



Experiment 3.1

Student Name: Sachin Maurya

Branch: BE-CSE

Semester: 6th

Subject Name: Cloud Computing and
Distributed System

UID: 21BCS1956

Section/Group: CC-615-B

Date of Performance: 26/03/24

Subject Code: 21CSP-378

1. Aim:

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

2. Objective:

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

3. Procedure:

Install Java

Configure and install hadoop

Test hadoop installation

Create wordcount program

Input file to mapreduce

Display the output

4. Steps to Install Hadoop and run simple applications:

Step 1: JAVA Installation

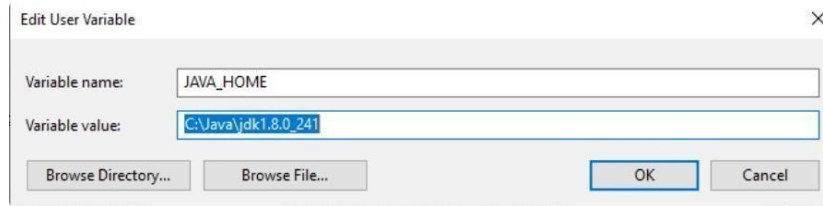
Go to official Java Downloading page <https://www.oracle.com/java/technologies/javase-jre8downloads.html>

After downloading java, run the jdk-8u241-windows-x64.exe file

Follow the instructions and click next

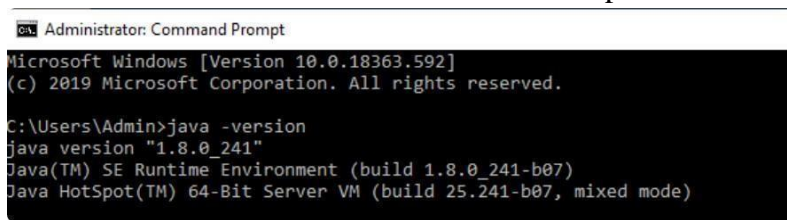
After finishing the installation it is need to set Java environment variable

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



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

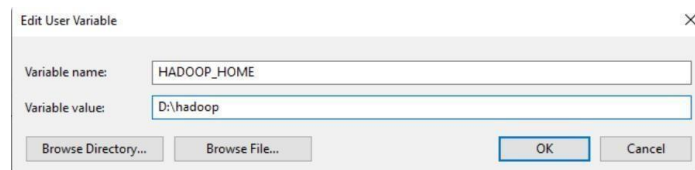
Then click Ok and Go to Command Prompt



Step 2: Configuring And Installing Hadoop

Download Hadoop 2.8.0 from <http://archive.apache.org/dist/hadoop/core/hadoop-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



```
<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 file D:\Hadoop\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>
```

- 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 completion you will get a message like namenode format successfully and shutdown message

```
hdfs namenode -format
```

Step 3: Testing Hadoop Installation

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

```
C:\Users\Admin>hadoop -version
java version "1.8.0_241"
Java(TM) SE Runtime Environment (build 1.8.0_241-b07)
Java HotSpot(TM) 64-Bit Server VM (build 25.241-b07, mixed mode)
```

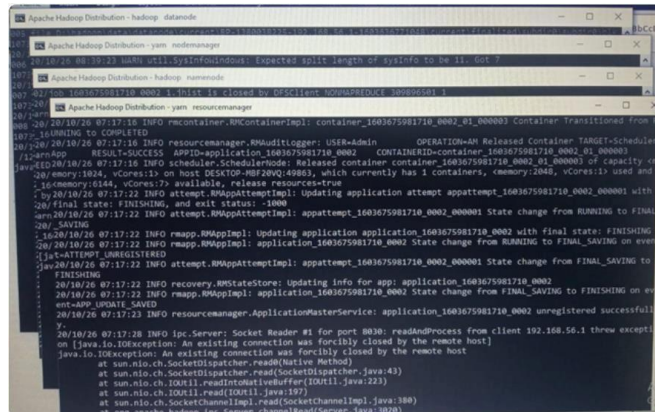
To start the hadoop locate to “D:\hadoop\sbin” via command prompt and press startall.cmd

```
Administrator: Command Prompt

C:\Users\Admin>D:
D:\>cd hadoop\sbin

D:\hadoop\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
```

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

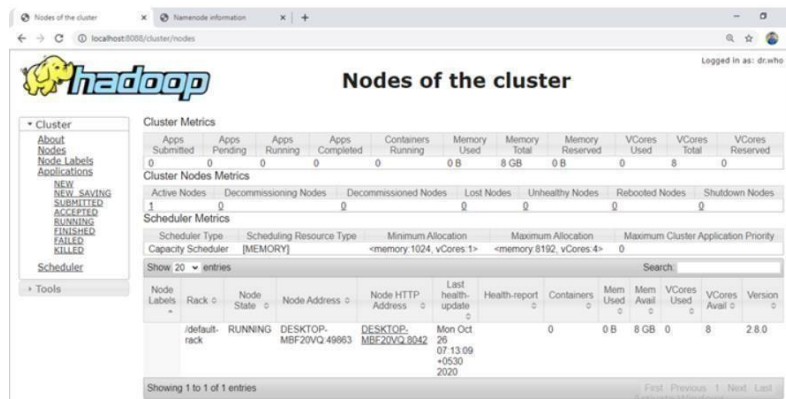


Now type “jps”. JPS (Java Virtual Machine Process Status Tool) is a command is used to check all the Hadoop daemons like NameNode, DataNode, ResourceManager, NodeManager etc.

