Practical 3

Aim: Installation of Hadoop and java in windows.

→

Installation on java jdk and jre

Step 1: Download SE Development Kit from the given link
https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html

Step 2: Download Java SE Runtime Environment 8u202 from the given link
https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html
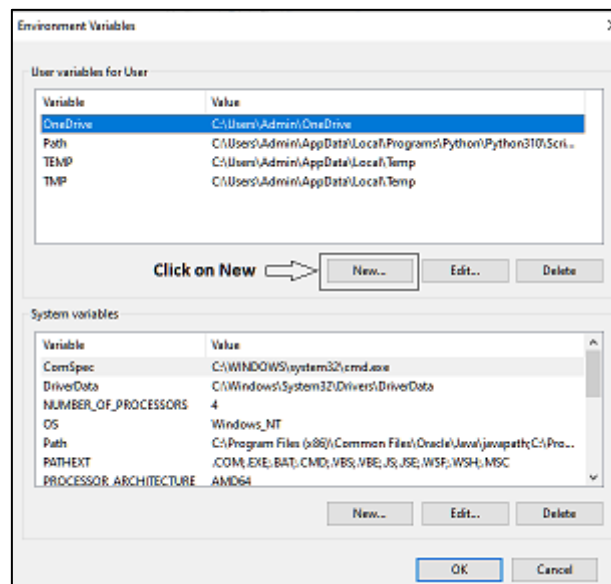

Setting up the JDK in Environmental Variable

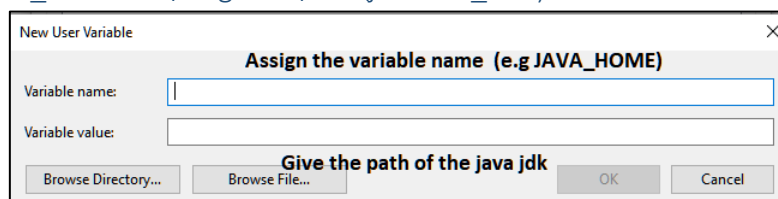Step 1: Start → Edit the system environment variables
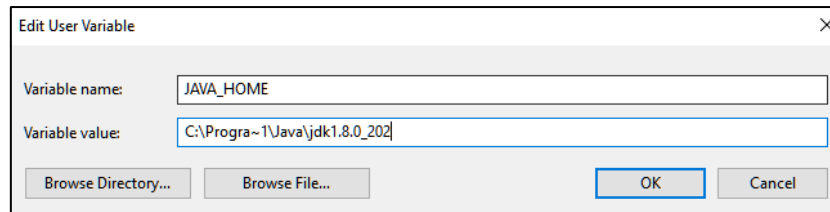
Step 2: Click on the environment Variables
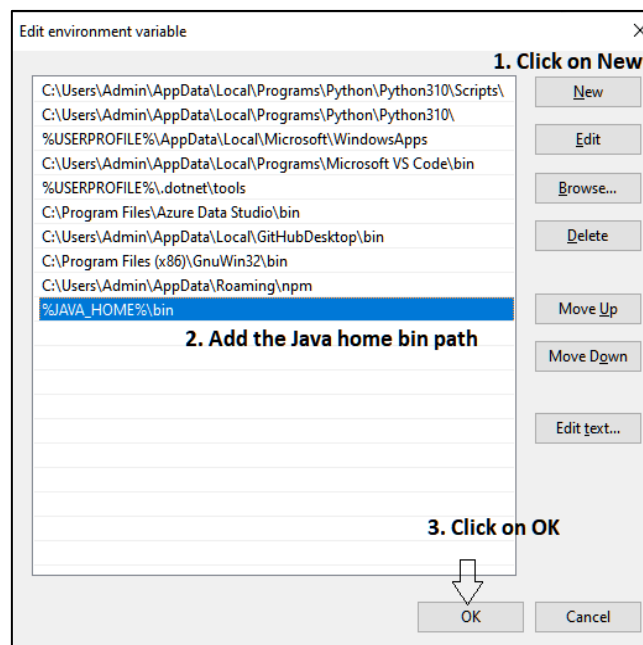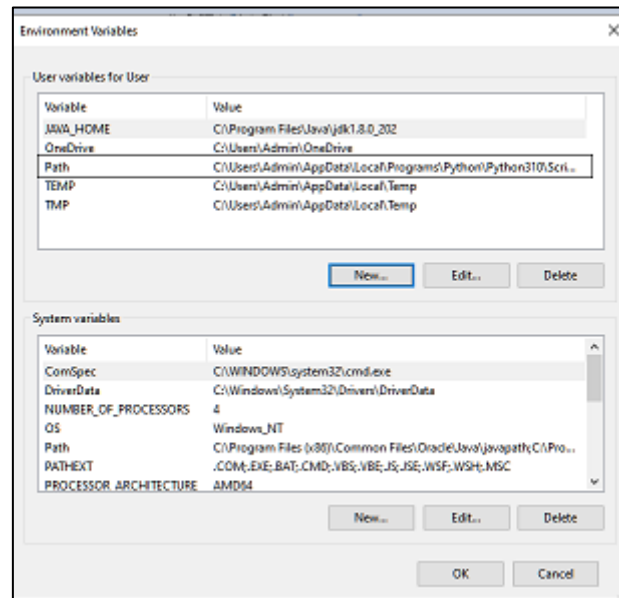

