

## 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>

Operating System	Size	Download Link
Solaris SPARC 64-bit (SVR4 package)	125.09 MB	jdk-8u202-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	88.1 MB	jdk-8u202-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	124.37 MB	jdk-8u202-solaris-x64.tar.Z
Solaris x64	85.38 MB	jdk-8u202-solaris-x64.tar.gz
Windows x86	201.64 MB	jdk-8u202-windows-i586.exe
<b>Windows x64</b>	<b>211.58 MB</b>	<b>jdk-8u202-windows-x64.exe</b>

Java SE Runtime Environment 8u202  
This software is licensed under the Oracle Binary Code License Agreement for Java SE Platform Products

Click on Download

Step 2: Download Java SE Runtime Environment 8u202 from the given link

<https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html>

Operating System	Size	Download Link
Mac OS X x64	69.37 MB	jre-8u202-macosx-x64.tar.gz
Solaris SPARC 64-bit	46.07 MB	jre-8u202-solaris-sparcv9.tar.gz
Solaris x64	43.36 MB	jre-8u202-solaris-x64.tar.gz
Windows x86 Online	1.83 MB	jre-8u202-windows-i586-iftw.exe
Windows x86 Offline	65.73 MB	jre-8u202-windows-i586.exe
Windows x86	68.4 MB	jre-8u202-windows-i586.tar.gz
<b>Windows x64</b>	<b>73.7 MB</b>	<b>jre-8u202-windows-x64.exe</b>
Windows x64	73.25 MB	jre-8u202-windows-x64.tar.gz

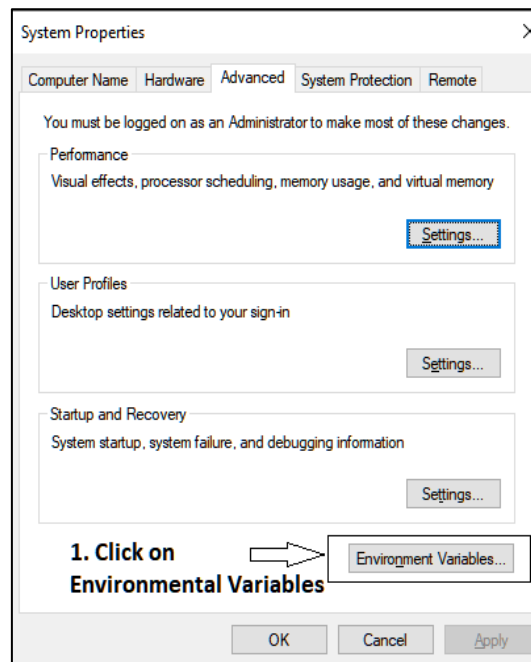
Server JRE (Java SE Runtime Environment) 8u202

Click on Download

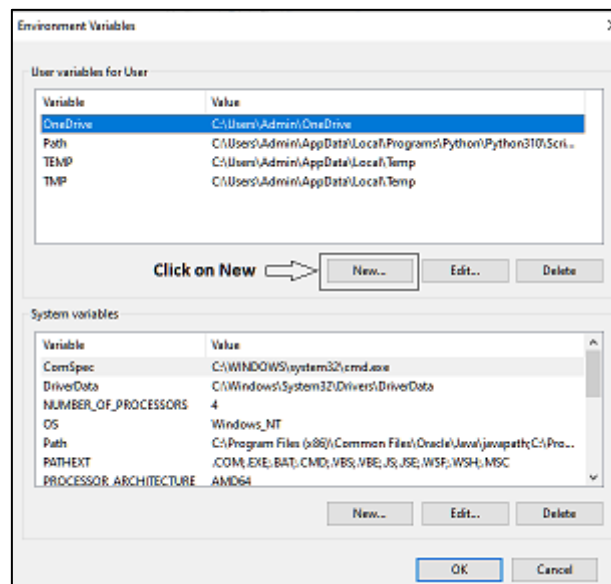
Setting up the JDK in Environmental Variable

Step 1: Start → Edit the system environment variables

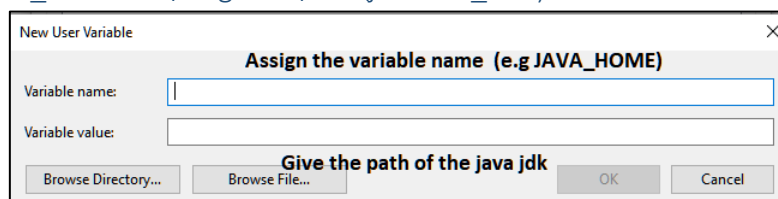
Step 2: Click on the environment Variables

