Experiment No:3.1

Student Name: Kunal Ghode **Branch: BE-CSE**

UID:21BCS2610 **Subject Name: CC&DS LAB**

Section/Group: CC_612-B **Date of Performance:18/03/24**

Subject Code: 21CSP-378

Aim: Install Hadoop single node cluster and run simple applications like word count.

Hadoop framework is well comportable in the Linux environment but for the users who are not familiar with Linux environment but want to use the hadoop framework can be make use of this article. This article is aim to Install hadoop single node cluster and run simple application like wordcout.

Procedure:

1. Install Java
2. Configure and install hadoop
3. Test hadoop installation
4. Create wordcount program
5. Input file to mapreduce
6. Display the output

I. JAVA Installation

1. Go to official Java Downloading page https://www.oracle.com/java/technologies/javase-jre8-downloads.html 1. After downloading java, run the **jdk-8u241-windows-x64.exe** file
2. Follow the instructions and click next.
3. After finishing the installation it is need to set Java environment variable
4. Go to Start->Edit the System environment variable->Environment variable
5. Then Click new and enter variable name as “JAVA_HOME” 6. In the value field Enter the java path such as

“C:\Java\jdk1.8.0_241”(Consider your installation folder)

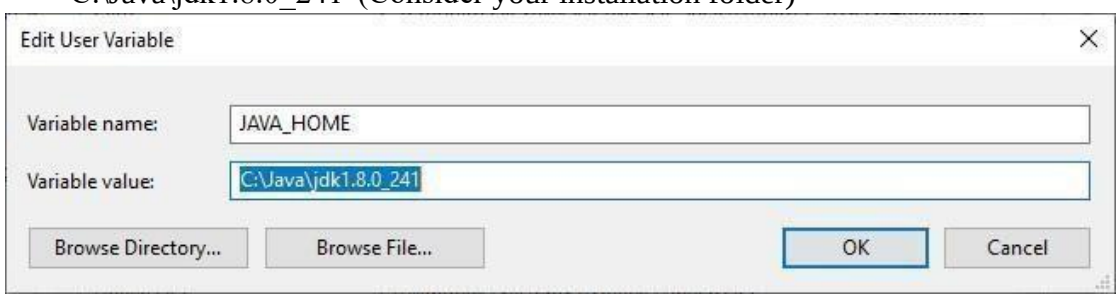

Fig-3.1

7. Go to path and click edit then type “%JAVA_HOME%\bin”

Fig-3.2

8 . Then click Ok and Go to Command Prompt

9. Type “Java -version”. If it prints the installed version of java, now java successfully installed in your System.

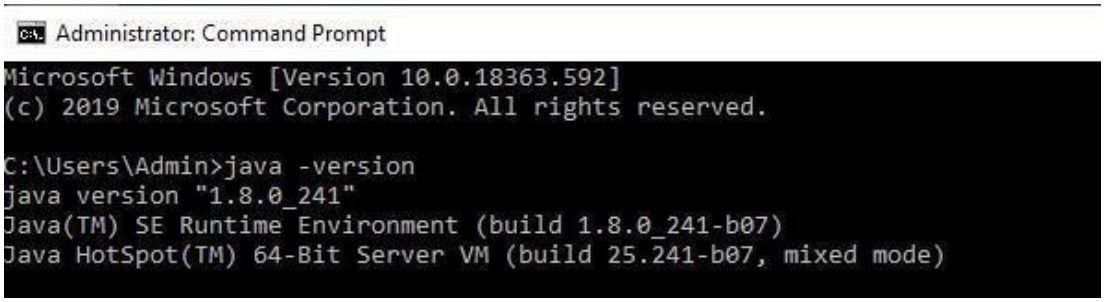

Fig-3.3

II Configuring And Installing Hadoop 1. Download Hadoop 2.8.0 from http://archive.apache.org/dist/hadoop/core//hadoop-2.8.0/hadoop- 2.8.0.tar.gz) 2. Extract the tar file (in my case I used **7-zip** to extract the file and I stored the extracted file in the **D:\hadoop**)

3. After finishing the extraction it is need to set Hadoop environment variable
4. Go to Start->Edit the System environment variable->Environment variable
5. Then Click new and enter variable name as “HADOOP_HOME”
6. In the value field Enter the java path such as “D:\hadoop”(Consider your installation folder)