Step 3: Create the new user variables

Step 4: Give the Variable Name and java jdk path (NOTE: Rename the path C:\Program files\Java\jdk1.8.0_202 to C:\Progra~1\Java\jdk1.8.0_202)
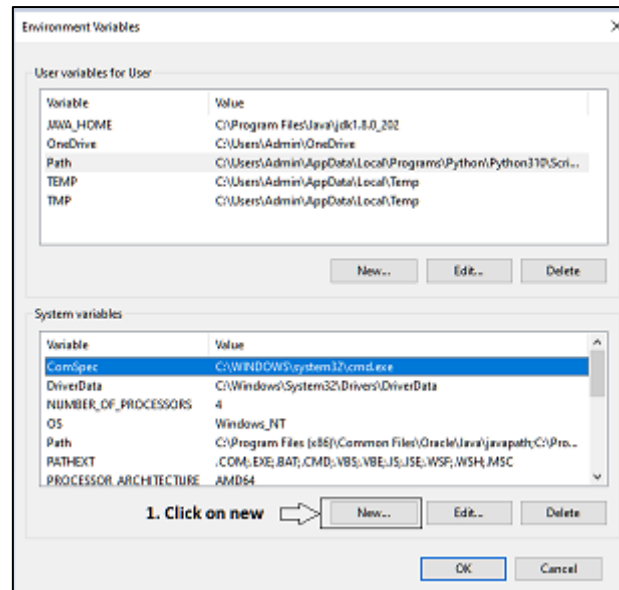
Edit User Variable   ✕

Variable name: JAVA_HOME

Variable value: C:\Progra~1\Java\jdk1.8.0_202

Browse Directory...   Browse File...   OK   Cancel

Step 5: Add the Variable into the Path

Environment Variables   ✕

User variables for User

Variable	Value
JAVA_HOME	C:\Program Files\Java\jdk1.8.0_202
OneDrive	C:\Users\Admin\OneDrive
Path	C:\Users\Admin\AppData\Local\Programs\Python\Python310\Scri...
TEMP	C:\Users\Admin\AppData\Local\Temp
TMP	C:\Users\Admin\AppData\Local\Temp

New...   Edit...   Delete

System variables

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\Pro...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64

New...   Edit...   Delete

OK   Cancel

Edit environment variable   ✕

1. Click on New

C:\Users\Admin\AppData\Local\Programs\Python\Python310\Scripts\   New

C:\Users\Admin\AppData\Local\Programs\Python\Python310\

%USERPROFILE%\AppData\Local\Microsoft\WindowsApps   Edit

C:\Users\Admin\AppData\Local\Programs\Microsoft VS Code\bin

%USERPROFILE%\.dotnet\tools   Browse...

C:\Program Files\Azure Data Studio\bin

C:\Users\Admin\AppData\Local\GitHubDesktop\bin   Delete

C:\Program Files (x86)\GnuWin32\bin

C:\Users\Admin\AppData\Roaming\npm

%JAVA_HOME%\bin   Move Up

2. Add the Java home bin path

Move Down

Edit text...

3. Click on OK

⬇

OK   Cancel
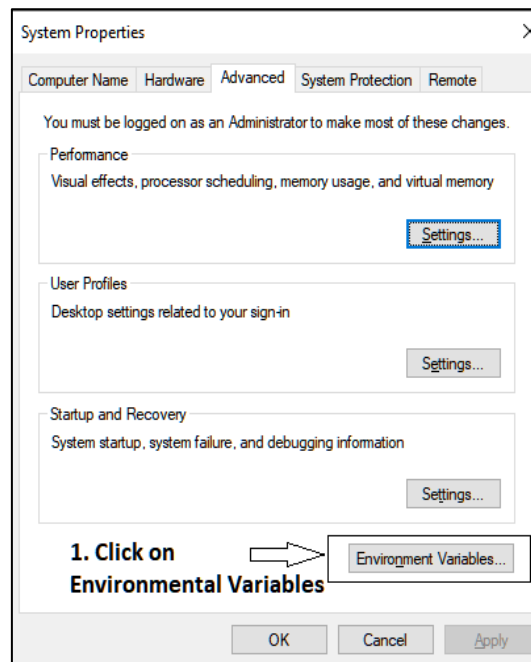
Step 6: Create the new system variables

