Installing Hadoop-2.6.x on Windows 10.

Shantanu Sharma
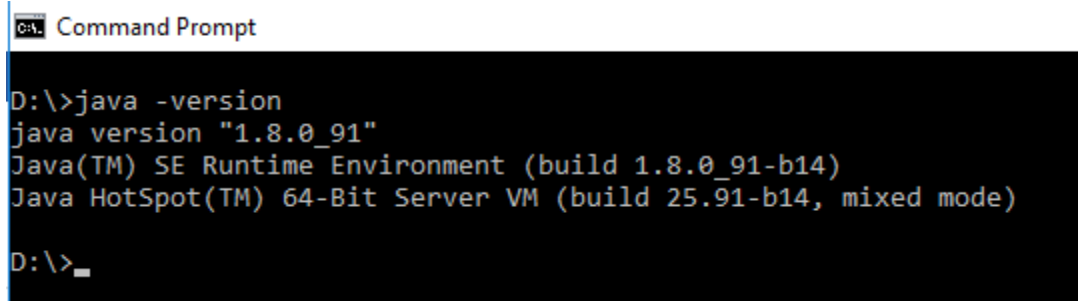
Department of Computer Science, Ben-Gurion University, Israel.

sharmas@cs.bgu.ac.il

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:

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 pate 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]

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
<?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
<?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
<?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>
```

```xml
    <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
<?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
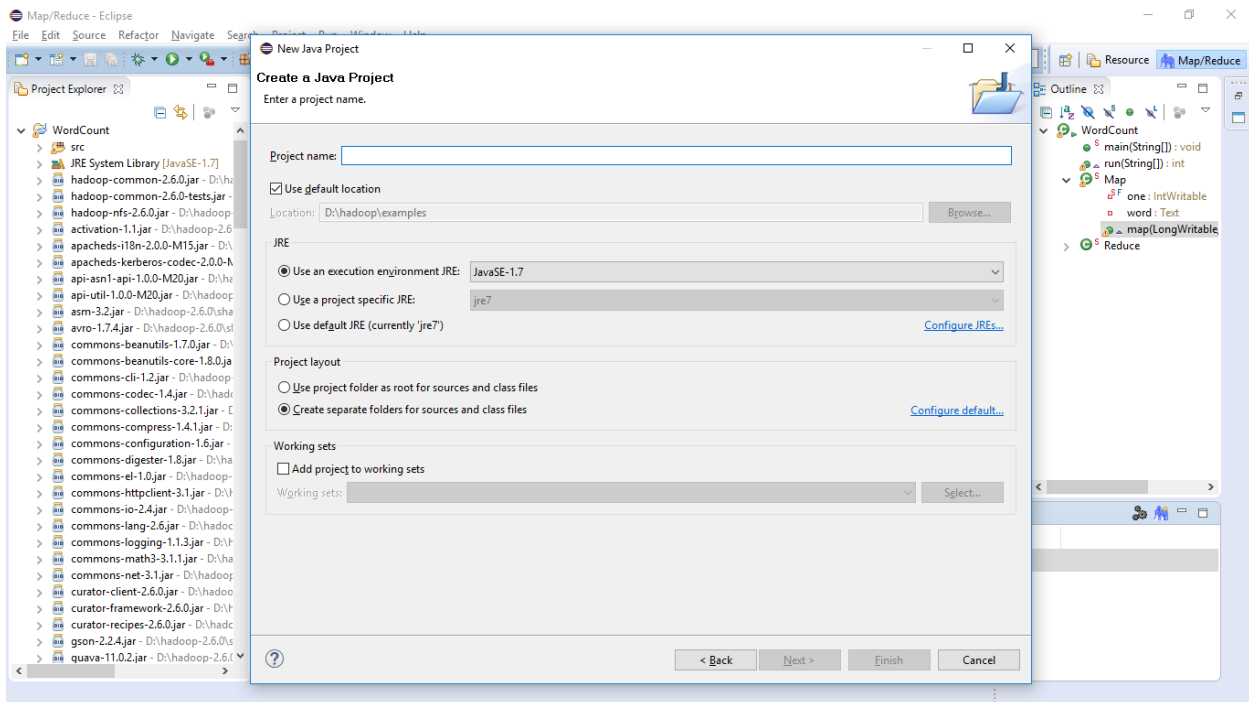   set JAVA_HOME=C:\java\jdk1.8.0_91
h. Set environmental variables: Do: My computer → Properties → Advance system settings →
   Advanced → Environmental variables
   i. User variables:
      • Variable: HADOOP_HOME
      • Value: D:\hadoop-2.6.0
   ii. System variable
      • Variable: Path
      • Value: 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*
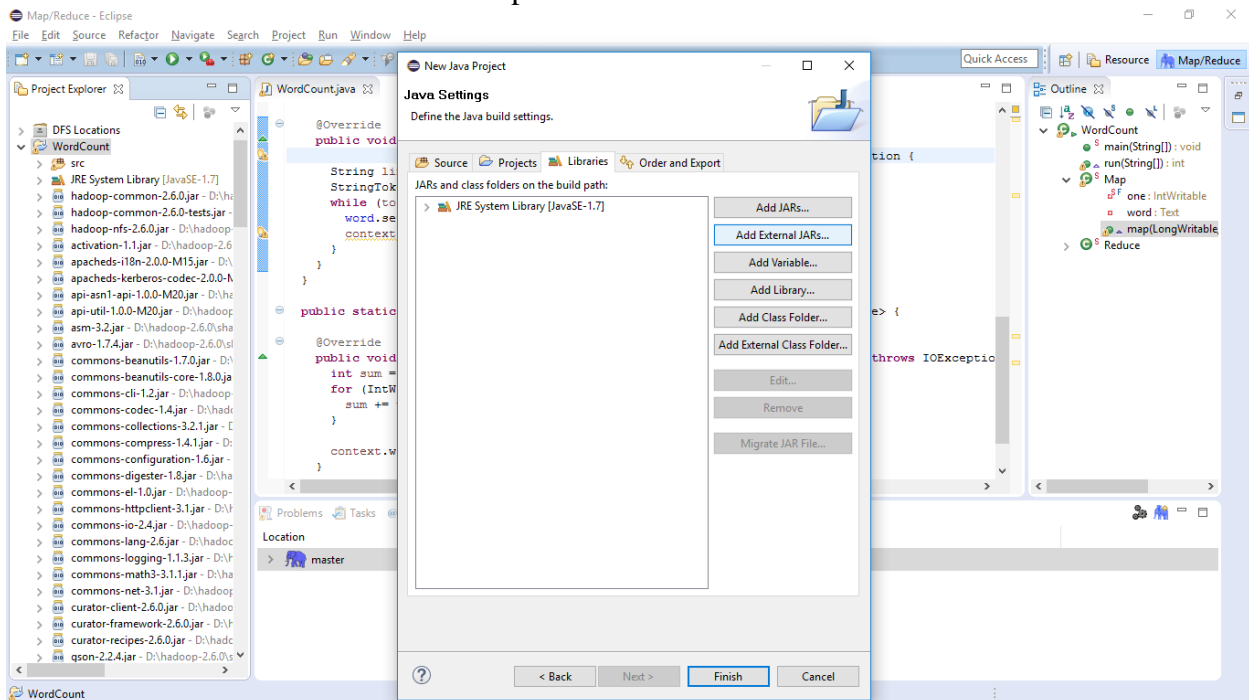i. Check on cmd; see below.

```
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"

