
Name: Muhammad Abeer

Roll no: 19P-0061

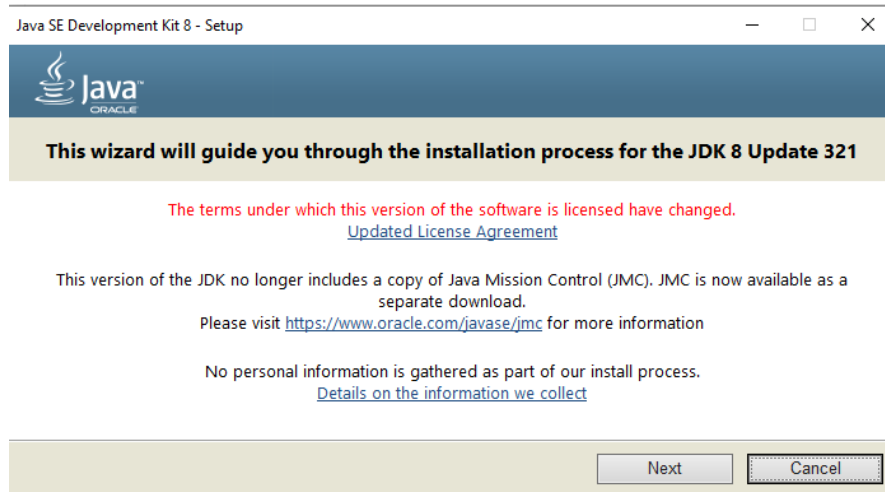
Section: 6-B

Deploying HADOOP on Windows

Step 1

Download and set up Java.

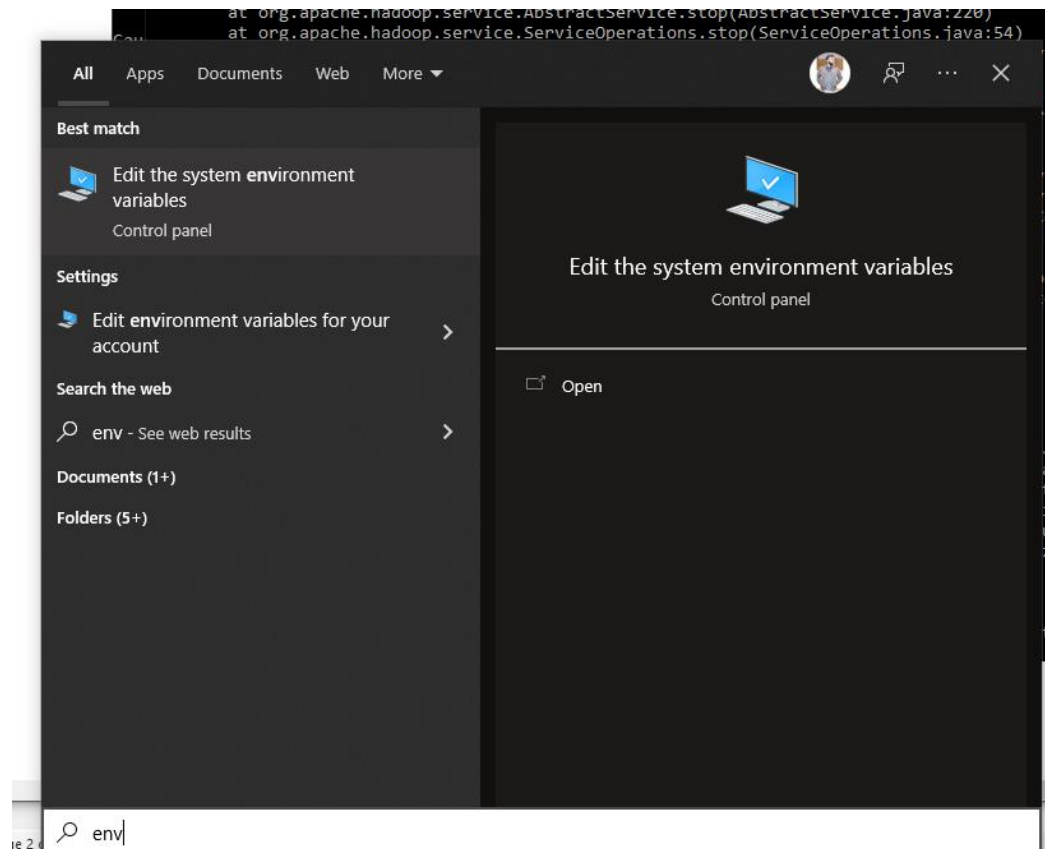
Go to: <https://www.oracle.com/java/technologies/downloads/#java8>



