Experiment 3.1

|  |  |
| --- | --- |
| **Student Name:** Aaryan Aneja | **UID:** 21BCS9186 |
| **Branch:** BE-CSE | **Section/Group:** CC-646-B |
| **Semester:** 6th | **Date of Performance:** 26/03/24 |
| **Subject Name:** Cloud Computing and | **Subject Code:** 21CSP-378 |

Distributed System

# Aim:

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

# Objective:

The Objective of this to Install Hadoop single node cluster and run simple applications like word count.

# Procedure:

Install Java

Configure and install hadoop Test hadoop installation Create wordcount program Input file to mapreduce Display the output

# Steps to Install Hadoop and run simple applications:

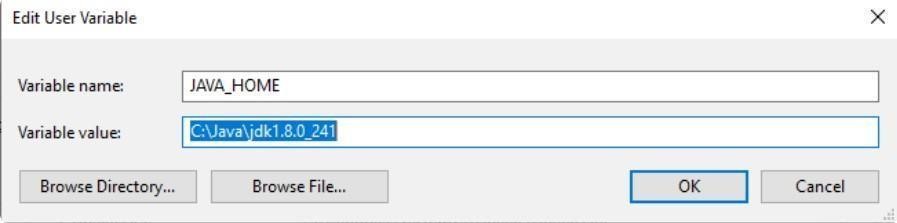
**Step 1:** JAVA Installation

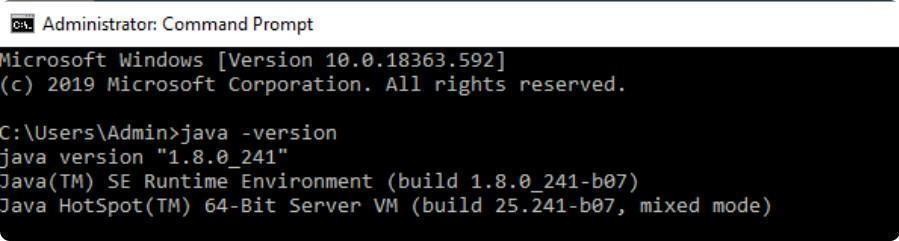
Go to official Java Downloading page [https://www.oracle.com/java/technologies/javase-](https://www.oracle.com/java/technologies/javase-jre8-downloads.html) [jre8downloads.html](https://www.oracle.com/java/technologies/javase-jre8-downloads.html)

After downloading java, run the jdk-8u241-windows-x64.exe file Follow the instructions and click next

After finishing the installation it is need to set Java environment variable

Go to Start->Edit the System environment variable->Environment variable

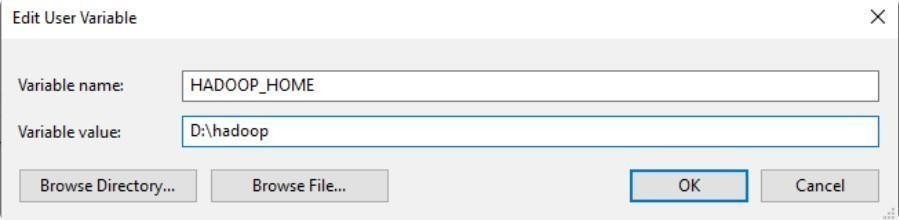


Go to path and click edit then type “%JAVA\_HOME%\bin” Then click Ok and Go to Command Prompt

**Step 2:** Configuring And Installing Hadoop

Download Hadoop 2.8.0 from [http://archive.apache.org/dist/hadoop/core//hadoop-](http://archive.apache.org/dist/hadoop/core/hadoop-) 2.8.0/hadoop- 2.8.0.tar.gz)

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)



Go to path and click edit then type “%HADOOP\_HOME%\bin” Now we have to configure the hadoop.

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

i.) Core.site.xml



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.



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

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

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

iv.) 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)

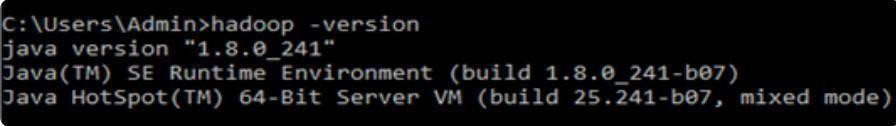
v.) Download file Hadoop Configuration.zip <https://github.com/Prithiviraj2503/hadoop-installation-windows>

