

M.C.A. SEMESTER-III

Sub:- Big Data Analytics and Visualization Lab

(Journal)

PCP CENTER: DTSS COLLEGE

University of Mumbai



Institute of Distance and Open Learning (IDOL)

Dr. Shankardayal Sharma bhavan, Vidyanagari, Santacruz(E)

M.C.A. SEMESTER-III



Institute of Distance and Open Learning
Vidya Nagari, Kalina, Santacruz East – 400098.

CERTIFICATE

This is to certify that **Shweta Mahendra Sutar (Application id- 171740)** of Master in Computer Application (MCA) Semester-III has completed the specified term work in the subject of **Big Data Analytics and Visualization Lab** satisfactorily within this institute as laid down by University of Mumbai during the academic year 2022 to 2023 .

Subject In-charge

External Examiner

Coordinator – M.C.A

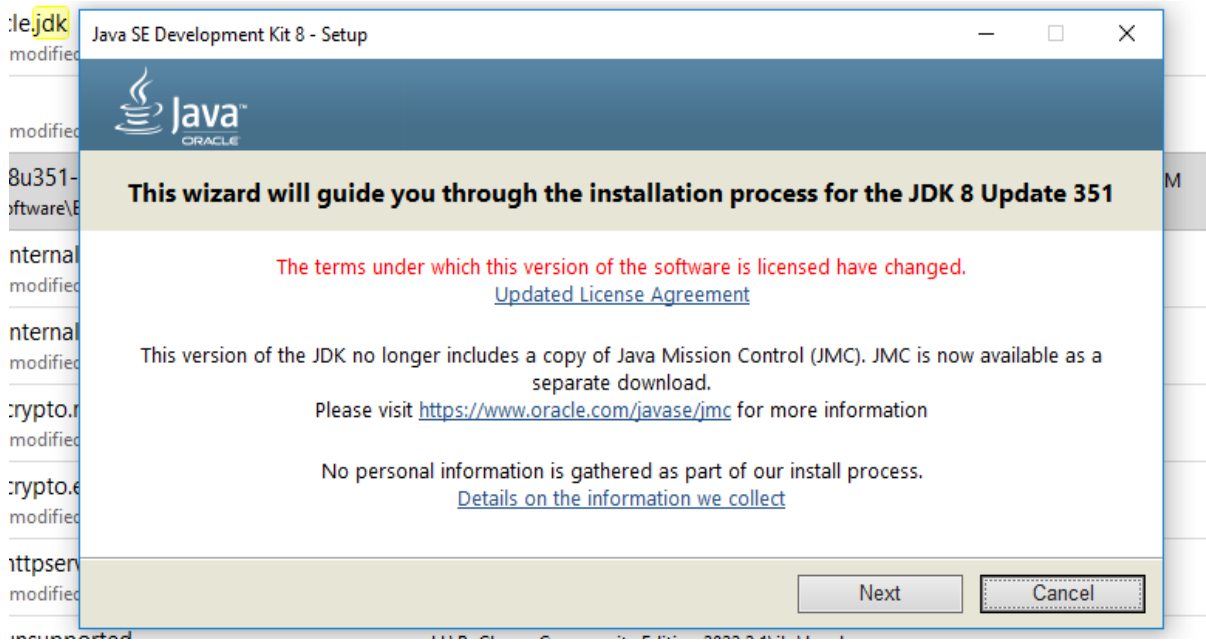
INDEX

SR	PRACTICAL NAME	REMAKS
1	Hadoop Installation	
2	Hdfs commands	
3	Mapreduce in hadoop-Word count program	
4	Mongo db Installation	
5	Mongo db operations	
6	Hive operations	
7	Analyse Data using Tableau	
8	Data visualization using Tableau	

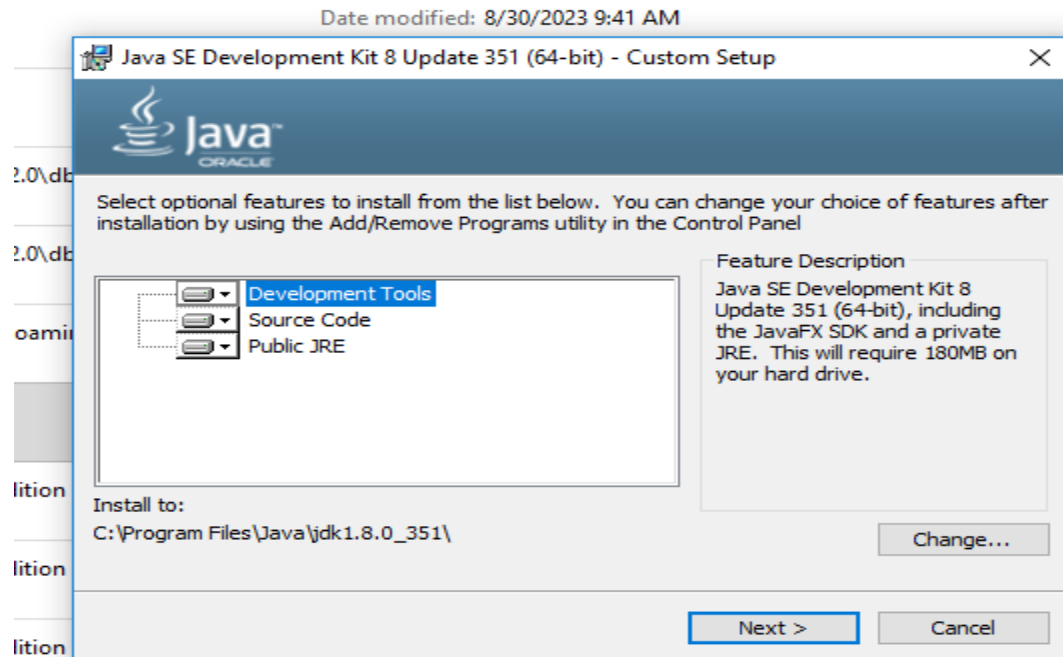
PRACTICAL NO. 1 : Hadoop Installation

Step 1 : first install the java(jdk)

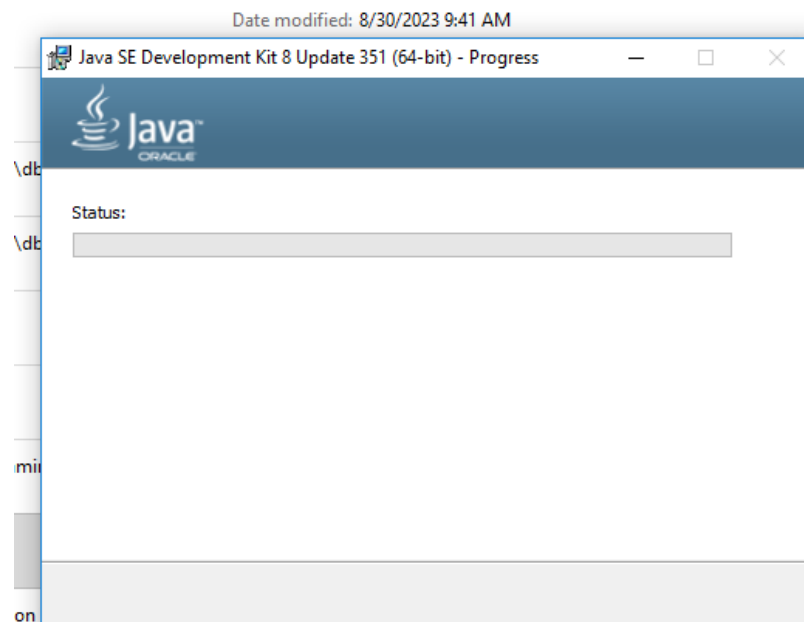
Here the step begin

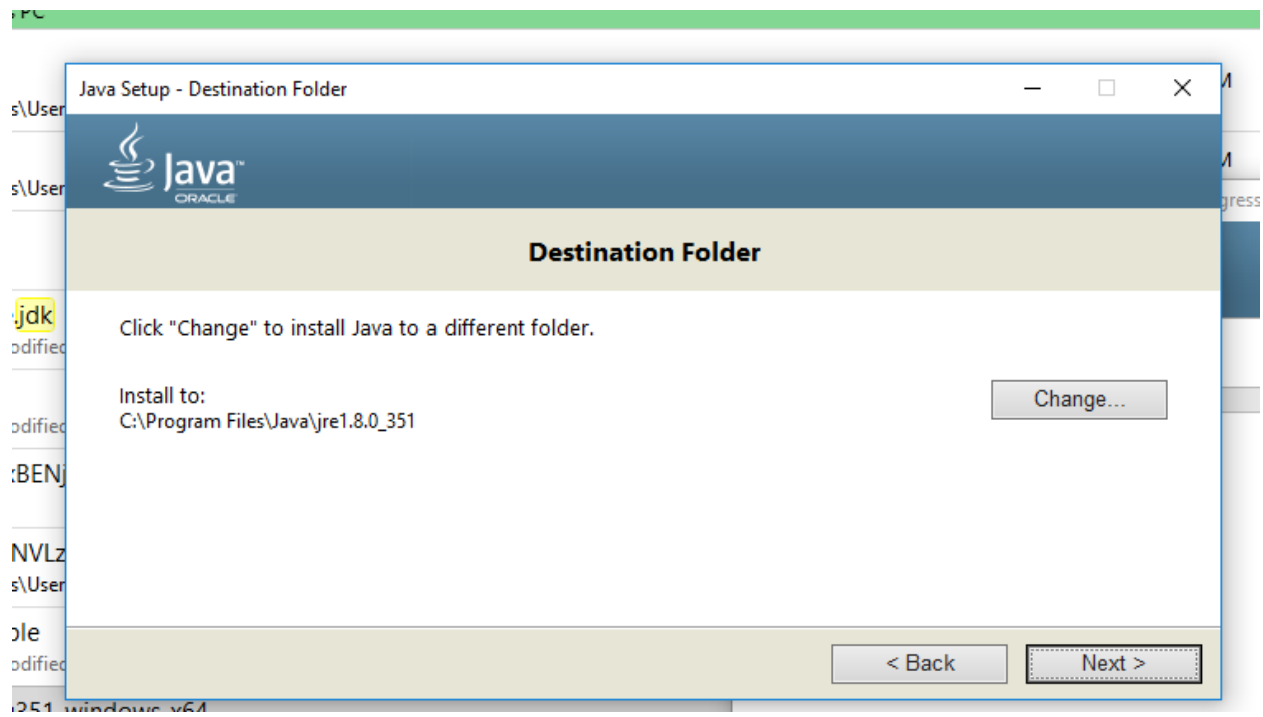


Then click on next button



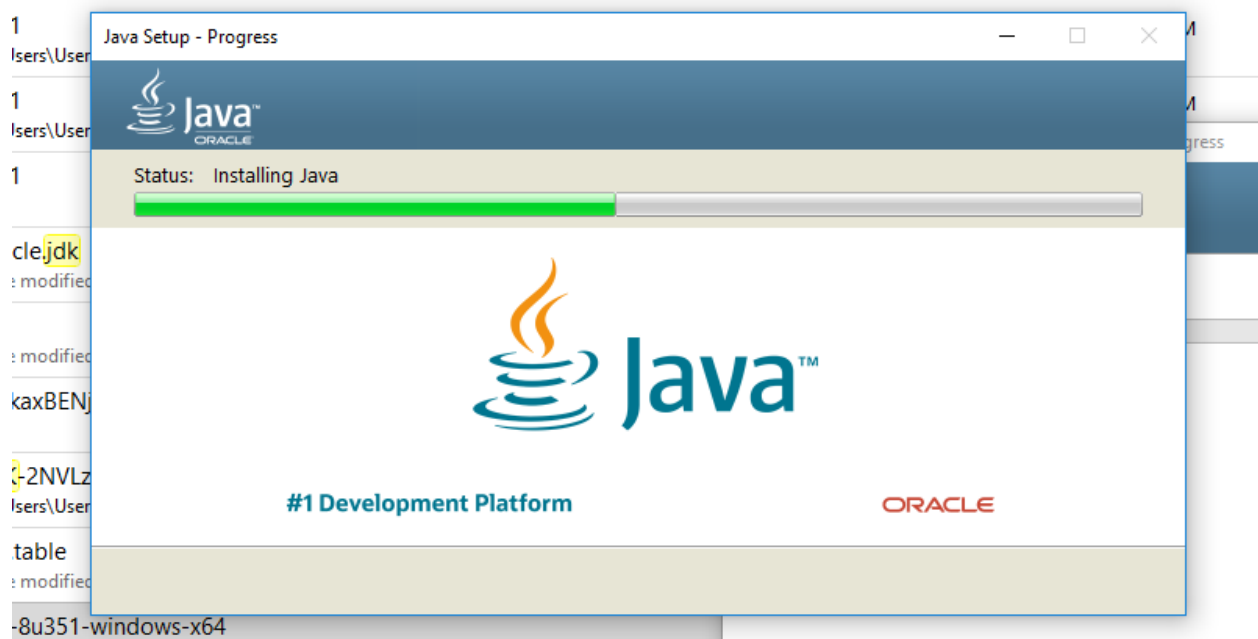
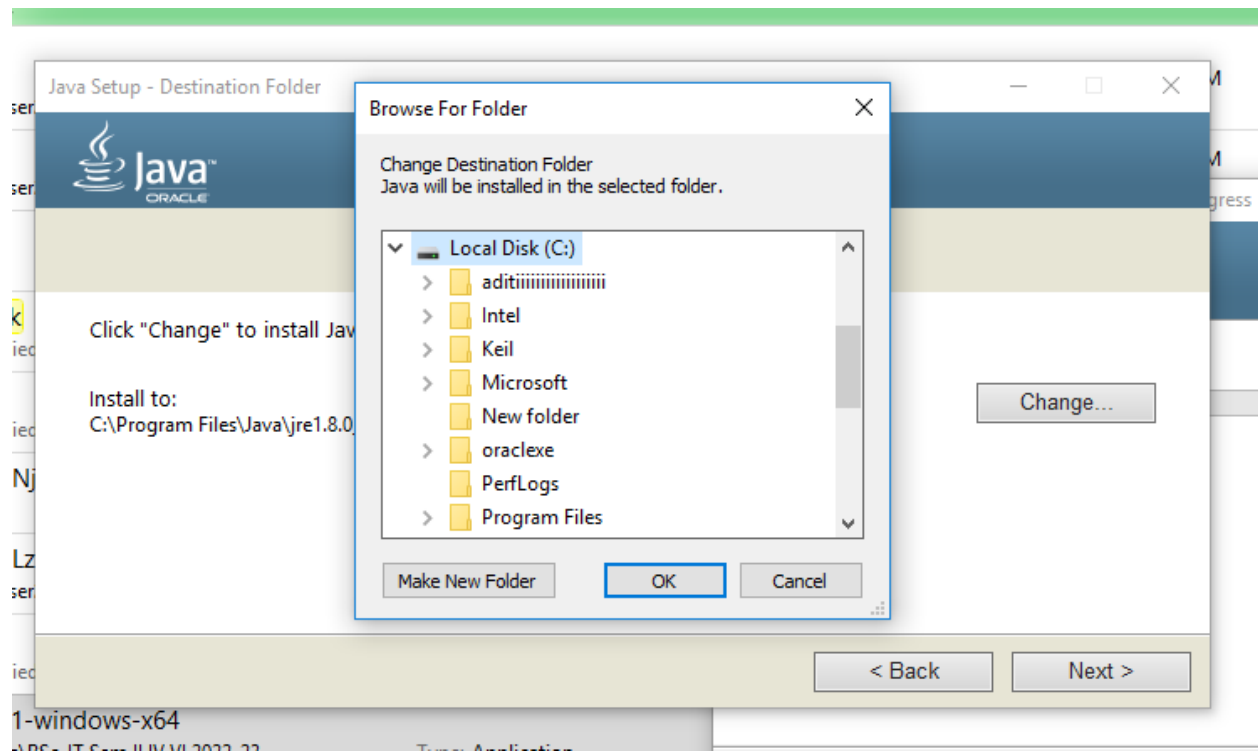
Click on Next button