Step 7: Give the Variable Name and java jdk path (NOTE: Rename the path C:\Program files\Java\jdk1.8.0_202 to C:\Progra~1\Java\jdk1.8.0_202)

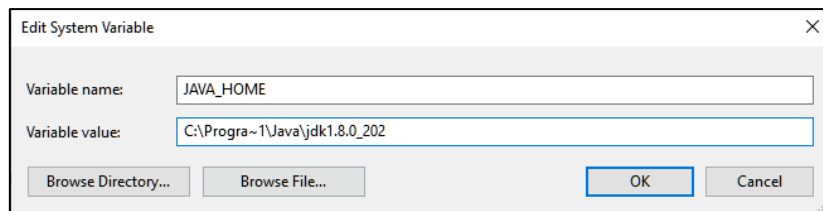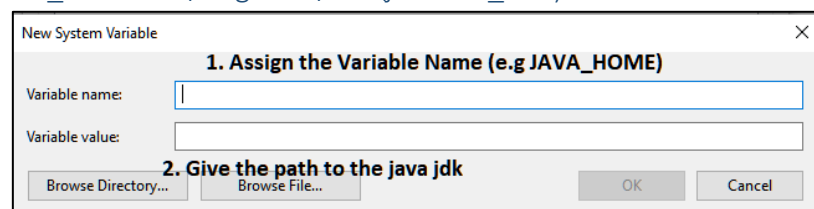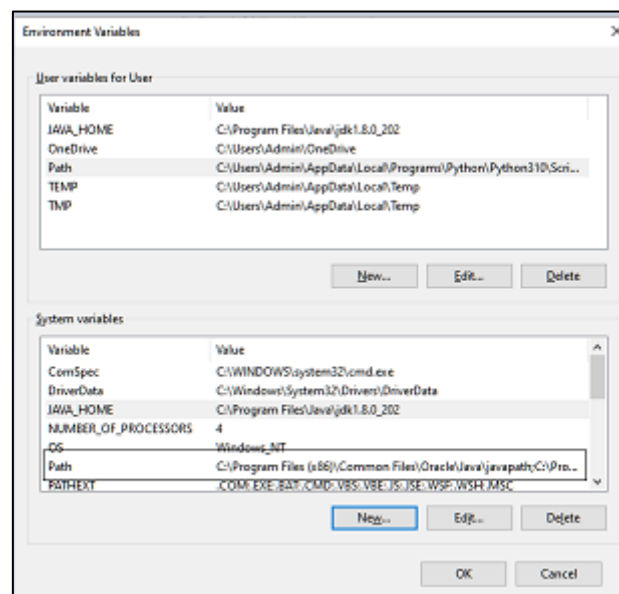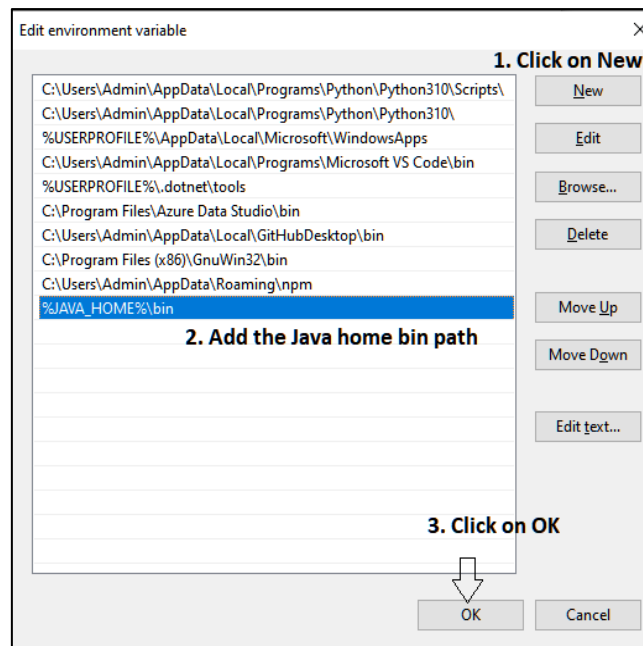Step 8: Add the Variable into the Path

Installation of Hadoop

Step 1: Download Apache Hadoop from the given link and click on the binary
https://hadoop.apache.org/releases.html