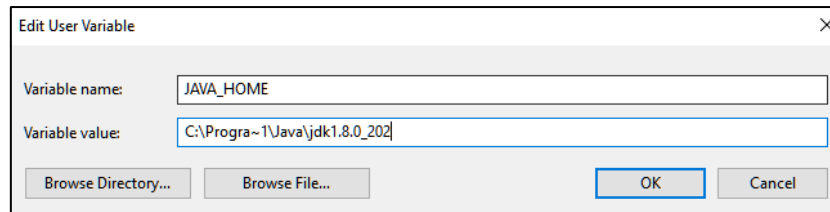


Step 3: Create the new user variables

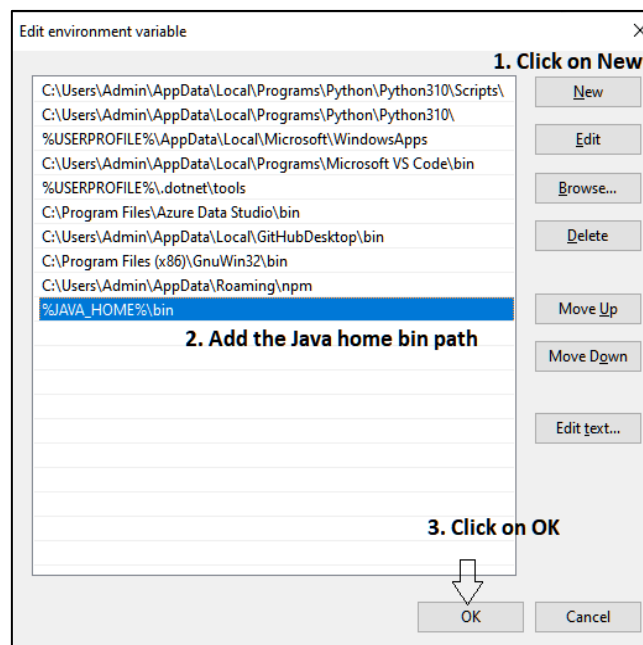
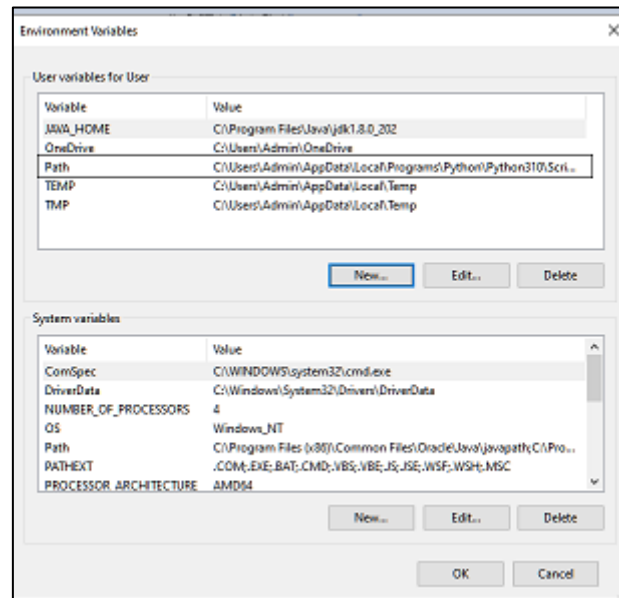


Step 4: Give the Variable Name and java jdk path (NOTE: Rename the path C:\Program files\Java\jdk1.8.0\_202 to C:\Progra~1\Java\jdk1.8.0\_202)

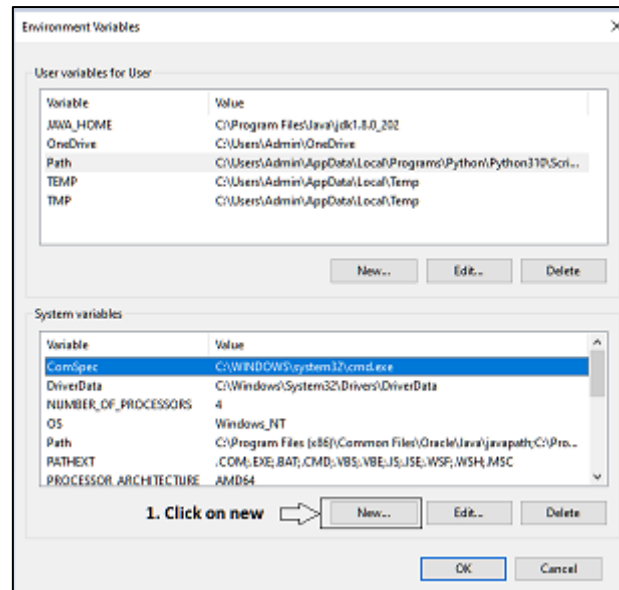




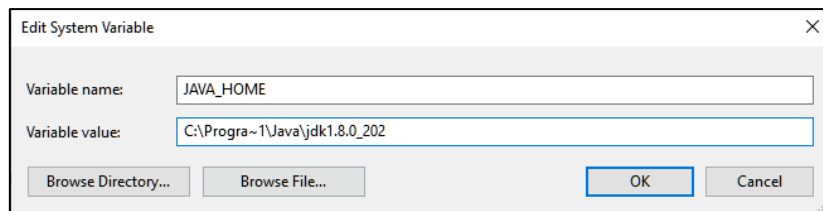
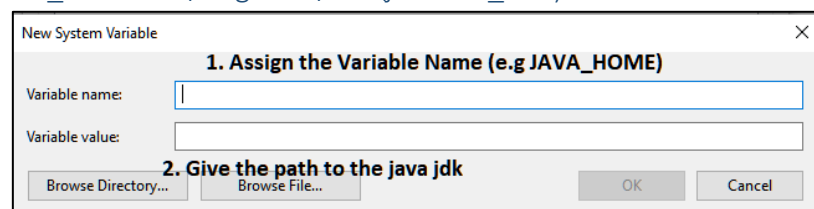
Step 5: Add the Variable into the Path