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

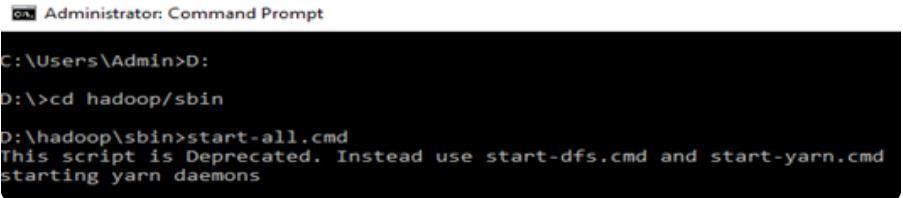
vii.) Open cmd and typing command "hdfs namenode –format" .You will see through command prompt which tasks are processing, after competeation you will get a massage like namenode format succesfully and shutdown message

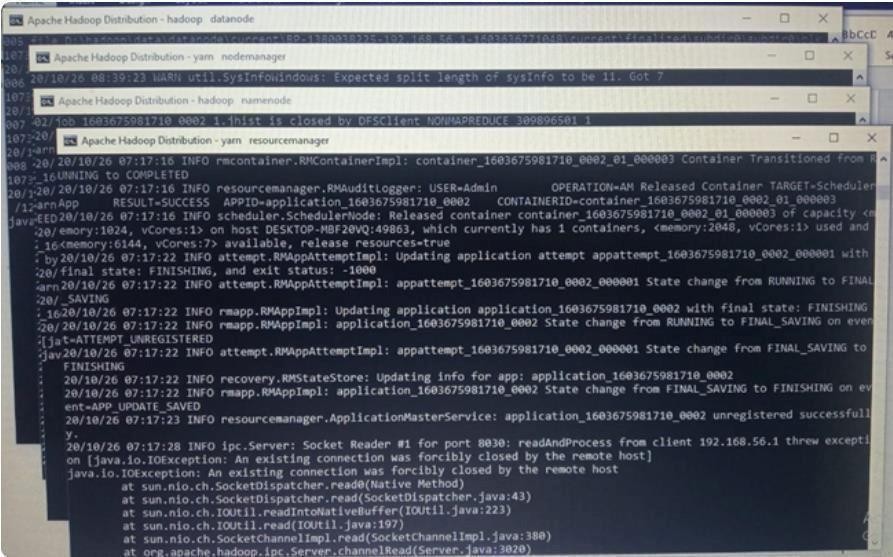


**Step 3:** Testing Hadoop Installation

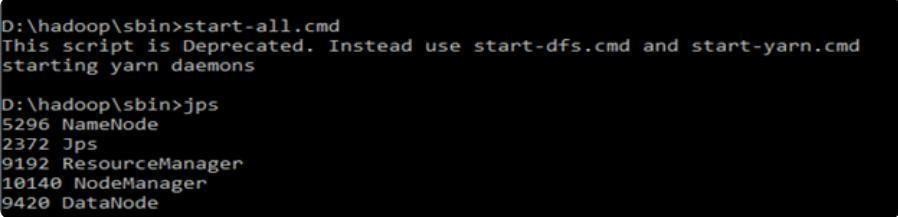
Open Cmd and type the following “Hadoop -version”.

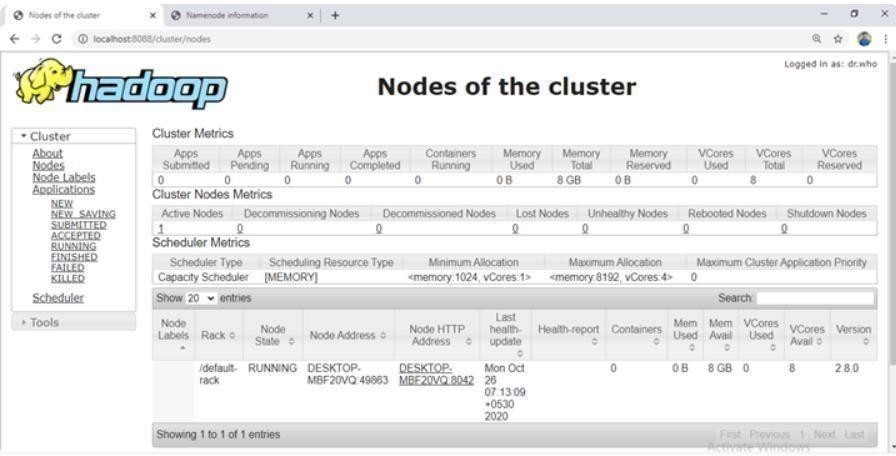
To start the hadoop locate to “D:\hadoop\sbin” via command prompt and press startall.cmd