Step 2: Click the link to download the zip
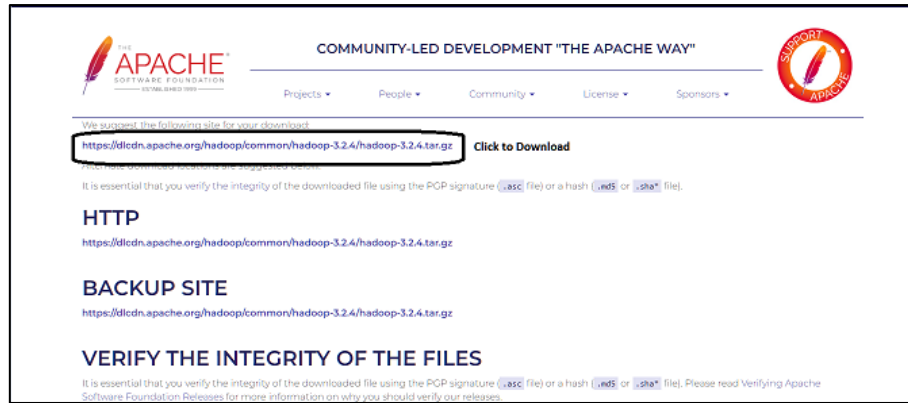
Installation of the 7zip

Step 1: Download the 7zip and install into the system https://www.7-zip.org/download.html
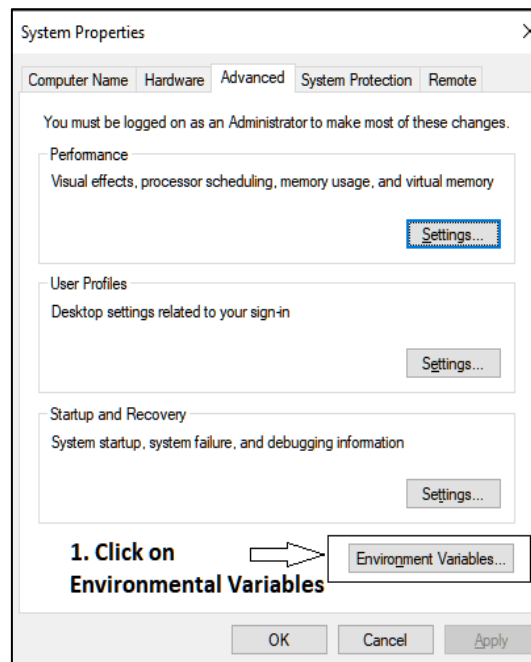
Extracting Hadoop zip using 7Zip