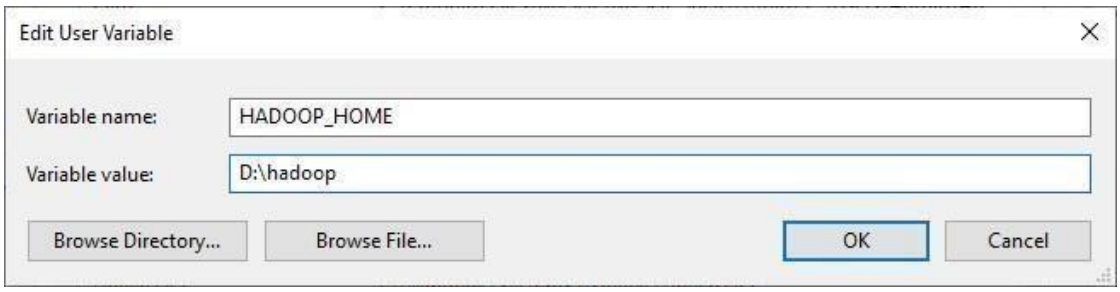

Fig-3.4

7. Go to path and click edit then type “%HADOOP_HOME%\bin”

Fig-3.6

8. Now we have to configure the hadoop.

9. Go to D:/hadoop/etc/hadoop/.. folder, find the below mentioned files and paste the following.

i. core-site.xml

```
<configuration> <property> <name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value> </property>
</configuration>
```

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

```
<configuration>  <property> &https://www.linkedin.com/redir/phishing-
page?url=lt%3Bname%26gt%3Bmapreduce%2eframework%2ename</name>
 <value>yarn</value>  </property>
  </configuration>
```

iii. Create folder "data" under "D:\Hadoop"

- Create folder "datanode" under "D:\Hadoop\data"
- Create folder "namenode" under "D:\Hadoop\data" data

iv. hdfs-site.xml

```
 <configuration>  <property> <name>dfs.replication</name>
 <value>1</value>  </property>  <property>
<name>dfs.namenode.name.dir</name>
 <value>D:\hadoop\data\namenode</value>  </property>  <property>
 <name>dfs.datanode.data.dir</name>
   <value>D:\hadoop\data\datanode</value>  </property>
   </configuration>
```

v. yarn-site.xml

```
 <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>
```

vi. Edit file D:\Hadoop\etc\hadoop**hadoop-env**.cmd by closing the command line "JAVA_HOME=%JAVA_HOME%" instead of set "JAVA_HOME= C:\Java\jdk1.8.0_241" (if your java file in Program Files

the instead of give **Progra~1** otherwise you will get JAVA_HOME incorrectly set error)

vii. Download file Hadoop Configuration.zip https://github.com/Prithiviraj2503/hadoop-installation-windows

viii. Delete file bin on D:\Hadoop\bin and replace it by the bin file of Downloaded configuration file (from Hadoop Configuration.zip).

ix. Open cmd and typing command "**hdfs namenode – format**" .You will see through command prompt which tasks are processing, after competeation you will get a massage like namenode format succesfully and shutdown message

