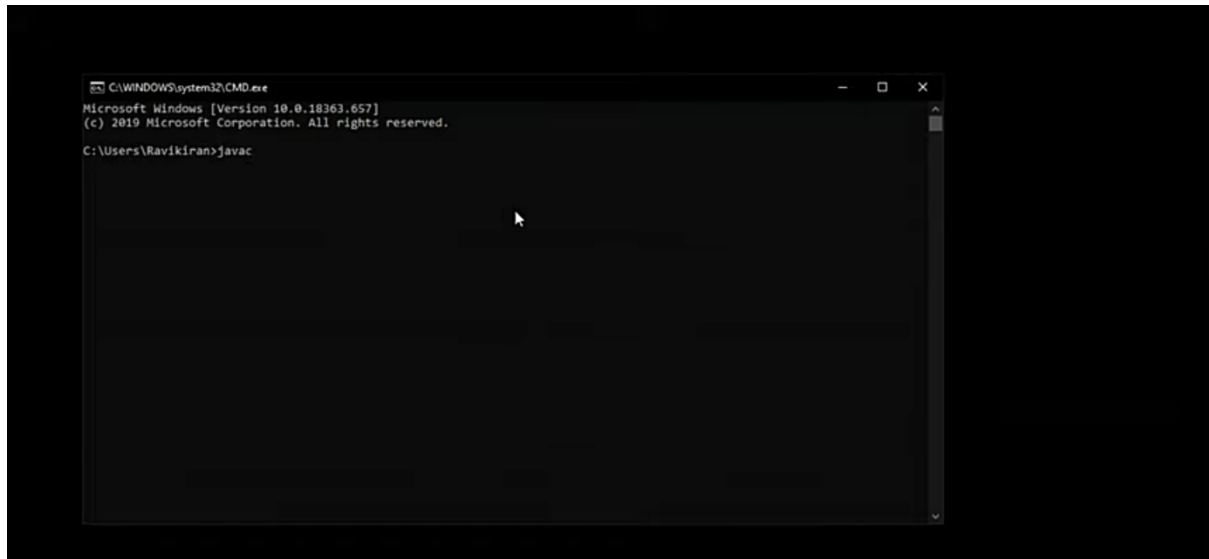


Step 7:

Now in System Variable section Select **System Variables** in that select **Path** and **Edit**
Here Set path to **jdk** bin location as shown below



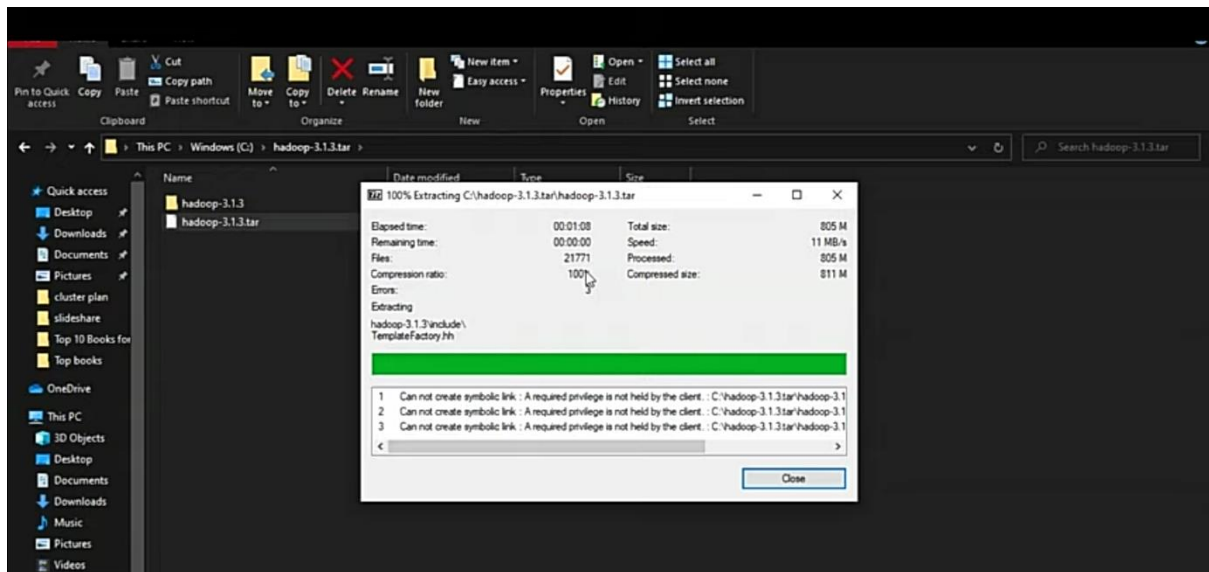
To check java installation type **javac** or **java -version** in Command Prompt



Step 8:

Installing Hadoop

Extract the tar file ,Ignore if Errors pop up as shown



Rename the extracted file to **“hadoop”**

Step 9:

Go to: "C:\hadoop-3.2.1\etc\hadoop"