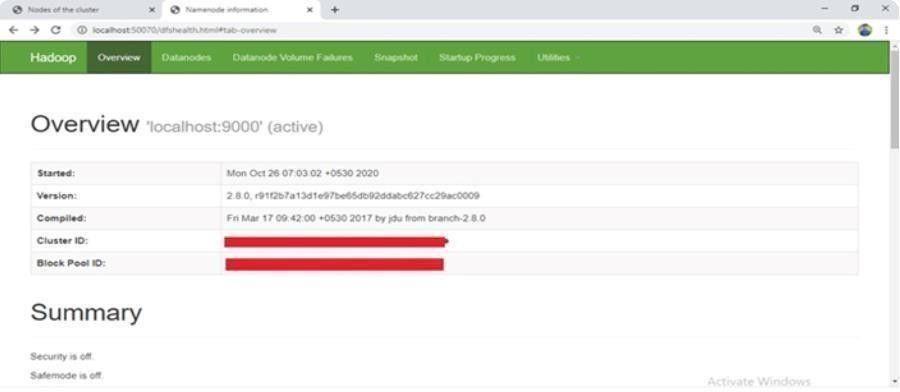
Now, you can see the namenode, datanode and yarn engines getting start,

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.



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

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



Now hadoop succesfully installed in your System.