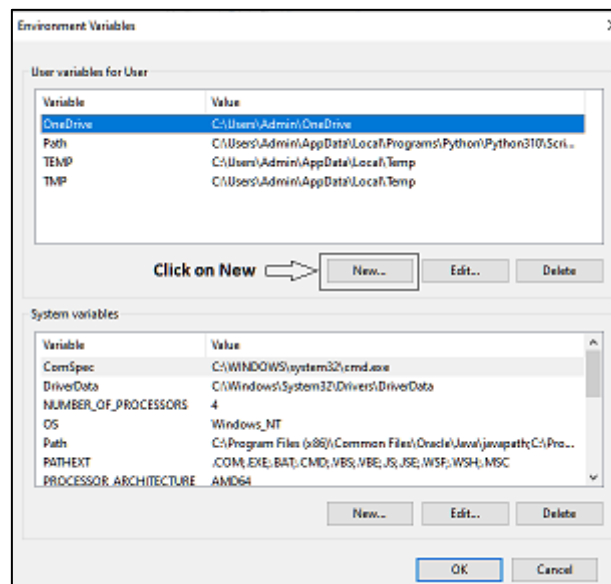Step 1 : Right Click on the zip → Click on 7-zip → Extract Here
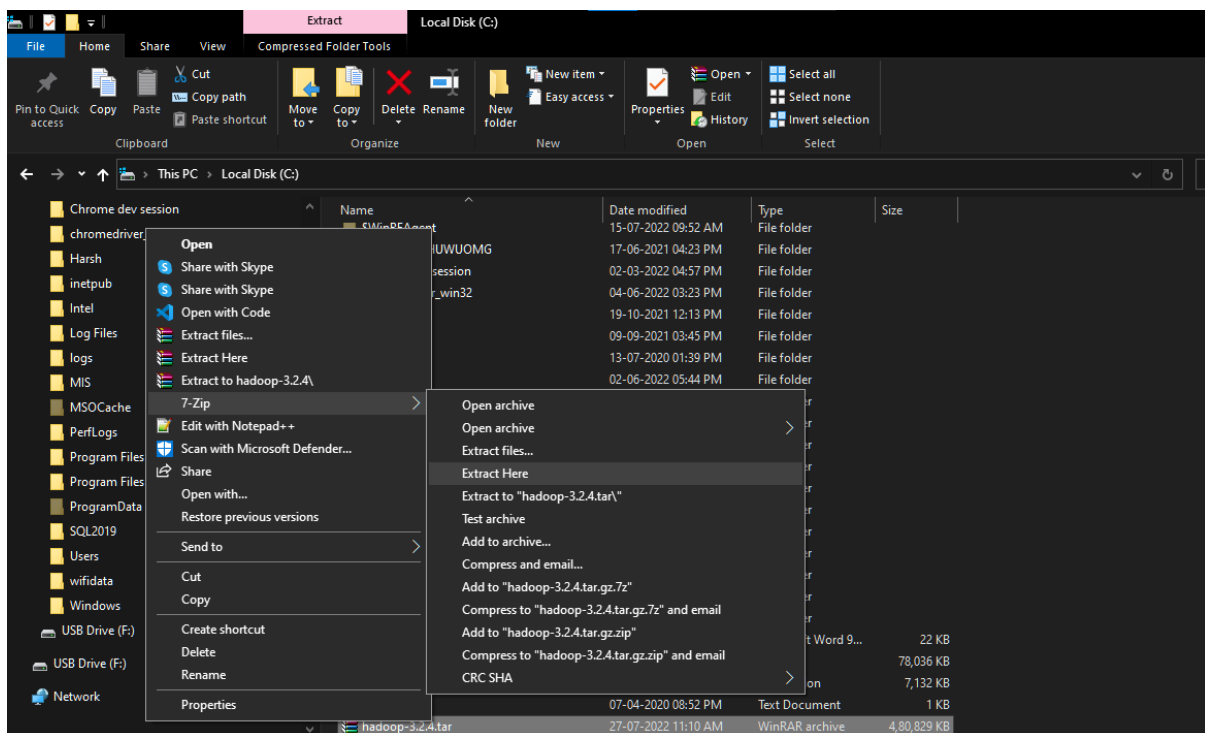
Setting up the Hadoop in Environmental Variable

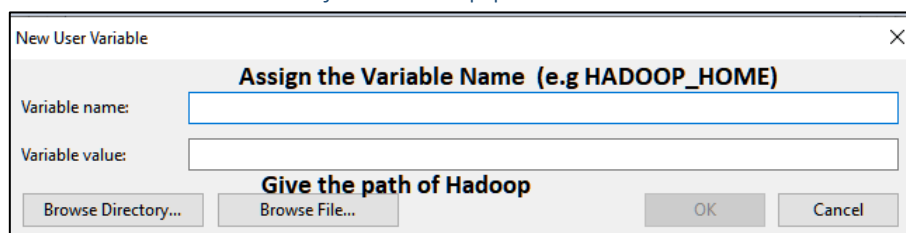Step 1: Start → Edit the system environment variables

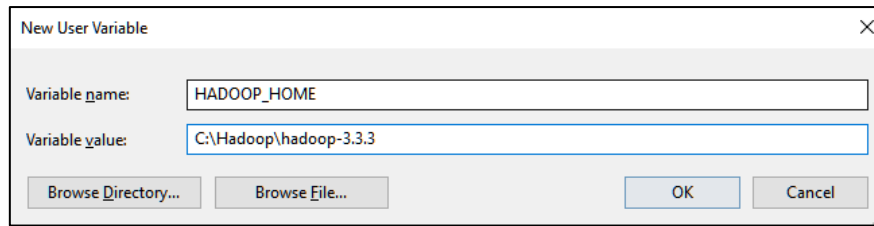Step 2: Click on the environment Variables
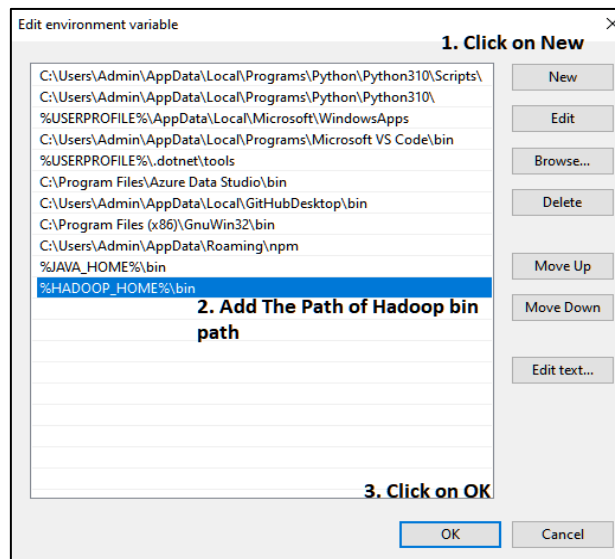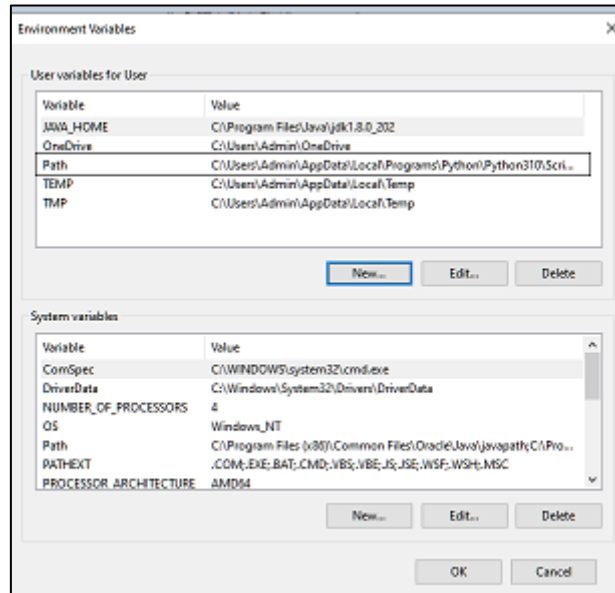