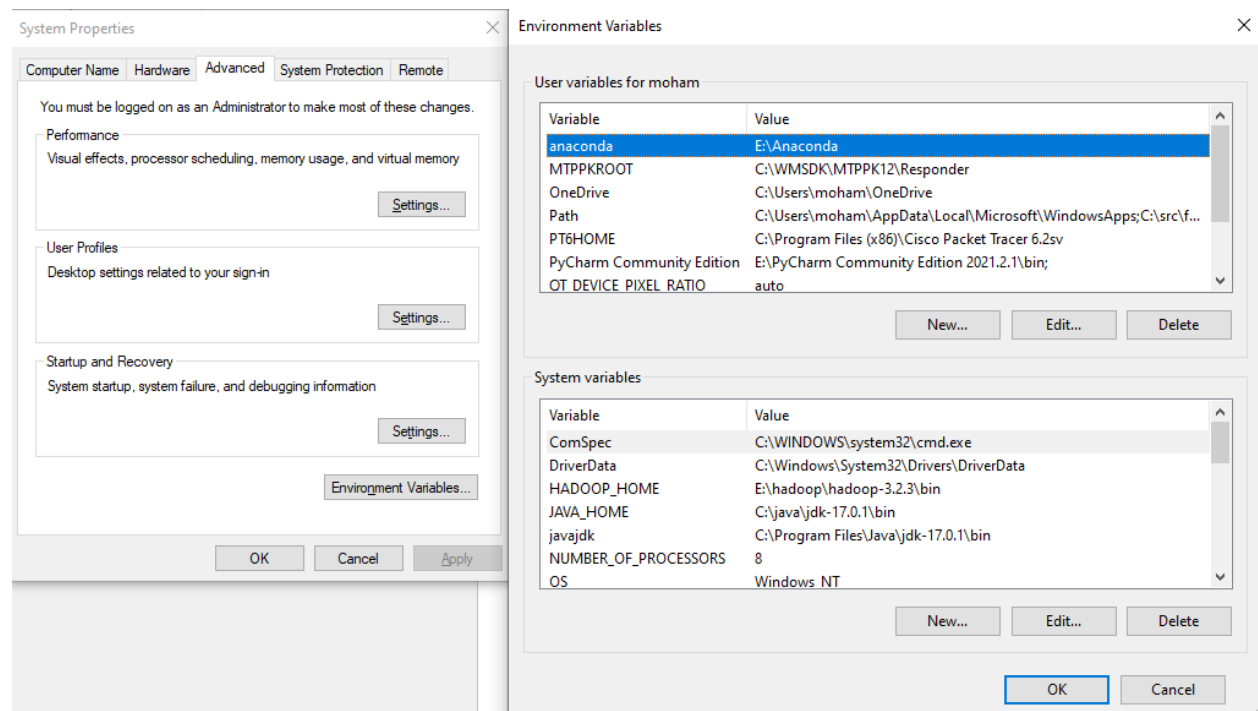
Make a directory in C naming Java and install it there. Then for installation of it's jdk files, keep the default directory i.e, **program files/java/jdk-17.0.1**

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 by typing env in start.