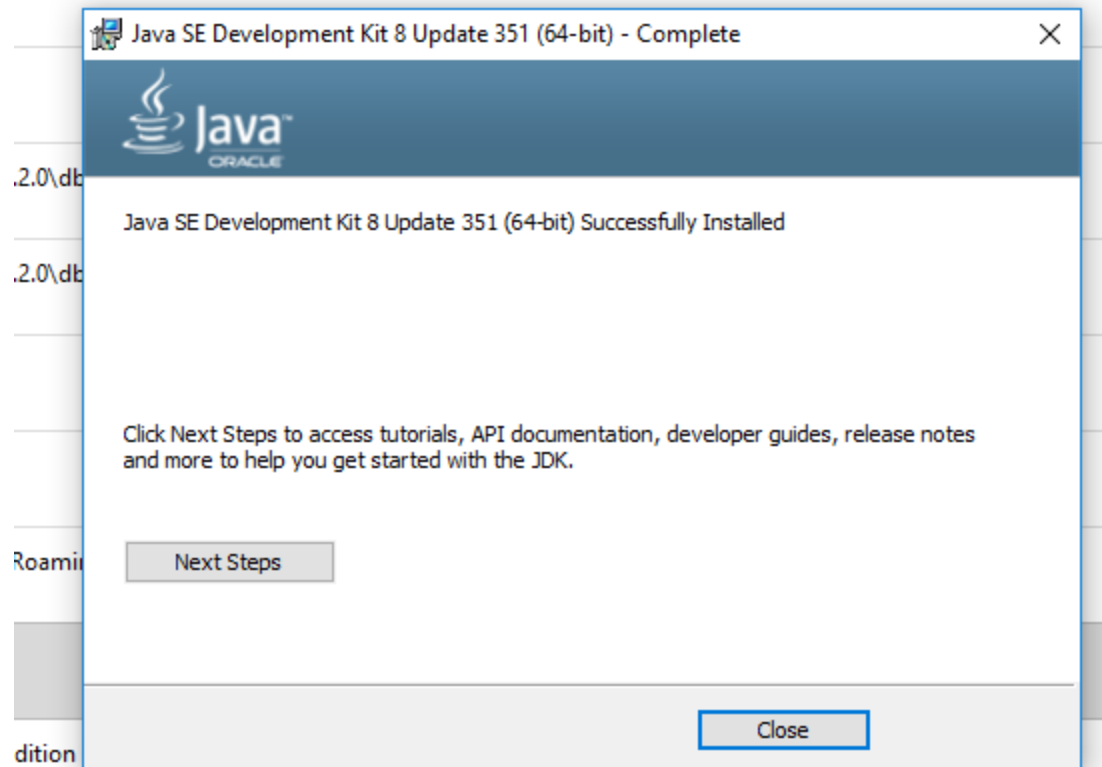
After that create a new folder

Name it java



Wait for installation

Date modified: 8/30/2023 9:41 AM



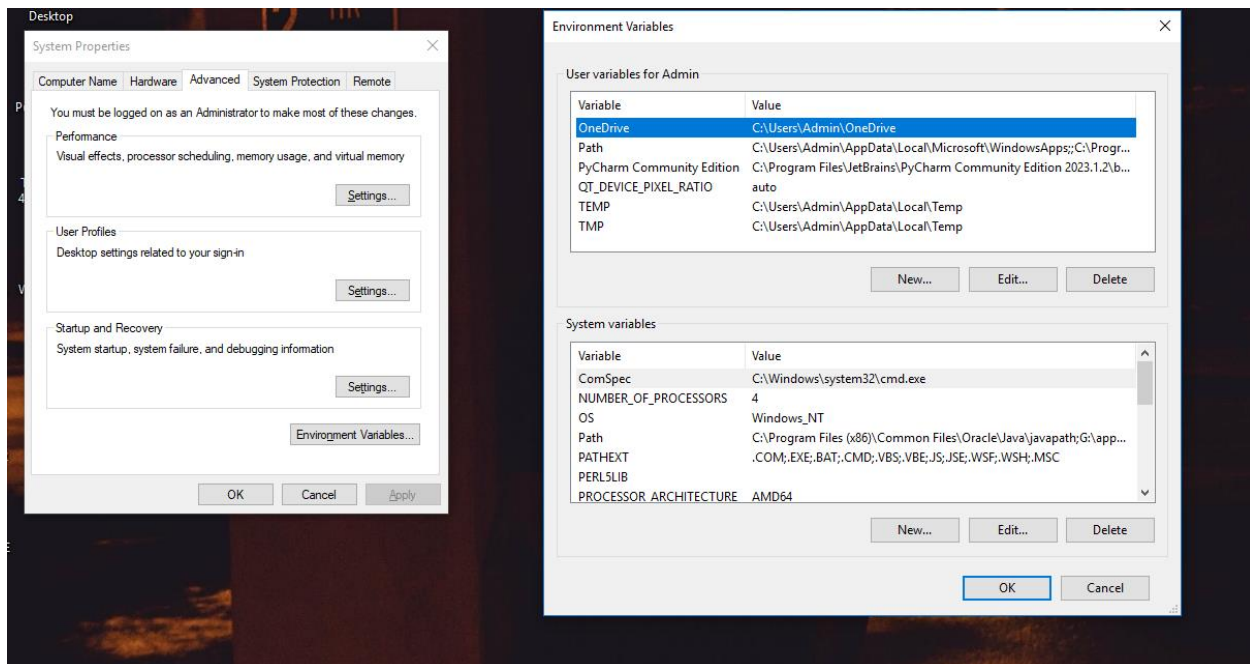
Set Environment variable for java

Go to

Edit the system environment variable

And

Click on edit environment variable

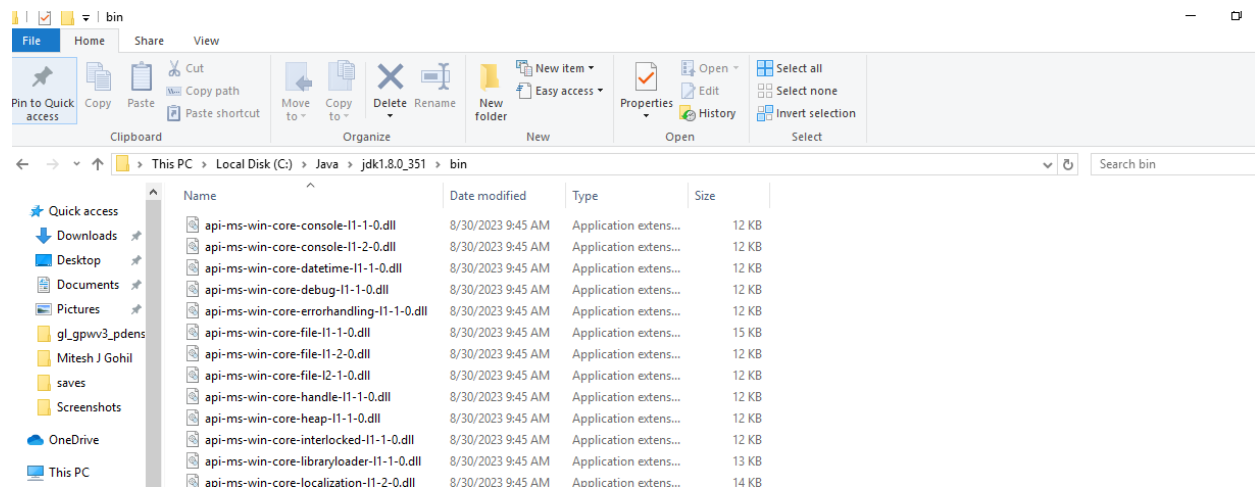


After that create a new environment

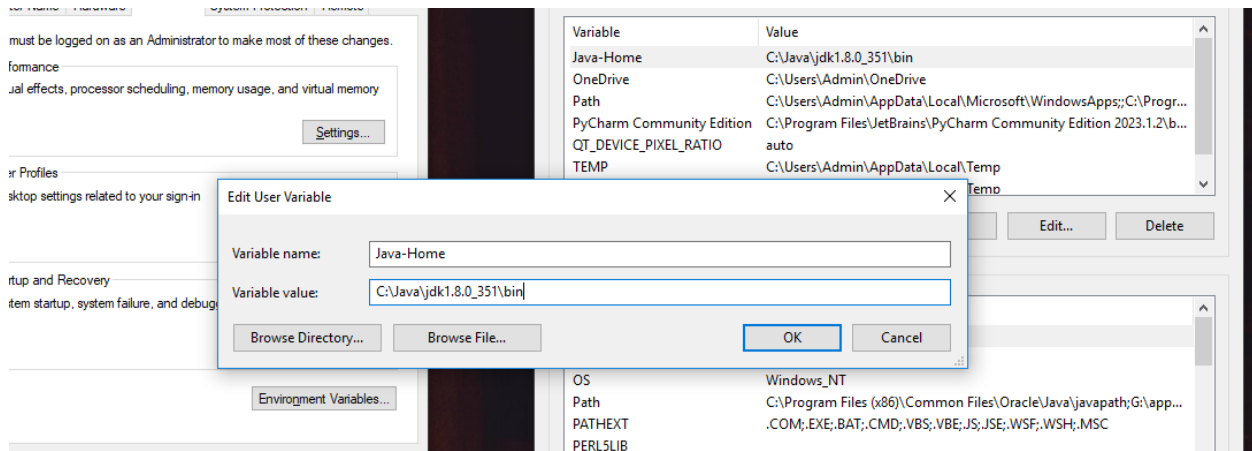
Click on new button and create it

And copy the path from the java folder which create on installation time

Copy the path from here



Paste in the Variable value

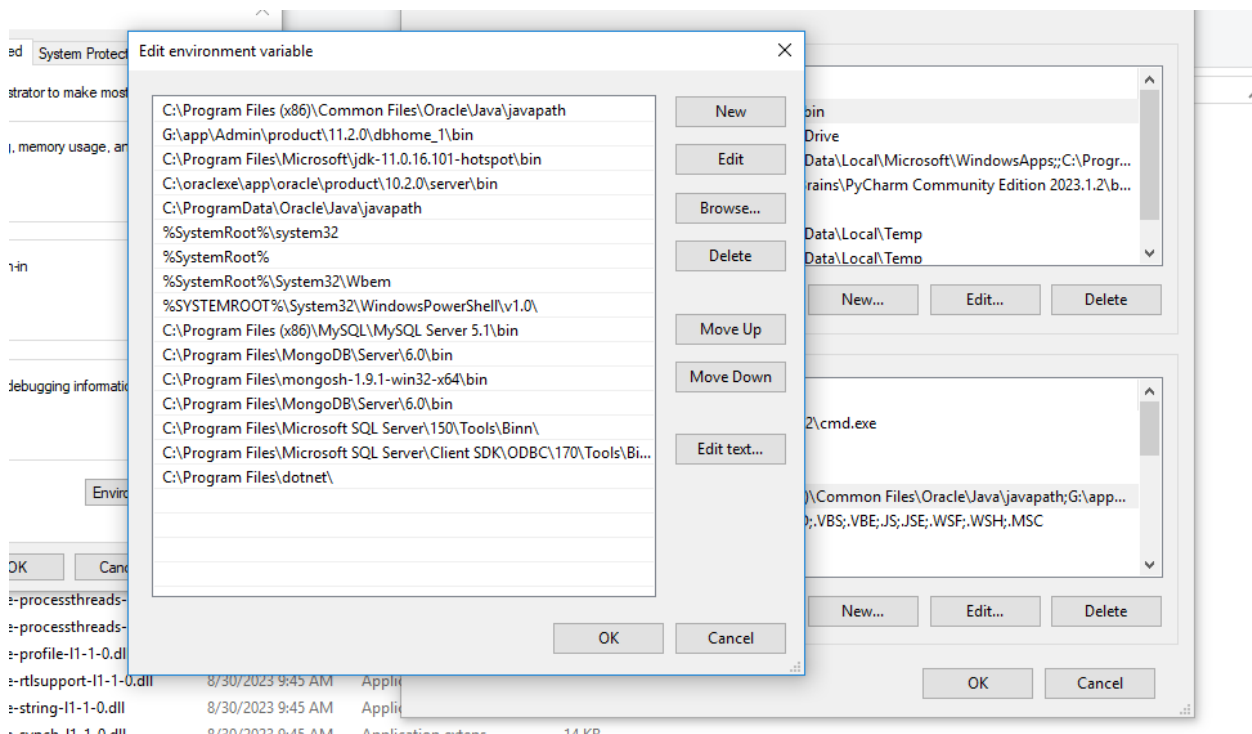


Then Click on ok

It will be created

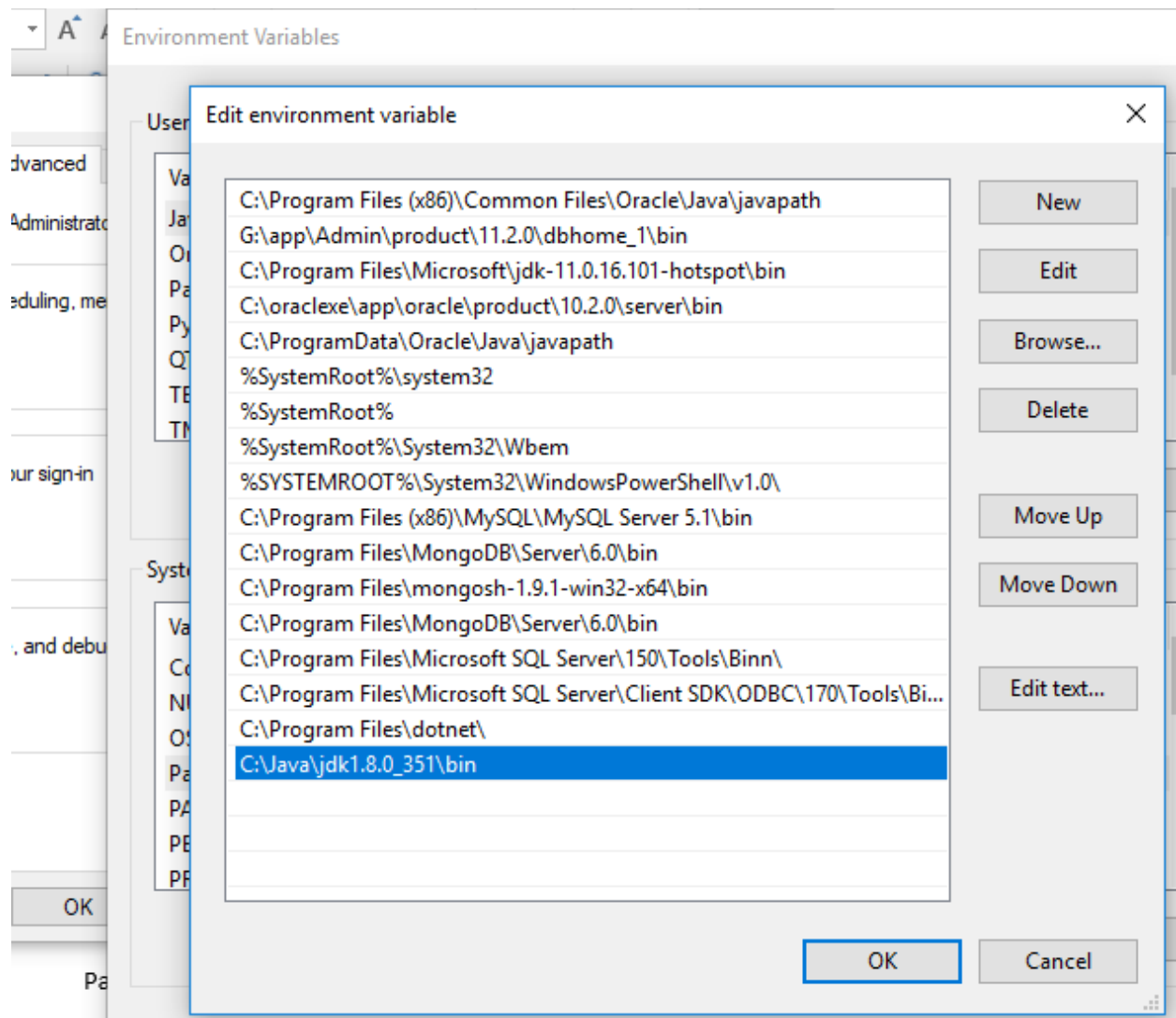
After the we have create a path in system variable

And create a path



Paste the of this path

"C:\Java\jdk1.8.0_351\bin"



Press the ok button

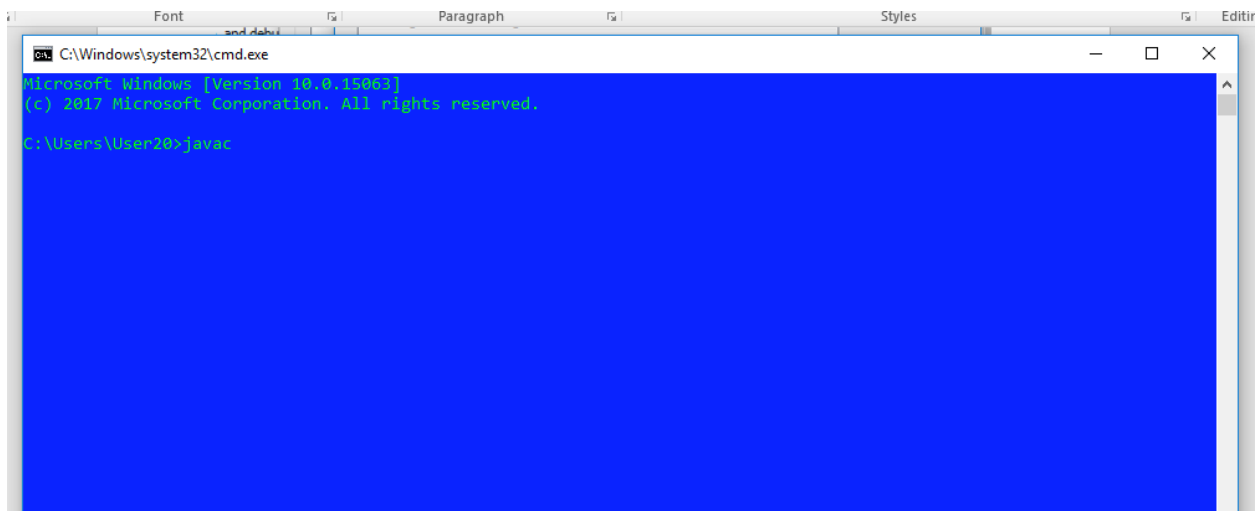
It will be created successfully

Here Java installation completed

Now we have whether java successfully or not

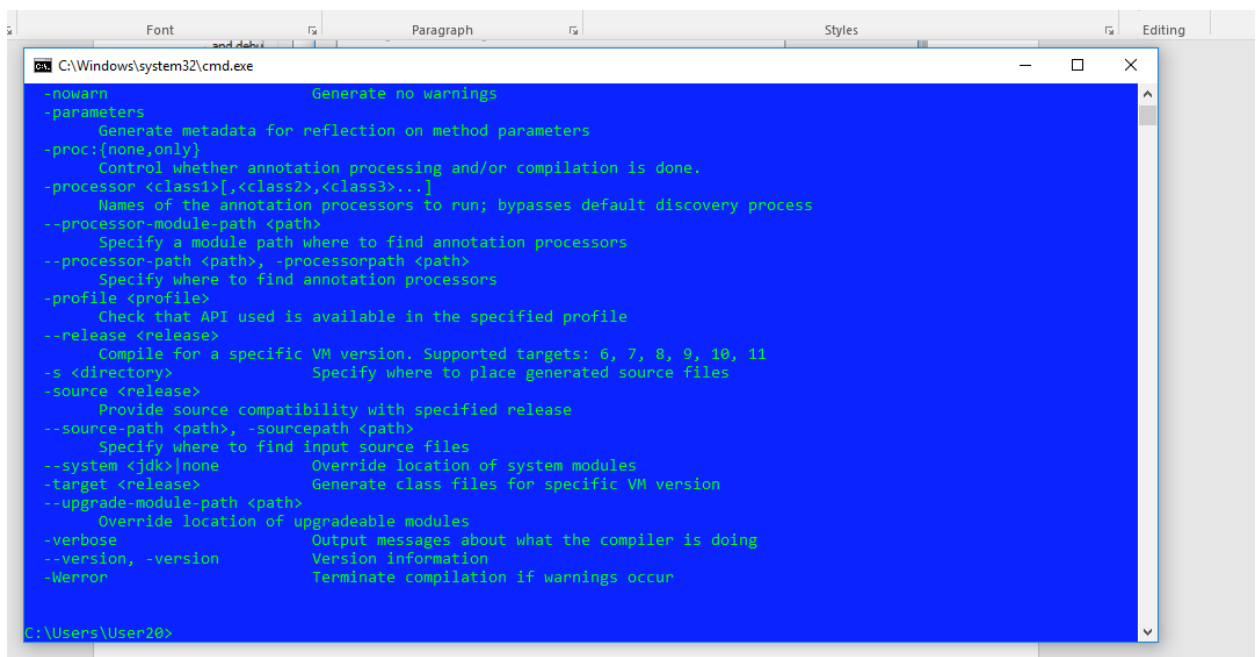
Go to command prompt

Type "javac" command



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 10.0.15063]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\User20>javac
```



```
C:\Windows\system32\cmd.exe
-nowrap          Generate no warnings
-parameters      Generate metadata for reflection on method parameters
-proc:{none,only} Control whether annotation processing and/or compilation is done.
-processor <class>[,<class>,<class>...]
                  Names of the annotation processors to run; bypasses default discovery process
--processor-module-path <path>
                  Specify a module path where to find annotation processors
--processor-path <path>, -processorpath <path>
                  Specify where to find annotation processors
-profile <profile>
                  Check that API used is available in the specified profile
--release <release>
                  Compile for a specific VM version. Supported targets: 6, 7, 8, 9, 10, 11
-s <directory>    Specify where to place generated source files
-source <release>
                  Provide source compatibility with specified release
--source-path <path>, -sourcepath <path>
                  Specify where to find input source files
--system <jdk>|none Override location of system modules
-target <release>   Generate class files for specific VM version
--upgrade-module-path <path>
                  Override location of upgradeable modules
-verbose          Output messages about what the compiler is doing
--version, -version
                  Version information
-Werror           Terminate compilation if warnings occur

C:\Users\User20>
```

Now it confirmed that it has install successfully

Now it confirmed that it has install successfully

Hadoop installation process

Download the Hadoop file form browser using internet

After download extract it and paste the file in c drive

This PC > Downloads >

Name	Date modified	Type	Size
hadoop-3.3.6.tar	8/30/2023 10:37 AM	WinRAR archive	712,996 KB
Application Report (2)	6/22/2023		
pexels-tnarg-2463584	8/29/2023		
bayes	8/29/2023		
pexels-photo-1097456	8/29/2023		
dataSet	8/29/2023		
hadoop-3.3.6	8/30/2023		

Extracting from hadoop-...

C:\Users\User20\Downloads\hadoop-3.3.6.tar.gz
extracting
package-summary.html 5%

Elapsed time 00:00:16
Time left 00:04:13

Processed 5%

Background Pause

Cancel Mode... Help

> This PC > Local Disk (C:)

Name	Date modified	Type	Size
aditi	8/11/2023 8:50 AM	File folder	
hadoop-3.3.6	8/30/2023 10:40 AM	File folder	
Intel	5/11/2023 1:21 PM	File folder	
Java	8/30/2023 9:57 AM	File folder	
Keil	5/27/2023 1:34 PM	File folder	
Microsoft	8/1/2023 3:13 PM	File folder	
New folder	8/22/2023 10:38 AM	File folder	
oraclexe	5/11/2023 1:22 PM	File folder	
PerfLogs	3/19/2017 2:33 AM	File folder	
Program Files	8/12/2023 9:26 AM	File folder	
Program Files (x86)	8/12/2023 9:26 AM	File folder	
TEMP	8/17/2023 9:14 AM	File folder	
TurboC4	5/11/2023 1:19 PM	File folder	
Users	5/14/2023 11:14 AM	File folder	
utk	5/27/2023 3:37 PM	File folder	
Windows	8/30/2023 9:32 AM	File folder	
install	6/30/2023 9:02 AM	Text Document	596 KB

Now go to Hadoop file

Inside Hadoop file go to the etc folder

Inside the etc folder edit five file

Which will be edit only in note++

Following are the file

1 core-site.xml

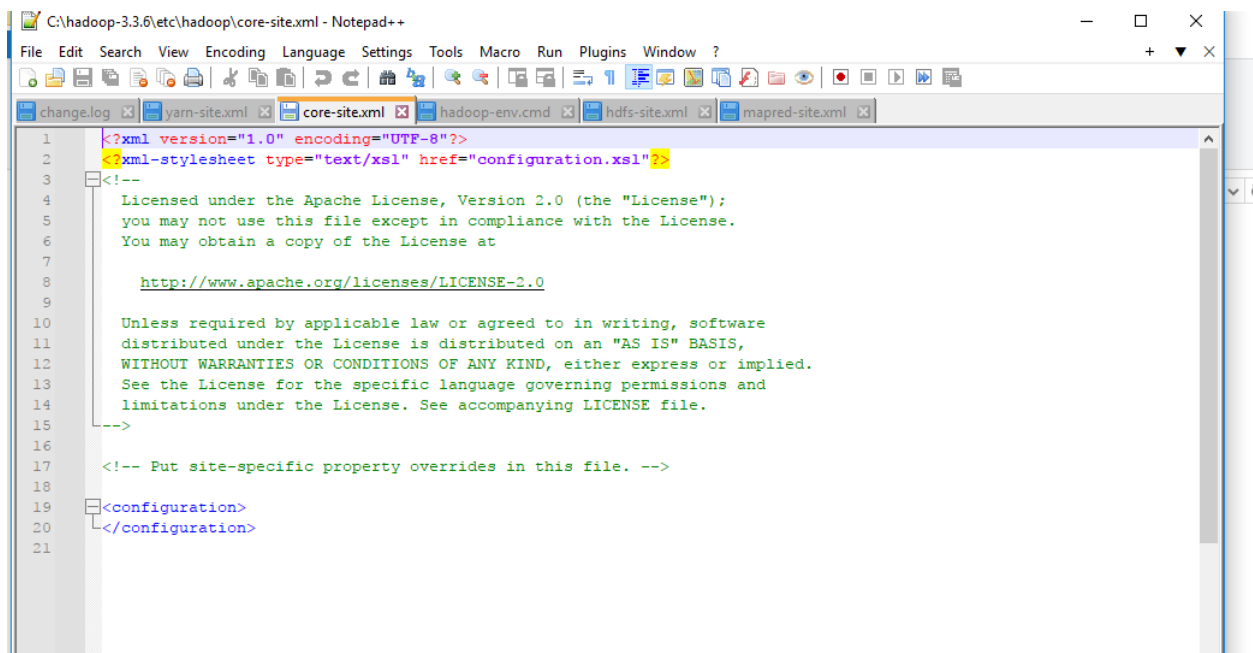
2 mapred-site.xml

3 yarn-site.xml

4 hdfs-site.xml

5 hadoop-env.cmd

Open this files in notepad++



```
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 </configuration>
21
```

All above file opened

Before that we have create folder "Data" inside the Hadoop folder

Inside the data folder we have to create two folder Namenode and datanode

After we go for edit the files

1 core-site.xml

Following are the code for configuration the core-site.xml file

Code:

```
<configuration>

<property>

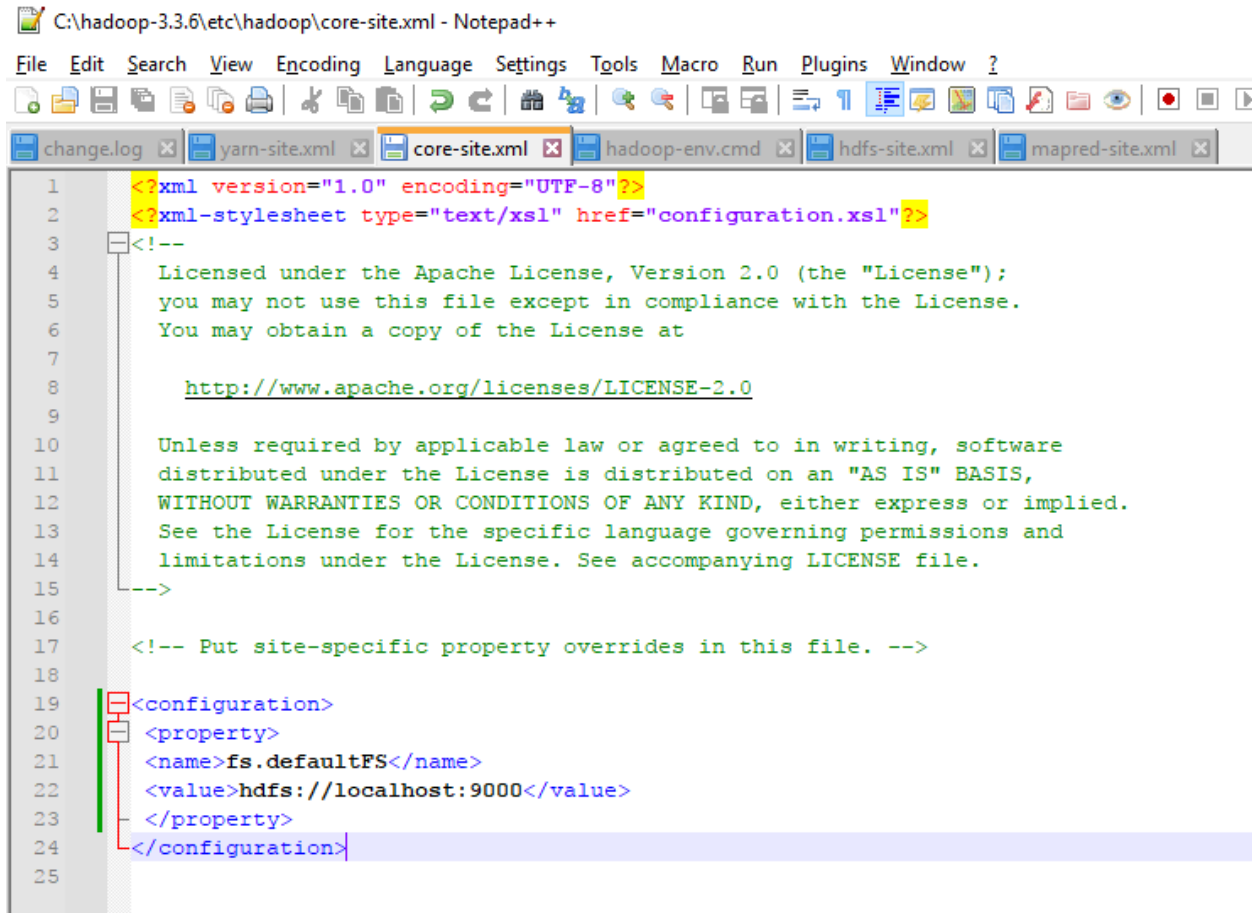
<name>fs.defaultFS</name>

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

</property>

</configuration>
```

After save it



2. mapred-site.xml

Following are the code for configuration the mapred-site.xml file

Code:

```
<configuration>

<property>

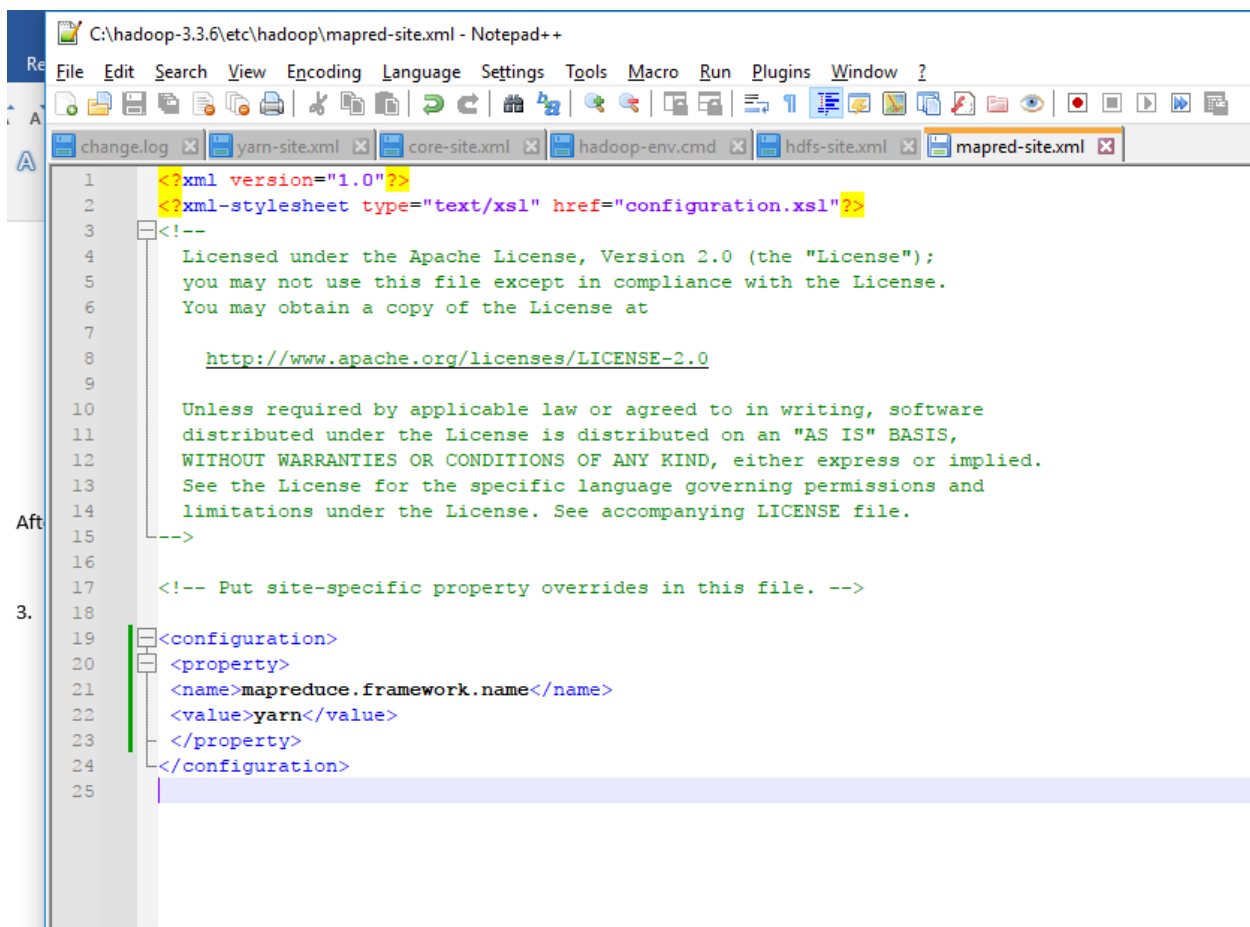
<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>
```

After write save it



```
C:\hadoop-3.3.6\etc\hadoop\mapred-site.xml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
change.log x yarn-site.xml x core-site.xml x hadoop-env.cmd x hdfs-site.xml x mapred-site.xml x
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
```

3. yarn-site.xml

Following are the code for configuration the yarn-site.xml file

Code:

```
<configuration>
```



```

<property>

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

<value>mapreduce_shuffle</value>

</property>

<property>

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

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

</property>

</configuration>

```

After write save it

```

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.auxservice.mapreduce.shuffle.class</name>
22 <value>org.apache.hadoop.mapred.ShuffleHandler</value>
23 </property>
24 </configuration>
25

```

4 hdfs-site.xml

Following are the code for configuration the hdfs-site.xml file

Code:

```
<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>C:\hadoop-3.3.6\Data\namenode</value>