Step 3: Create the new user variables

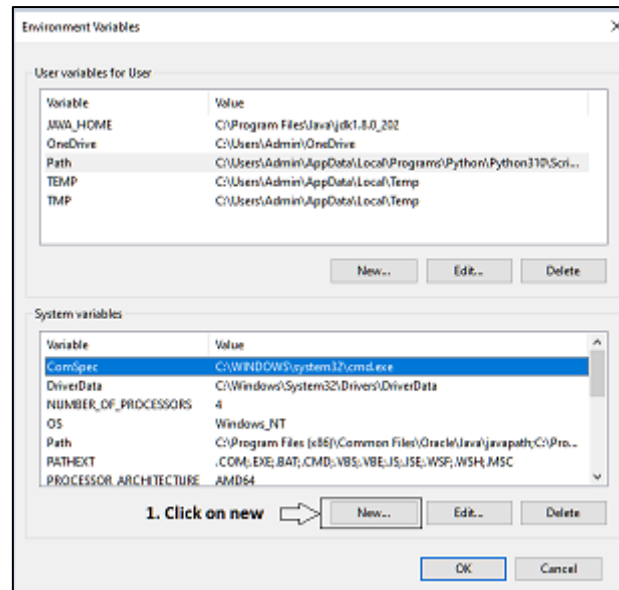Step 4: Give the Variable Name and java Hadoop path

New User Variable   ✕

Variable name: HADOOP_HOME

Variable value: C:\Hadoop\hadoop-3.3.3

[Browse Directory...]   [Browse File...]   [OK]   [Cancel]

Step 5: Add the Variable into the Path

Environment Variables   ✕

User variables for User

Variable	Value
JAVA_HOME	C:\Program Files\Java\jdk1.8.0_202
OneDrive	C:\Users\Admin\OneDrive
Path	C:\Users\Admin\AppData\Local\Programs\Python\Python310\Scri...
TEMP	C:\Users\Admin\AppData\Local\Temp
TMP	C:\Users\Admin\AppData\Local\Temp

[New...]   [Edit...]   [Delete]

System variables

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\Pro...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64

[New...]   [Edit...]   [Delete]

[OK]   [Cancel]

Edit environment variable   ✕

                                                                                   1. Click on New

C:\Users\Admin\AppData\Local\Programs\Python\Python310\Scripts\	[New]
C:\Users\Admin\AppData\Local\Programs\Python\Python310\	[Edit]
%USERPROFILE%\AppData\Local\Microsoft\WindowsApps	
C:\Users\Admin\AppData\Local\Programs\Microsoft VS Code\bin	[Browse...]
%USERPROFILE%\.dotnet\tools	
C:\Program Files\Azure Data Studio\bin	[Delete]
C:\Users\Admin\AppData\Local\GitHubDesktop\bin	
C:\Program Files (x86)\GnuWin32\bin	
C:\Users\Admin\AppData\Roaming\npm	
%JAVA_HOME%\bin	[Move Up]
%HADOOP_HOME%\bin	[Move Down]

2. Add The Path of Hadoop bin path

[Edit text...]