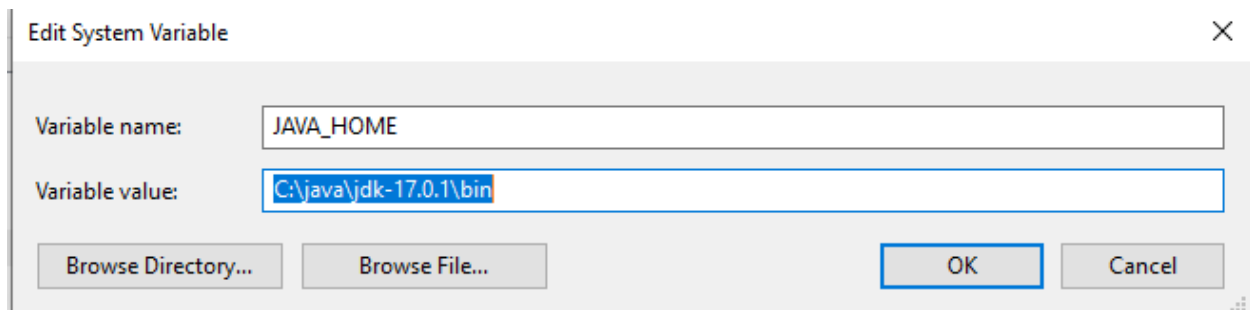


It will look like this



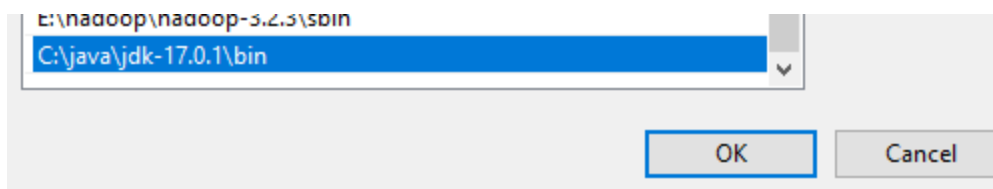
In system Variables, click on new and type new variable JAVA_HOME and set it's value to