hdfs namenode –format

III. Testing Hadoop Installation

1. Open Cmd and type the following "Hadoop -version"

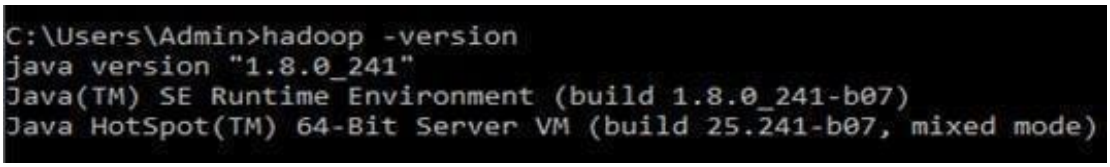

Fig-3.7

2. To start the hadoop locate to "D:\hadoop\sbin" via command prompt and press **start-all.cmd**

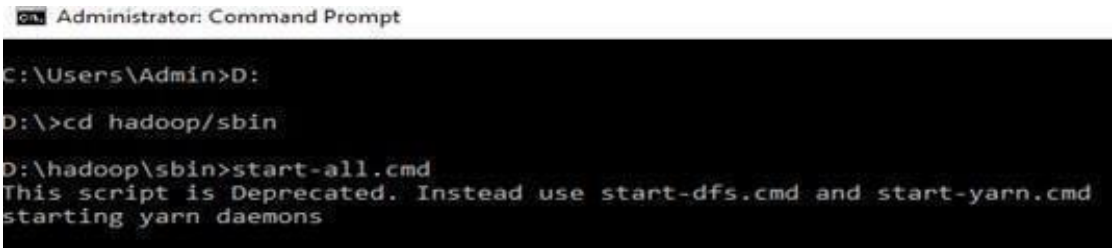

Fig-3.8

Now, you can see the namenode, datanode and yarn engines getting start,

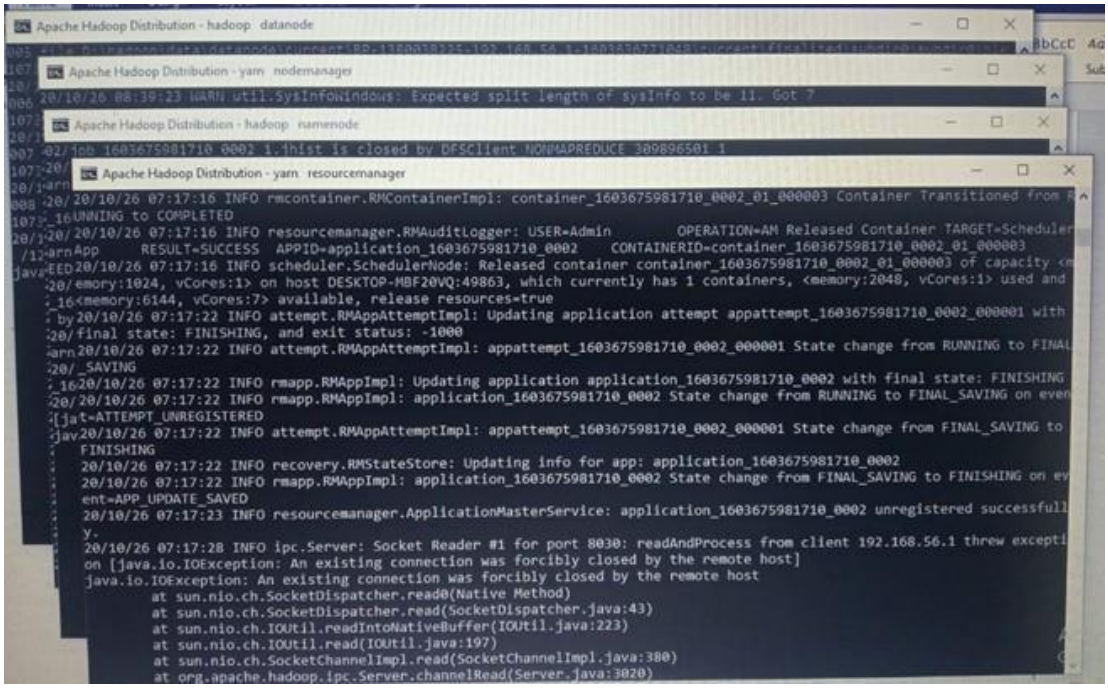

Fig-3.9

3. Now type “jps”. JPS (Java Virtual Machine Process Status Tool) is a command is
used to check all the Hadoop daemons like NameNode, DataNode, ResourceManager, NodeManager etc.

