




# HADOOP-DEPLOYING ON WINDOWS

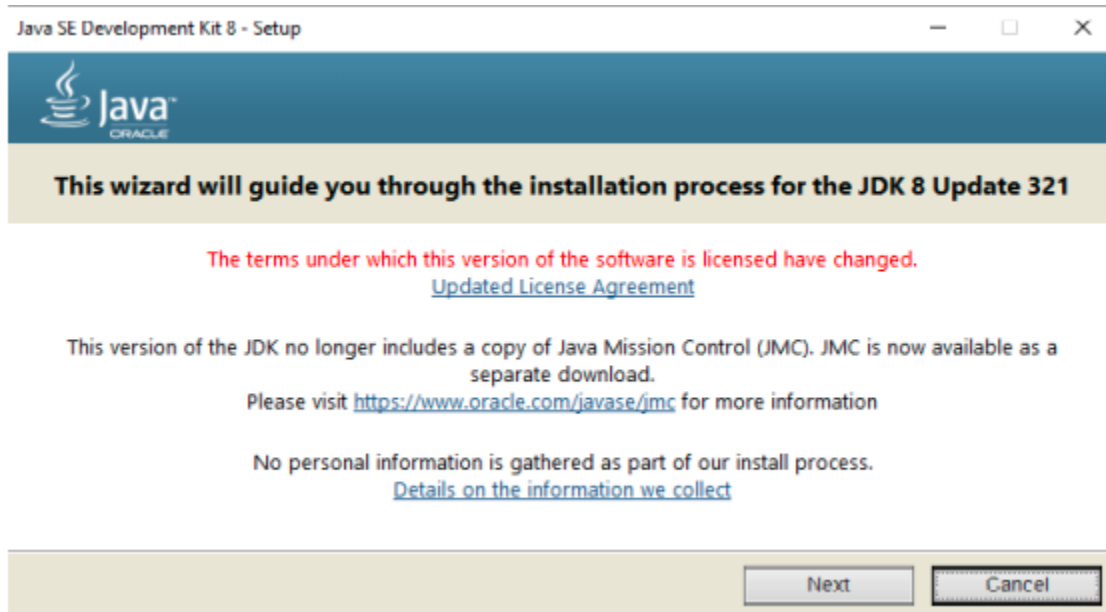
Rao Nauman

P19-0073

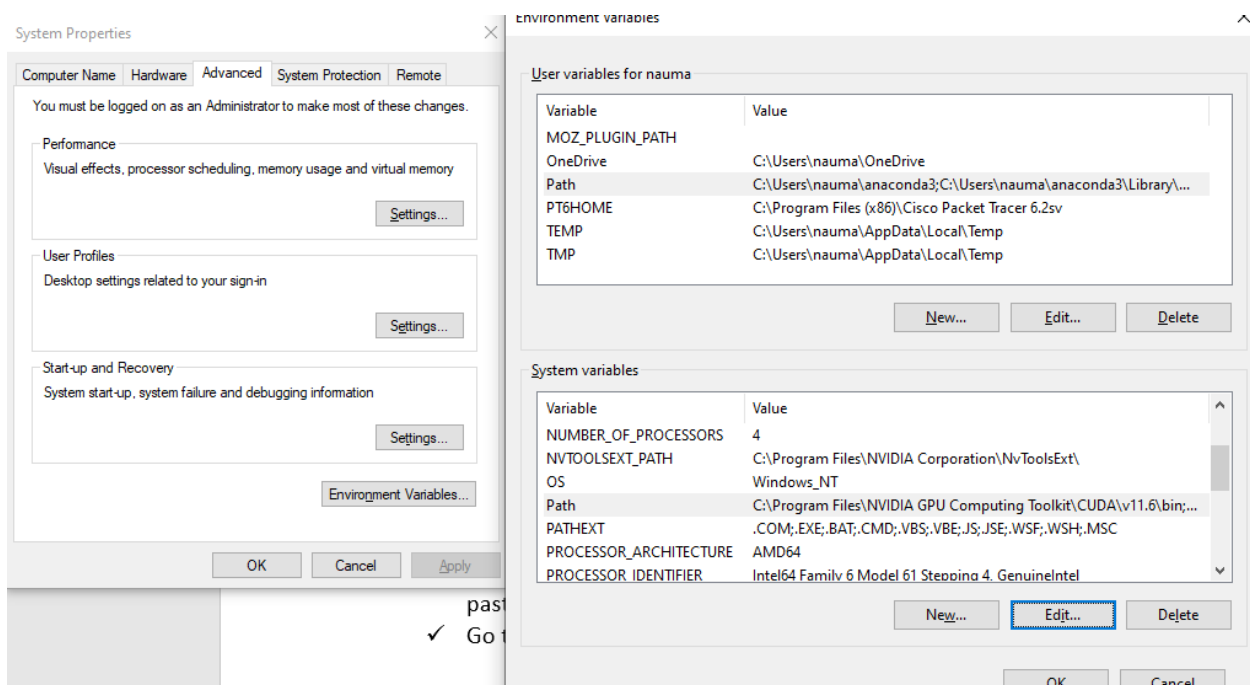


