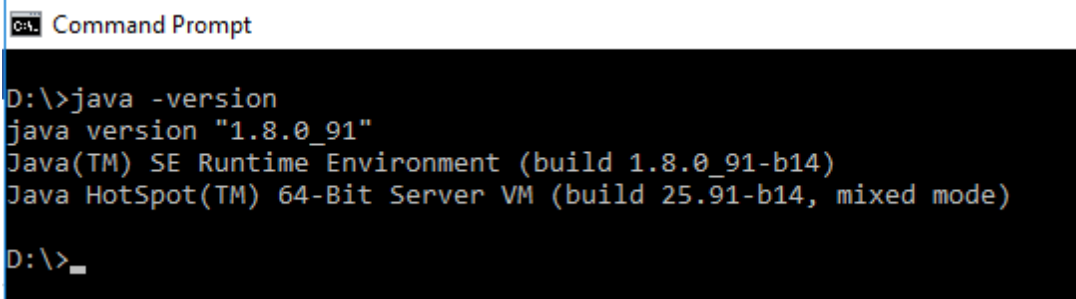


Installing Hadoop-2.6.x on Windows 10.

1. Install Java 8: Download Java 8 from the link:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

- a. Set environmental variables:
 - i. User variable:
 - Variable: JAVA_HOME
 - Value: C:\java
 - ii. System variable:
 - Variable: PATH
 - Value: C:\java\bin
- b. Check on cmd, see below:



```

C:\> Command Prompt

D:\>java -version
java version "1.8.0_91"
Java(TM) SE Runtime Environment (build 1.8.0_91-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.91-b14, mixed mode)

D:\>_
  
```

2. Install Eclipse Mars. Download it from the link: <https://eclipse.org/downloads/> and extract it into C drive.

- a. Set environmental variables:
 - i. User variable:
 - Variable: ECLIPSE_HOME
 - Value: C:\eclipse
 - ii. System variable:
 - Variable: PATH
 - Value: C:\eclipse\bin
- b. **Download “hadoop2x-eclipse-plugin-master.”** You will see three Jar files on the path “hadoop2x-eclipse-plugin-master\release.” Copy these three jar files and paste them into “C:\eclipse\dropins.”
- c. **Download “slf4j-1.7.21.”** Copy Jar files from this folder and paste them to “C:\eclipse\plugins”. This step may create errors; when you will execute Eclipse, you will see errors like org.apa.....jar file in multiple places. So, now delete these files from all the places except one.

Errors
SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/C:/eclipse/plugins/org.slf4j.impl.log4j12_1.7.2.v20131105-2200.jar!/org/slf4j/impl/StaticLoggerBinder.class] SLF4J: Found binding in [jar:file:/C:/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

3. **Download Apache-ant-1.9.6:** (optional step) Download Apache-ant from the link: <http://ant.apache.org/bindownload.cgi> and extract it into a folder in C drive.
4. **Download Hadoop-2.6.x:** download Hadoop 2.6.x from the link: <http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.6.2/hadoop-2.6.2.tar.gz>
 - a. Put extracted Hadoop-2.6.x files into D drive. Note that do not put these extracted files into C drive, where you installed your Windows.
 - b. **Download “hadoop-common-2.6.0-bin-master”** from the link: <https://github.com/amihalik/hadoop-common-2.6.0-bin/tree/master/bin>. You will see 11 files there. Paste all these files into the “bin” folder of Hadoop-2.6.x.
 - c. Create a “data” folder inside Hadoop-2.6.x, and also create two more folders in the “data” folder as “data” and “name.”
 - d. Create a folder to store temporary data during execution of a project, such as “D:\hadoop\temp.”
 - e. Create a log folder, such as “D:\hadoop\userlog”
 - f. Go to Hadoop-2.6.x → etc → Hadoop and edit four files:
 - i. core-site.xml
 - ii. hdfs-site.xml
 - iii. mapred.xml
 - iv. yarn.xml

core-site.xml

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

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

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

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

<configuration>
<property>
  <name>hadoop.tmp.dir</name>
  <value>D:\hadoop\temp</value>
</property>