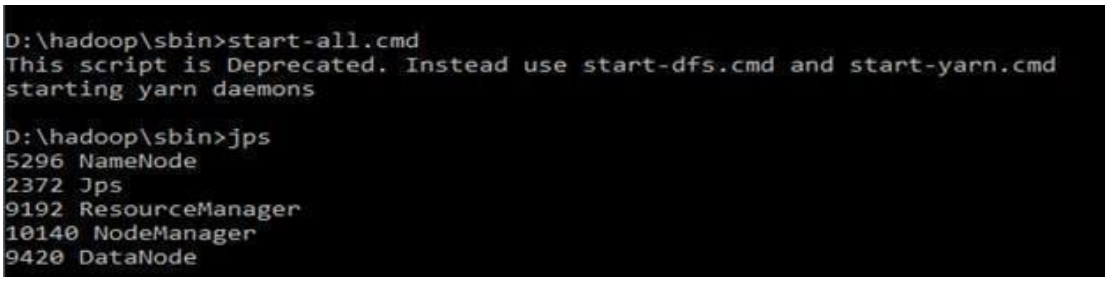

Fig-3.10

4. Open: http://localhost:8088 in any browser

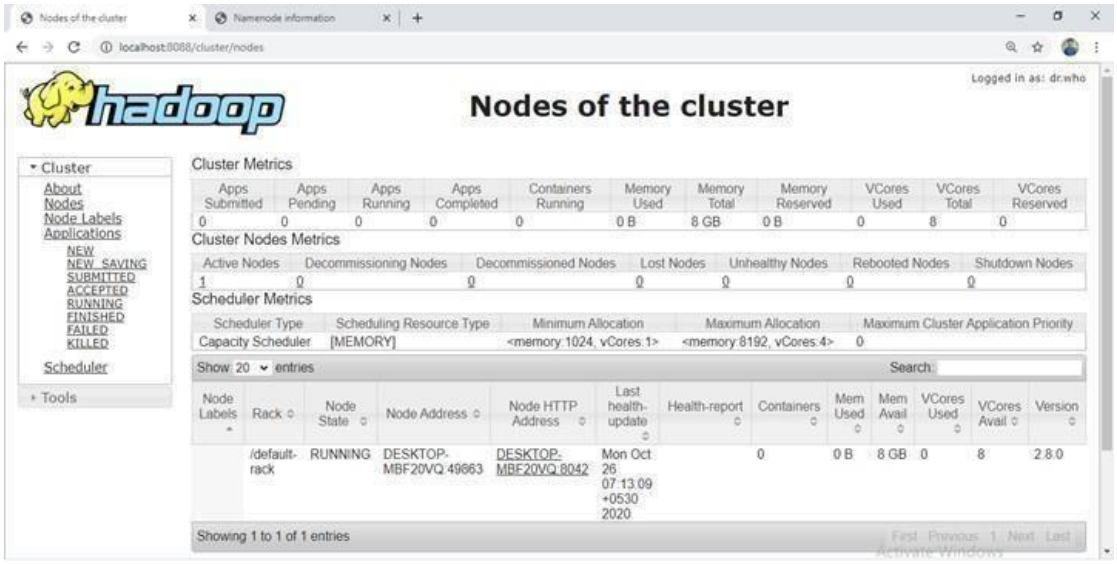

Fig-3.11

5. Open: http://localhost:50070 in any browser

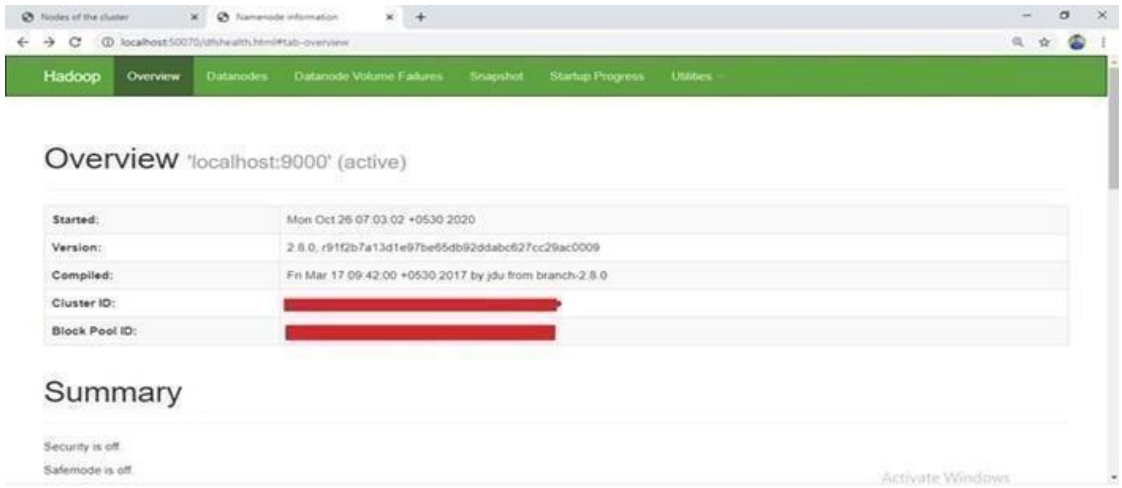

Fig-3.12

Now hadoop succesfully installed in your System.