C:\Program Files\Java\jdk-17.0.1\bin



Also add new value for path variable

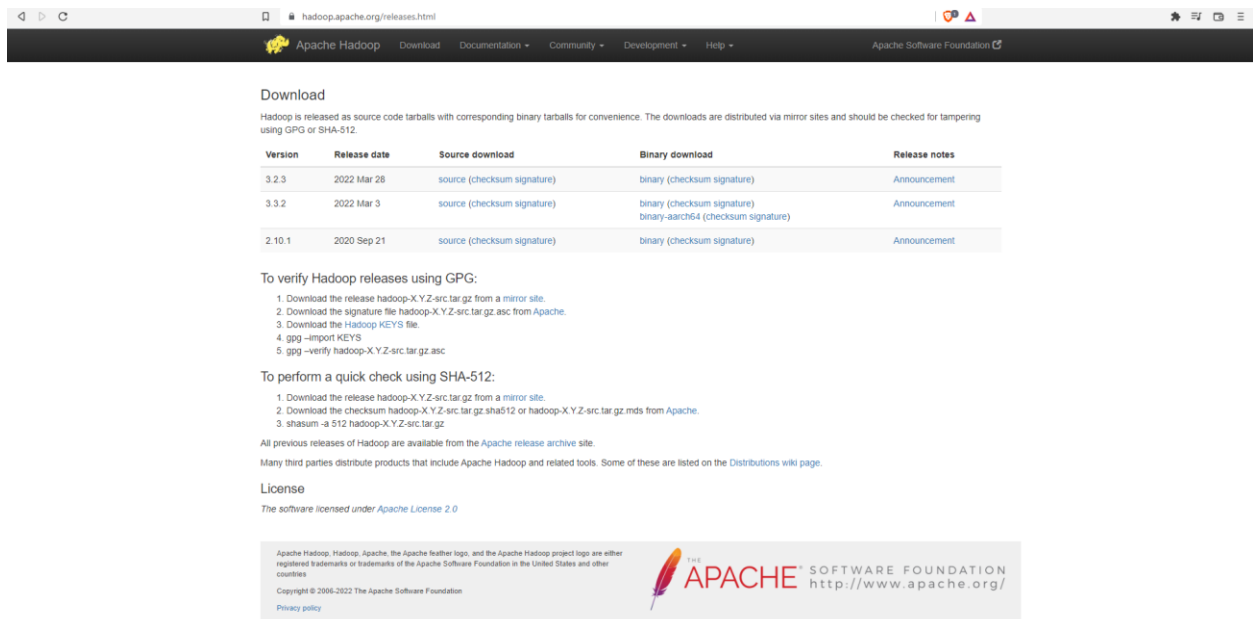
C:\Program Files\Java\jdk-17.0.1\bin



Step 2

Download and Set up HADOOP

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



The screenshot shows the Apache Hadoop releases page. At the top, there's a navigation bar with links for Download, Documentation, Community, Development, and Help. Below this, the 'Download' section is highlighted. It contains a table with columns for Version, Release date, Source download, Binary download, and Release notes. The table lists three versions: 3.2.3 (released Mar 28, 2022), 3.3.2 (released Mar 3, 2022), and 2.10.1 (released Sep 21, 2020). Each version has links for source and binary downloads, and a link to release notes. Below the table, there are instructions on how to verify releases using GPG and SHA-512. At the bottom, there's a footer with the Apache Software Foundation logo and copyright information.

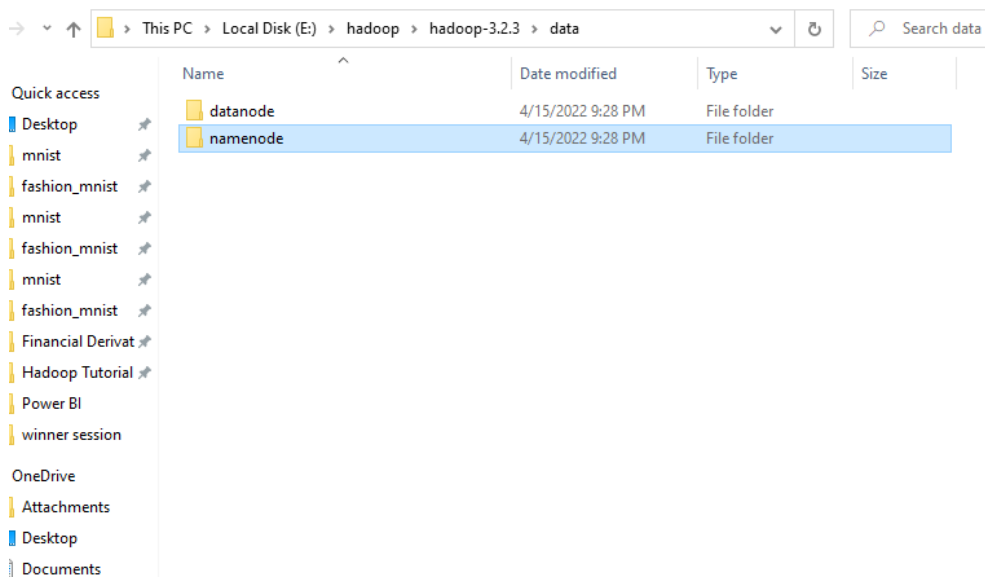
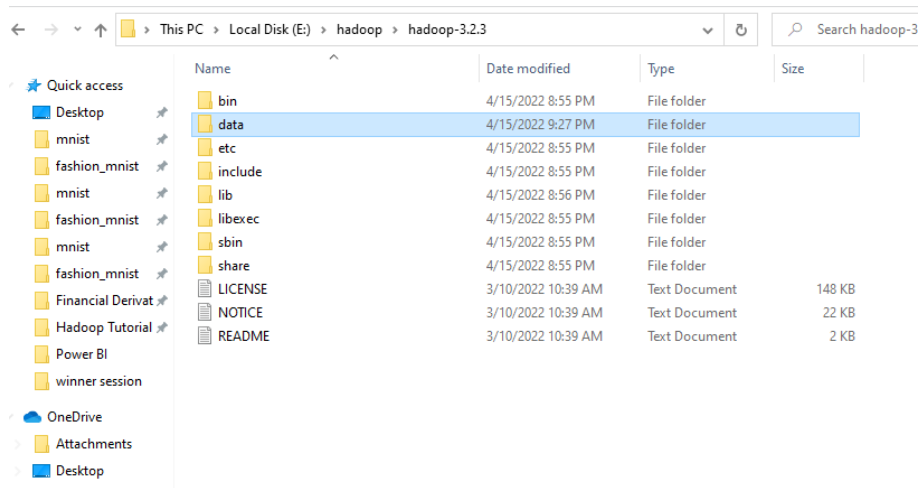
Version	Release date	Source download	Binary download	Release notes
3.2.3	2022 Mar 28	source (checksum signature)	binary (checksum signature)	Announcement
3.3.2	2022 Mar 3	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.1	2020 Sep 21	source (checksum signature)	binary (checksum signature)	Announcement

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**.