<property>
<name>fs.default.name</name>
  <value>hdfs://localhost:50071</value>
</property>
```

```
</configuration>
```

hdfs-site.xml

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

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

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

<!-- 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>/hadoop-
2.6.0/data/name</value><final>true</final></property>

<property><name>dfs.datanode.data.dir</name><value>/hadoop-
2.6.0/data/data</value><final>true</final> </property>

</configuration>
```

mapred.xml

```
<?xml version="1.0"?>
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>

  <property>
    <name>mapred.job.tracker</name>
    <value>localhost:9001</value>
  </property>

  <property>
    <name>mapreduce.application.classpath</name>
```

```

<value>/hadoop-2.6.0/share/hadoop/mapreduce/*,
/hadoop-2.6.0/share/hadoop/mapreduce/lib/*,
/hadoop-2.6.0/share/hadoop/common/*,
/hadoop-2.6.0/share/hadoop/common/lib/*,
/hadoop-2.6.0/share/hadoop/yarn/*,
/hadoop-2.6.0/share/hadoop/yarn/lib/*,
/hadoop-2.6.0/share/hadoop/hdfs/*,
/hadoop-2.6.0/share/hadoop/hdfs/lib/*,
</value>
</property>

</configuration>

```

yarn-site.xml

```

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

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

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

<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>

  <property>
    <name>yarn.nodemanager.log-dirs</name>
    <value>D:\hadoop\userlog</value><final>true</final>
  </property>

  <property><name>yarn.nodemanager.local-dirs</name><value>D:\hadoop\temp\nm-local-
dir</value></property>

  <property>
    <name>yarn.nodemanager.delete.debug-delay-sec</name>
    <value>600</value>

```

```

</property>

<property><name>yarn.application.classpath</name>
  <value>/hadoop-2.6.0/./hadoop-2.6.0/share/hadoop/common/*,/hadoop-
2.6.0/share/hadoop/common/lib/*,/hadoop-2.6.0/share/hadoop/hdfs/*,/hadoop-
2.6.0/share/hadoop/hdfs/lib/*,/hadoop-2.6.0/share/hadoop/mapreduce/*,/hadoop-
2.6.0/share/hadoop/mapreduce/lib/*,/hadoop-2.6.0/share/hadoop/yarn/*,/hadoop-
2.6.0/share/hadoop/yarn/lib/*</value>
</property>

</configuration>

```

- g. Go to the location: “Hadoop-2.6.0→etc→hadoop,” and edit “hadoop-env.cmd” by writing
set JAVA_HOME=C:\java\jdk1.8.0_91
- h. Set environmental variables: Do: My computer → Properties → Advance system settings →
Advanced → Environmental variables
 - i. User variables:
 - Variable: HADOOP_HOME
 - Value: D:\hadoop-2.6.0
 - ii. System variable
 - Variable: Path
 - Value: D:\hadoop-2.6.0\bin
D:\hadoop-2.6.0\sbin
D:\hadoop-2.6.0\share\hadoop\common*
D:\hadoop-2.6.0\share\hadoop\hdfs
D:\hadoop-2.6.0\share\hadoop\hdfs\lib*
D:\hadoop-2.6.0\share\hadoop\hdfs*
D:\hadoop-2.6.0\share\hadoop\yarn\lib*
D:\hadoop-2.6.0\share\hadoop\yarn*
D:\hadoop-2.6.0\share\hadoop\mapreduce\lib*
D:\hadoop-2.6.0\share\hadoop\mapreduce*
D:\hadoop-2.6.0\share\hadoop\common\lib*
- i. Check on cmd; see below.