We have to edit 5 files namely

1 core-site.xml

2 hdfs-site.xml

3 mapred-site.xml

4 yarn-site.xml

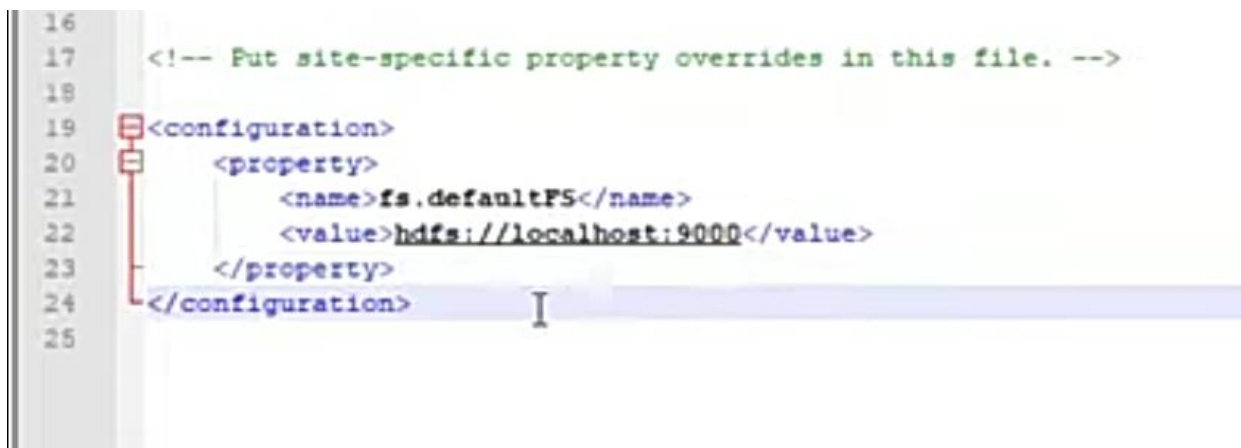
5 hadoop-env

open all in Notepad++

Step 10 :

Editing Core-site.xml

SEE THE LINE NUMBERS TO BE EDITED



```
16
17  <!-- Put site-specific property overrides in this file. -->
18
19  <configuration>
20    <property>
21      <name>fs.defaultFS</name>
22      <value>hdfs://localhost:9000</value>
23    </property>
24  </configuration>
25
```

Step 11 :

Editing Mapred-site.xml

SEE THE LINE NUMBERS TO BE EDITED



```
16
17  <!-- Put site-specific property overrides in this file. -->
18
19  <configuration>
20    <property>
21      <name>mapreduce.framework.name</name>
22      <value>yarn</value>
23    </property>
24  </configuration>
25
```


Step 12 :

Editing Yarn-site.xml

SEE THE LINE NUMBERS TO BE EDITED

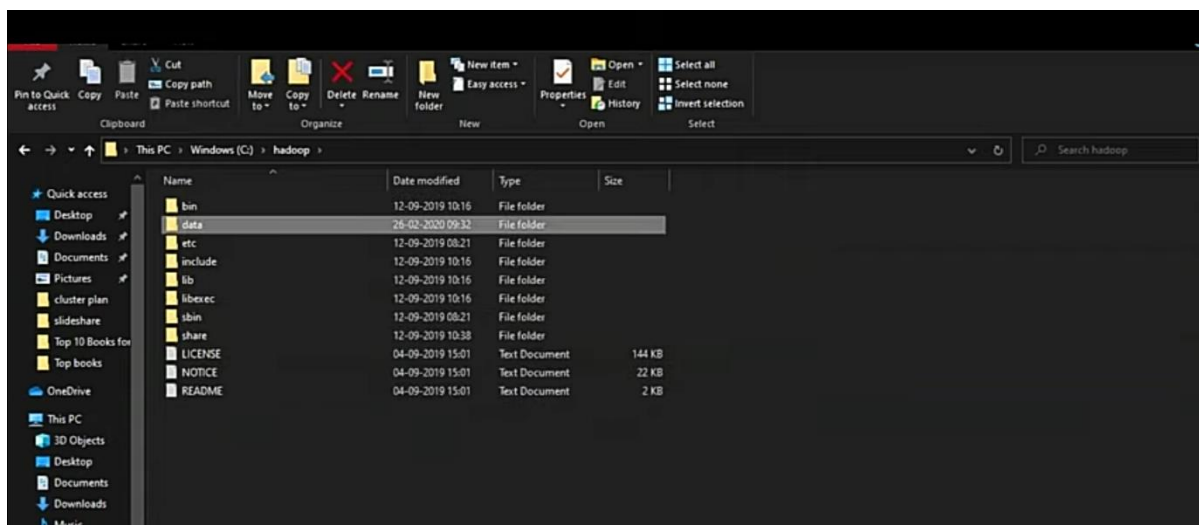
```

12      See the License for the specific language governing permissions and
13      limitations under the License. See accompanying LICENSE file.
14
15      <configuration>
16      <property>
17          <name>yarn.nodemanager.aux-services</name>
18          <value>mapreduce_shuffle</value>
19      </property>
20      <property>
21          <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
22          <value>org.apache.hadoop.mapred.ShuffleHandler</value>
23      </property>
24      </configuration>
25

```

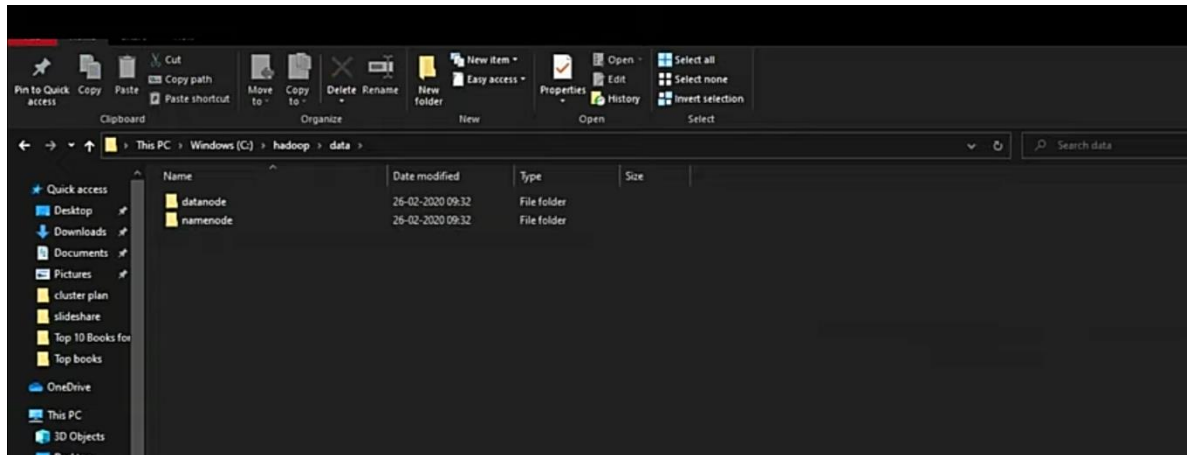
Step 13:

Create a new folder named **data** in C:\hadoop as shown



Step 14 :

Create 2 more folders in data named as **namenode** and **datanode**

**Step 15 :**

Editing hdfs-site.xml

SEE THE LINE NUMBERS TO BE EDITED



Make sure you Paste the paths of name node and data node

SAVE ALL THE EDITED FILES

Step 16 :

Edit hadoop-env.cmd

Paste the path of jdk in front of “set JAVA_HOME ” line (**E.g.: set JAVA_HOME=c:\Java\jdk...**)

SEE THE LINE NUMBERS TO BE EDITED

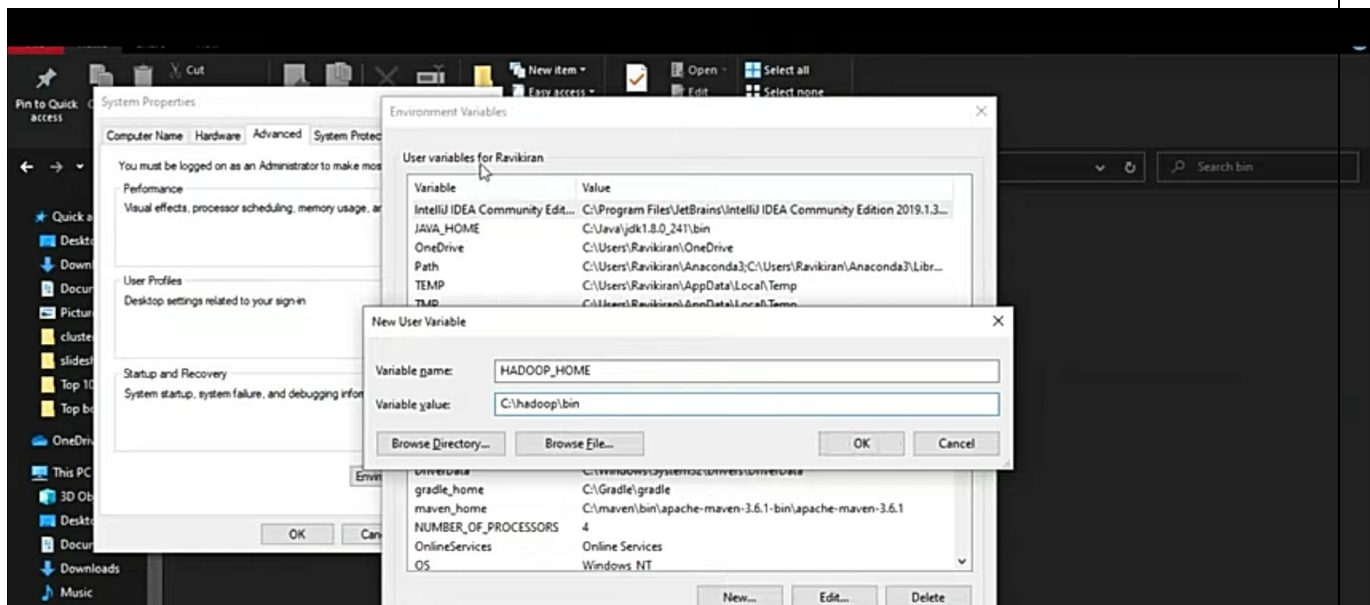
```

16
17 #rem Set Hadoop-specific environment variables here.
18
19 #rem The only required environment variable is JAVA_HOME.
20 #rem optional. When running a distributed configuration it
21 #rem set JAVA_HOME in this file, so that it is correctly de
22 #rem remote nodes.
23
24 #rem The java implementation to use. Required.
25 set JAVA_HOME=C:\Java\jdk1.8.0_241
26
27 #rem The jsvc implementation to use. Jsvc is required to run
28 #rem set JSVC_HOME=%JSVC_HOME%

```

Step 17 :