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**

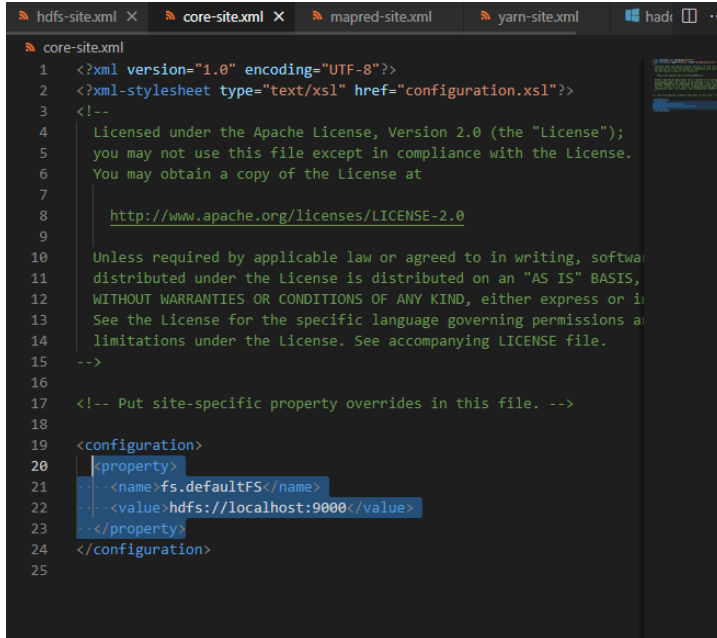
In core-site.xml file commit following changes under <configuration>

<property>

<name>fs.defaultFS</name>

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

```
</property>
```



```
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.defaultFS</name>
22     <value>hdfs://localhost:9000</value>
23   </property>
24 </configuration>
25
```

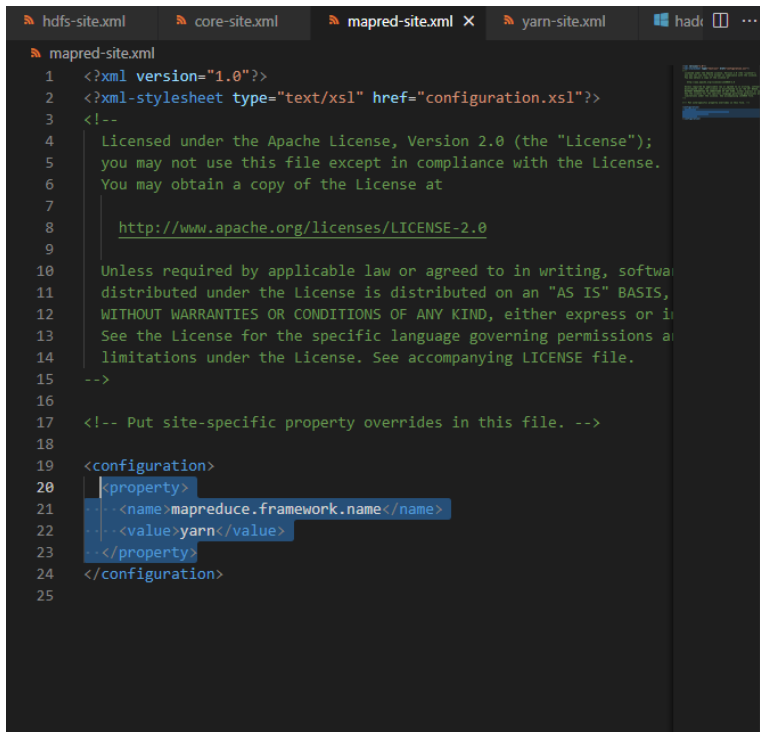
In Mapred-site.xml commit following changes under <configuration>

```
<property>
```