Command Prompt

```

D:\>hadoop version
Hadoop 2.6.0
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r e3496499ecb8d220fba99dc5ed4c99c8f9e33bb1
Compiled by jenkins on 2014-11-13T21:10Z
Compiled with protoc 2.5.0
From source with checksum 18e43357c8f927c0695f1e9522859d6a
This command was run using /D:/hadoop-2.6.0/share/hadoop/common/hadoop-common-2.6.0.jar

```

- j. **Format name-node:** On cmd go to the location “Hadoop-2.6.0→bin” by writing on cmd “cd
hadoop-2.6.0\bin” and then “hdfs namenode –format”
- k. Start Hadoop. Go to the location: “D:\hadoop-2.6.0\sbin.” Run the following files as
administrator “start-dfs.cmd” and “start-yarn.cmd”

```

Apache Hadoop Distribution - hadoop - namenode
16/05/22 12:43:29 INFO blockmanagement.CacheReplie
16/05/22 12:43:29 INFO blockmanagement.CacheReplie
16/05/22 12:43:33 INFO namenode.FSNamesystem: the
16/05/22 12:43:37 INFO namenode.FSNamesystem: the
16/05/22 12:43:59 INFO blockmanagement.CacheReplie
16/05/22 12:43:59 INFO blockmanagement.CacheReplie
16/05/22 12:44:29 INFO blockmanagement.CacheReplie
16/05/22 12:44:29 INFO blockmanagement.CacheReplie
16/05/22 12:44:59 INFO blockmanagement.CacheReplie
16/05/22 12:44:59 INFO blockmanagement.CacheReplie

Apache Hadoop Distribution - hadoop - datanode
16/05/22 12:41:01 INFO impl.FsDatasetImpl: Total time to scan all replicas for block pool BP-99127786-132.72.225.79-1460371444670: 592ms
16/05/22 12:41:01 INFO impl.FsDatasetImpl: Adding replicas to map for block pool BP-99127786-132.72.225.79-1460371444670: 224ms
16/05/22 12:41:01 INFO impl.FsDatasetImpl: Time to add replicas to map for block pool BP-99127786-132.72.225.79-1460371444670: 224ms
16/05/22 12:41:01 INFO impl.FsDatasetImpl: Total time to add all replicas to map: 246ms
16/05/22 12:41:01 INFO datanode.DataNode: Block pool BP-99127786-132.72.225.79-1460371444670 (Datanode Uuid null) service to localhost/127.0.0.1:50071 beginning handshake with NN
16/05/22 12:41:01 INFO datanode.DataNode: Block pool BP-99127786-132.72.225.79-1460371444670 (Datanode Uuid null) service to localhost/127.0.0.1:50071 successfully registered with NN
16/05/22 12:41:01 INFO datanode.DataNode: For namenode localhost/127.0.0.1:50071 using DELETEREPORT_INTERVAL of 300000 msec
16/05/22 12:41:01 INFO datanode.DataNode: BLOCKREPORT_INTERVAL of 21600000msec CACHEREPORT_INTERVAL of 10000msec Initial delay: 0msec; heartbeatInterval=300000msec

Apache Hadoop Distribution - yarn - resourcemanager
16/05/22 12:42:08 INFO delegation.AbstractDelegati
16/05/22 12:42:08 INFO delegation.AbstractDelegati