IV. Simple WordCount Program

1) After successful hadoop installation we need to create an directory in the hadoop file system

2) Start the hadoop via command prompt **$ start-all.cmd**

3) By using **$jps** command Ensure hadoop nodes are running

4) To create a directory, use: **$ hadoop fs –mkdir /inputdir**

5) To input a file within a directory, use: **$ hadoop fs – put D:/input_file.txt/inputdir**

6) To ensure wether your file succesfully imported, use: **$ hadoop fs –ls /inputdir/**

7) To view the content of the file, use: **$ hadoop dfs –cat /inputdir/input_file.txt**

Link for input file : https://github.com/Prithiviraj2503/hadoop-installation- windows

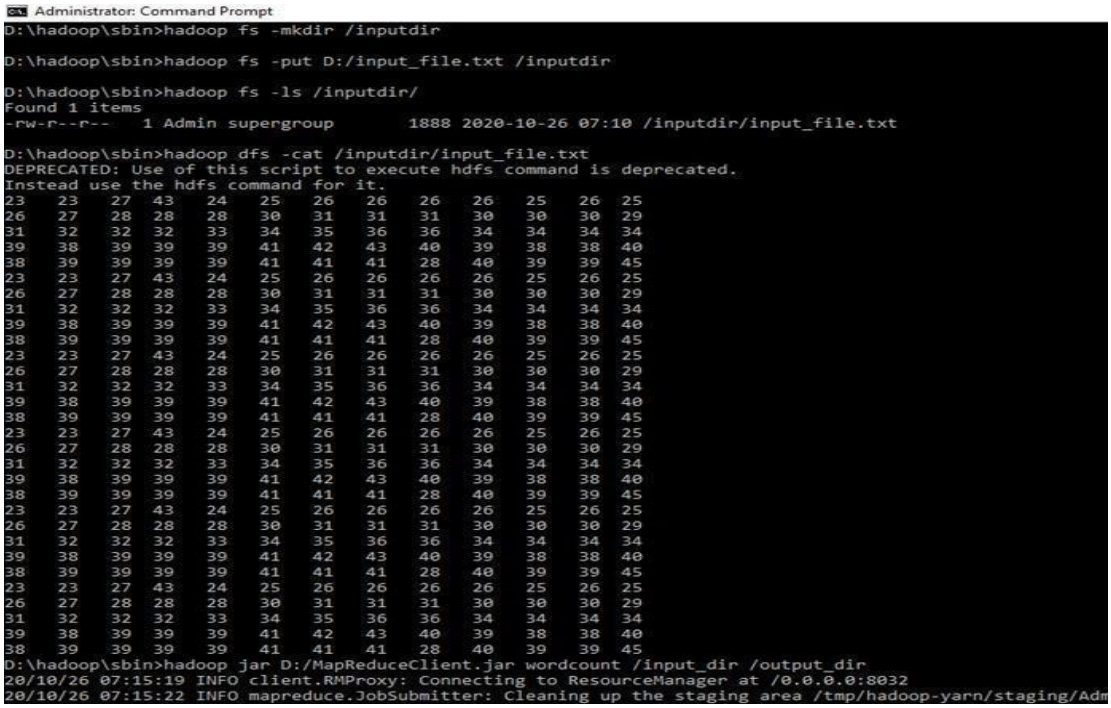

Fig-3.13

8) Now appy mapreduce program to the input file. We have a **mapReduceClient.jar** which contain java mapper and reducer programs. After applying the jar file you can see the task performed in the mapreduce phase.All the resuts of completed tasks will be printed in the command prompt. Link for mapReduceClient.jar : https://github.com/Prithiviraj2503/hadoop-installationwindows