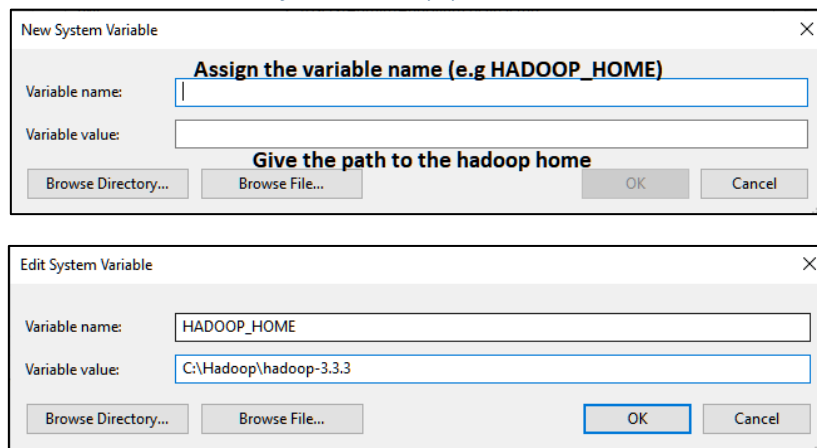**3. Click on OK**

[OK]   [Cancel]
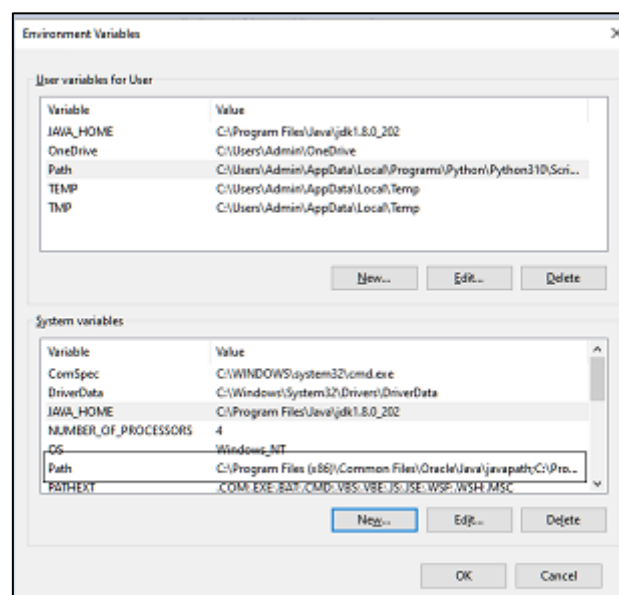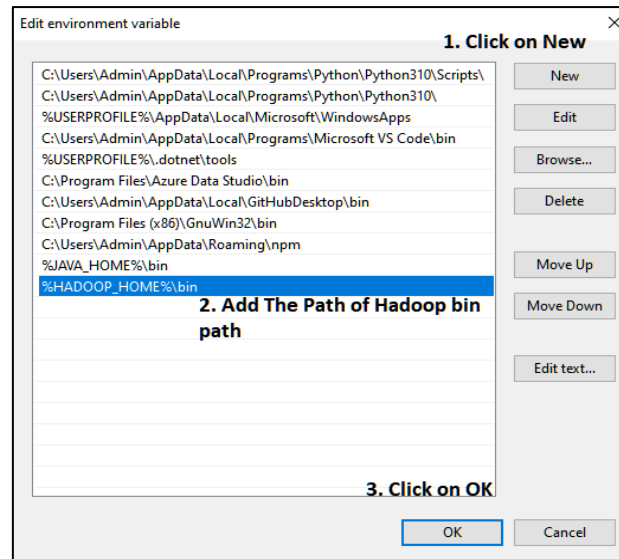
Step 6: Create the new system variables

Step 7: Give the Variable Name and java Hadoop path

Step 8: Add the Variable into the Path

Setting up Hadoop

Step 1: Create the new Folder with the name DATA

Step 2: Go inside the data folder and create 2 folder i.e. NameNode and DataNode

Step 3: open the file **hdfs-site.xml** C:\Hadoop\hadoop-3.3.3\etc\hadoop

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>


<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
<name>dfs.namenode.name.dir</name>
<value>C:\Hadoop\hadoop-3.3.3\DATA\NameNode</value>
</property>

<property>
<name>dfs.datanode.data.dir</name>
<value>C:\Hadoop\hadoop-3.3.3\DATA\DataNode</value>
</property>

<property>
  <name>dfs.permissions</name>
  <value>false</value>
</property>
</configuration>
```

Step 3: open the file **hdfs-site.xml** C:\Hadoop\hadoop-3.3.3\etc\hadoop