16/05/22 12:42:08 INFO delegation.AbstractDelegati
16/05/22 12:42:08 INFO delegation.AbstractDelegati
16/05/22 12:42:09 INFO delegation.AbstractDelegati
16/05/22 12:42:09 INFO delegation.AbstractDelegati
16/05/22 12:42:09 INFO delegation.AbstractDelegati
16/05/22 12:42:09 INFO delegation.AbstractDelegati
16/05/22 12:42:09 INFO delegation.AbstractDelegati
16/05/22 12:42:09 INFO delegation.AbstractDelegati
16/05/22 12:42:11 INFO webapp.WebApps: Registered
16/05/22 12:42:11 INFO ipc.CallQueueManager: Using
16/05/22 12:42:11 INFO pb.RpcServerFactoryPBImpl:
16/05/22 12:42:11 INFO ipc.Server: Starting Socket
16/05/22 12:42:11 INFO ipc.Server: IPC Server list
16/05/22 12:42:11 INFO ipc.Server: IPC Server Resp
16/05/22 12:42:12 INFO util.RackResolver: Resolved
16/05/22 12:42:12 INFO resourcemanager.ResourceTrac
16/05/22 12:42:12 INFO rmmode.RMNodeImpl: 132.72.216
16/05/22 12:42:12 INFO capacity.CapacityScheduler:

Apache Hadoop Distribution - yarn - nodemanager
16/05/22 12:42:09 INFO http.HttpRequestLog: Http request log for http.requests.nodemanager is not defined
16/05/22 12:42:09 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingInputFilter)
16/05/22 12:42:09 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context node
16/05/22 12:42:09 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context static
16/05/22 12:42:09 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
16/05/22 12:42:09 INFO http.HttpServer2: adding path spec: /node/*
16/05/22 12:42:09 INFO http.HttpServer2: adding path spec: /ws/*
16/05/22 12:42:09 INFO http.HttpServer2: Jetty bound to port 8042
16/05/22 12:42:09 INFO mortbay.log: Jetty-6.1.26
16/05/22 12:42:09 INFO mortbay.log: Extract jar:file:/D:/hadoop-2.6.0/share/hadoop/yarn/hadoop-yarn-common-2.6.0.jar!/webapps/node to C:/Users/shantanu/AppData/Local/Temp/Jetty_0_0_0_8042_node_19tj0x/webapp
16/05/22 12:42:10 INFO mortbay.log: Started HttpServer2$SelectChannelConnectorWithSafeStartup@0.0.0.0:8042
16/05/22 12:42:10 INFO webapp.WebApps: Web app /node started at 8042
16/05/22 12:42:12 INFO webapp.WebApps: Registered webapp guice modules
16/05/22 12:42:12 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8031
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers: []
16/05/22 12:42:12 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id
16/05/22 12:42:12 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id 20514
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as 132.72.225.79:57060 with to
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Notifying ContainerManager to unblock new container-requests

