Assignment – 7

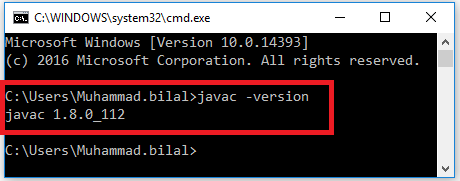
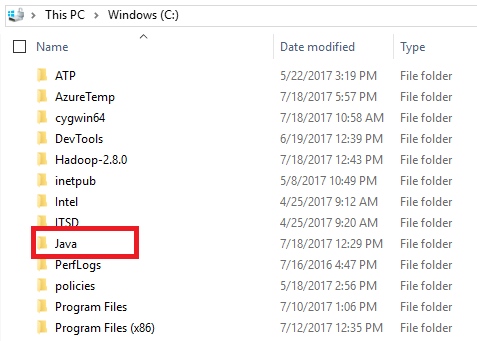
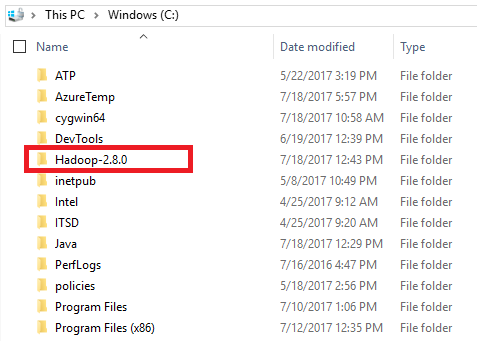
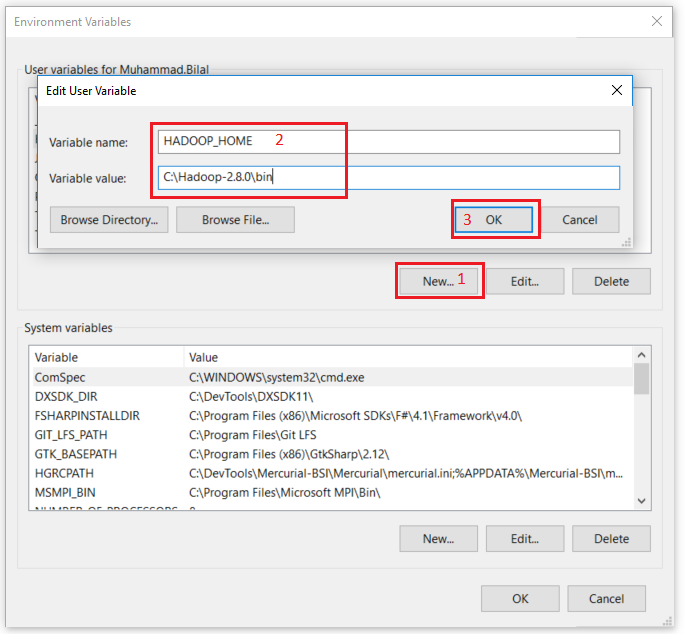
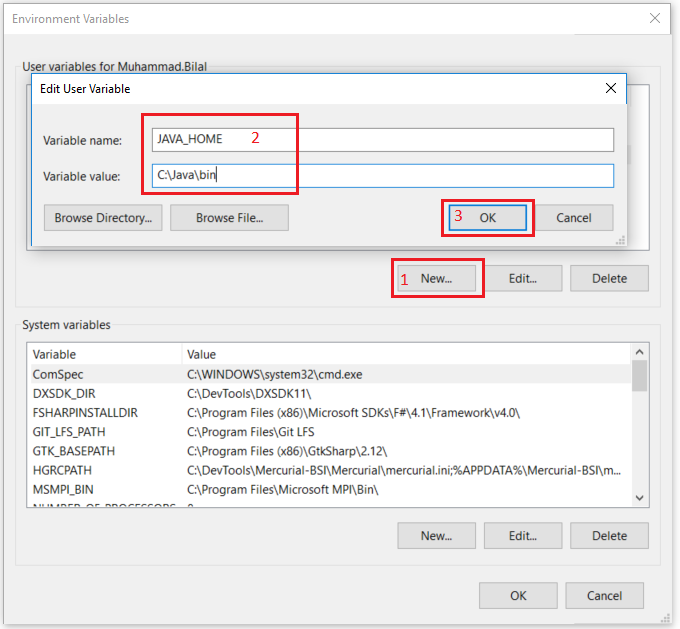
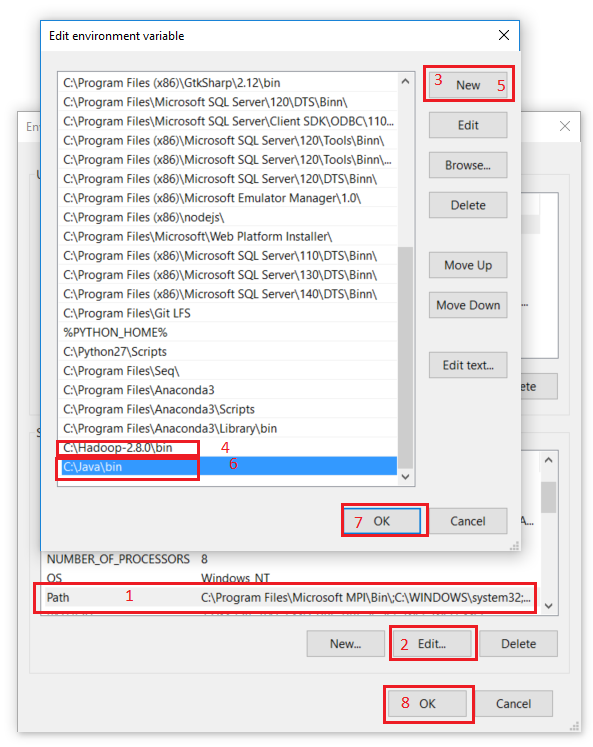
Step by step Hadoop 2.8.0 installation on Window 10

**Prepare:**

These softwares should be prepared to install Hadoop 2.8.0 on window 10 64bit

1. Download Hadoop 2.8.0 (Link: <http://www-eu.apache.org/dist/hadoop/common/hadoop-2.8.0/hadoop-2.8.0.tar.gz> OR [http://archive.apache.org/dist/hadoop/core//hadoop-2.8.0/hadoop-2.8.0.tar.gz](http://archive.apache.org/dist/hadoop/core/hadoop-2.8.0/hadoop-2.8.0.tar.gz))
2. Java JDK 1.8.0.zip (Link: <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>)

**Set up**

1. Check either Java 1.8.0 is already installed on your system or not, use **"Javac -version"** to check. 
2. If Java is not installed on your system then first install java under **"C:\JAVA"** 
3. Extract file Hadoop 2.8.0.tar.gz or Hadoop-2.8.0.zip and place under **"C:\Hadoop-2.8.0"**. 
4. Set the path HADOOP\_HOME Environment variable on windows 10(see Step 1,2,3 and 4 below). 
5. Set the path JAVA\_HOME Environment variable on windows 10(see Step 1,2,3 and 4 below). 
6. Next we set the Hadoop bin directory path and JAVA bin directory path. 

**Configuration**

1. Edit file **C:/Hadoop-2.8.0/etc/hadoop/core-site.xml**, paste below xml paragraph and save this file.

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

1. Rename "mapred-site.xml.template" to "mapred-site.xml" and edit this file **C:/Hadoop-2.8.0/etc/hadoop/mapred-site.xml**, paste below xml paragraph and save this file.

<configuration>

<property>

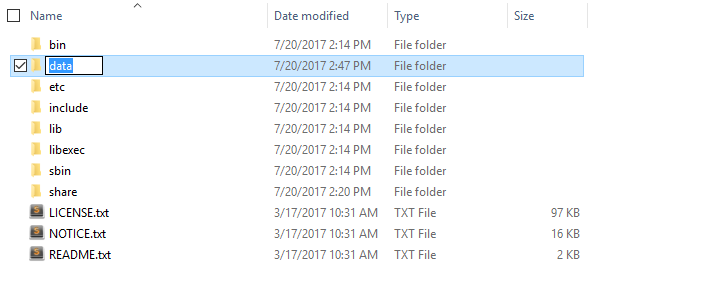
<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

1. Create folder **"data"** under **"C:\Hadoop-2.8.0"**

* Create folder **"datanode"** under **"C:\Hadoop-2.8.0\data"**
* Create folder **"namenode"** under **"C:\Hadoop-2.8.0\data"** 

1. Edit file **C:\Hadoop-2.8.0/etc/hadoop/hdfs-site.xml**, paste below xml paragraph and save this file.

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>C:\hadoop-2.8.0\data\namenode</value>