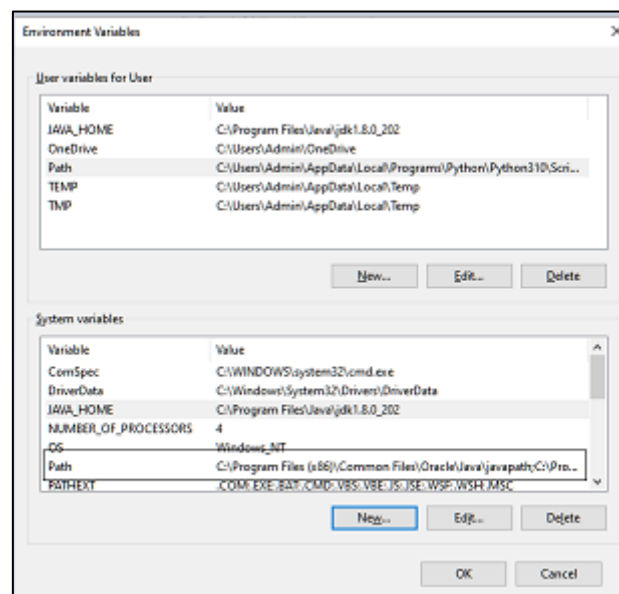
Step 6: Create the new system variables

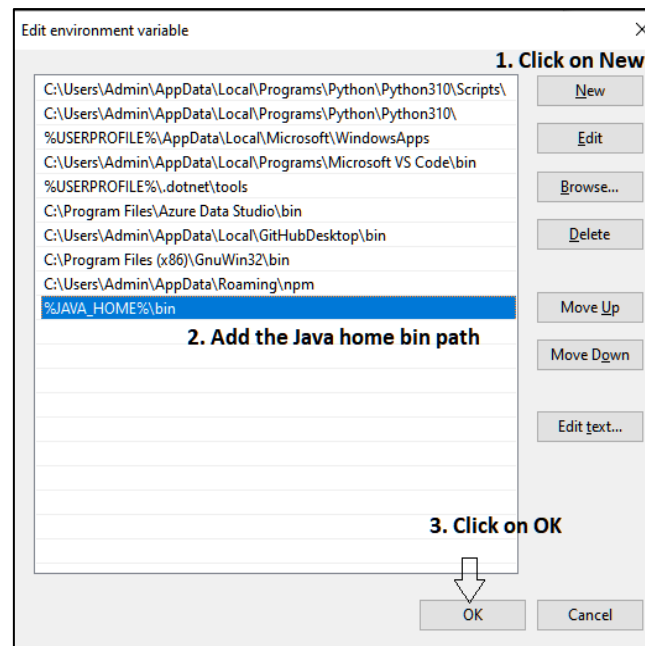


Step 7: Give the Variable Name and java jdk path (NOTE: Rename the path C:\Program files\Java\jdk1.8.0\_202 to C:\Progra~1\Java\jdk1.8.0\_202)



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>

**Download**

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
3.2.4	2022 Jul 22	source (checksum signature)	binary (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement
3.3.3	2022 May 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement

**Click to Download**

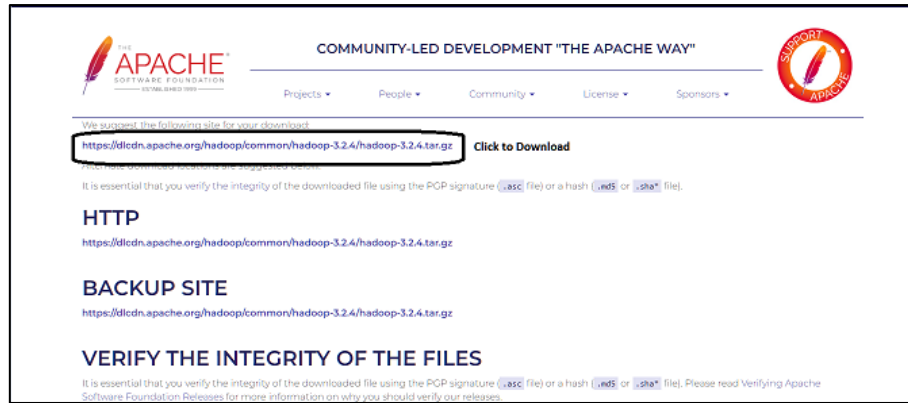
**To verify Hadoop releases using GPG:**

1. Download the release hadoop-X.Y.Z-src.tar.gz from a mirror site.
2. Download the signature file hadoop-X.Y.Z-src.tar.gz.asc from Apache.
3. Download the Hadoop KEYS file.
4. gpg -import KEYS
5. gpg --verify hadoop-X.Y.Z-src.tar.gz.asc

**To perform a quick check using SHA-512:**

1. Download the release hadoop-X.Y.Z-src.tar.gz from a mirror site.
2. Download the checksum hadoop-X.Y.Z-src.tar.gz.sha512 or hadoop-X.Y.Z-src.tar.gz.mds from Apache.
3. shasum -a 512 hadoop-X.Y.Z-src.tar.gz