```

5. **Execute a project (Word Count).** You can download „word count“ code (say WordCount.java) from the Internet. Suppose the code is at the location “D:\hadoop-2.6.0\Hadoop-WordCount.”

- Now, we have to store a text file (input.txt) in HDFS open command prompt, and to do so, write on cmd: “cd D:\hadoop-2.6.0\Hadoop-WordCount” and then
hadoop fs -put <input file name> /<location in HDFS>/<file name in hdfs>
“hadoop fs -put input.txt input1.txt”

```

C:\> Command Prompt
D:\hadoop-2.6.0\Hadoop-WordCount>hadoop fs -put input.txt /wc/input1.txt
D:\hadoop-2.6.0\Hadoop-WordCount>_

```

- Now, execute WordCount.java for obtaining the result. On cmd
hadoop jar <Jar file name> <Class name> <Input file in HDFS> <Output file in HDFS>
hadoop jar wordcount.jar WordCount /wc/input1.txt /wc/answer.txt

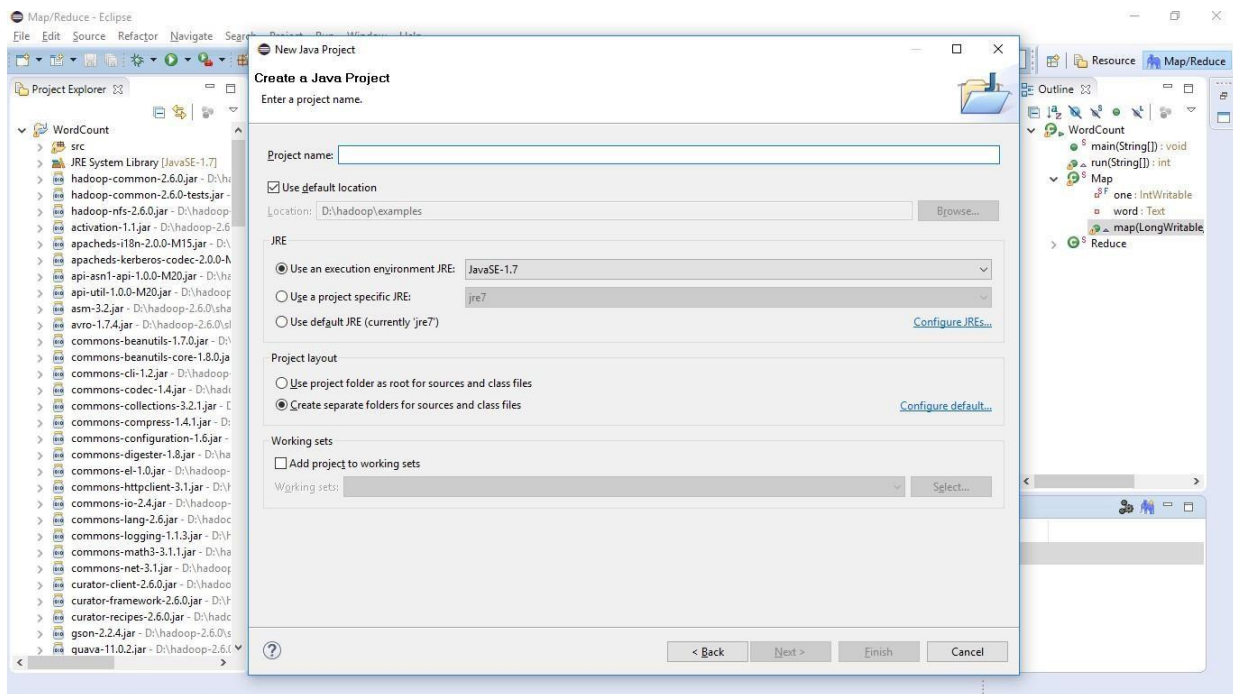
```

C:\> Command Prompt - hadoop jar wordcount.jar WordCount /wc/input1.txt /wc/answer.txt
D:\hadoop-2.6.0\Hadoop-WordCount>hadoop jar wordcount.jar WordCount /wc/input1.txt /wc/answer.txt
16/05/22 13:18:58 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/05/22 13:19:00 INFO input.FileInputFormat: Total input paths to process : 1
16/05/22 13:19:01 INFO mapreduce.JobSubmitter: number of splits:1
16/05/22 13:19:01 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1463910126038_0003
16/05/22 13:19:03 INFO impl.YarnClientImpl: Submitted application application_1463910126038_0003
16/05/22 13:19:03 INFO mapreduce.Job: The url to track the job: http://DESKTOP-KRIBNFT:8088/proxy/application_1463910126038_0003
16/05/22 13:19:03 INFO mapreduce.Job: Running job: job_1463910126038_0003
16/05/22 13:19:29 INFO mapreduce.Job: Job job_1463910126038_0003 running in uber mode : false
16/05/22 13:19:29 INFO mapreduce.Job: map 0% reduce 0%
16/05/22 13:19:44 INFO mapreduce.Job: map 100% reduce 0%
16/05/22 13:19:56 INFO mapreduce.Job: map 100% reduce 100%

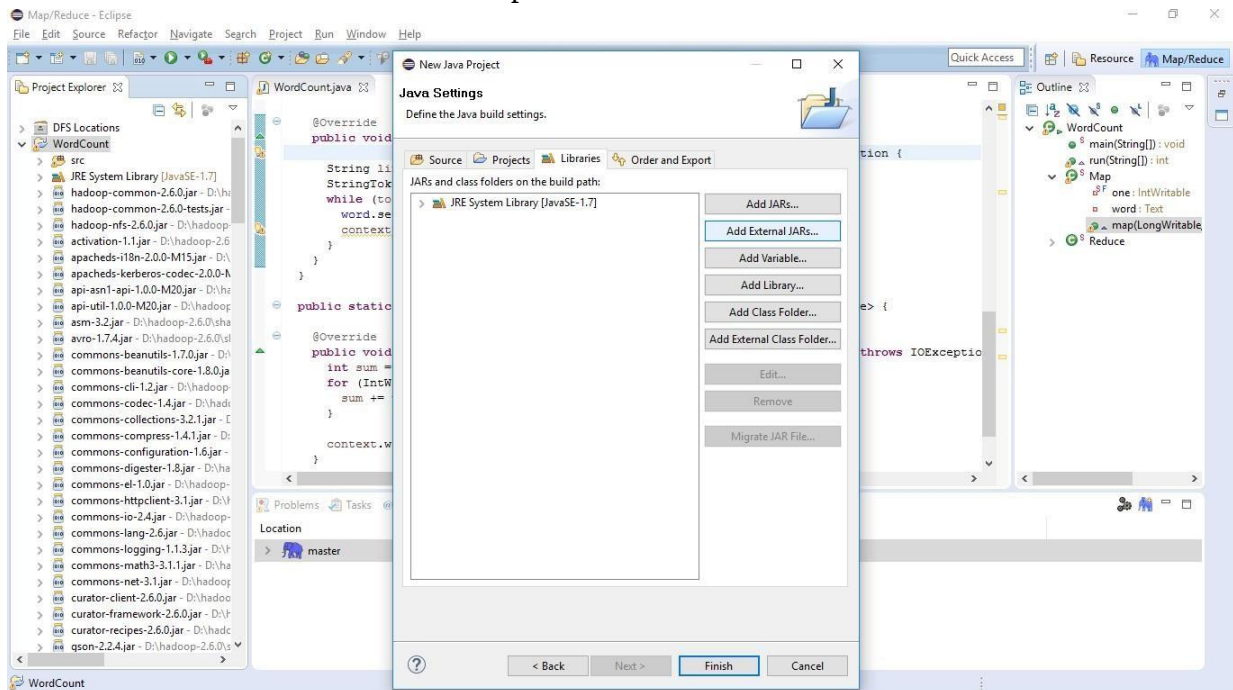
```