</property>

<property>

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

<value>C:\hadoop-2.8.0\data\datanode</value>

</property>

</configuration>

1. Edit file **C:/Hadoop-2.8.0/etc/hadoop/yarn-site.xml**, paste below xml paragraph and save this file.

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

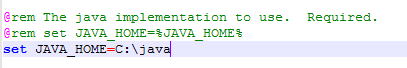
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

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

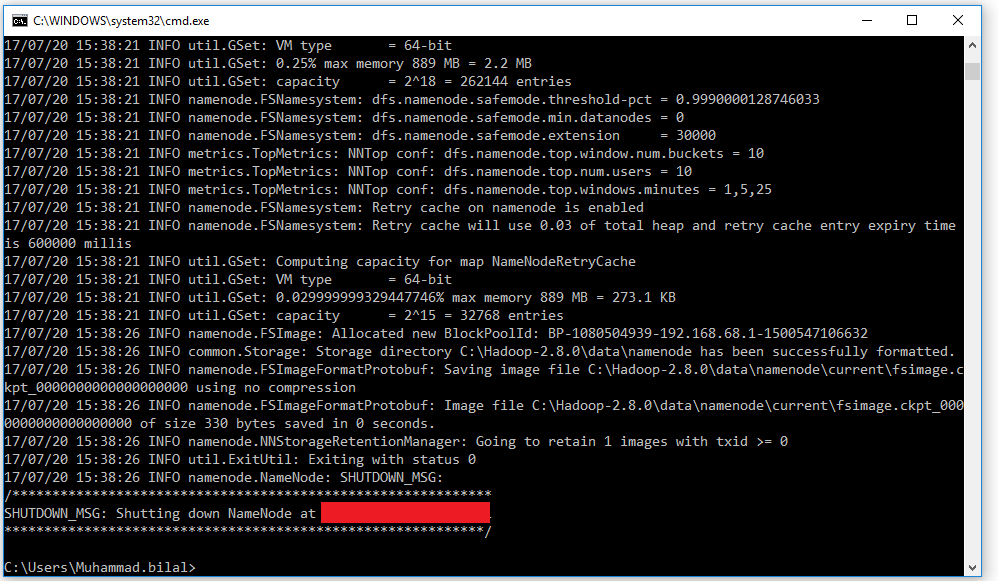
</property>

</configuration>

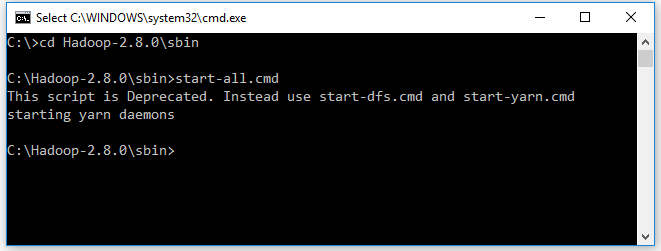
1. Edit file **C:/Hadoop-2.8.0/etc/hadoop/hadoop-env.cmd** by closing the command line**"JAVA\_HOME=%JAVA\_HOME%"** instead of set **"JAVA\_HOME=C:\Java"** (On C:\java this is path to file jdk.18.0)

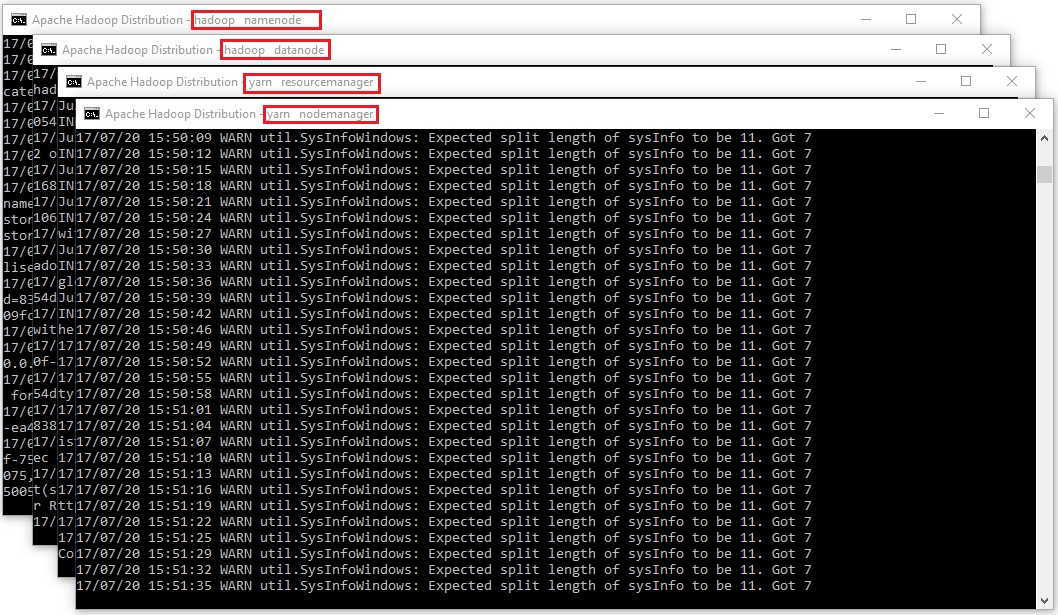


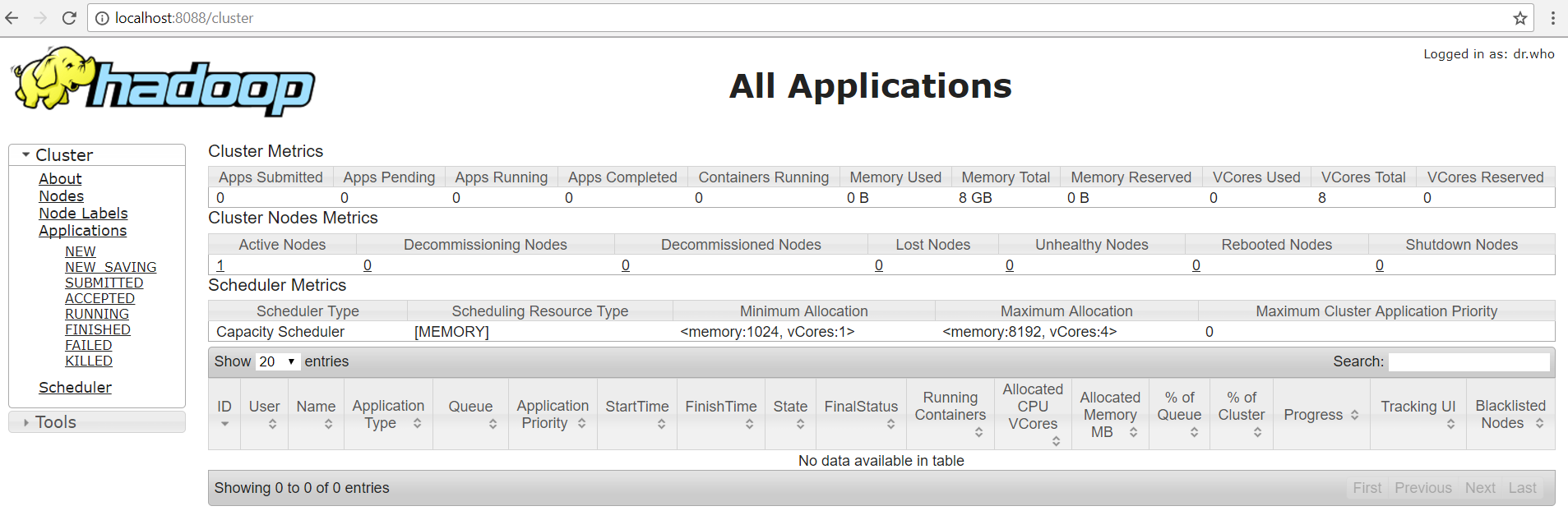
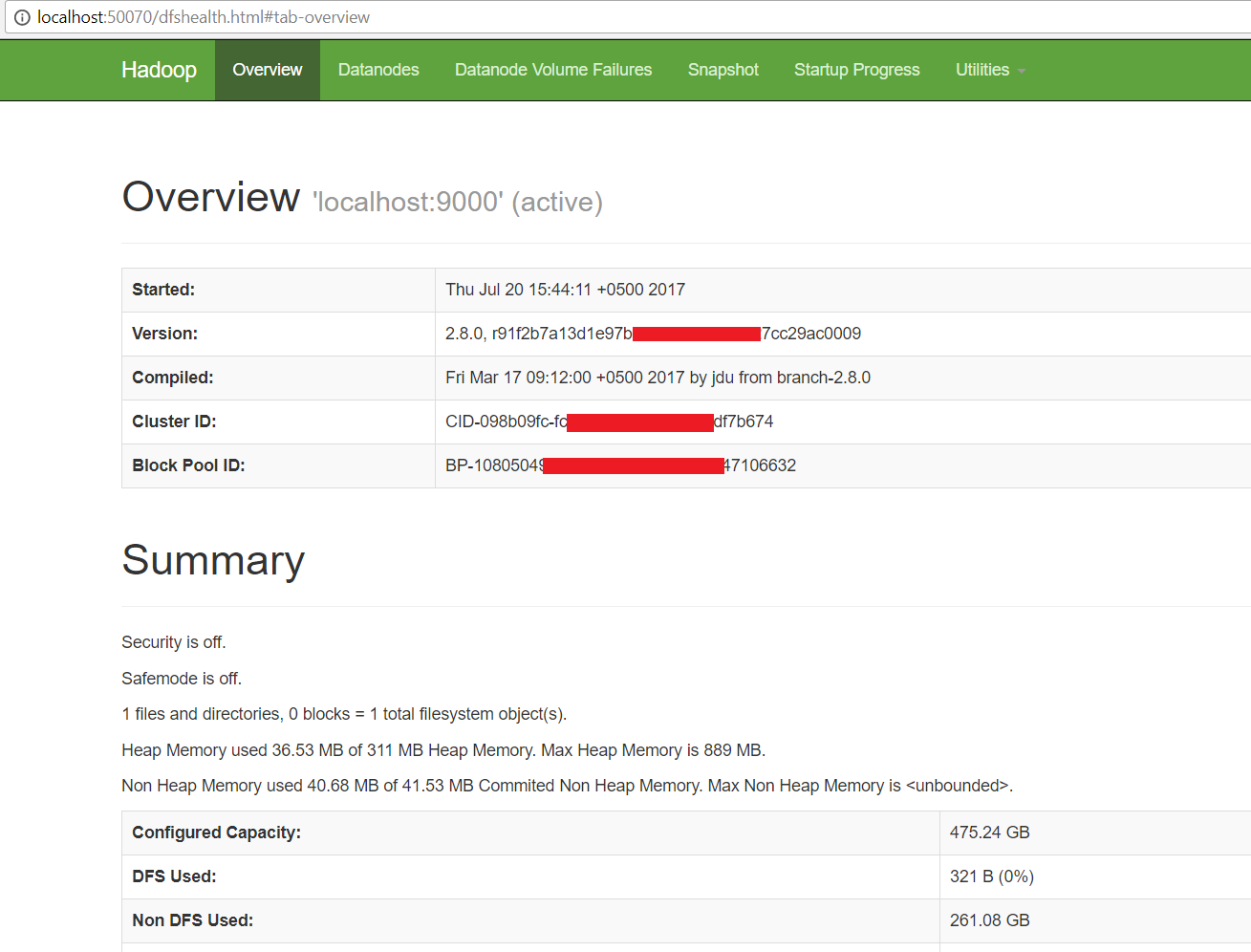
**Hadoop Configuration**

1. Dowload file Hadoop Configuration.zip (Link: <https://github.com/MuhammadBilalYar/HADOOP-INSTALLATION-ON-WINDOW-10/blob/master/Hadoop%20Configuration.zip>)
2. Delete file bin on C:\Hadoop-2.8.0\bin, replaced by file bin on file just download (from Hadoop Configuration.zip).
3. Open cmd and typing command **"hdfs namenode –format"** . You will see 

**Testing**

1. Open cmd and change directory to "C:\Hadoop-2.8.0\sbin" and type **"start-all.cmd"** to start apache. 
2. Make sure these apps are running

* Hadoop Namenode
* Hadoop datanode
* YARN Resourc Manager
* YARN Node Manager 

1. Open: [http://localhost:8088](http://localhost:8088/) 
2. Open: [http://localhost:50070](http://localhost:50070/) 

**Congratulations, Hadoop installed.**

# Basic Hadoop Admin Commands

(Taken from [Hadoop Wiki](http://wiki.apache.org/hadoop/GettingStartedWithHadoop)'s *Getting Started with Hadoop*):

The ~/hadoop/bin directory contains some scripts used to launch Hadoop DFS and Hadoop Map/Reduce daemons. These are:

* **start-all.sh** - Starts all Hadoop daemons, the namenode, datanodes, the jobtracker and tasktrackers.
* **stop-all.sh** - Stops all Hadoop daemons.
* **start-mapred.sh** - Starts the Hadoop Map/Reduce daemons, the jobtracker and tasktrackers.
* **stop-mapred.sh** - Stops the Hadoop Map/Reduce daemons.
* **start-dfs.sh** - Starts the Hadoop DFS daemons, the namenode and datanodes.
* **stop-dfs.sh** - Stops the Hadoop DFS daemons.