Step 2: Click the link to download the zip

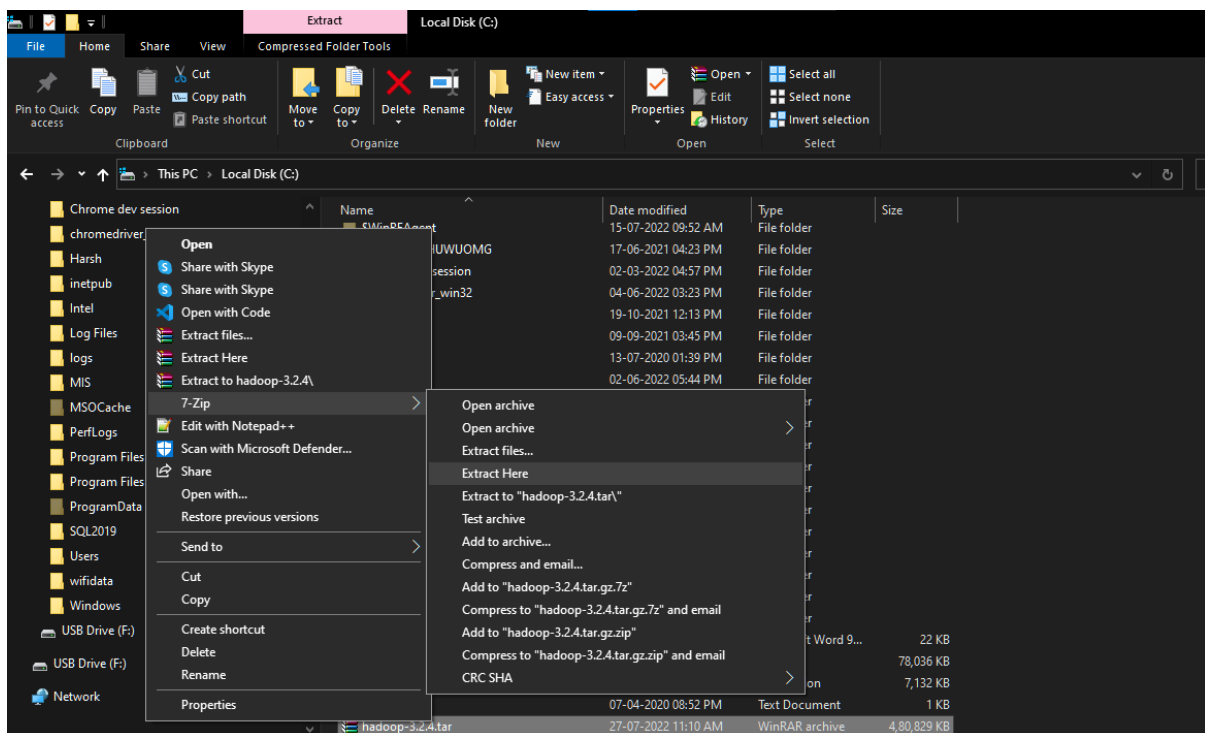


## Installation of the 7zip

Step 1: Download the 7zip and install into the system <https://www.7-zip.org/download.html>

## Extracting Hadoop zip using 7Zip

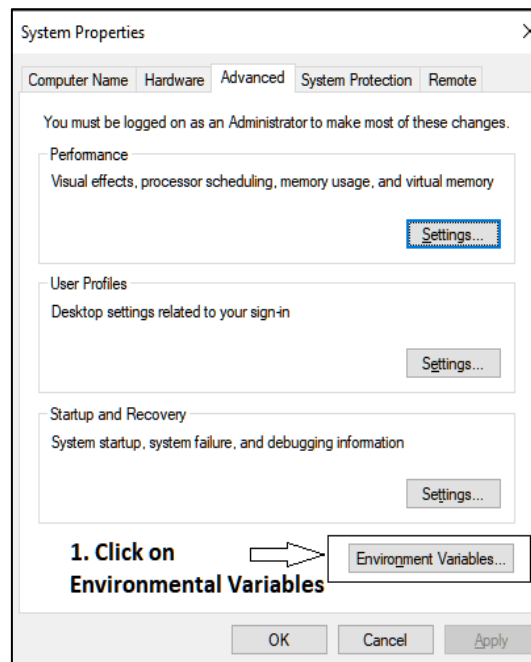
Step 1 : Right Click on the zip → Click on 7-zip → Extract Here



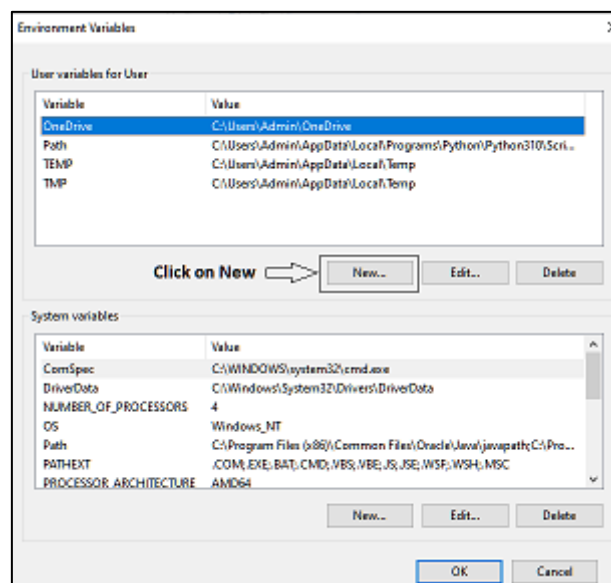
## Setting up the Hadoop in Environmental Variable

Step 1: Start → Edit the system environment variables

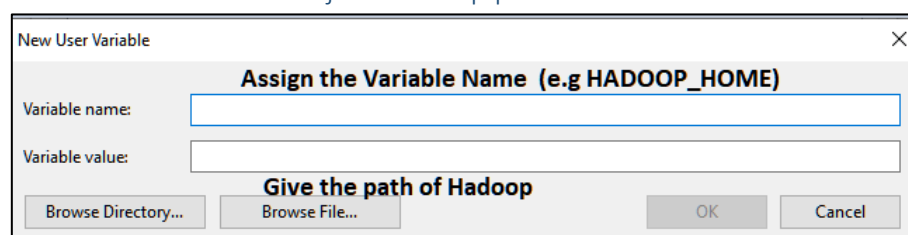
Step 2: Click on the environment Variables