</property>

<property>

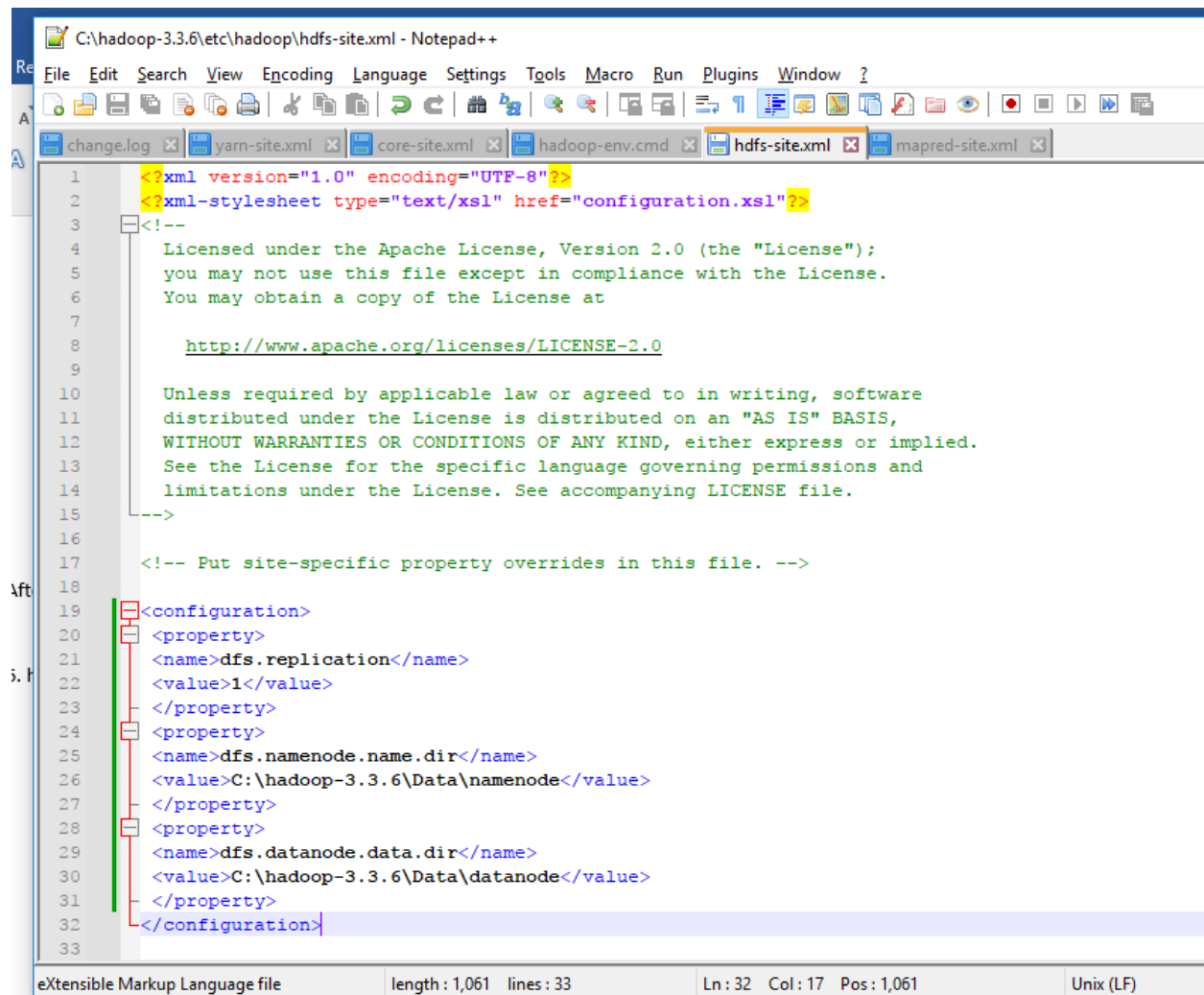
<name>dfs.datanode.data.dir</name>

<value>C:\hadoop-3.3.6\Data\datanode</value>

</property>

</configuration>
```

After write save it



```
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:\hadoop-3.3.6\Data\namenode</value>
27 </property>
28 <property>
29 <name>dfs.datanode.data.dir</name>
30 <value>C:\hadoop-3.3.6\Data\datanode</value>
31 </property>
32 </configuration>
33
```

Extensible Markup Language file length : 1,061 lines : 33 Ln : 32 Col : 17 Pos : 1,061 Unix (LF)

5. hadoop-env.cmd

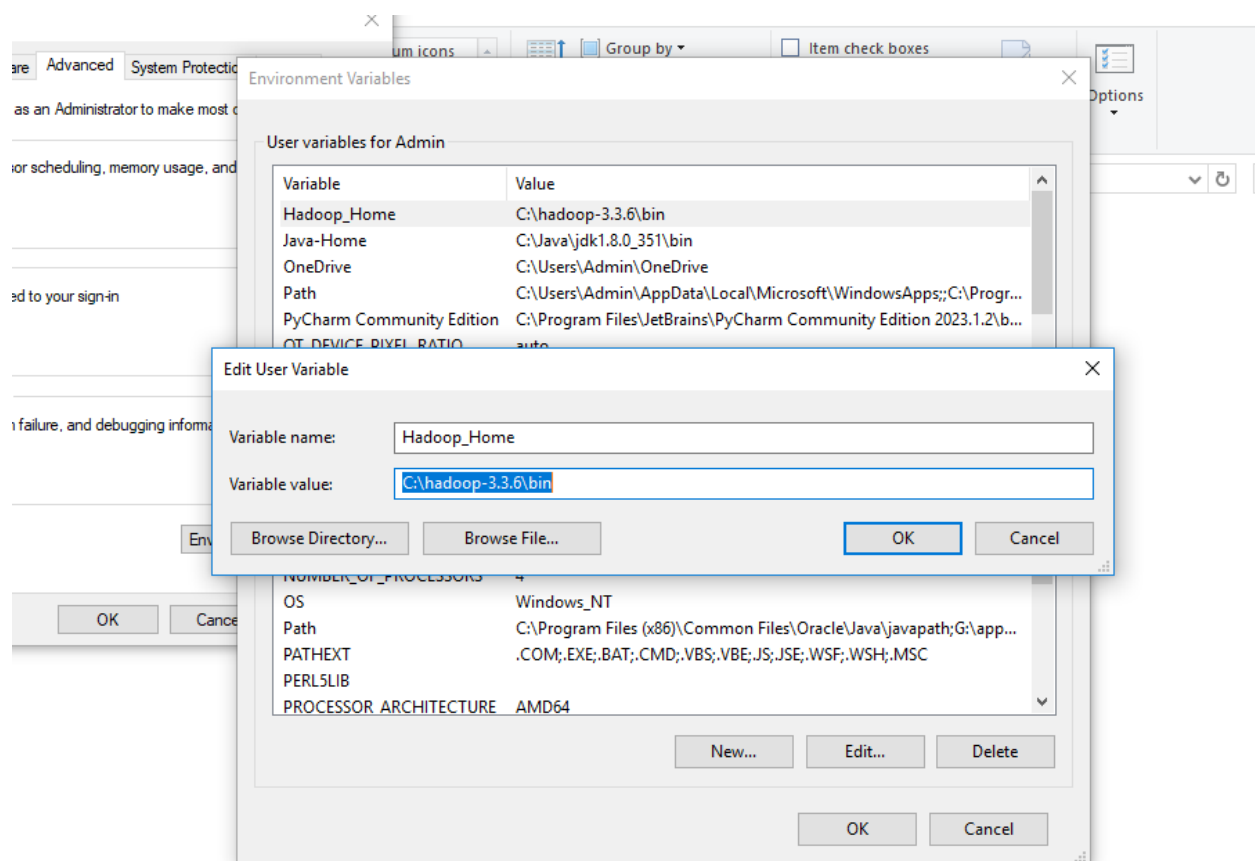
Following are the code for configuration the Hadoop-env-cmd file

Here we java environment variable path

The path is "C:\Java\jdk1.8.0_351"

```
*C:\hadoop-3.3.6\etc\hadoop\hadoop-env.cmd - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
change.log x yarn-site.xml x core-site.xml x hadoop-env.cmd x hdfs-site.xml x mapred-site.xml x
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\jdk1.8.0_351
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=
```

After that we have to create a Hadoop new environment variable and then give the path here as well



After edit save it exit from here

Use this Hadoop command “hdfs namenode -format”

```
C:\Windows\system32\cmd.exe
Microsoft Windows [version 10.0.15003]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\User2B>hdfs namenode -format
2023-09-20 12:41:56,328 INFO namenode.NameNode: STARTUP_MSG:
*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = DESKTOP-POTPUHS/192.168.60.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.3.6
STARTUP_MSG: classpath = C:\hadoop-3.3.6\etc\hadoop\C:\hadoop-3.3.6\share\hadoop\common\C:\hadoop-3.3.6\share\hadoop\common\lib\animal-sniffer-annotations-
1.17.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\audience-annotations-0.5.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\avro-1.7.7.jar;C:\hadoop-3.3.6\share\
hadoop\common\lib\checker-qual-2.5.2.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-beanutils-1.9.4.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-
cli-1.2.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-coder-1.45.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\hadoop-3.
3.6\share\hadoop\common\lib\commons-compress-1.21.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-configuration-2.8.0.jar;C:\hadoop-3.3.6\share\hadoop\com
mon\lib\commons-daemon-1.0.13.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-io-2.8.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-lang-3.12.0
.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-logging-1.1.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-math3-3.1.1.jar;C:\hadoop-3.3.6\share\
hadoop\common\lib\commons-net-3.9.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\commons-text-1.10.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\curator-client-
5.2.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\curator-framework-5.2.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\curator-p2s-5.2.0.jar;C:\hadoop-
3.3.6\share\hadoop\common\lib\hdinsight-java-2.3.2.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\h4x-hive-serdes-1.9.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\isozn-2.3
9.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\guava-27.0-jre.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-annotations-3.3.6.jar;C:\hadoop-3.3.6\share\
hadoop\common\lib\hadoop-auth-3.3.6.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-guava-k12-1.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-mapred-
shared-profile-3.7.1-1.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-remote-4.5.12.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-secure-4.4.1.jar;C:\ha
dop-3.3.6\share\hadoop\common\lib\hadoop2lib-annotations-1.1.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-jackson-annotations-2.12.7.jar;C:\hadoop-3.3.6\share\ha
dop\common\lib\hadoop-jackson-core-2.12.7.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-jackson-core-asl-4.0.13.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-jackso
n-stax2-12.7.4.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-mapper-asl-1.9.13.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\jakarta-activation-api-1.4
.1.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-jaxrs-servlet-api-3.1.0.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-jaxrs-api-2.2.11.jar;C:\hadoop-3.3.6\share\
hadoop\common\lib\hadoop-jaxrs-impl-2.2.11.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-ipc-annotations-1.0-1.jar;C:\hadoop-3.3.6\share\hadoop\common\lib\jersey-con
```

```
C:\Windows\system32\cmd.exe
2023-08-30 12:41:57.703 INFO util.GSet: Capacity = 2^15 = 32768 entries
2023-08-30 12:42:02.313 INFO namenode.FSImage: Allocated new BlockPoolID: BP-2123295340-192.168.60.1-1603379522235
2023-08-30 12:42:02.392 INFO common.Storage: Storage directory C:\hadoop\3.3.0\data\namenode has been successfully formatted.
2023-08-30 12:42:02.407 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop\3.3.0\data\namenode\current\fsimage.chkpt.000000000000000000 using n
+ compression
2023-08-30 12:42:02.501 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop\3.3.0\data\namenode\current\fsimage.chkpt.00000000000000000000 of size 401 by
tes saved in 6 seconds
2023-08-30 12:42:02.563 INFO namenode.NMStorageRetentionManager: Going to retain 1 images with txid >= 0
2023-08-30 12:42:02.610 INFO namenode.FSNameSystem: Stopping services started for active state
2023-08-30 12:42:02.610 INFO namenode.FSNameSystem: Stopping services started for standby state
2023-08-30 12:42:02.626 INFO namenode.FSImageSaver: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2023-08-30 12:42:02.626 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-POTRUBS/192.168.60.1
*****/
C:\Users\User20>
```

Inside the Hadoop folder go to the sbin folder after the excute the hdfs commands

- 1)start-dfs
- 2)start-yarn.cmd
- 3)jps

C:\Windows\system32\cmd.exe

```
C:\>dir
Volume in drive C has no label.
Volume Serial Number is 3CCD-5216

Directory of C:\

08/11/2023  08:58 AM    <DIR>          aditiiiiiiiiiiiiiiii
08/30/2023  12:41 PM    <DIR>          hadoop-3.3.6
06/30/2023  09:02 AM        609,794  install.log
05/11/2023  01:21 PM    <DIR>          Intel
08/30/2023  09:57 AM    <DIR>          Java
05/27/2023  01:34 PM    <DIR>          Keil
08/01/2023  03:13 PM    <DIR>          Microsoft
08/12/2023  10:38 AM    <DIR>          New folder
05/11/2023  01:22 PM    <DIR>          oraclexe
03/19/2017  02:33 AM    <DIR>          Perflogs
08/30/2023  10:50 AM    <DIR>          Program Files
08/12/2023  09:26 AM    <DIR>          Program Files (x86)
08/17/2023  09:14 AM    <DIR>          TEMP
05/11/2023  01:19 PM    <DIR>          TurboCd
05/14/2023  11:14 AM    <DIR>          Users
05/27/2023  03:37 PM    <DIR>          utk
08/30/2023  09:32 AM    <DIR>          Windows
               1 File(s)          609,794 bytes
               16 Dir(s)  23,546,871,008 bytes free
```

```
C:\>cd hadoop-3.3.6
```

```
C:\hadoop-3.3.6>cd sbin
```

```
C:\hadoop-3.3.6\sbin>start-dfs
```

```
C:\hadoop-3.3.6\sbin>start-yarn.cmd
starting yarn daemons
```

```
C:\hadoop-3.3.6\sbin>jps
3440 ResourceManager
10216 NodeManager
6536 Jps
9384 NameNode
2012 DataNode
```

```
C:\hadoop-3.3.6\sbin>
```

Activat
Go to Set



PRACTICAL NO.2 : Hdfs commands

Starting HDFS

Format the configured HDFS file system and then open the namenode(HDFS server) and execute the following command.

```
$ hadoop namenode -format
```

Start the distributed file system and follow the command listed below to start the namenode as well as the data nodes in cluster.

```
$ start-dfs.sh
```

Read & Write Operations in HDFS

You can execute almost all operations on Hadoop Distributed File Systems that can be executed on the local file system. You can execute various reading, writing operations such as creating a directory, providing permissions, copying files, updating files, deleting, etc. You can add access rights and browse the file system to get the cluster information like the number of dead nodes, live nodes, spaces used, etc.

HDFS Operations to Read the file

To read any file from the HDFS, you have to interact with the NameNode as it stores the metadata about the DataNodes. The user gets a token from the NameNode and that specifies the address where the data is stored.

You can put a read request to NameNode for a particular block location through distributed file systems. The NameNode will then check your privilege to access the DataNode and allows you to read the address block if the access is valid.

```
$ hadoop fs -cat <file>
```

HDFS Operations to write in file

Similar to the read operation, the HDFS Write operation is used to write the file on a particular address through the NameNode. This NameNode provides the slave address where the client/user can write or add data. After writing on the block location, the slave replicates that block and copies to another slave location using the factor 3 replication. The slave is then reverted back to the client for authentication.

The process for accessing a NameNode is pretty similar to that of a reading operation. Below is the HDFS write commence:

```
bin/hdfs dfs -ls <path>
```

Listing Files in HDFS

Finding the list of files in a directory and the status of a file using 'ls' command in the terminal. Syntax of ls can be passed to a directory or a filename as an argument which are displayed as

follows:

```
$ $HADOOP_HOME/bin/hadoop fs -ls <args>
```

Inserting Data into HDFS

Below mentioned steps are followed to insert the required file in the Hadoopfile system.

Step1: Create an input directory

```
$ $HADOOP_HOME/bin/hadoop fs -mkdir /user/input
```

Step2: Use the put command transfer and store the data file from the localsystems to the HDFS using the following commands in the terminal.

```
$ $HADOOP_HOME/bin/hadoop fs -put /home/intellipaat.txt /user/input
```

Step3: Verify the file using ls command.

```
$ $HADOOP_HOME/bin/hadoop fs -ls /user/input
```

Retrieving Data from HDFS

For instance, if you have a file in HDFS called Intellipaat. Then retrieve therequired file from the Hadoop file system by carrying out:

Step1: View the data from HDFS using the cat command.

```
$ $HADOOP_HOME/bin/hadoop fs -cat /user/output/intellipaat
```

Step2: Gets the file from HDFS to the local file system using get commandas shown below

```
$ $HADOOP_HOME/bin/hadoop fs -get /user/output/ /home/hadoop_tp/
```

Shutting Down the HDFS

Shut down the HDFS files by following the below command

```
$ stop-dfs.sh
```

Multi-Node Cluster

[Installing Java](#)

Syntax of java version command

```
99$ java -version
```

Following output is presented.

```
java version "1.7.0_71"
```


Java(TM) SE Runtime Environment (build 1.7.0_71-b13) Java

HotSpot(TM) Client VM (build 25.0-b02, mixed mode)**Bottom of Form**

Creating User Account

System user account is used on both master and slave systems for the [Hadoop installation](#).

```
# useradd hadoop# passwd
```

```
hadoop
```

Mapping the nodes

Hosts files should be edited in /etc/ folder on each and every nodes and IP address of each system followed by their host names must be specified mandatorily.

```
# vi /etc/hosts
```

Enter the following lines in the /etc/hosts file.

```
192.168.1.109 hadoop-master
```

```
192.168.1.145 hadoop-slave-1
```

```
192.168.56.1 hadoop-slave-2
```

Configuring Key Based Login

Ssh should be set up in each node so they can easily converse with one another without any prompt for a password.

```
# su hadoop
```

```
$ ssh-keygen -t rsa
```

```
$ ssh-copy-id -i ~/.ssh/id_rsa.pub tutorialspoint@hadoop-master
```

```
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop_tp1@hadoop-slave-1
```

```
$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop_tp2@hadoop-slave-2
```

```
$ chmod 0600 ~/.ssh/authorized_keys
```

```
$ exit
```

PRACTICAL NO.3 : Mapreduce in hadoop-Word count program

```
public class WordCount {
public static void main(String [] args) throws Exception
{
Configuration c=new Configuration();
String[] files=new
GenericOptionsParser(c,args).getRemainingArgs();Path
input=new Path(files[0]);
Path output=new Path(files[1]); Job
j=new Job(c,"wordcount");
j.setJarByClass(WordCount.class);
j.setMapperClass(MapForWordCount.class);
j.setReducerClass(ReduceForWordCount.class);
j.setOutputKeyClass(Text.class);
j.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(j, input);
FileOutputFormat.setOutputPath(j, output);
System.exit(j.waitForCompletion(true)?0:1);
}

public static class MapForWordCount extends
Mapper<LongWritable, Text, Text, IntWritable>{

public void map(LongWritable key, Text value, Context con) throws
IOException, InterruptedException
{
String line = value.toString();
String[] words=line.split(",");
for(String word: words )
{
Text outputKey = new Text(word.toUpperCase().trim());IntWritable
outputValue = new IntWritable(1); con.write(outputKey, outputValue);
}
}
}
```

```

public static class ReduceForWordCount extends Reducer<Text,IntWritable, Text,
IntWritable{

void reduce(Text word, Iterable<IntWritable> values,Context con) throws
IOException, InterruptedException

{
int sum = 0;
for(IntWritable value : values)
{
sum += value.get();
}
con.write(word, new IntWritable(sum));
}
}
}

```

The above program consists of three classes:

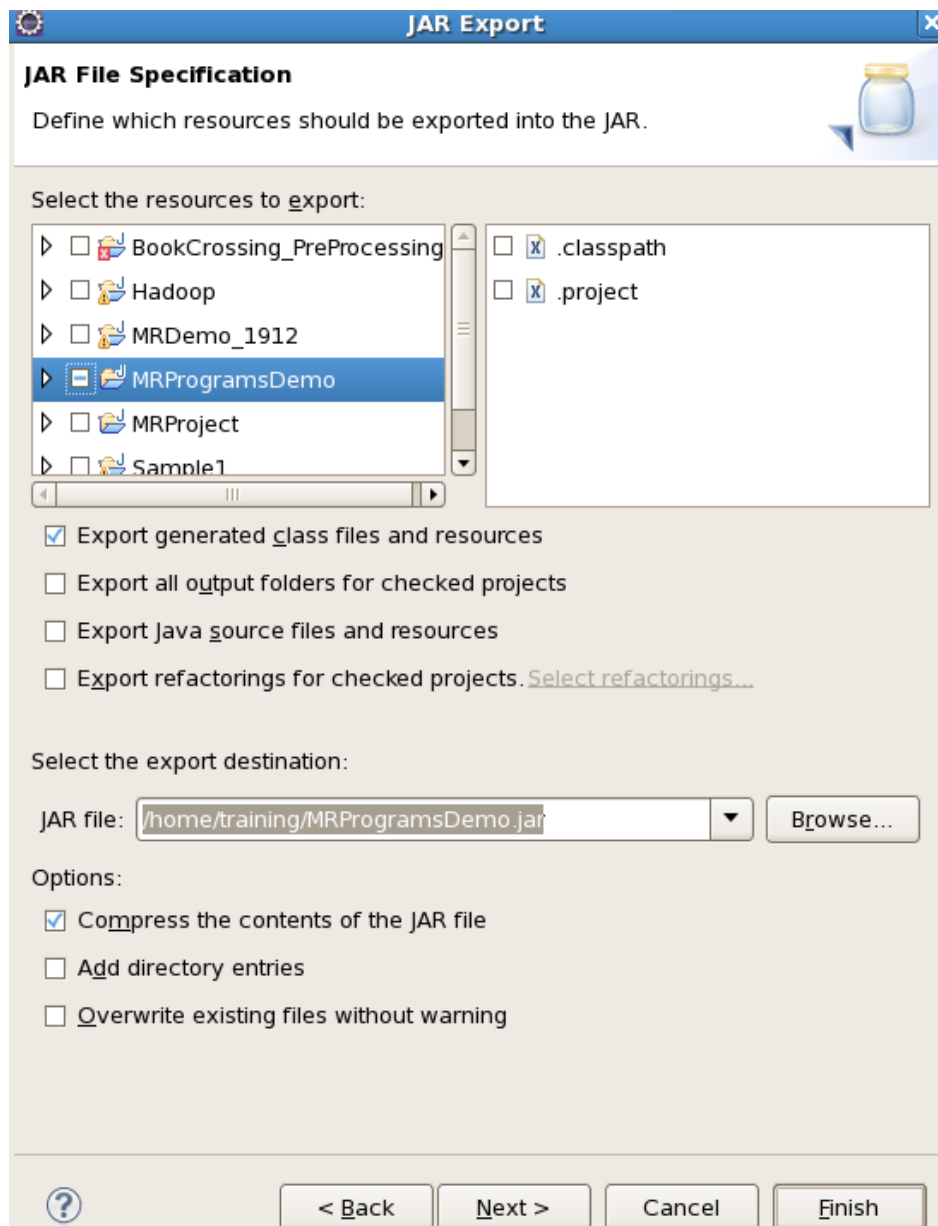
Driver class (Public, void, static, or main; this is the entrypoint).

The Map class which **extends** the public class
Mapper<KEYIN,VALUEIN,KEYOUT,VALUEOUT> and implements the Map function.

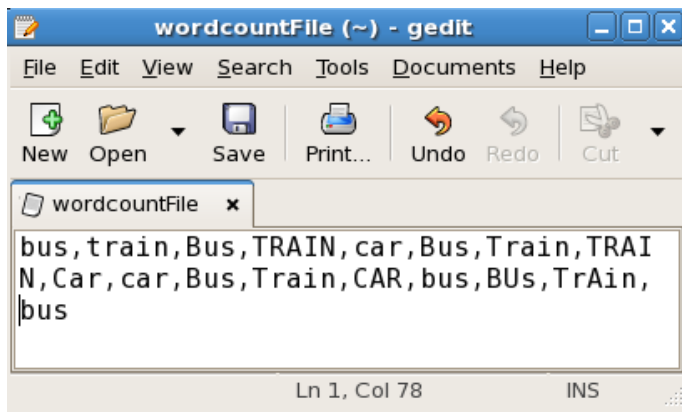
The Reduce class which extends the public class
Reducer<KEYIN,VALUEIN,KEYOUT,VALUEOUT> and implements the Reduce function.

Make a jar file

Right Click on Project> Export> Select export destination as **JarFile** > next> Finish



Take a text file and move it into HDFS format:



To move this into Hadoop directly, open the terminal and enter the following commands:

9. Open the result:

```
[training@localhost ~]$ hadoop fs -ls MRDir1
```

Practical No 4 : Mongo db Installation

Perform the following practical in mongod

- Create Database
- Create Collection
- Drop collection
- Insert Data And Insert many Data
- Show Collection
- Update Document

```
amol> use College
switched to db College
College> db.createCollection('Students');
{ ok: 1 }
College> db.Students.insert({'Name':'Prit','Age':20,'RollNo':'10'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId("650295303c1c9a03f6b797a3") }
}
College> db.Students.insert({'Name':'Shyam','Age':20,'RollNo':'11'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId("6502954f3c1c9a03f6b797a4") }
}
College> db.Students.insert({'Name':'Krushna','Age':20,'RollNo':'6'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId("650295633c1c9a03f6b797a5") }
}
College> db.Students.insert({'Name':'Karan','Age':20,'RollNo':'7'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId("650295703c1c9a03f6b797a6") }
}
```

```
College> db.Students.find()
[
  {
    _id: ObjectId("650295303c1c9a03f6b797a3"),
    Name: 'Prit',
    Age: 20,
    RollNo: '10'
  },
  {
    _id: ObjectId("6502954f3c1c9a03f6b797a4"),
    Name: 'Shyam',
    Age: 20,
    RollNo: '11'
  },
  {
    _id: ObjectId("650295633c1c9a03f6b797a5"),
    Name: 'Krushna',
    Age: 20,
    RollNo: '6'
  },
]
```

```

College> db.Students.insertMany([{'Name':'Tushar','Age':20,'RollNo':'5'},{'Name':'Pallavi','Age':25,'RollNo':'1'},{'Name':'Aditya','Age':20,'RollNo':'2'},{'Name':'Prem','Age':20,'RollNo':'9'}])
{
  acknowledged: true,
  insertedIds: {
    '0': ObjectId("650296983c1c9a03f6b797a7"),
    '1': ObjectId("650296983c1c9a03f6b797a8"),
    '2': ObjectId("650296983c1c9a03f6b797a9"),
    '3': ObjectId("650296983c1c9a03f6b797aa")
  }
}

```

```

College> db.Students.find()
[
  {
    _id: ObjectId("650295303c1c9a03f6b797a3"),
    Name: 'Prit',
    Age: 20,
    RollNo: '10'
  },
  {
    _id: ObjectId("6502954f3c1c9a03f6b797a4"),
    Name: 'Shyam',
    Age: 20,
    RollNo: '11'
  },
  {
    _id: ObjectId("650295633c1c9a03f6b797a5"),
    Name: 'Krushna',
    Age: 20,
    RollNo: '6'
  },
  {
    _id: ObjectId("650295703c1c9a03f6b797a6"),
    Name: 'Karan',
    Age: 20,
    RollNo: '7'
  },
  {
    _id: ObjectId("650296983c1c9a03f6b797a7"),
    Name: 'Tushar',
    Age: 20,

```

```

TypeError: db.Students.update is not a function
College> db.Students.update({'Age':25},{ $set:{'Age':20}})
DeprecationWarning: Collection.update() is deprecated. Use updateOne()
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
College>

```

PRACTICAL NO. 5 : Mongo db operations

To build a database in MongoDB, first construct a MongoClient object, and then supply a connection URL with the right IP address and the database name. If the database does not already exist, MongoDB will create it and connect to it.

Create a database called "mydb"

```
var MongoClient = require('mongodb').MongoClient; var url
"mongodb://localhost:27017/mydb";
```

```
MongoClient.connect(url, function(err, db) { if (err) throw
err;
console.log("Database created!"); db.close();
});
```

Save the code above in a file called "demo_create_mongo_db.js" and run the file:

Run "demo_create_mongo_db.js"

C:\Users\Your Name>node demo_create_mongo_db.js This will

give you this result:

Database created!

Note: MongoDB waits until you have created a collection (table), with at least one document (record) before it actually creates the database (and collection).

The use Command

MongoDB use DATABASE_NAME is used to create database. The command will create a new database if it doesn't exist, otherwise it will return the existing database.

Syntax

Basic syntax of use DATABASE statement is as follows –
use DATABASE_NAME

Example

If you want to use a database with name <mydb>, then use DATABASE statement would be as follows –

>use mydb

switched to db mydb

To check your currently selected database, use the command db

>db Mydb

If you want to check your databases list, use the command show dbs.

>show dbs

Local 078125GB

test 0.23012GB

Your created database (mydb) is not present in list. To display database, you need to insert at least one document into it.

```
>db.movie.insert({"name":"tutorials point"})
```

```
>show dbs
```

Local 0.78125GB

mydb 0.23012GB

test 0.23012GB

In MongoDB

default database is

test. If you didn't

create any database,

then collections

will be stored in

test database.

[DROP DATABASE](#)

The **dropDatabase()** Method

MongoDB db.dropDatabase() command is used to drop a existing database.

[Syntax](#)

Basic syntax of dropDatabase() command is as follows –

```
db.dropDatabase()
```

This will delete the selected database. If you have not selected any database, then it will delete default 'test' database.

[Example](#)

First, check the list of available databases by using the command, show dbs.

```
>show dbs
```

Local 0.78125GB

mydb 0.23012GB

test 0.23012GB

```
>
```

If you want to delete new database <mydb>, then dropDatabase() command would be as follows –

```
>use mydb
```

switched to db mydb

```
>db.dropDatabase()
```

```
>{ "dropped" : "mydb", "ok" : 1 }
```

```
>
```

Now check list of databases.

```
>show dbs
```

Local 0.78125GBtest 0.23012GB

```
>
```

PRACTICAL NO. 6 : Hive operations

Here we are explaining the operators available in Hive. There are types of operators in Hive:

Relational Operators

Arithmetic Operators

Logical Operators

Complex Operators

Log in to a master node of the EMR cluster, switch to the Hadoop user, go to the Hive directory, and connect to Hive:

```
[hadoop@172 hive]$ su hadoop
[hadoop@172 hive]$ cd /usr/local/service/hive/bin
[hadoop@172 bin]$ hive
```

You can use the -h parameter to get basic information on Hive commands. You can also use the Beeline mode to connect to a database. To do so, you also need to log in to a master node in EMR, switch to the Hadoop user, and go to the Hive directory. In the conf/hive-site.xml configuration file, get the connection port \$port and host address \$host of Hive server 2

```
<property>
  <name>hive.server2.thrift.bind.host</name>
  <value>$host</value>
</property>
<property>
  <name>hive.server2.thrift.port</name>
  <value>$port</value>
</property>
```

In the bin directory, run the following statement to connect to Hive

```
[hadoop@172 hive]$ cd bin
[hadoop@172 bin]$ ./beeline -u "jdbc:hive2:// $host: $port " -n hadoop -p hadoop
```

PRACTICAL NO. 7 : Analyse Data using Tableau

To analyze data using Tableau practically, follow these steps:

1. Prepare Your Data

Format your data: Ensure the data is in a format Tableau can read, such as Excel (.xls/.xlsx), CSV, or a database connection (SQL, etc.).

Check for missing values: Before importing into Tableau, it's important that the data is clean and structured.

Load the data: Open Tableau and choose "File" to import your dataset (Excel, CSV, Google Sheets, or connect to a server/database).

2. Import Data into Tableau

Connect to your data source: After opening Tableau, choose the "Data" menu and select the type of data connection (Excel, CSV, or connect directly to a database).

Choose the correct data sheet: If you have multiple sheets in your Excel or CSV file, select the appropriate one.

3. Understand the Data Structure

Dimensions: These are qualitative fields (e.g., Category, Product Name, Region).

Measures: These are quantitative fields (e.g., Sales, Profit, Quantity).

Drag and Drop: Drag dimensions and measures into "Rows" or "Columns" to begin visualizing the data.

4. Data Exploration and Visualization

Create basic charts:

Bar Chart: Drag a dimension (e.g., Category) to Columns and a measure (e.g., Sales) to Rows to generate a bar chart.

Line Chart: Use line charts for trend analysis by placing time-based data (e.g., Date) on the X-axis and a measure like Sales on the Y-axis.

Scatter Plot: Drag two measures (e.g., Profit and Sales) into Columns and Rows to create a scatter plot, which helps identify correlations.

Filter Data: Use the filter pane to limit the data shown in your visualizations, based on dimensions or measures.

Example: Filter by Region to only show data for a specific region.

Add Color and Size: Drag a dimension or measure into the "Color" shelf to visually distinguish data, or use "Size" to vary the size of marks based on a measure.

5. Create Calculated Fields

Example Calculation: If you need to calculate profit margins, you can create a calculated field:

Go to **Analysis > Create Calculated Field**.

Enter a formula like Profit Margin = $\text{SUM}([\text{Profit}]) / \text{SUM}([\text{Sales}])$.

6. Create Dashboards

Combine multiple charts: You can combine various views (bar charts, line charts, maps, etc.) on a dashboard.

Add interactivity: Use filters and actions (e.g., click on a bar in a bar chart to filter data in another visualization).

Example: Drag a map (showing sales by region) and a bar chart (showing sales by product category) onto a dashboard. Use filter actions so selecting a region on the map filters the product category data.

7. Data Analysis Techniques

Trend Analysis: Using line charts, moving averages, or other time-based visualizations.

Segmentation: Group your data using dimensions like Region, Category, or Customer Segment to perform comparative analysis.

Aggregation: Tableau automatically aggregates data, but you can also adjust the aggregation level (SUM, AVG, etc.) in the field settings.

8. Insights and Storytelling

Annotations and Tooltips: Add annotations to highlight key insights or trends in the data. Customize tooltips to provide additional information when hovering over data points.

Story Points: Create a "story" to communicate the analysis to others. A story consists of a sequence of sheets that tell a narrative.

9. Export and Share Results

Export Visualizations: You can export your Tableau visualizations into image files (PNG) or PDF formats.

Publish to Tableau Server/Online: If you're working in a team or want to share dashboards with others, publish your workbook to Tableau Server or Tableau Online.

10. Best Practices

Keep it simple: Avoid overloading dashboards with too many visuals. Focus on the key insights.

Use color effectively: Use color to draw attention to important data points or trends, but avoid using too many colors that may confuse viewers.

Interactive filters: Allow users to interact with the data through filters for a more personalized analysis.

Example Practical Scenario:

Let's say you are analyzing sales data across different regions and products:

Load the data: Import the sales data CSV into Tableau, which contains fields like Region, Product Category, Sales, Profit, and Date.

Create a Bar Chart:

Drag Product Category to the Columns shelf.

Drag Sales to the Rows shelf.

Tableau will create a bar chart showing total sales by product category.

Filter by Region:

Drag Region to the Filters shelf.

Select one or more regions to view sales data for those regions.

Create a Line Chart for Trends:

Drag Date to Columns and Sales to Rows.

Change the chart type to a line chart to visualize sales over time.

Add Calculated Field:

Create a calculated field for **Profit Margin**: $\text{Profit Margin} = [\text{Profit}] / [\text{Sales}]$.

Drag the new Profit Margin field to the Tooltip shelf for additional context in your charts.

Create a Dashboard:

Combine the bar chart and line chart on a single dashboard.

Add a filter action that lets users select a region in the bar chart and filter the line chart to show sales trends for that region.

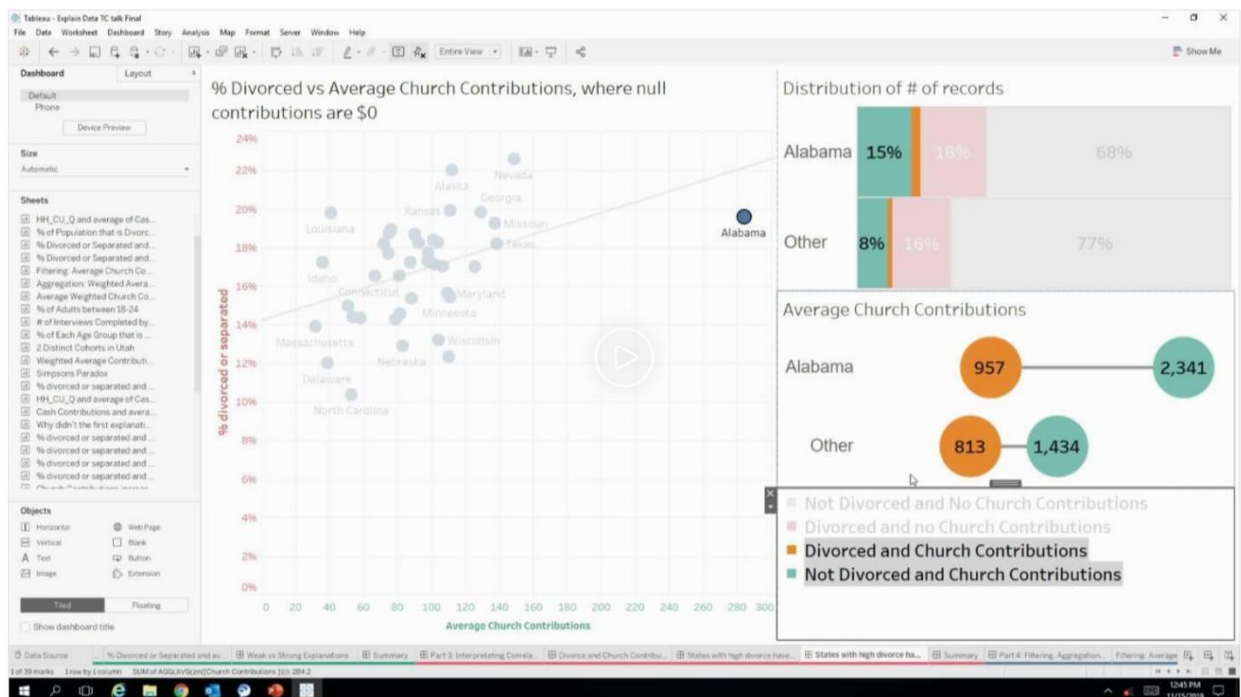
Publish and Share:

After the analysis is done, publish the dashboard to Tableau Server or export it as a PDF for sharing.

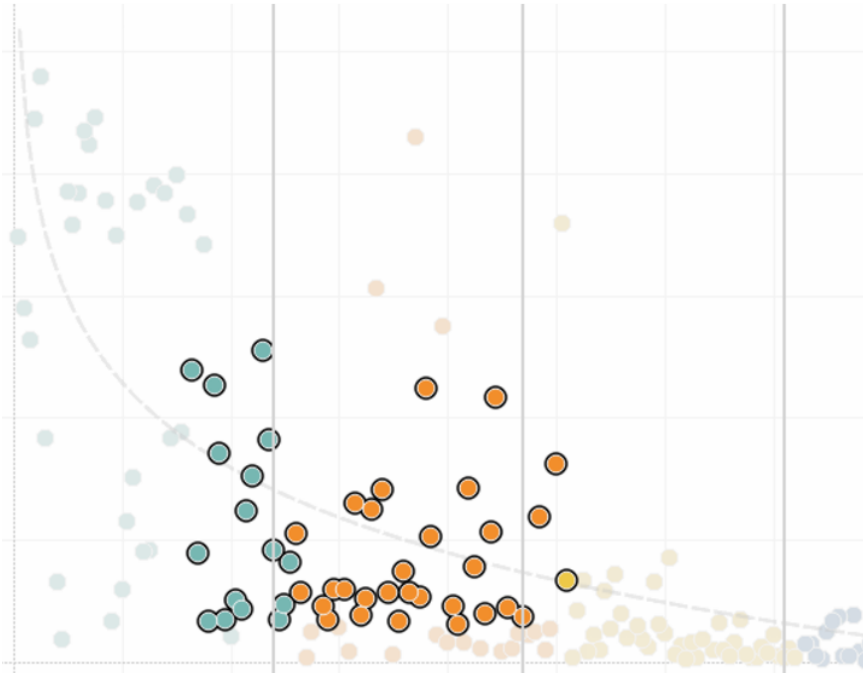
Conclusion:

Using Tableau for practical data analysis involves importing data, creating visualizations, and applying various techniques (filters, calculations, and dashboard creation) to extract valuable insights. Through interactive dashboards and clear visual storytelling, Tableau helps users uncover trends, patterns, and relationships in their data efficiently.

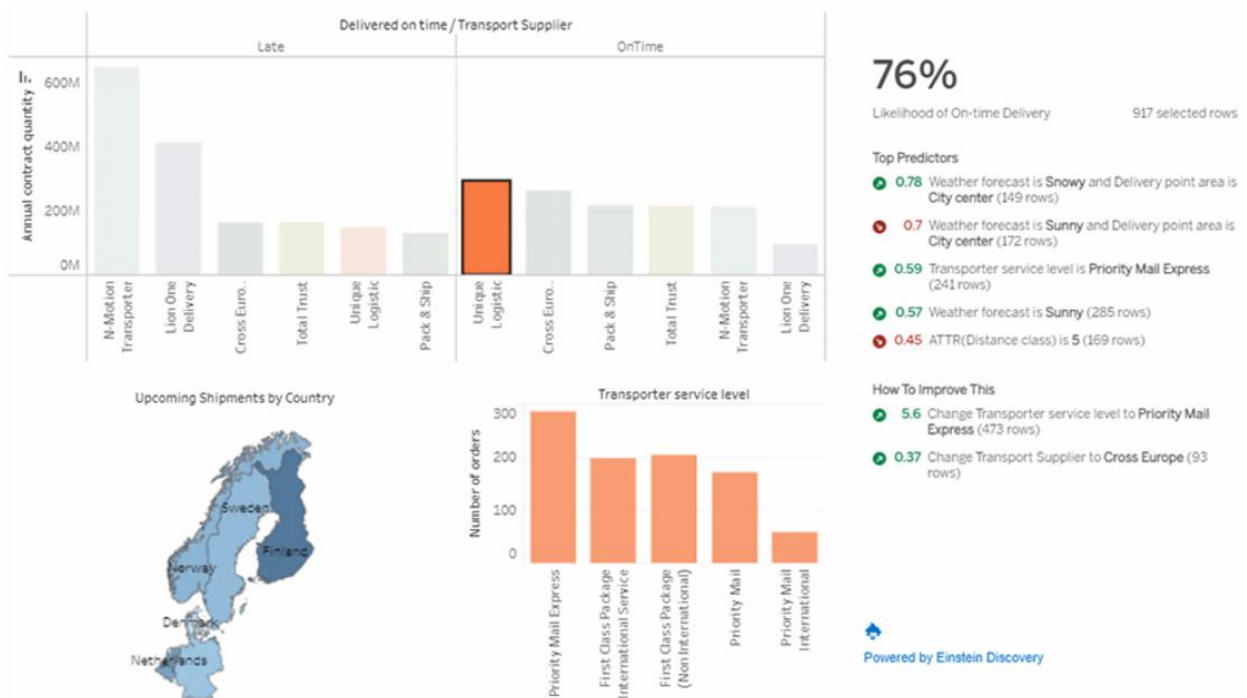
Explain Data helps you uncover unknown unknowns and answer why faster. Pairing your domain knowledge with Explain Data's computational power will supercharge your skills from analyst to statistician.



You can select multiple marks by holding down the Ctrl key on your keyboard (Command key on Mac) while you select each mark. You can also use one of the advanced selection tools to select marks within a specific shape.



Integrate on-demand, interpretable predictions from Einstein Discovery in Tableau dashboards.



PRACTICAL NO. 8 : Data visualization using Tableau

Step 1: Set Up Your Tableau Environment

Download Tableau: If you haven't already, download and install Tableau Desktop. You can use the Tableau Public version if you're just starting out (free but with limitations on privacy).

Open Tableau: Launch Tableau Desktop and start a new project by clicking on File > New.

Step 2: Load Your Data into Tableau

Connect to Data:

On the left sidebar, under Connections, click the data source type (e.g., Microsoft Excel, Text File, Web Data Connector, SQL Server).

Select your file (Excel, CSV, etc.) and click Open. Tableau will load the data into the workspace.

Preview the Data:

Tableau will display a preview of the data. It automatically identifies the dimensions (qualitative variables like Region, Category) and measures (quantitative variables like Sales, Profit).

Drag and Drop to Workspace:

Drag the data table (sheet) you want to analyze to the canvas.

Step 3: Create Basic Visualizations

In Tableau, visualizations are built by dragging dimensions (e.g., Category, Region) and measures (e.g., Sales, Profit) onto rows, columns, and the "marks" area.

Bar Chart (Sales by Category)

Drag 'Category' to Columns: This will create categories along the x-axis.

Drag 'Sales' to Rows: Tableau will generate a bar chart showing sales across different categories.

Sort Bars: To sort bars by sales, click the Sort icon on the chart or right-click on the axis and choose "Sort".

Line Chart (Sales Over Time)

Drag 'Date' to Columns: Tableau will automatically recognize that this is a time-based field and will create a time axis.

Drag 'Sales' to Rows: A line chart will be generated showing sales trends over time.

Adjust Time Granularity: You can adjust the time scale (e.g., from year to month or day) by right-clicking on the Date field and selecting the level of granularity.

Pie Chart (Sales by Region)

Drag 'Region' to Columns: This will set the regions along the pie chart.

Drag 'Sales' to Rows: Tableau will create a pie chart where each slice represents sales by region.

Show Labels: Click on the pie chart and enable labels to show the sales values on each segment.

Scatter Plot (Profit vs. Sales)

Drag 'Profit' to Columns: This will be one axis.

Drag 'Sales' to Rows: This will be the other axis.

Adjust Size and Color: Use the Size and Color shelves to adjust the size of the scatter points based on a measure (like Quantity) and color based on another (like Region).

Step 4: Refine Your Visualizations

Filters:

Drag a dimension (e.g., Region) to the Filters shelf to filter data by a specific region.

You can also filter by measures (e.g., only show sales greater than \$100,000).

Color:

Drag a dimension (like Category) to the Color shelf to differentiate segments using color.

Adjust the color scheme by clicking on the Color legend and selecting "Edit Colors" to choose a color palette that fits the data.

Tooltips:

Customize tooltips to provide more detailed information when hovering over data points.

Click on Tooltip in the Marks card to modify what is displayed.

Add Annotations:

Right-click on a data point or part of the visualization and choose Annotate to add notes or explanations.

Step 5: Create a Dashboard

To combine multiple visualizations into a single dashboard:

Create a New Dashboard:

Click on the Dashboard tab at the bottom of Tableau and then select New Dashboard.

Drag Worksheets to Dashboard:

Drag each worksheet (e.g., bar chart, line chart, pie chart) from the left sidebar into the dashboard area.

Size Your Dashboard: Adjust the size of the dashboard by choosing Fixed Size or Automatic under the Size section on the left.

Add Interactivity:

To add interactivity between visualizations, use Actions. For example, you can set up a filter action where clicking on a bar in the bar chart filters data in the line chart.

Go to Dashboard > Actions > Add Action > Filter to set up this feature.

Step 6: Final Touches and Sharing

Customize Titles and Labels:

Edit titles and labels to make your visualizations clearer.

Right-click the title area and choose Edit to change the text, fonts, and alignment.

Publish the Dashboard:

Once satisfied with your visualizations, you can publish the dashboard.

Click File > Save As to save the workbook locally or Server to upload it to Tableau Online or Tableau Server.

Export:

Export the dashboard as a PDF, image, or PowerPoint by going to File > Export.

Practical Example Scenario: Sales and Profit Analysis

Let's assume you have a dataset with the following fields: Region, Category, Product, Sales, Profit, and Date. Here's how you would visualize this data:

Bar Chart: Show total sales for each product category.

Drag Category to Columns.

Drag Sales to Rows.

Sort the chart to display categories with the highest sales first.

Line Chart: Show sales trends over time (e.g., monthly sales).

Drag Date to Columns (ensure it's set to a monthly level).

Drag Sales to Rows to create a line chart showing sales over the months.

Scatter Plot: Show the relationship between sales and profit.

Drag Sales to Columns.

Drag Profit to Rows.

Color the scatter points by Region to see the geographical distribution of sales and profit.

Pie Chart: Show the sales distribution by region.

Drag Region to Columns.

Drag Sales to Rows to create a pie chart.

Label the slices with the sales values for better clarity.

Dashboard:

Combine the bar chart, line chart, scatter plot, and pie chart into a single dashboard.

Add filter actions, e.g., clicking on a region in the pie chart filters the data in the bar and line charts.

Conclusion

By following this process, you will be able to create interactive, insightful data visualizations using

Tableau. From importing data to refining visualizations and creating dashboards, Tableau enables you to tell a compelling story with your data. With these basic skills, you can analyze trends, identify patterns, and make informed decisions based on your data.

The downloaded file is a zip file that contains an excel that looks like the given picture below:

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Shipping Cost	Order Priority
2	United State	10024	US	East	TEC-AC-1000	Technology	Accessories	Plantronics C	2309.65	7	0	762.1845	933.57	Critical
3	Australia		APAC	Oceania	FUR-CH-1000	Furniture	Chairs	Novimex Exe	3709.395	9	0.1	-288.765	923.63	Critical
4	Australia		APAC	Oceania	TEC-PH-1000	Technology	Phones	Nokia Smart	5175.171	9	0.1	919.971	915.49	Medium
5	Germany		EU	Central	TEC-PH-1000	Technology	Phones	Motorola Sr	2892.51	5	0.1	-96.54	910.16	Medium
6	Senegal		Africa	Africa	TEC-SHA-100	Technology	Copiers	Sharp Wirele	2832.96	8	0	311.52	903.04	Critical
7	Australia		APAC	Oceania	TEC-PH-1000	Technology	Phones	Samsung Sm	2862.675	5	0.1	763.275	897.35	Critical
8	New Zealand		APAC	Oceania	FUR-CH-1000	Furniture	Chairs	Novimex Exe	1822.08	4	0	564.84	894.77	Critical
9	New Zealand		APAC	Oceania	FUR-TA-1000	Furniture	Tables	Chromcraft C	5244.84	6	0	996.48	878.38	High
10	United State	95823	US	West	OFF-BI-1000	Office Suppli	Binders	Fellowes PBS	5083.96	5	0.2	1906.485	867.69	Low
11	United State	28027	US	South	FUR-TA-1000	Furniture	Tables	Chromcraft T	4297.644	13	0.4	-1862.3124	865.74	Critical
12	United State	22304	US	South	OFF-SU-1000	Office Suppli	Supplies	Martin Yale i	4164.05	5	0	83.281	846.54	High
13	Afghanistan		APAC	Central Asia	FUR-TA-1000	Furniture	Tables	Bevis Confer	4626.15	5	0	647.55	835.57	High
14	Saudi Arabia		EMEA	EMEA	TEC-CIS-1000	Technology	Phones	Cisco Smart I	2616.96	4	0	1151.4	832.41	Critical
15	Brazil		LATAM	South	FUR-CH-1000	Furniture	Chairs	Harbour Cre	2221.8	7	0	622.02	810.25	Critical
16	China		APAC	North Asia	OFF-AP-1000	Office Suppli	Appliances	KitchenAid N	3701.52	12	0	1036.08	804.54	Critical
17	France		EU	Central	OFF-AP-1000	Office Suppli	Appliances	Breville Refri	1869.588	4	0.1	186.948	801.66	Critical
18	United State	42420	US	South	TEC-AC-1000	Technology	Accessories	Logitech diN	2249.91	9	0	517.4793	780.70	Critical
19	Italy		EU	South	OFF-AP-1000	Office Suppli	Appliances	Hoover Stovi	7958.58	14	0	3979.08	778.32	Low
20	Australia		APAC	Oceania	TEC-CO-1000	Technology	Copiers	Brother Fax I	2565.594	9	0.1	28.404	766.93	Critical
21	Tanzania		Africa	Africa	OFF-KIT-1000	Office Suppli	Appliances	KitchenAid S	3409.74	6	0	818.28	763.38	High
22	Poland		EMEA	EMEA	FUR-HON-10	Furniture	Tables	Hon Comput	1977.72	4	0	276.84	759.47	Critical
23	United State	60610	US	Central	TEC-PH-1000	Technology	Phones	Apple iPhon	2735.952	6	0.2	341.994	752.51	High
24	China		APAC	North Asia	FUR-CH-1000	Furniture	Chairs	SAFCO Execu	2754	6	0	358.02	752.47	Critical
25	United Kingdom		EU	North	OFF-AP-1000	Office Suppli	Appliances	KitchenAid R	5273.7	10	0	1898.4	730.91	High
26	Mexico		LATAM	North	TEC-PH-1000	Technology	Phones	Motorola Sr	1713.84	4	0	445.52	728.97	Critical
27	El Salvador		LATAM	Central	FUR-TA-1000	Furniture	Tables	Hon Comput	2106.496	8	0.2	526.496	728.39	Critical
28	Taiwan		APAC	North Asia	FUR-TA-1000	Furniture	Tables	Lesro Confer	1715.16	2	0	720.36	725.57	Critical

will demonstrate to you how to load data and perform some preprocessing.

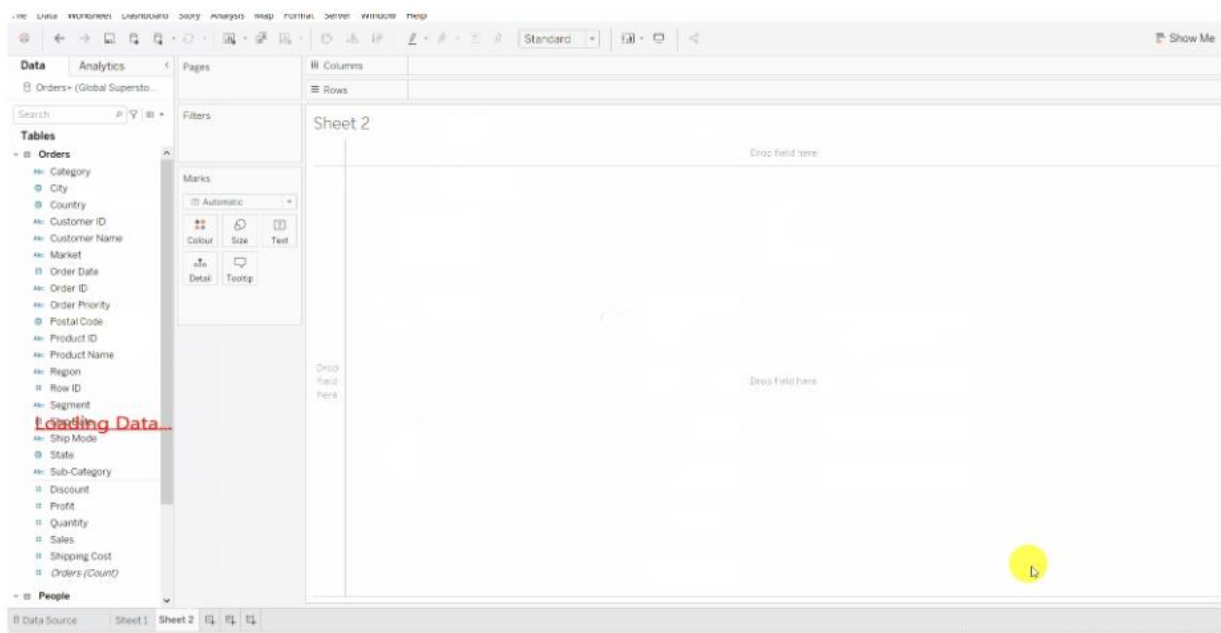


Tableau work-page consist of different section. Let's understand them first before plotting our graphs.