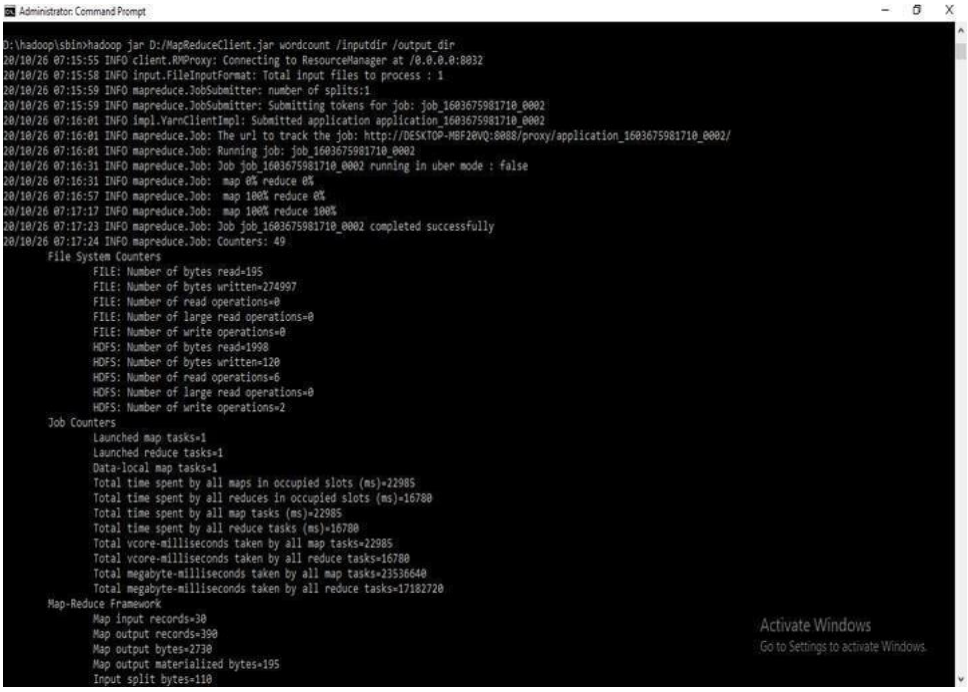

Fig-3.14

9) After completed the mapreduce tasks the output will be stored in the **output_dir** directory To see the output, use: **$ hadoop dfs –cat /output_dir**/

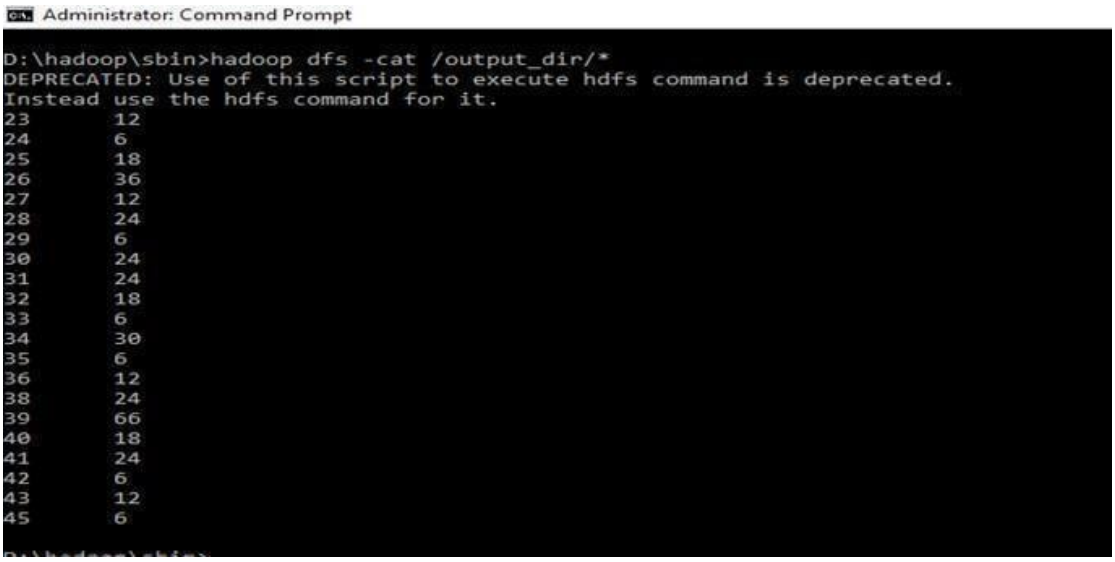

Fig-3.15

10) To stop the hadoop type **$stop-all.cmd**

Now the hadoop single node cluster was installed succesfully and the simple word count program were executed succesfully in your windows system.

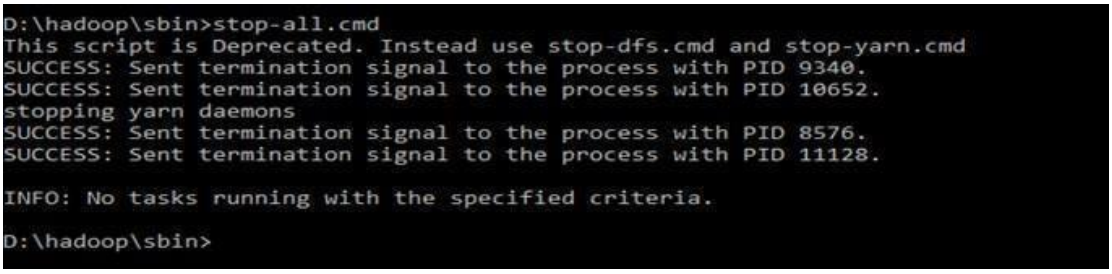

Fig-3.16

Analysis: