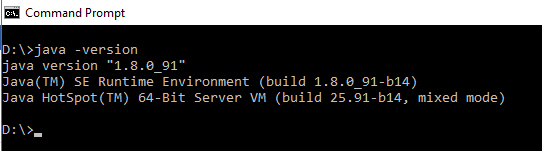
**1.Install, configure and run Hadoop and HDFS**

1. **Install Java 8:** Download Java 8 from the link: <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
   1. Set environmental variables:
      1. User variable:
         * Variable: JAVA\_HOME
         * Value: C:\java
      2. System variable:
         * Variable: PATH
         * Value: C:\java\bin
   2. Check on cmd, see below:



1. **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.
2. **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>
   1. 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.
   2. **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.
   3. Create a “data” folder inside Hadoop-2.6.x, and also create two more folders in the “data” folder as “data” and “name.”
   4. Create a folder to store temporary data during execution of a project, such as “D:\hadoop\temp.”
   5. Create a log folder, such as “D:\hadoop\userlog”
   6. Go to Hadoop-2.6.x etc  Hadoop and edit four files:
      1. core-site.xml
      2. hdfs-site.xml
      3. mapred.xml
      4. 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>

* 1. 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
  2. Set environmental variables: Do: My computer  Properties  Advance system settings 

Advanced  Environmental variables

* 1. User variables:
* Variable: HADOOP\_HOME
* Value: D:\hadoop-2.6.0

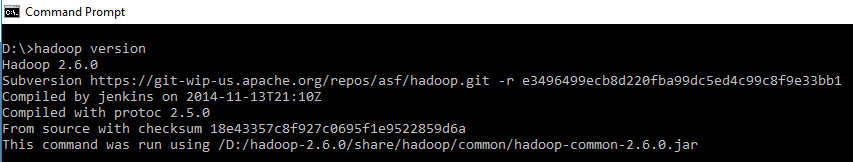
1. System variable
   * Variable: Path
   * Value: D:\hadoop-2.6.0\bin

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\\*

1. Check on cmd; see below.



1. **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”
2. 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”