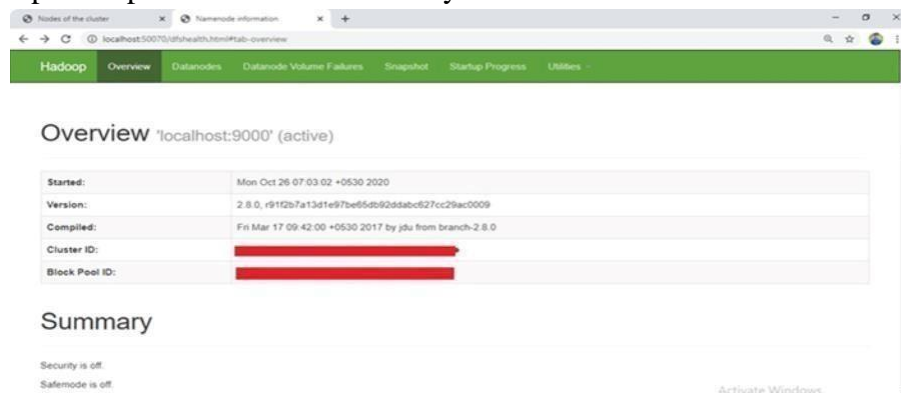
```
D:\hadoop\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop\sbin>jps
5296 NameNode
2372 Jps
9192 ResourceManager
10140 NodeManager
9420 DataNode
```

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>

```
D:\hadoop\sbin>hadoop fs -ls /inputdir/
Found 1 items
-rw-r--r-- 1 Admin supergroup 1888 2020-10-26 07:10 /inputdir/input_file.txt

D:\hadoop\sbin>hadoop dfs -cat /inputdir/input_file.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45
23 23 27 43 24 25 26 26 26 26 25 26 25
26 27 28 28 28 30 31 31 31 30 30 30 29
31 32 32 32 33 34 35 36 36 34 34 34 34
39 38 39 39 39 41 42 43 40 39 38 38 40
38 39 39 39 39 41 41 41 28 40 39 39 45

D:\hadoop\sbin>hadoop jar D:\MapReduceClient.jar wordcount /input_dir /output_dir
20/10/26 07:15:19 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/10/26 07:15:22 INFO mapreduce.JobSubmitter: Cleaning up the staging area /tmp/hadoop-varn/staging/area
```

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

Link for mapReduceClient.jar : <https://github.com/Prithiviraj2503/hadoop-installationwindows>

```

20/10/26 07:15:55 INFO ClientRMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/10/26 07:15:58 INFO InputFileInputFormat: Total input files to process : 1
20/10/26 07:15:59 INFO mapreduce.JobSubmitter: number of splits:1
20/10/26 07:15:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603675981710_0002
20/10/26 07:16:01 INFO Impl.VarnClientImpl: Submitted application application_1603675981710_0002
20/10/26 07:16:02 INFO mapreduce.Job: The url to track the job: http://GDSKTOP-WBF2N0VQ-0080/proxy/application_1603675981710_0002/
20/10/26 07:16:01 INFO mapreduce.Job: Running job: job_1603675981710_0002
20/10/26 07:16:31 INFO mapreduce.Job: Job job_1603675981710_0002 running in uber mode : false
20/10/26 07:16:31 INFO mapreduce.Job: map 0% reduce 0%
20/10/26 07:16:57 INFO mapreduce.Job: map 100% reduce 0%
20/10/26 07:17:17 INFO mapreduce.Job: map 100% reduce 100%
20/10/26 07:17:23 INFO mapreduce.Job: Job job_1603675981710_0002 completed successfully
20/10/26 07:17:24 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=195
  FILE: Number of bytes written=274997
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1598
  HDFS: Number of bytes written=100
  HDFS: Number of read operations=0
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Write-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=22995
  Total time spent by all reduces in occupied slots (ms)=16170
  Total time spent by all map tasks (ms)=22995
  Total time spent by all reduce tasks=16170
  Total score-milliseonds taken by all map tasks=22995
  Total score-milliseonds taken by all reduce tasks=16170
  Total megabyte-milliseonds taken by all map tasks=23536648
  Total megabyte-milliseonds taken by all reduce tasks=17182728

MapReduce Framework
  Map input records=30

```


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/

```
Administrator: Command Prompt
D:\hadoop\sbin>hadoop dfs -cat /output_dir/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
23      12
24      6
25      18
26      36
27      12
28      24
29      6
30      24
31      24
32      18
33      6
34      30
35      6
36      12
38      24
39      66
40      18
41      24
42      6
43      12
45      6
```

To stop the hadoop type \$stop-all.cmd

```
D:\hadoop\sbin>stop-all.cmd
This script is Deprecated. Instead use stop-dfs.cmd and stop-yarn.cmd
SUCCESS: Sent termination signal to the process with PID 9340.
SUCCESS: Sent termination signal to the process with PID 10652.
stopping yarn daemons
SUCCESS: Sent termination signal to the process with PID 8576.
SUCCESS: Sent termination signal to the process with PID 11128.

INFO: No tasks running with the specified criteria.
D:\hadoop\sbin>
```

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.