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

```
<value>yarn</value>
```

```
</property>
```



```
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
19 <configuration>
20 <property>
21 <name>mapreduce.framework.name</name>
22 <value>yarn</value>
23 </property>
24 </configuration>
25
```

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

<property>

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

<value>mapreduce_shuffle</value>

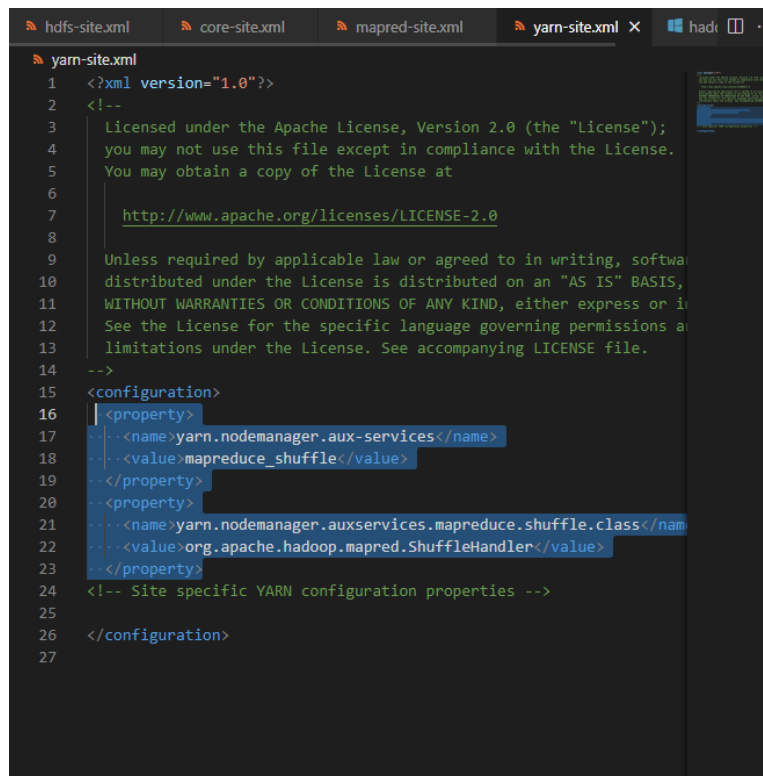
</property>

<property>

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

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

</property>



```
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 <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 <!-- Site specific YARN configuration properties -->
25
26 </configuration>
27
```

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>E:\hadoop\hadoop-3.2.3\data\namenode</value>

</property>

<property>

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

<value>E:\hadoop\hadoop-3.2.3\data\datanode</value>

</property>

```
hdfs-site.xml x core-site.xml mapred-site.xml yarn-site.xml hadi ...
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>E:\hadoop\hadoop-3.2.3\data\namenode</value>
27 </property>
28 <property>
29 <name>dfs.datanode.data.dir</name>
30 <value>E:\hadoop\hadoop-3.2.3\data\datanode</value>
31 </property>
32 </configuration>
33
```

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

```
hadoop-env.cmd
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
20 @rem optional. When running a distributed configuration it is best
21 @rem to 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
28 @rem set JSVC_HOME=%JSVC_HOME%
29
30 @rem set HADOOP_CONF_DIR=
31
32 @rem Extra Java CLASSPATH elements. Automatically insert capacity-
33 if exist %HADOOP_HOME%\contrib\capacity-scheduler (
34 if exist %HADOOP_HOME%\contrib\capacity-scheduler (