How to create a new MapReduce project in Eclipse

- Open Ellipse
- Click File → New Project → Java project



3. Click next and add external Jars for MapReduce.



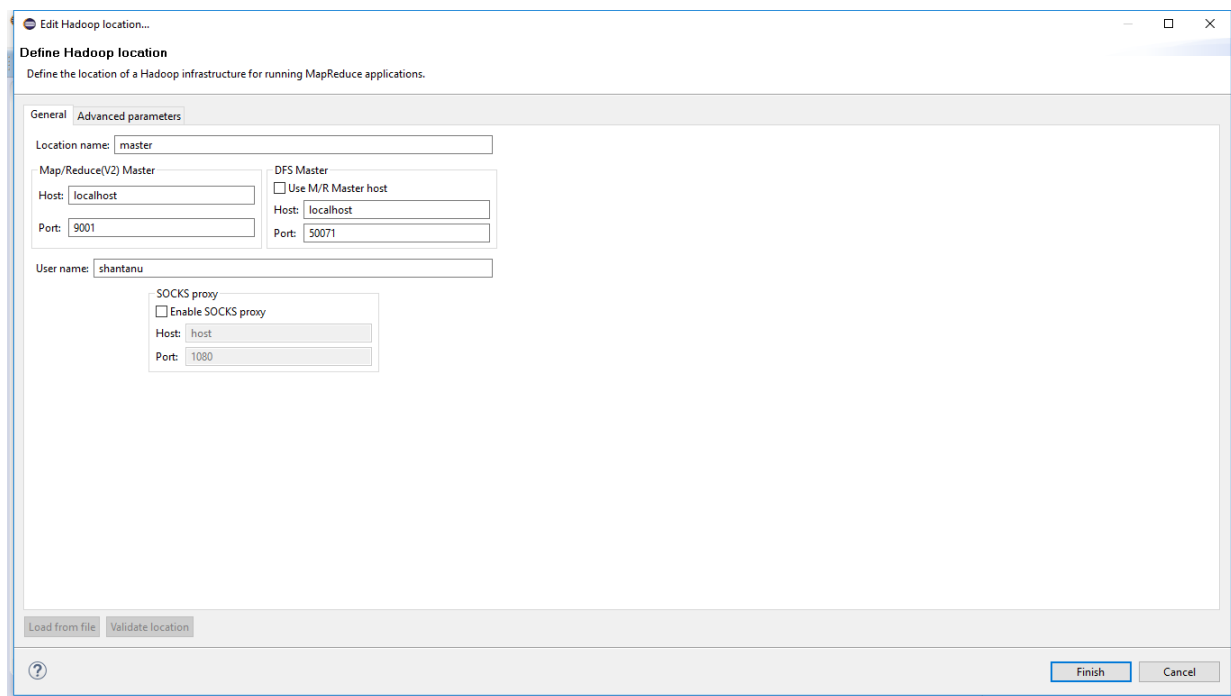
Copy all the Jar files from the locations “D:\hadoop-2.6.0\”

- \share\hadoop\common\lib
- \share\hadoop\mapreduce
- \share\hadoop\mapreduce\lib share\hadoop\yarn
- \share\hadoop\yarn\lib

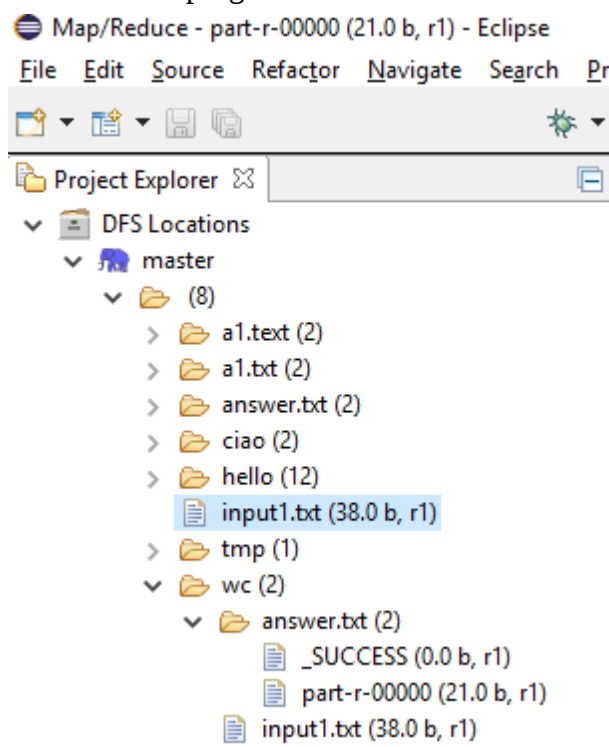
4. Connect DFS in Eclipse

Eclipse → Window → Perspective → Open Perspective → Other → MapReduce → Click OK.
See a bar at the bottom. Click on Map/Reduce locations.

Right click on blank space, then click on “Edit setting,” and you will see the following screen.



- a. Set the following:
 - i. MapReduce (V2) Master
 - Host: localhost
 - Port: 9001
 - ii. DFS Master
 - Host: localhost
 - Port: 50071
- b. Click finish, and you will see the following. Note that you can see the input and output files of the above mentioned program execution.



You can also progress on <http://localhost:8088> and <http://localhost:50070>