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."

   a. 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"

   b. 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

How to create a new MapReduce project in Eclipse

   1. Open Ellipse
   2. Click File → New Project → Java project

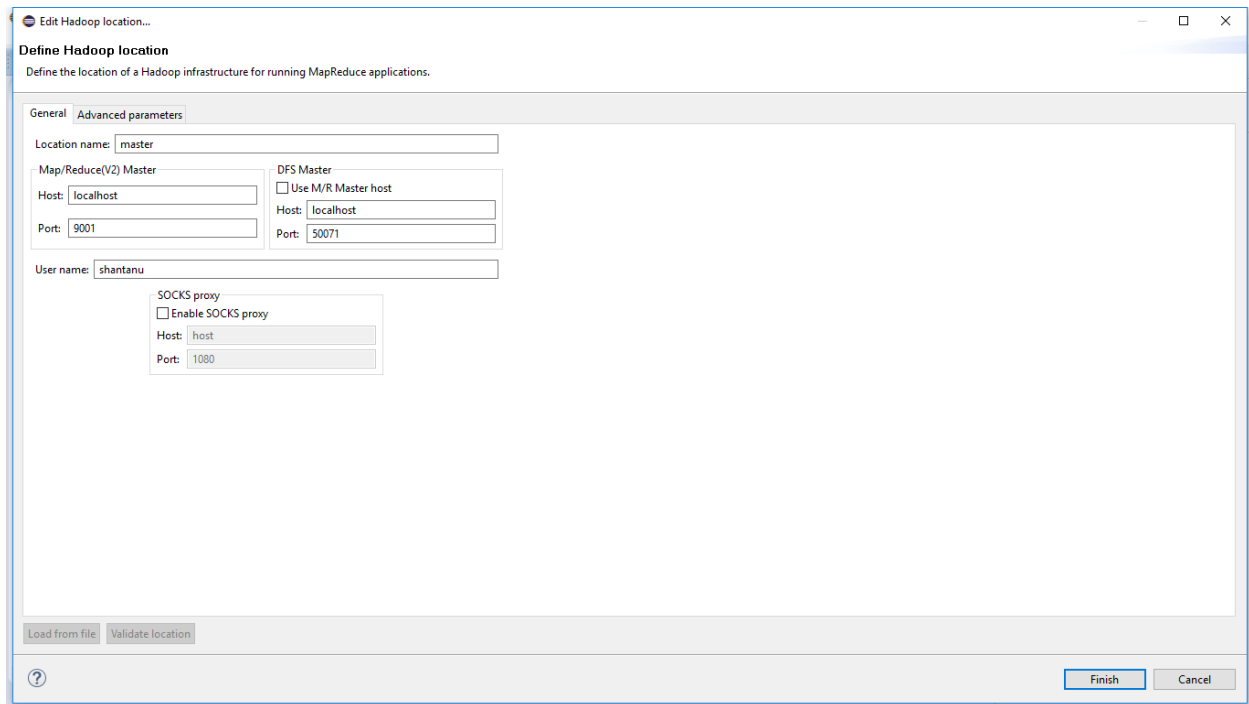3. Click next and add external Jars for MapReduce.

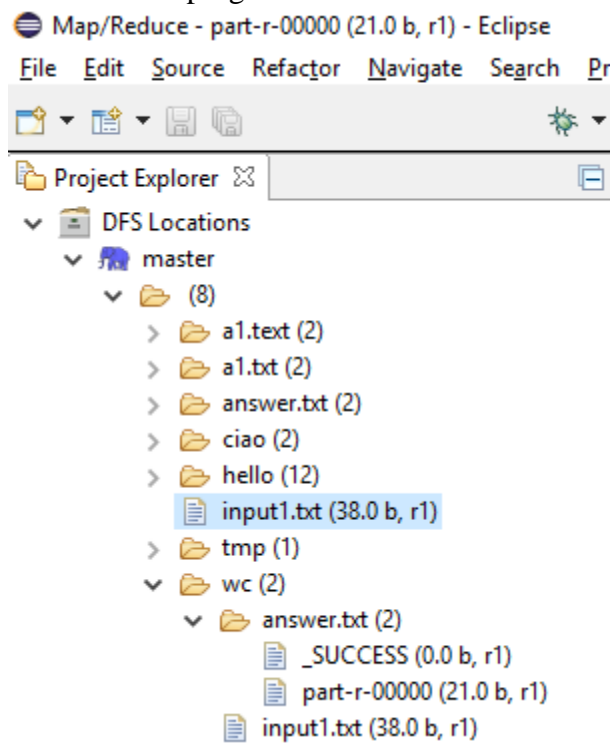Copy all the Jar files from the locations "D:\hadoop-2.6.0\"
   a. \share\hadoop\common\lib
   b. \share\hadoop\mapreduce
   c. \share\hadoop\mapreduce\lib share\hadoop\yarn
   d. \share\hadoop\yarn\lib

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