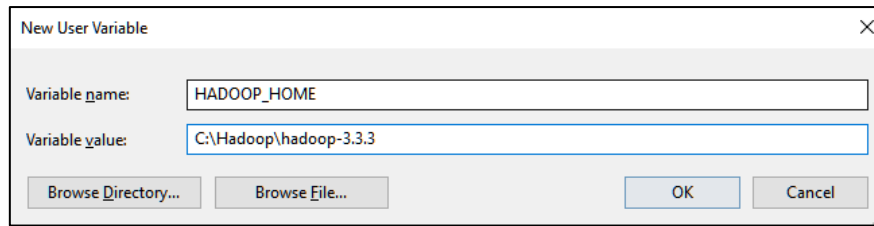


Step 3: Create the new user variables

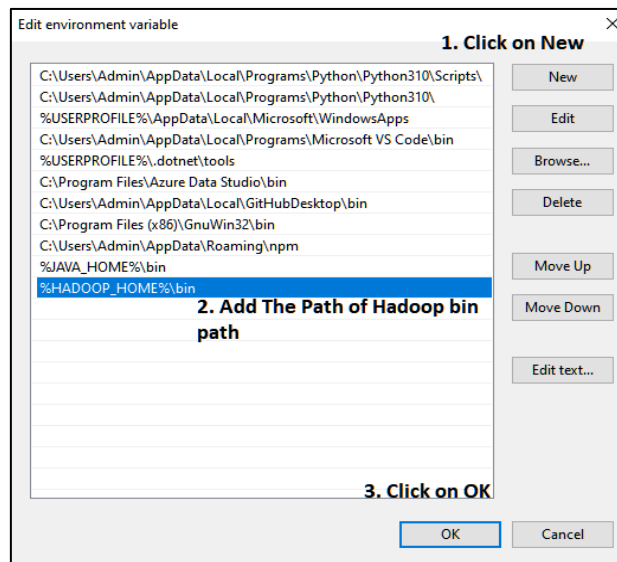
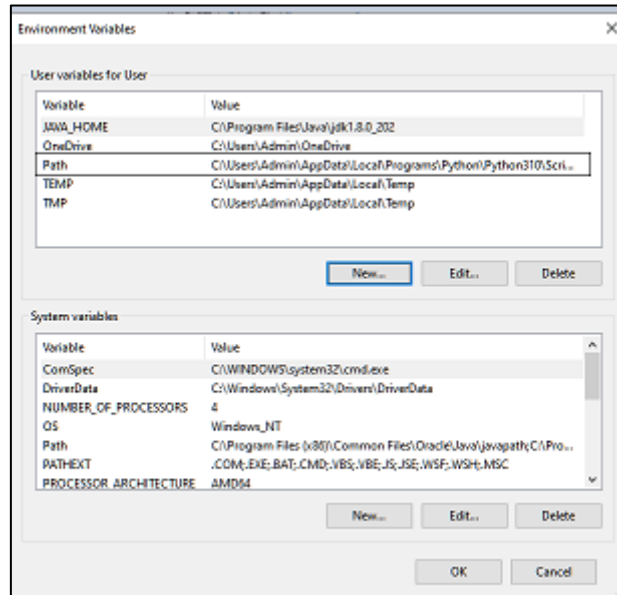


Step 4: Give the Variable Name and java Hadoop path



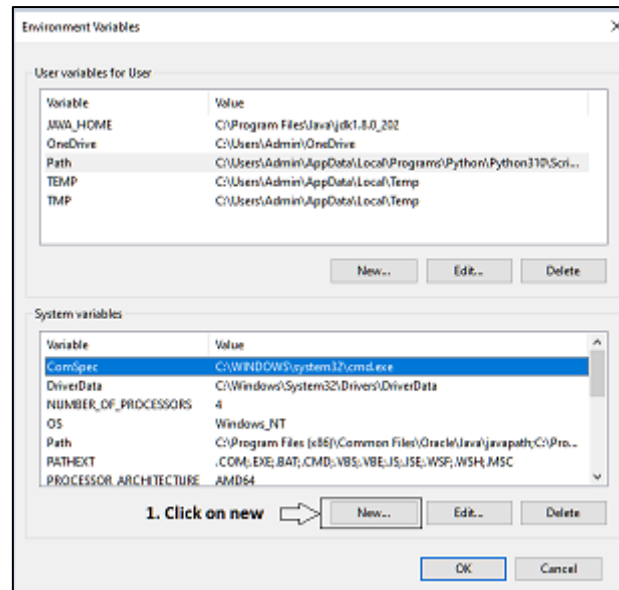


Step 5: Add the Variable into the Path

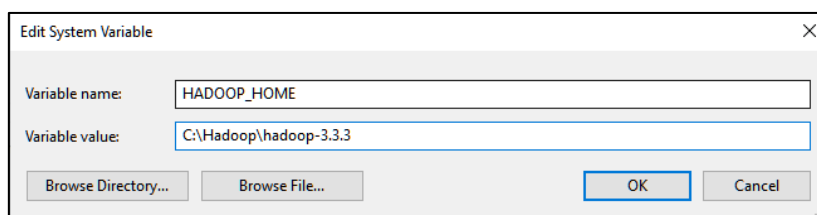
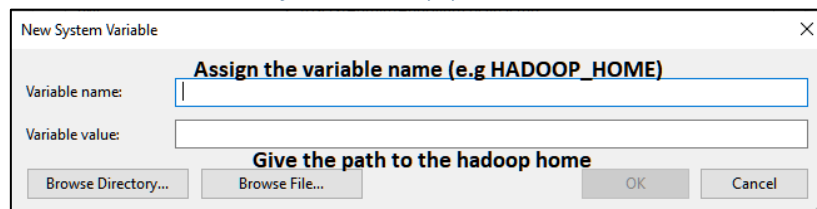