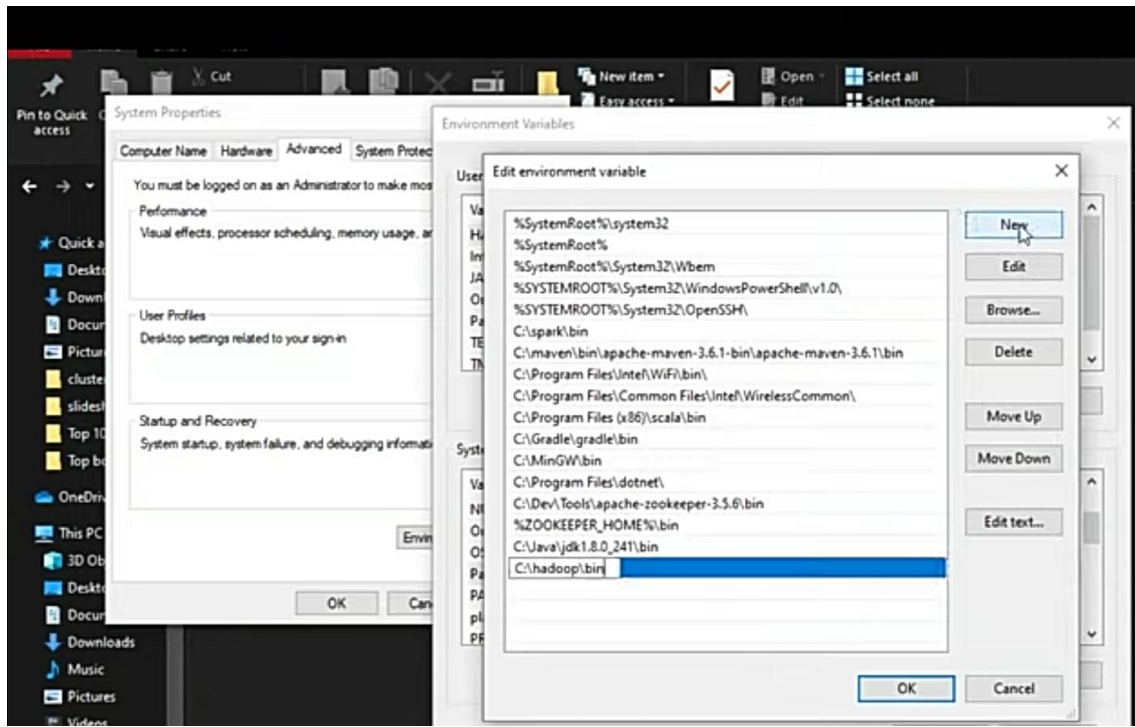
Set Environment Variables for Hadoop Home as shown below (**Set it in User Variable Section**)



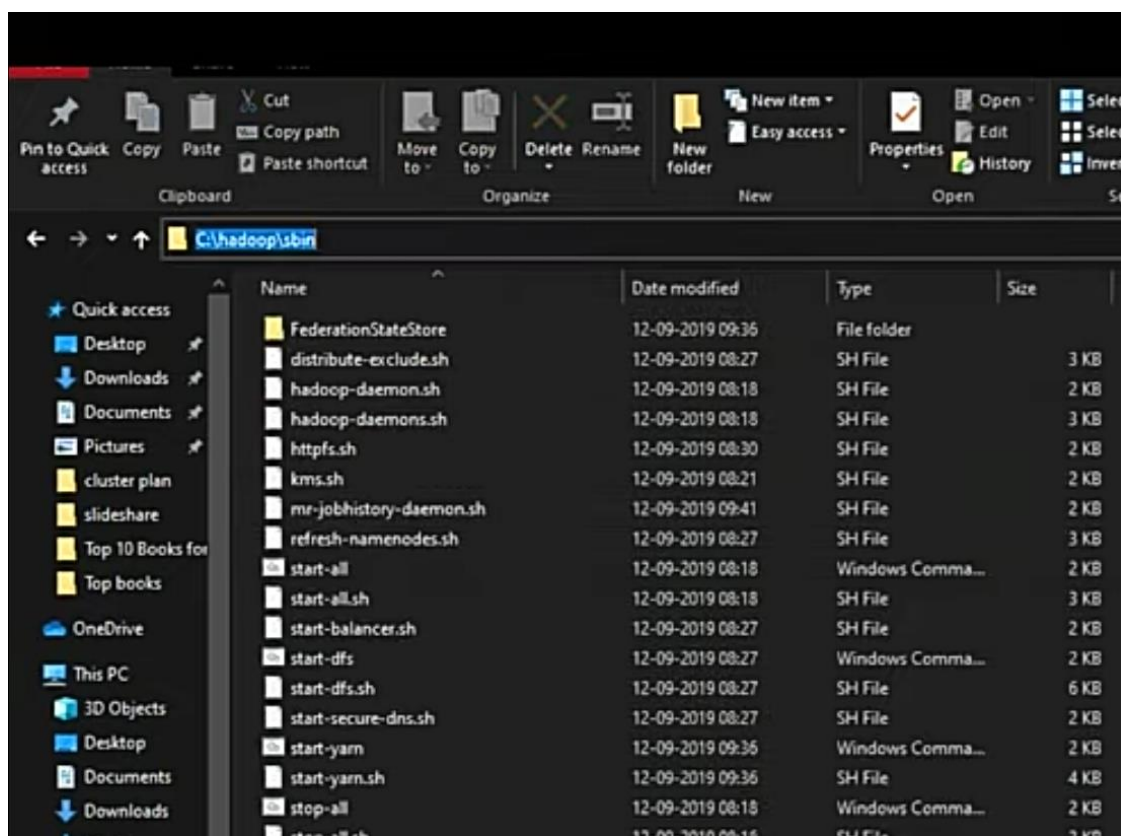
Step 18 :

Now in System Variable section Select **System Variables** in that select **Path** and **Edit**

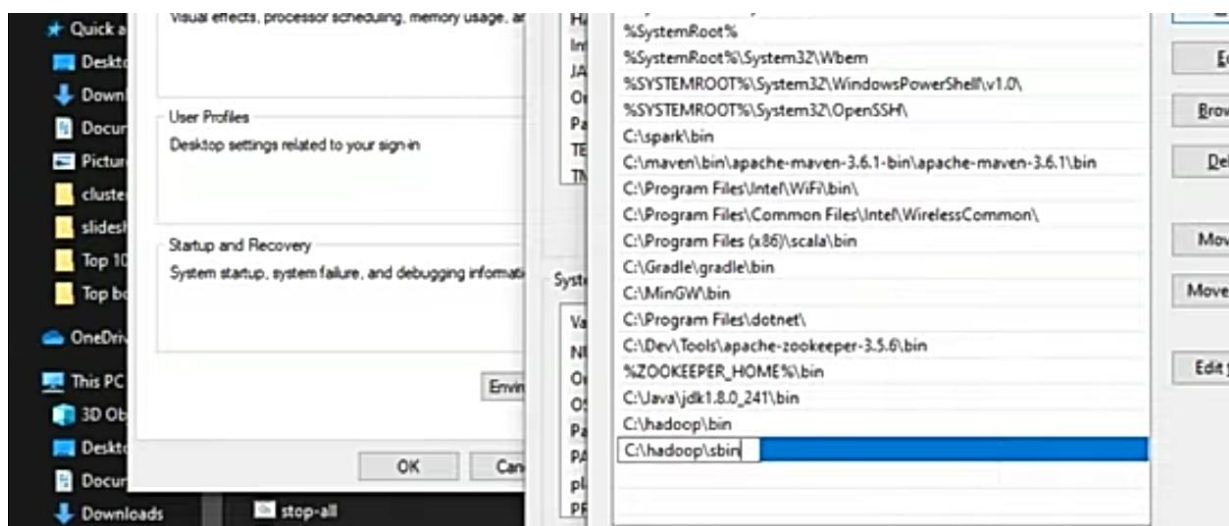
Paste the bin location of hadoop as shown



Click and Paste the sbn location also as shown



Click on New and Paste the sbin location also as shown



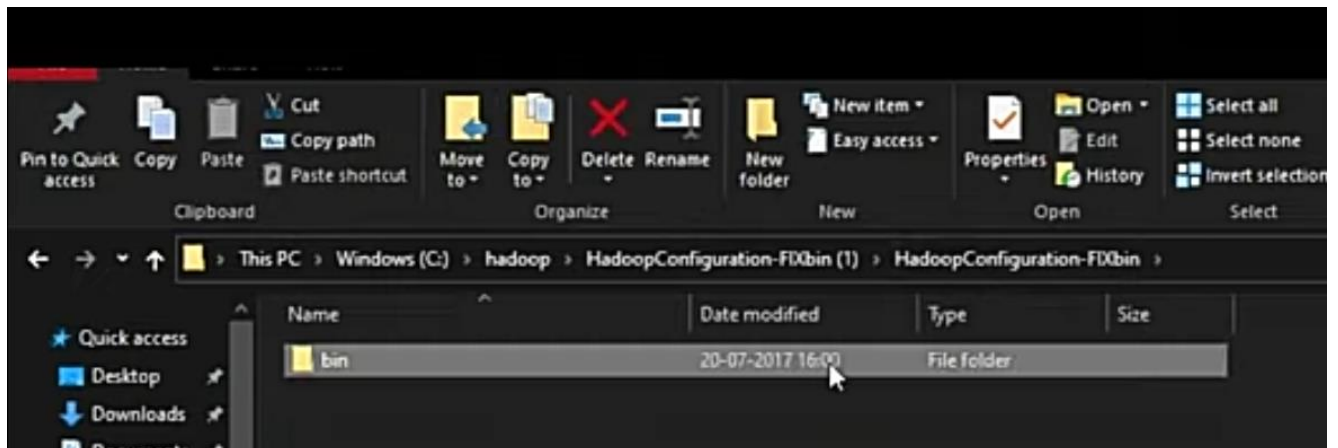
Step 19 :

Download the configuration file from below link

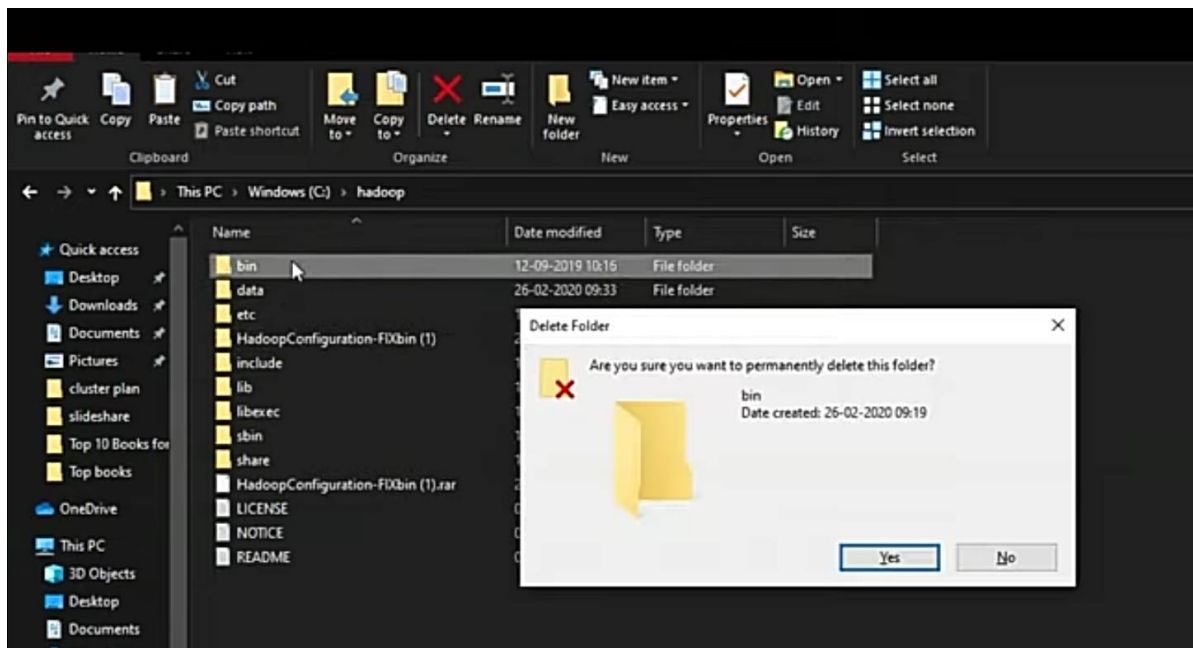
<https://drive.google.com/file/d/1AMqV4F5ybPF4ab4CeK8B3AsjdGtQCdvy/view>

Extract this folder of Configuration

In this folder there will be a (new) **bin** folder as shown

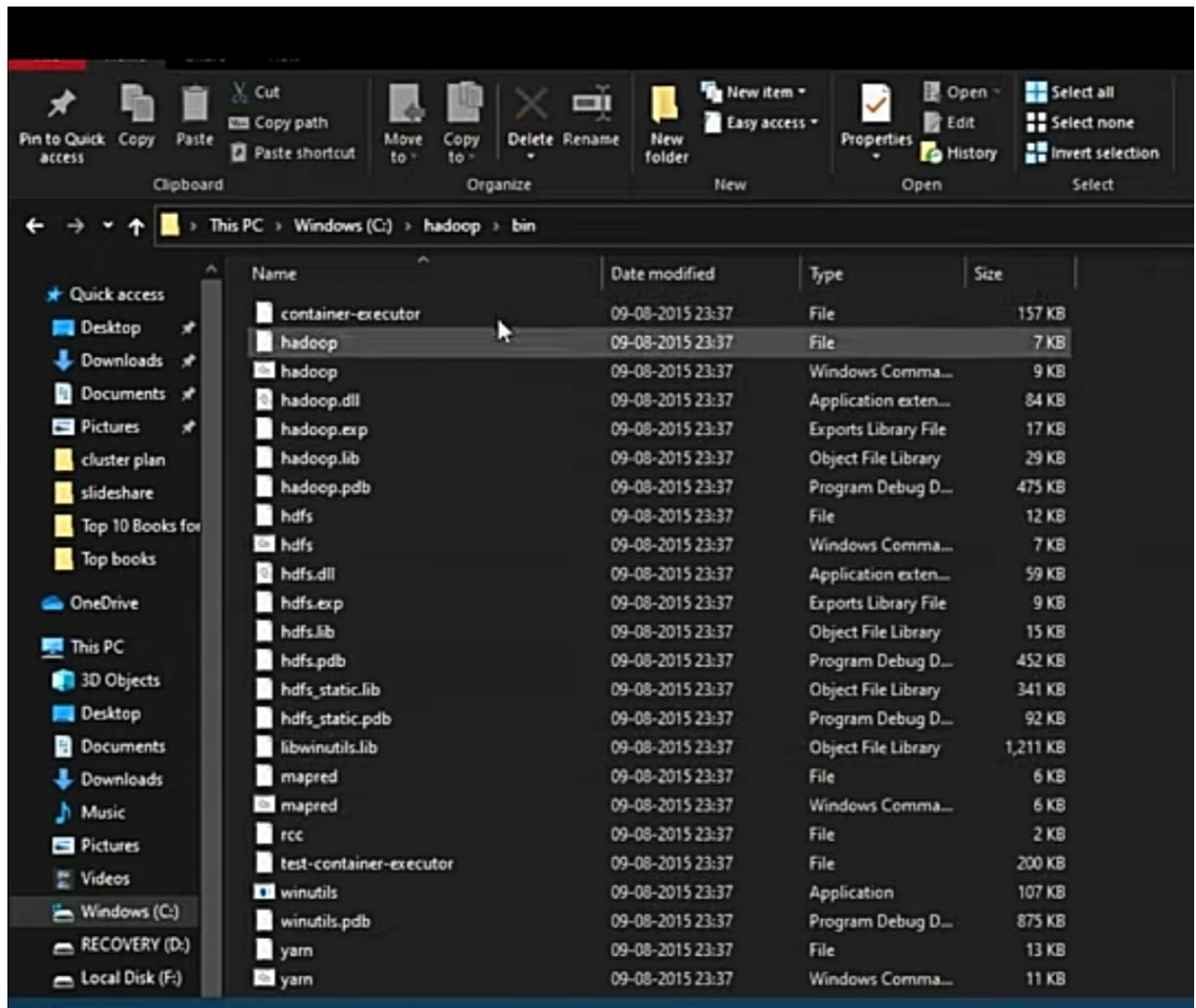
**Step 20 :**

Delete the previously present **bin** folder



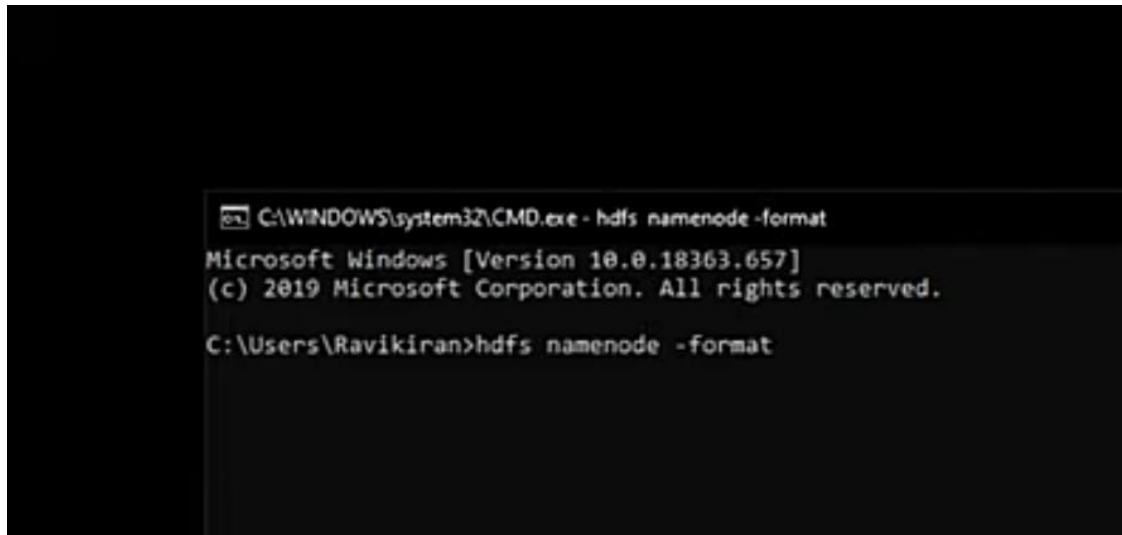
Simply paste the (new) bin folder which is CONFIGURATION ZIP WHICH WE HAVE DOWNLOADED:

You will see below files in your new bin folder



Step 21:

To check if Hadoop is correctly installed type in Command Prompt type **hdfs namenode -format**

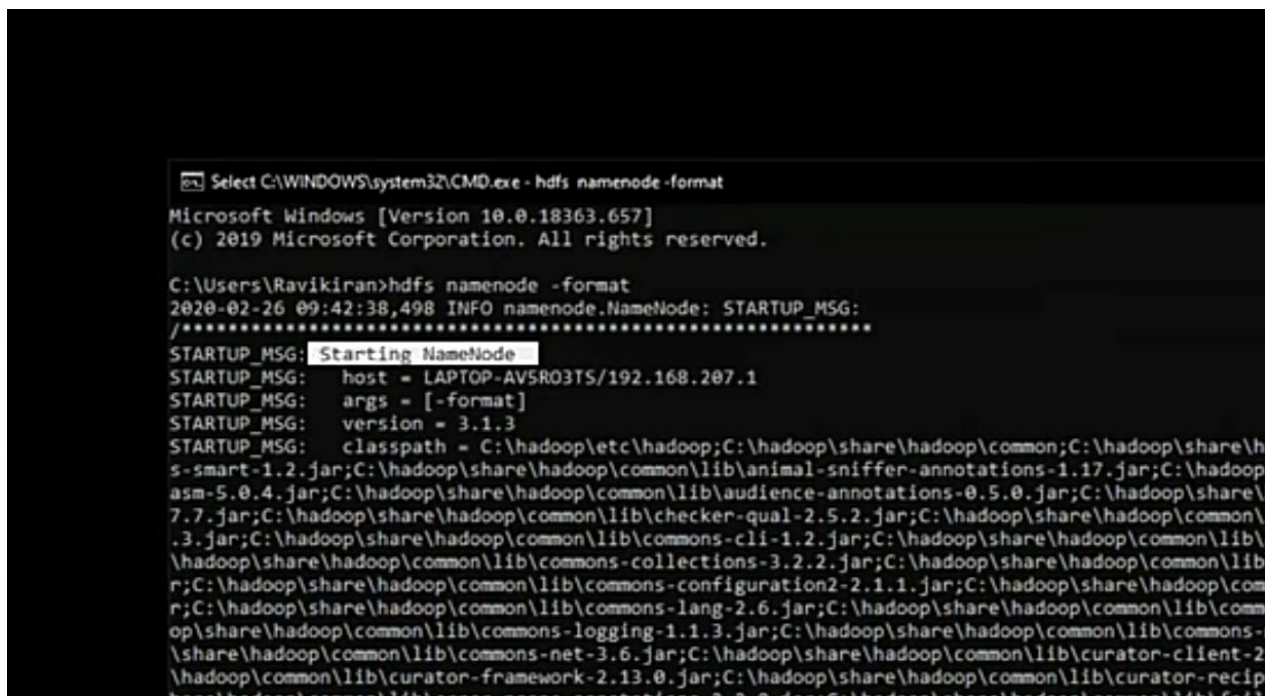


```

C:\WINDOWS\system32\CMD.exe - hdfs namenode -format
Microsoft Windows [Version 10.0.18363.657]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Ravikiran>hdfs namenode -format
  
```

IF SET OF FILES POP UP AS SHOWN THEN HADOOP IS CORRECTLY INSTALLED...



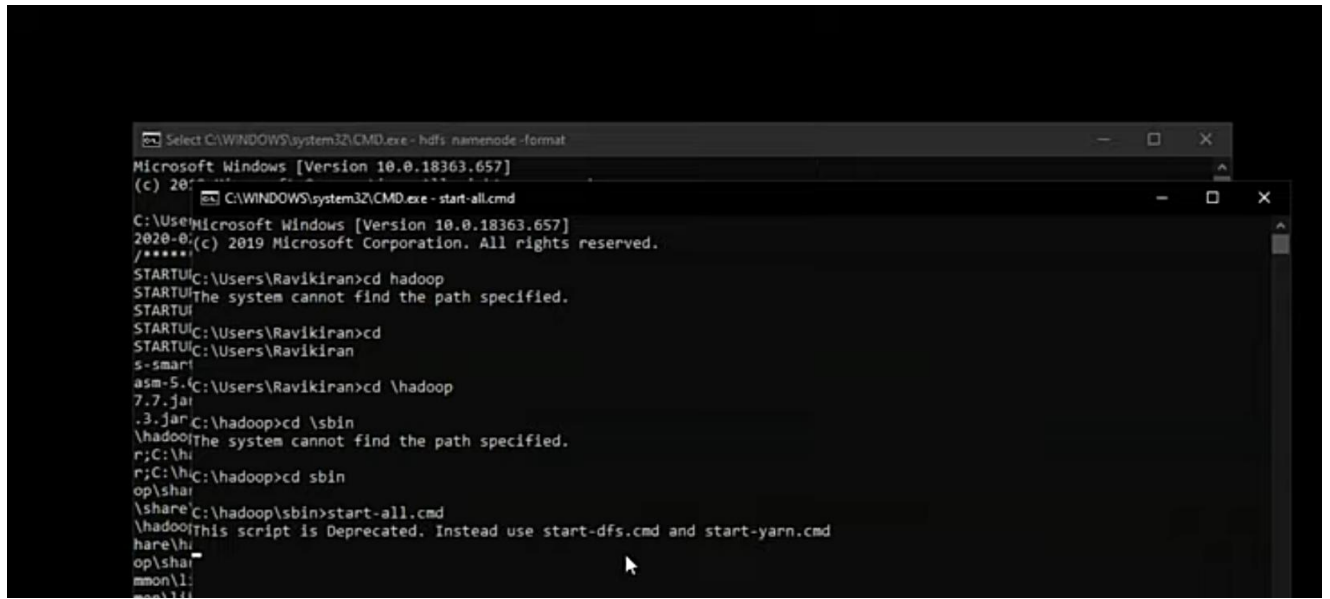
```

Select C:\WINDOWS\system32\CMD.exe - hdfs namenode -format
Microsoft Windows [Version 10.0.18363.657]
(c) 2019 Microsoft Corporation. All rights reserved.

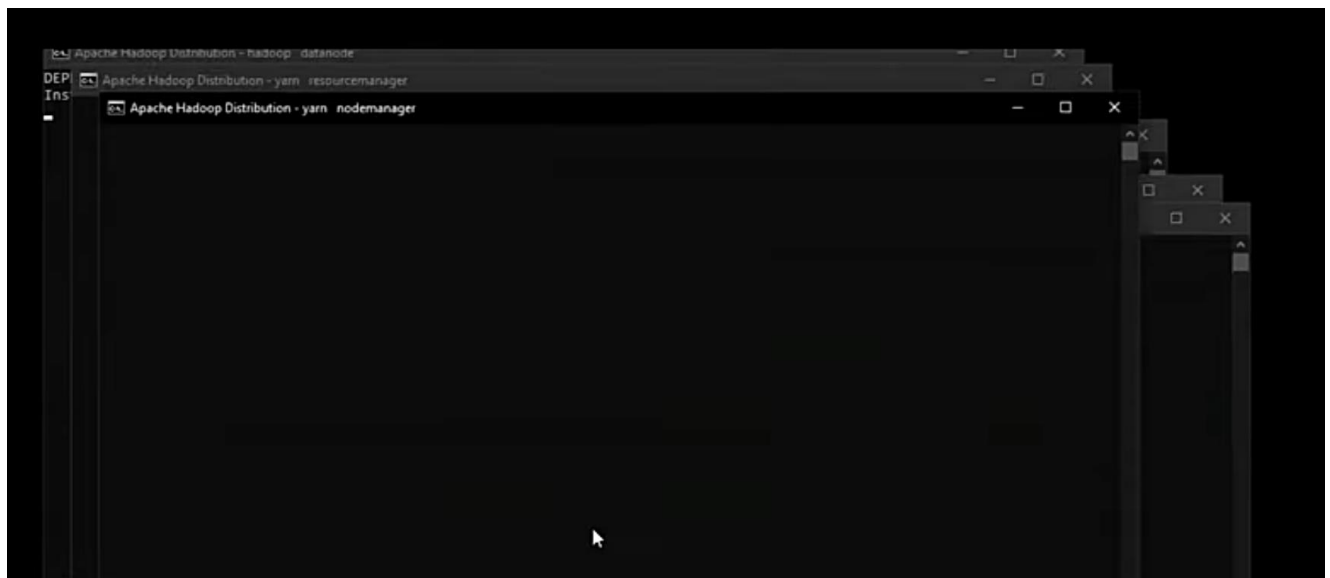
C:\Users\Ravikiran>hdfs namenode -format
2020-02-26 09:42:38,498 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = LAPTOP-AV5R03TS/192.168.207.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.1.3
STARTUP_MSG:   classpath = C:\hadoop\etc\hadoop;C:\hadoop\share\hadoop\common;C:\hadoop\share\h
s-smart-1.2.jar;C:\hadoop\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;C:\hadoop
asm-5.0.4.jar;C:\hadoop\share\hadoop\common\lib\audience-annotations-0.5.0.jar;C:\hadoop\share\
7.7.jar;C:\hadoop\share\hadoop\common\lib\checker-qual-2.5.2.jar;C:\hadoop\share\hadoop\common\
.3.jar;C:\hadoop\share\hadoop\common\lib\commons-cli-1.2.jar;C:\hadoop\share\hadoop\common\lib\
\hadoop\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\hadoop\share\hadoop\common\lib
r;C:\hadoop\share\hadoop\common\lib\commons-configuration2-2.1.1.jar;C:\hadoop\share\hadoop\com
r;C:\hadoop\share\hadoop\common\lib\commons-lang-2.6.jar;C:\hadoop\share\hadoop\common\lib\comm
op\share\hadoop\common\lib\commons-logging-1.1.3.jar;C:\hadoop\share\hadoop\common\lib\commons-
\share\hadoop\common\lib\commons-net-3.6.jar;C:\hadoop\share\hadoop\common\lib\curator-client-2
\hadoop\common\lib\curator-framework-2.13.0.jar;C:\hadoop\share\hadoop\common\lib\curator-recip
hadoop\common\lib\commons-logging-1.1.3.jar;C:\hadoop\share\hadoop\common\lib\curator-recip
  
```


Step 21:

Open a new Command prompt go to location **C:\HADOOP\SBIN** as shown and type **start-all.cmd**



All your daemon files are started.



That's how you install Hadoop.....

FOR ANY QUERIES REFER THE BELOW YOU TUBE VIDEO

References:

<https://www.youtube.com/watch?v=g7Qpnmi0Q-s>

-Documented By Tushar Kalyankar

CSE, ADCET, Ashta