```

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

E:\hadoop\hadoop-3.2.3\bin

E:\hadoop\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.

```

C:\> Command Prompt

Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

C:\Users\moham>hdfs namenode -format
2022-04-15 22:25:49,926 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = DESKTOP-AII2QDM/192.168.100.52
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.2.3
STARTUP_MSG: classpath = E:\hadoop\hadoop-3.2.3\etc\hadoop;E:\hadoop\hadoop-3.2.3\share\hadoop\common;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\accessors-smart-2.4.7.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\asm-5.0.4.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\audience-annotations-0.5.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\avro-1.7.7.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\checker-qual-2.5.2.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-beanutils-1.9.4.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-cli-1.2.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-codec-1.11.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-collections-3.2.2.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-compress-1.21.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-configuration2-2.1.1.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-io-2.8.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-lang3-3.7.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-logging-1.1.3.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-math3-3.1.1.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-net-3.6.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\commons-text-1.4.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\curator-client-2.13.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\curator-framework-2.13.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\curator-recipes-2.13.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\dnsjava-2.1.7.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\error-prone-annotations-2.2.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\failureaccess-1.0.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\gson-2.2.4.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\guava-27.0-jre.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\hadoop-annotations-3.2.3.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\hadoop-auth-3.2.3.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\htrace-core4-4.1.0-incubating.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\httpclient-4.5.13.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\httpcore-4.4.13.jar;E:\hadoop\hadoop-3.2.3\share\hadoop\common\lib\j2objc-annotations-1.1.jar;E:\hadoop
*****/

```

Now to Run Hadoop, open **hadoop\hadoop-3.2.3\sbin** folder in cmd and type

start-all.cmd

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

E:\hadoop\hadoop-3.2.3>cd sbin

E:\hadoop\hadoop-3.2.3\sbin>start-all.cmd
```

Hadoop running will look like this

```
Apache Hadoop Distribution
ResourceLocalizationService.java:426)
    at org.apache.hadoop.service.AbstractService.stop(AbstractService.java:220)
    at org.apache.hadoop.service.ServiceOperations.stop(ServiceOperations.java:54)
    at org.apache.hadoop.service.ServiceOperations.stopQuietly(ServiceOperations.java:102)
    at org.apache.hadoop.service.CompositeService.stop(CompositeService.java:158)
    at org.apache.hadoop.service.CompositeService.serviceStop(CompositeService.java:132)
    at org.apache.hadoop.yarn.server.nodemanager.containermanager.ContainerManagerImpl.serviceStop(ContainerManagerImpl.java:720)
    at org.apache.hadoop.service.AbstractService.stop(AbstractService.java:220)
    at org.apache.hadoop.service.ServiceOperations.stop(ServiceOperations.java:54)
    at org.apache.hadoop.service.ServiceOperations.stopQuietly(ServiceOperations.java:102)
    at org.apache.hadoop.service.CompositeService.stop(CompositeService.java:158)
    at org.apache.hadoop.service.CompositeService.serviceStop(CompositeService.java:132)
    at org.apache.hadoop.yarn.server.nodemanager.NodeManager.serviceStop(NodeManager.java:529)
    at org.apache.hadoop.service.AbstractService.stop(AbstractService.java:220)
    at org.apache.hadoop.service.ServiceOperations.stop(ServiceOperations.java:54)
    at org.apache.hadoop.service.ServiceOperations.stopQuietly(ServiceOperations.java:102)
    at org.apache.hadoop.service.CompositeService$CompositeServiceShutdownHook.run(CompositeService.java:184)
    at java.base/java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:539)
    at java.base/java.util.concurrent.FutureTask.run(FutureTask.java:264)
    at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1136)
    at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:635)
    at java.base/java.lang.Thread.run(Thread.java:833)
SHUTDOWN_MSG: Shutting down NodeManager at 2022-04-15 22:28:15.257
****/*****
SHUTDOWN_MSG: Shutting down NodeManager at 2022-04-15 22:28:15.257
****/*****

E:\hadoop\hadoop-3.2.3\sbin>
2022-04-15 22:28:17,281 INFO ipc.Server: IPC Server Responder: starting
2022-04-15 22:28:17,284 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000
2022-04-15 22:28:17,288 INFO namenode.FSNamesystem: Starting services required for active state
2022-04-15 22:28:17,288 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)
2022-04-15 22:28:17,296 INFO namenode.FSDirectory: Quota initialization completed in 8 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2022-04-15 22:28:17,300 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds
```