Step 6: Create the new system variables

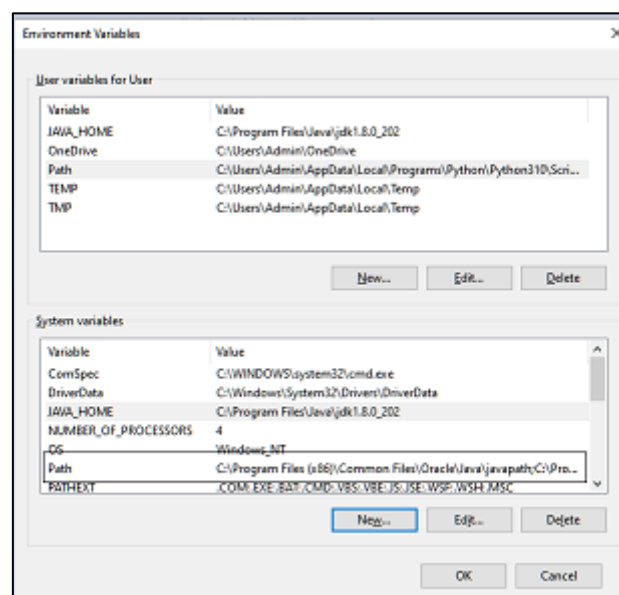


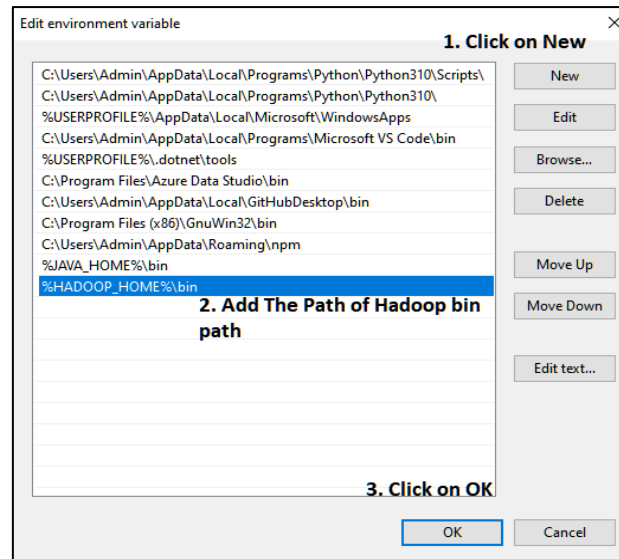


Step 7: Give the Variable Name and java Hadoop path



Step 8: Add the Variable into the Path





## Setting up Hadoop

Step 1: Create the new Folder with the name DATA

Name	Date modified	Type	Size
bin	09-05-2022 11:14 PM	File folder	
DATA	27-07-2022 02:19 PM	File folder	
etc	09-05-2022 10:09 PM	File folder	
include	09-05-2022 11:14 PM	File folder	
lib	09-05-2022 11:14 PM	File folder	
libexec	09-05-2022 11:14 PM	File folder	
licenses-binary	09-05-2022 11:14 PM	File folder	
sbin	09-05-2022 10:09 PM	File folder	
share	09-05-2022 11:50 PM	File folder	
LICENSE	09-05-2022 09:54 PM	Text Document	15 KB
LICENSE-binary	09-05-2022 09:54 PM	File	23 KB
NOTICE	22-04-2022 08:28 PM	Text Document	2 KB
NOTICE-binary	09-05-2022 09:54 PM	File	29 KB
README	22-04-2022 08:28 PM	Text Document	1 KB

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

← → ↶ ↷	📁 This PC > Local Disk (C:) > Hadoop > hadoop-3.3.3 > DATA >	Name	Date modified	Type	Size
★ Quick access		📁 DataNode	27-07-2022 02:21 PM	File folder	
☁ OneDrive		📁 NameNode	27-07-2022 02:21 PM	File folder	
💻 This PC					
📁 3D Objects					
🖥 Desktop					
📁 Documents					
⬇ Downloads					
🎵 Music					
🖼 Pictures					
📺 Videos					
📁 Local Disk (C:)					
📁 USB Drive (F:)					

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

```
<?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.
-->

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

<configuration>
<property>
<name>dfs.replication</name>
<value>1</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>
</configuration>
```

```

hdfs-site.xml X
C: > Hadoop > hadoop-3.3.3 > etc > hadoop > hdfs-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>dfs.replication</name>
22 <value>1</value>
23 </property>
24
25 <property>
26 <name>dfs.namenode.name.dir</name>
27 <value>C:\Hadoop\hadoop-3.3.3\DATA\NameNode</value>
28 </property>
29
30 <property>
31 <name>dfs.datanode.data.dir</name>
32 <value>C:\Hadoop\hadoop-3.3.3\DATA\DataNode</value>
33 </property>
34 </configuration>
35

```

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

```

<?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

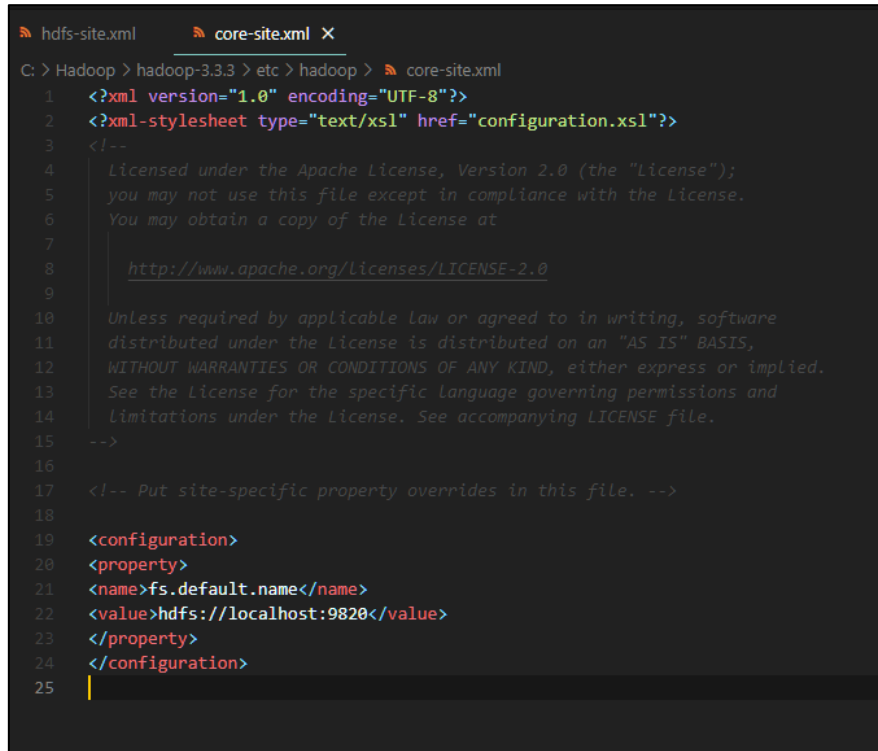
        http://www.apache.org/licenses/LICENSE-2.0

    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.
-->

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

```

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



```
hdfs-site.xml  core-site.xml X
C: > Hadoop > hadoop-3.3.3 > etc > hadoop > 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:9820</value>
23 </property>
24 </configuration>
25
```

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

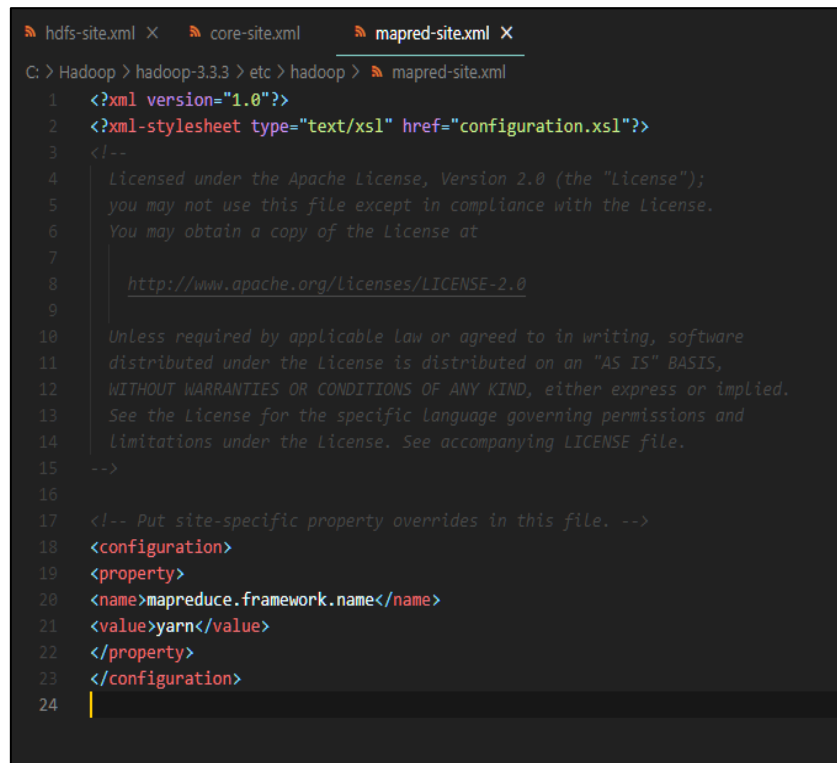
```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
    Licensed under the Apache License, Version 2.0 (the "License");
    you may not use this file except in compliance with the License.
    You may obtain a copy of the License at

        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>
</configuration>
```



```
hdfs-site.xml X core-site.xml mapred-site.xml X
C: > Hadoop > hadoop-3.3.3 > etc > hadoop > mapred-site.xml
1 <?xml version="1.0"?>
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 <configuration>
19 <property>
20 <name>mapreduce.framework.name</name>
21 <value>yarn</value>
22 </property>
23 </configuration>
24
```

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

```
<?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

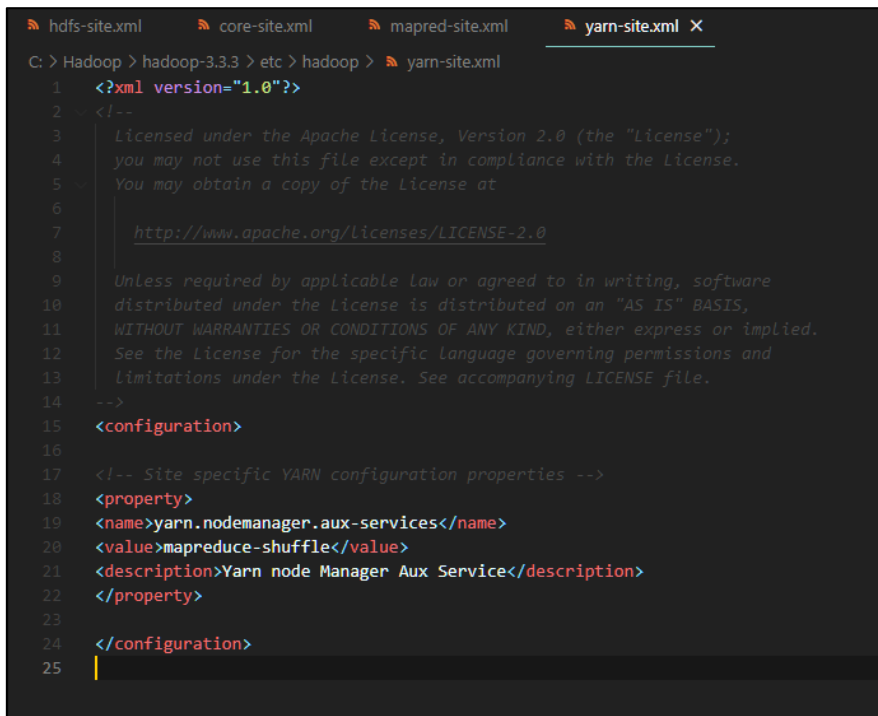
http://www.apache.org/licenses/LICENSE-2.0

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 -->
<property>
```

```
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce-shuffle</value>
<description>Yarn node Manager Aux Service</description>
</property>

</configuration>
```



```
hdfs-site.xml  core-site.xml  mapred-site.xml  yarn-site.xml X
C: > Hadoop > hadoop-3.3.3 > etc > hadoop > 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 <property>
19 <name>yarn.nodemanager.aux-services</name>
20 <value>mapreduce-shuffle</value>
21 <description>Yarn node Manager Aux Service</description>
22 </property>
23
24 </configuration>
25 |
```

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

```
C:\Windows\System32\cmd.exe
2022-07-27 14:56:57,410 INFO util.GSet: Computing capacity for map INodeMap
2022-07-27 14:56:57,411 INFO util.GSet: VM type = 64-bit
2022-07-27 14:56:57,413 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
2022-07-27 14:56:57,415 INFO util.GSet: capacity = 2^20 = 1048576 entries
2022-07-27 14:56:57,417 INFO namenode.FSDirectory: ACLs enabled? true
2022-07-27 14:56:57,417 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2022-07-27 14:56:57,419 INFO namenode.FSDirectory: XAttrs enabled? true
2022-07-27 14:56:57,420 INFO namenode.NameNode: Caching file names occurring more than 10 times
2022-07-27 14:56:57,433 INFO snapshot.SnapshotManager: Loaded config captureOpenfiles: false, skipCaptureAccessTimeOnlyChange: false, snapshotDiffAllowSnapRootDescendan
t: true, maxSnapshotLength: 65536
2022-07-27 14:56:57,438 INFO snapshot.SnapshotManager: SkipList is disabled
2022-07-27 14:56:57,440 INFO util.GSet: Computing capacity for map cachedBlocks
2022-07-27 14:56:57,448 INFO util.GSet: VM type = 64-bit
2022-07-27 14:56:57,452 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
2022-07-27 14:56:57,452 INFO util.GSet: capacity = 2^18 = 262144 entries
2022-07-27 14:56:57,475 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-07-27 14:56:57,476 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-07-27 14:56:57,476 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-07-27 14:56:57,490 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-07-27 14:56:57,492 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2022-07-27 14:56:57,497 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-07-27 14:56:57,497 INFO util.GSet: VM type = 64-bit
2022-07-27 14:56:57,499 INFO util.GSet: 0.0299999999329447746% max memory 889 MB = 273.1 KB
2022-07-27 14:56:57,501 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory root= C:\Hadoop\hadoop-3.3.3\DATA\NameNode; location= null ? (Y or N) Y
2022-07-27 14:58:09,984 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1977693252-172.28.48.1-1658914d889970
2022-07-27 14:58:09,985 INFO common.Storage: Will remove files: [C:\Hadoop\hadoop-3.3.3\DATA\NameNode\current\fsimage_000000000000000000, C:\Hadoop\hadoop-3.3.3\DATA\N
ameNode\current\fsimage_000000000000000000.mds, C:\Hadoop\hadoop-3.3.3\DATA\NameNode\current\seen_txid, C:\Hadoop\hadoop-3.3.3\DATA\NameNode\current\VERSION]
2022-07-27 14:58:10,036 INFO common.Storage: Storage directory C:\Hadoop\hadoop-3.3.3\DATA\NameNode has been successfully formatted.
2022-07-27 14:58:10,081 INFO namenode.FSImageFormatProtobuf: Saving image file C:\Hadoop\hadoop-3.3.3\DATA\NameNode\current\fsimage_cpkt_000000000000000000 using no co
mpression
2022-07-27 14:58:10,296 INFO namenode.FSImageFormatProtobuf: Image file C:\Hadoop\hadoop-3.3.3\DATA\NameNode\current\fsimage_000000000000000000 of size 399 bytes
saved in 0 seconds .
2022-07-27 14:58:10,310 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-07-27 14:58:10,341 INFO namenode.FSNamesystem: Stopping services started for active state
2022-07-27 14:58:10,341 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-07-27 14:58:10,346 INFO namenode.FSImage: FSImage save clean checkpoint: txid=0 when meet shutdown.
2022-07-27 14:58:10,346 INFO namenode.NameNode: SHUTDOWN_MSG:
*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-D4U1K1L/172.28.48.1
*****
C:\Hadoop\hadoop-3.3.3\bin>
```

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



```
Select C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.1826]
(c) Microsoft Corporation. All rights reserved.

C:\Hadoop\hadoop-3.3.3\sbin>.start-dfs.cmd

C:\Hadoop\hadoop-3.3.3\sbin>_
```

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

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.1826]
(c) Microsoft Corporation. All rights reserved.

C:\Hadoop\hadoop-3.3.3\sbin>.start-yarn.cmd
starting yarn daemons

C:\Hadoop\hadoop-3.3.3\sbin>_
```

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

The screenshot shows a web browser window displaying the Hadoop NameNode Overview page. The browser's address bar shows the URL `localhost:9870/dfshealth.html#tab-overview`. The page has a green header with navigation tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Overview 'localhost:9820' (✓active)". Below the title is a table with the following information:

Started:	Wed Jul 27 15:00:02 +0530 2022
Version:	3.3.3, rd37586cbda38c338d9fe481adda5a05fb516771
Compiled:	Mon May 09 22:06:00 +0530 2022 by steevel from branch-3.3.3
Cluster ID:	CID-421309ee-a178-42fc-a6a4-ec5c50084dce
Block Pool ID:	BP-1977693252-172.28.48.1-1658914089970

Below the table is a "Summary" section. It states: "Security is off.", "Safemode is off.", "1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).", and "Heap Memory used 70.68 MB of 191.5 MB Heap Memory. Max Heap Memory is 889 MB."