```xml
C:\Hadoop\hadoop-3.3.3\etc\hadoop\hdfs-site.xml
hdfs-site.xml
 1    <?xml version="1.0" encoding="UTF-8"?>
 2    <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
 3
 4
 5    <configuration>
 6      <property>
 7        <name>dfs.replication</name>
 8        <value>3</value>
 9      </property>
10      <property>
11    <name>dfs.namenode.name.dir</name>
12    <value>C:\Hadoop\hadoop-3.3.3\DATA\NameNode</value>
13    </property>
14
15    <property>
16    <name>dfs.datanode.data.dir</name>
17    <value>C:\Hadoop\hadoop-3.3.3\DATA\DataNode</value>
18    </property>
19
20    <property>
21      <name>dfs.permissions</name>
22      <value>false</value>
23    </property>
24    </configuration>
```

Step 4: open the file **core-site.xml** C:\Hadoop\hadoop-3.3.3\etc\hadoop

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
```

```xml
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
<!-- <value>hdfs://localhost:9820</value> -->
</property>
</configuration>
```

```
</> hdfs-site.xml        </> core-site.xml  X

</> core-site.xml
    1      <?xml version="1.0" encoding="UTF-8"?>
    2      <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
    3      <!--
    4        Licensed under the Apache License, Version 2.0 (the "License");
    5        you may not use this file except in compliance with the License.
    6        You may obtain a copy of the License at
    7
    8          http://www.apache.org/licenses/LICENSE-2.0
    9
   10        Unless required by applicable law or agreed to in writing, software
   11        distributed under the License is distributed on an "AS IS" BASIS,
   12        WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
   13        See the License for the specific language governing permissions and
   14        limitations under the License. See accompanying LICENSE file.
   15      -->
   16
   17      <!-- Put site-specific property overrides in this file. -->
   18
   19      <configuration>
   20      <property>
   21      <name>fs.default.name</name>
   22        <value>hdfs://localhost:9000</value>
   23      <!-- <value>hdfs://localhost:9820</value> -->
   24      </property>
   25      </configuration>
   26
```

Step 5: open the file **mapred-site.xml** C:\Hadoop\hadoop-3.3.3\etc\hadoop

```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at
```

```
  http://www.apache.org/licenses/LICENSE-2.0

 Unless required by applicable law or agreed to in writing, software
 distributed under the License is distributed on an "AS IS" BASIS,
 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 See the License for the specific language governing permissions and
 limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
<property>
<name>mapreduce.cluster.local.dir</name>
<value>${hadoop.tmp.dir}/mapred/local</value>
</property>

</configuration>
```

```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
<property>
<name>mapreduce.cluster.local.dir</name>
<value>${hadoop.tmp.dir}/mapred/local</value>
</property>

</configuration>
```

Step 6: open the file **yarn-site -site.xml** C:\Hadoop\hadoop-3.3.3\etc\hadoop

```xml
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>

<!-- Site specific YARN configuration properties -->
```

```xml
<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>
```

```
</> hdfs-site.xml      </> core-site.xml      </> mapred-site.xml      </> yarn-site.xml  X

</> yarn-site.xml
    1    <?xml version="1.0"?>
    2    <!--
    3      Licensed under the Apache License, Version 2.0 (the "License");
    4      you may not use this file except in compliance with the License.
    5      You may obtain a copy of the License at
    6
    7        http://www.apache.org/licenses/LICENSE-2.0
    8
    9      Unless required by applicable law or agreed to in writing, software
   10      distributed under the License is distributed on an "AS IS" BASIS,
   11      WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
   12      See the License for the specific language governing permissions and
   13      limitations under the License. See accompanying LICENSE file.
   14    -->
   15    <configuration>
   16
   17    <!-- Site specific YARN configuration properties -->
   18    <configuration>
   19    <property>
   20    <name>yarn.nodemanager.aux-services</name>
   21    <value>mapreduce_shuffle</value>
   22    </property>
   23    <property>
   24    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
   25    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
   26    </property>
   27
   28    </configuration>
   29
```

Step 7: Go to GitHub https://github.com/Selfgrowth/Apache-hadoop-3.1.1-winutils and download the bin folder and replace all the files with the C:\Hadoop\hadoop-3.3.3\bin

Step 8: Open the command prompt and change the location to C:\Hadoop\hadoop-3.3.3\bin

Step 9: hdfs namenode -format

Step 10: Go to sbin folder. Type cmd in the address bar and 2 different open command prompt

First cmd: **.\start-dfs.cmd**

Second cmd: **.\start-yarn.cmd**

Step 11: now open chrome or any other browser and type http://localhost:9870/