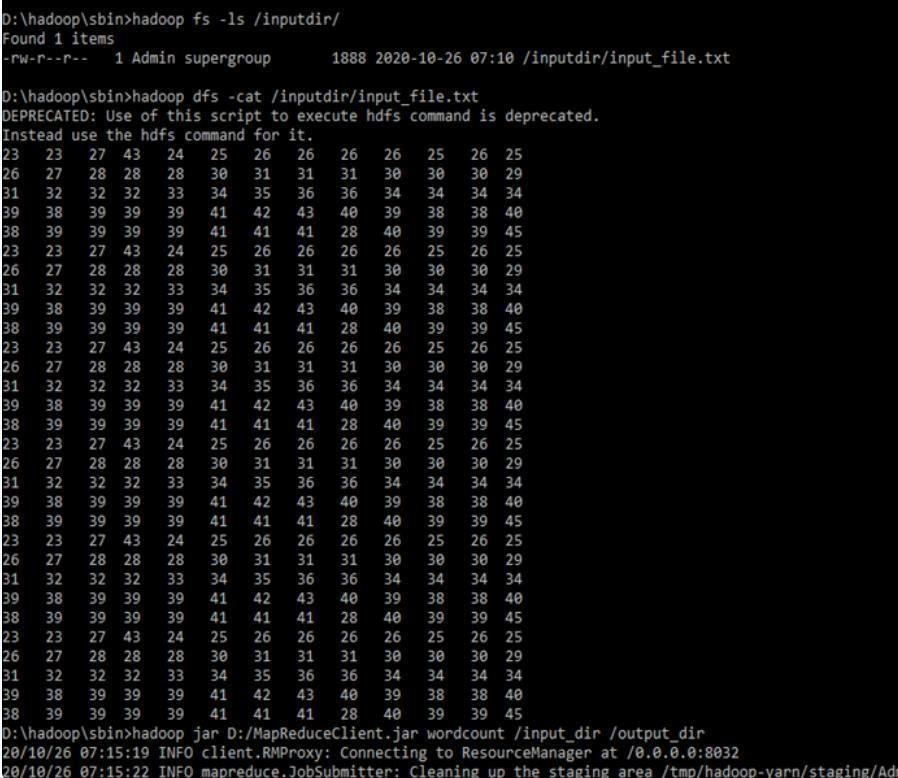
**Step 4:** Simple WordCount Program

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

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

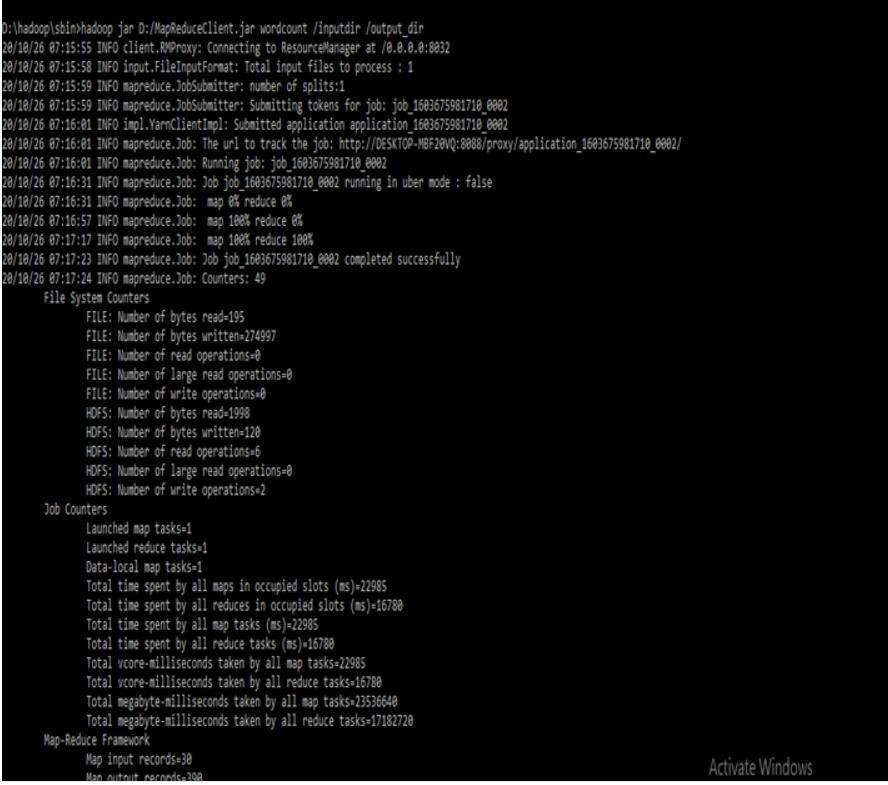
To input a file within a directory, use: $ hadoop fs –put D:/input\_file.txt/inputdir To ensure wether your file succesfully imported, use: $ hadoop fs –ls /inputdir/

To view the content of the file, use: $ hadoop dfs –cat /inputdir/input\_file.txt

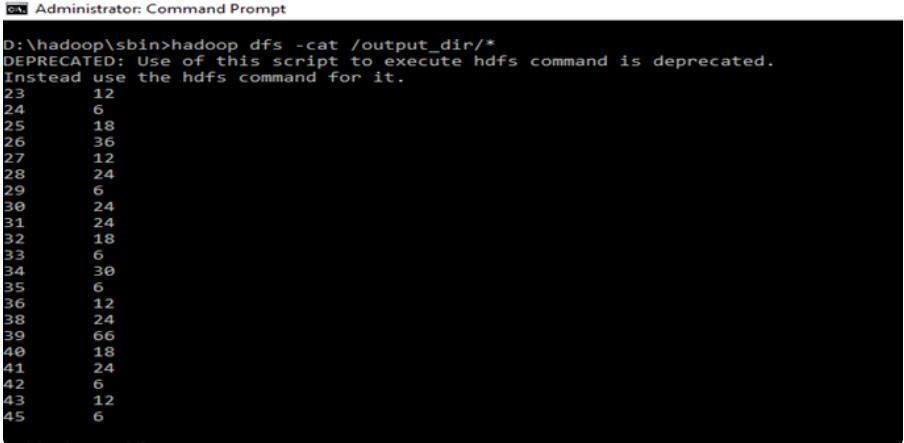
Link for input file : <https://github.com/Prithiviraj2503/hadoop-installation-windows>

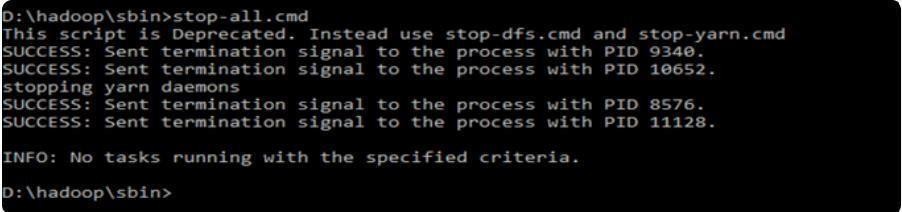
Now appy mapreduce program to the input file. We have a mapReduceClient.jar which contain java mapper and reducer programs. After applying the jar file you can see the task performed in the mapreduce phase.All the resuts of completed tasks will be printed in the command prompt.

Link for mapReduceClient.jar : [https://github.com/Prithiviraj2503/hadoop-](https://github.com/Prithiviraj2503/hadoop-installation-windows) [installationwindows](https://github.com/Prithiviraj2503/hadoop-installation-windows)



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/



To stop the hadoop type $stop-all.cmd

**Result:** Now the Hadoop single node cluster was installed successfully and the simple word count program were executed successfully in your windows system.

# Learning Outcomes:

* 1. Learnt about Hadoop.
  2. Learnt about its Hadoop’s implementation.