## Steps to follow:

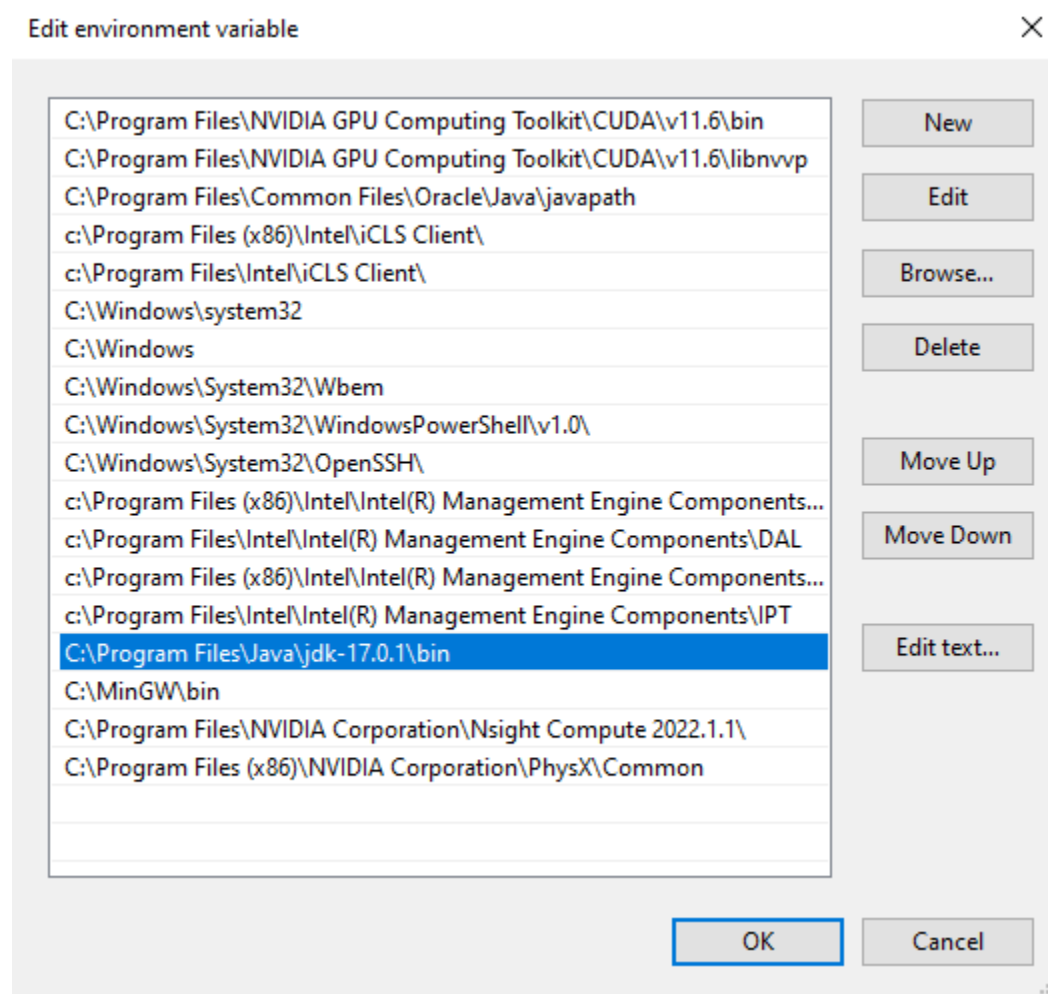
- ✓ Download and set up Java. Go to:  
<https://www.oracle.com/java/technologies/downloads/#java8>




- ✓ Make a directory in C naming Java and install it there. Then for installation of it's jdk files, keep the default directory i.e, **program files/java/jdk-17.0.1**
- ✓ After successful completion of installation, go to **program files/java** and cut that **jdk-17.0.1** and paste it in **C://java**.
- ✓ Go to system environment variables



I have already set it up for something else:



- ✓ **Download and Set up HADOOP**
- ✓ Go to <https://hadoop.apache.org/releases.html>


[Download](#)
[Documentation](#)
[Community](#)
[Development](#)
[Help](#)
[Apache Software Foundation](#)

## 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
3.2.3	2022 Mar 28	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>
3.3.2	2022 Mar 3	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a> <a href="#">binary-aarch64 (checksum signature)</a>	<a href="#">Announcement</a>
2.10.1	2020 Sep 21	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>

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`

All previous releases of Hadoop are available from the [Apache release archive site](#).



Many third parties distribute products that include Apache Hadoop and related tools. Some of these are listed on the [Distributions wiki page](#).

## License

The software licensed under [Apache License 2.0](#)

- ✓ Select your version and click on **Binary to download it**.
- ✓ After downloading, Install it in your desired directory,
- ✓ After installing go to Hadoop directory and create a new folder named data. **Open data folder** and create 2 new folders named **namenode & datanode**.

bin	18/04/2022 5:32 am	File folder
data	18/04/2022 5:36 am	File folder
etc	18/04/2022 5:32 am	File folder
include	18/04/2022 5:32 am	File folder
lib	18/04/2022 5:34 am	File folder
libexec	18/04/2022 5:32 am	File folder
sbin	18/04/2022 5:32 am	File folder
share	18/04/2022 5:32 am	File folder
LICENSE.txt	10/03/2022 10:39 am	Text Document
NOTICE.txt	10/03/2022 10:39 am	Text Document
README.txt	10/03/2022 10:39 am	Text Document

Name	Date modified	Type	Size
 datanode	18/04/2022 5:35 am	File folder	
 namenode	18/04/2022 5:35 am	File folder	

✓ Go to **hadoop\hadoop-3.2.3\etc\hadoop** and open following files in VSCODE Editor:

1. **Core-site.xml**
2. **Mapred-site.xml**
3. **Yarn-site.xml**
4. **Hdfs-site.xml**
5. **Hadoop-env.cmd**

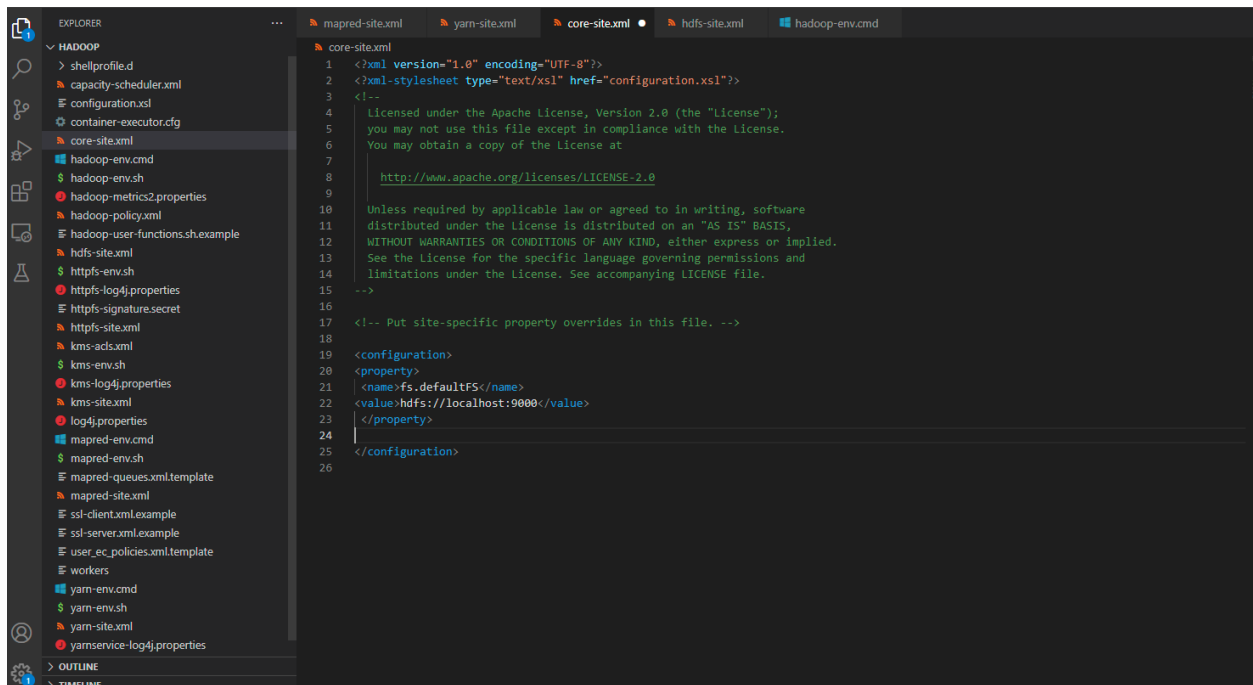
In core-site.xml file commit following changes under <configuration>

<property>

<name>fs.defaultFS</name>

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

</property>



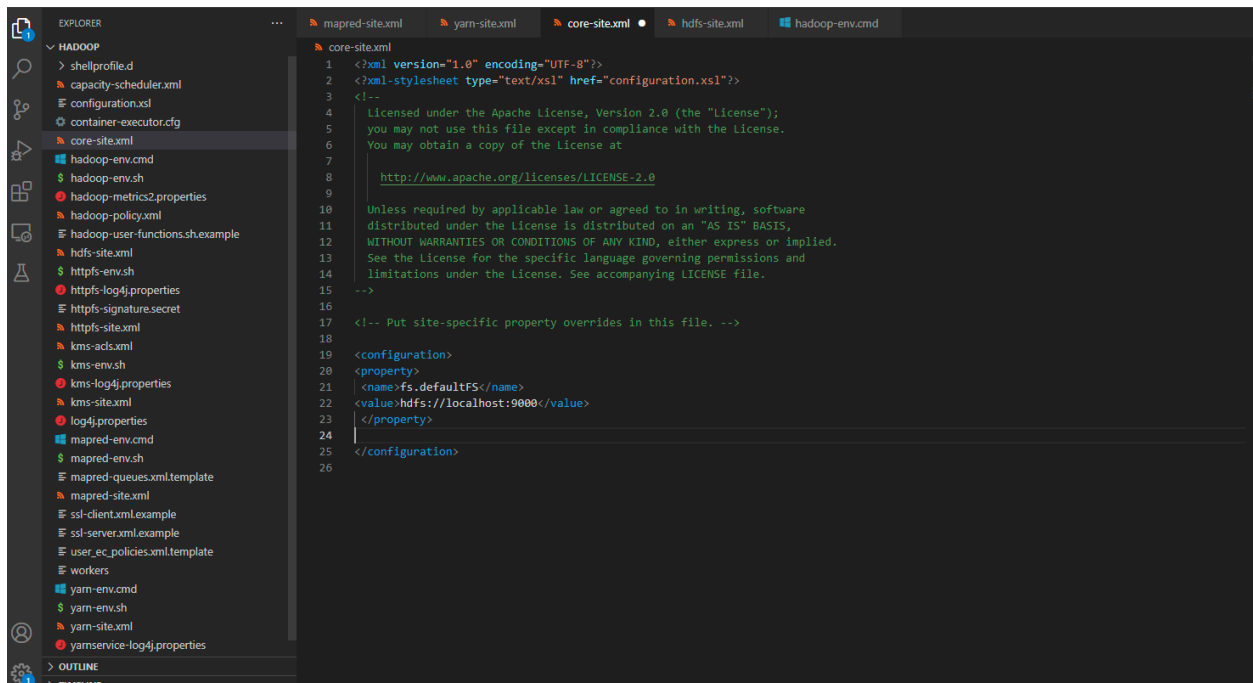
In Mapred-site.xml commit following changes under <configuration>

<property>

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

<value>yarn</value>

</property>



In yarn-site.xml commit following changes under <configuration>

```
<property>
```

```
<name>yarn.nodemanager.aux-services</name>
```

```
<value>mapreduce_shuffle</value>
```

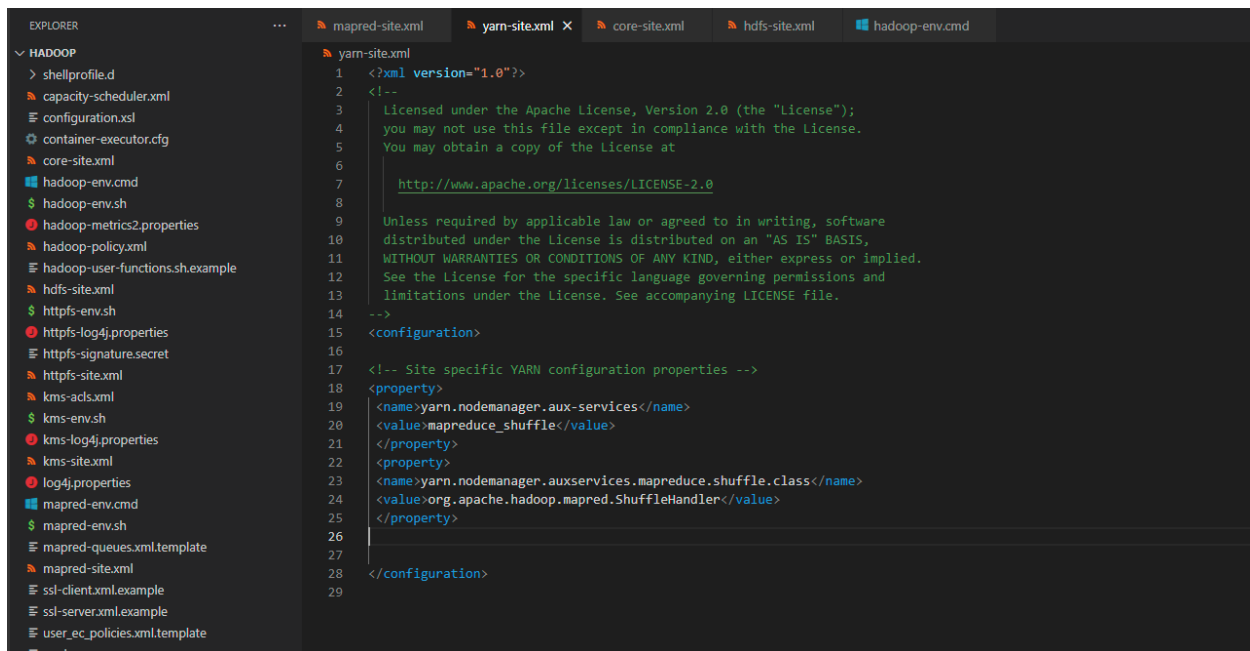
```
</property>
```

```
<property>
```

```
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
```

```
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
```

```
</property>
```



In hdfs-site.xml commit following changes under <configuration>

```
<property>
```

```
<name>dfs.replication</name>
```

```
<value>1</value>
```

```
</property>
```

```
<property>
```

```
<name>dfs.namenode.name.dir</name>
```

```
<value>C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\namenode</value>
```

```
</property>
```

```
<property>
```

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

```
<value>C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\datanode</value>
```

```
</property>
```



```
mapred-site.xml  yarn-site.xml  core-site.xml  hdfs-site.xml X  hadoop-env.cmd
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 <property>
25 <name>dfs.namenode.name.dir</name>
26 <value>C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\namenode</value>
27 </property>
28 <property>
29 <name>dfs.datanode.data.dir</name>
30 <value>C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\datanode</value>
31 </property>
32
33 </configuration>
34
35
```

In Hadoop-env.cmd set JAVA\_HOME=C:\java\jdk-17.0.1

```
mapred-site.xml  yarn-site.xml  core-site.xml  hdfs-site.xml  hadoop-env.cmd
hadoop-env.cmd
9  @rem http://www.apache.org/licenses/LICENSE-2.0
10 @rem
11 @rem Unless required by applicable law or agreed to in writing, software
12 @rem distributed under the License is distributed on an "AS IS" BASIS,
13 @rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 @rem See the License for the specific language governing permissions and
15 @rem limitations under the License.
16
17 @rem Set Hadoop-specific environment variables here.
18
19 @rem The only required environment variable is JAVA_HOME. All others are
20 @rem optional. When running a distributed configuration it is best to
21 @rem set JAVA_HOME in this file, so that it is correctly defined on
22 @rem remote nodes.
23
24 @rem The java implementation to use. Required.
25 set JAVA_HOME=C:\java\jdk-17.0.1
26
27 @rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
28 @rem set JSVC_HOME=%JSVC_HOME%
29
30 @rem set HADOOP_CONF_DIR=
31
32 @rem Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.
33 if exist %HADOOP_HOME%\contrib\capacity-scheduler (
34     if not defined HADOOP_CLASSPATH (
35         set HADOOP_CLASSPATH=%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
36     ) else (
37         set HADOOP_CLASSPATH=%HADOOP_CLASSPATH%;%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
38     )
39 )
40
41 @rem The maximum amount of heap to use, in MB. Default is 1000.
42 @rem set HADOOP_HEAPSIZE=
```

Now go again to System environment variables and add the following values to PATH variable

- ✓ C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\bin
- ✓ C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\sbin
- ✓ Now go to <https://drive.google.com/file/d/1AMqV4F5ybPF4ab4CeK8B3AsjdGtQCdvy/view> to download a bin file of Hadoop which you will replace with the old Hadoop bin file. The reason is that the old bin contains some missing files by default. After that open cmd and enter **hdfs namenode -format** to check the successful implementation of path values

```
Command Prompt

2022-04-18 06:04:10,580 INFO util.GSet: VM type = 64-bit
2022-04-18 06:04:10,596 INFO util.GSet: 0.25% max memory 1000 MB = 2.5 MB
2022-04-18 06:04:10,596 INFO util.GSet: capacity = 2^18 = 262144 entries
2022-04-18 06:04:10,611 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-04-18 06:04:10,611 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-04-18 06:04:10,611 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-04-18 06:04:10,611 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-04-18 06:04:10,611 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2022-04-18 06:04:10,627 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-04-18 06:04:10,627 INFO util.GSet: VM type = 64-bit
2022-04-18 06:04:10,627 INFO util.GSet: 0.029999999329447746% max memory 1000 MB = 307.2 KB
2022-04-18 06:04:10,627 INFO util.GSet: capacity = 2^15 = 32768 entries
2022-04-18 06:04:10,689 INFO namenode.FSImage: Allocated new BlockPoolId: BP-962276228-192.168.18.141-1650243850674
2022-04-18 06:04:10,767 INFO common.Storage: Storage directory C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\namenode has been successfully formatted.
2022-04-18 06:04:10,830 INFO namenode.FSImageFormatProtobuf: Saving image file C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\namenode\current\fsimage.ckpt_00000000000000000000 using no compression
2022-04-18 06:04:10,986 INFO namenode.FSImageFormatProtobuf: Image file C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\namenode\current\fsimage.ckpt_00000000000000000000 of size 397 bytes saved in 0 seconds .
2022-04-18 06:04:11,002 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-04-18 06:04:11,033 INFO namenode.FSNamesystem: Stopping services started for active state
2022-04-18 06:04:11,033 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-04-18 06:04:11,049 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2022-04-18 06:04:11,049 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at Nauman/192.168.18.141
*****/

C:\Users\nauma>
```

Now to Run Hadoop, open `hadoop\hadoop-3.2.3\sbin` folder in cmd and type `start-all.cmd`

Hadoop running will look like this

```
Apache Hadoop Distribution
at org.apache.hadoop.util.DiskChecker.checkDirInternal(DiskChecker.java:141)
at org.apache.hadoop.util.DiskChecker.checkDir(DiskChecker.java:116)
at org.apache.hadoop.hdfs.server.datanode.StorageLocation.check(StorageLocation.java:
at org.apache.hadoop.hdfs.server.datanode.checker.ThrottledAsyncChecker$1.call(Thrott
at com.google.common.util.concurrent.TrustedListenableFutureTask$TrustedFutureInter
dListenableFutureTask.java:125)
2022-04-18 06:05:54,505 INFO namenode.LeaseManager: Number of blocks under construction: 0
2022-04-18 06:05:54,521 INFO blockmanagement.BlockManager: Initializing replication queues
2022-04-18 06:05:54,522 INFO hdfs.StateChange: STATE* Leaving safe mode after 0 secs
2022-04-18 06:05:54,523 INFO hdfs.StateChange: STATE* Network topology has 0 racks and 0 datanodes
2022-04-18 06:05:54,523 INFO hdfs.StateChange: STATE* UnderReplicatedBlocks has 0 blocks
2022-04-18 06:05:54,536 INFO blockmanagement.BlockManager: Total number of blocks = 0
2022-04-18 06:05:54,537 INFO blockmanagement.BlockManager: Number of invalid blocks = 0
2022-04-18 06:05:54,538 INFO blockmanagement.BlockManager: Number of under-replicated blocks = 0
2022-04-18 06:05:54,539 INFO blockmanagement.BlockManager: Number of over-replicated blocks = 0
2022-04-18 06:05:54,540 INFO blockmanagement.BlockManager: Number of blocks being written = 0
2022-04-18 06:05:54,541 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over-
-replicated blocks completed in 18 msec
2022-04-18 06:05:54,592 INFO ipc.Server: IPC Server listener on 9000: starting
2022-04-18 06:05:54,592 INFO ipc.Server: IPC Server Responder: starting
2022-04-18 06:05:54,597 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000
2022-04-18 06:05:54,602 INFO namenode.FSNamesystem: Starting services required for active state
2022-04-18 06:05:54,603 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)
2022-04-18 06:05:54,613 INFO namenode.FSDirectory: Quota initialization completed in 10 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2022-04-18 06:05:54,620 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with inte
00 milliseconds

-18 06:05:47,735 ERROR datanode.DataNode: Exception in secureMain
che.hadoop.util.DiskChecker$DiskErrorException: Too many failed volumes - current valid
1, volumes failed: 1, volume failures tolerated: 0
at org.apache.hadoop.hdfs.server.datanode.DataNode.makeInstance(DataNode.java:2806)
at org.apache.hadoop.hdfs.server.datanode.DataNode.instantiateDataNode(DataNode.java:2763)
at org.apache.hadoop.hdfs.server.datanode.DataNode.createDataNode(DataNode.java:2763)
at org.apache.hadoop.hdfs.server.datanode.DataNode.secureMain(DataNode.java:2907)
at org.apache.hadoop.hdfs.server.datanode.DataNode.main(DataNode.java:2931)
-18 06:05:47,745 INFO util.ExitUtil: Exiting with status 1: org.apache.hadoop.util.Disk
many failed volumes - current valid volumes: 0, volumes configured: 1, volumes failed:

-18 06:05:47,800 INFO datanode.DataNode: SHUTDOWN_MSG:
*****
IN_MSG: Shutting down DataNode at Nauman/192.168.18.141
*****/

C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\sbin>

Loader.defineClass(java.lang.String,byte[],int,int,java.security.ProtectionDomain) throws java.lang.ClassFor
essible: module java.base does not "opens java.lang" to unnamed module @4cc451f2
at java.base/java.lang.reflect.AccessibleObject.checkCanSetAccessible(AccessibleObject.java:354)
at java.base/java.lang.reflect.AccessibleObject.checkCanSetAccessible(AccessibleObject.java:297)
at java.base/java.lang.reflect.Method.checkCanSetAccessible(Method.java:193)
at java.base/java.lang.reflect.Method.setAccessible(Method.java:193)
at com.google.inject.internal.cglib.core.$ReflectUtils$2.run(ReflectUtils.java:56)
at java.base/java.security.AccessController.doPrivileged(AccessController.java:318)
at com.google.inject.internal.cglib.core.$ReflectUtils.<clinit>(ReflectUtils.java:46)
... 29 more
2022-04-18 06:05:50,755 INFO resourcemanager.ResourceManager: Transitioning to standby state
2022-04-18 06:05:50,787 INFO resourcemanager.ResourceManager: Transitioned to standby state
2022-04-18 06:05:50,803 INFO resourcemanager.ResourceManager: SHUTDOWN_MSG:
*****
SHUTDOWN_MSG: Shutting down ResourceManager at Nauman/192.168.18.141
*****/

C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\sbin>
```

