

Experiment 3.1

Student Name: Manish Singh Barolia **UID:** 21BCS5712
Branch: BE-CSE **Section/Group:** NTPP_CC_601/A
Semester: 6 **Date of Performance:** 25/03/24
Subject Name: CC & DS lab **Subject Code:** 21CSP-378

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

2. Objectives:

- To install the Hadoop single one cluster and execute simple applications successfully.

3. Software Requirements:

- Java 8 Package
- Hadoop

4. Description:

Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model.

Single Node Cluster – It Has one DataNode running and setting up all the NameNode, DataNode, Resource Manager, and NodeManager on a single machine. This is used for studying and testing purposes.

5. Procedure:

1. Download the Java 8 Package. Save this file in the home directory and Extract the Java Tar File.

Command: `tar -xvf jdk-8u101-linux-i586.tar.gz`

Download the Hadoop 2.7.3 Package and extract the Hadoop tar File.

2. Setting up the HADOOP_HOME variable and JAVA_HOME variable Use windows environment variable setting for Hadoop Path setting.

Edit User Variable

Variable name:

Variable value:

Edit User Variable

Variable name:

Variable value:

3. Set Hadoop and Java bin directory path

Environment Variables

User variables for CAIN

Variable	Value
HADOOP_HOME	C:\hadoop-3.2.1
JAVA_HOME	C:\Progra~1\Java\jdk1.8.0_321
OneDrive	C:\Users\CAIN\OneDrive
Path	C:\Users\CAIN\anaconda3;C:\Users\CAIN\anaconda3\Library\min...
PhpStorm	C:\Program Files\JetBrains\PhpStorm 2021.1.2\bin;
PyCharm Community Edition	C:\Program Files\JetBrains\PyCharm Community Edition 2021.2.3...
TEMP	C:\Users\CAIN\AppData\Local\Temp
TMP	C:\Users\CAIN\AppData\Local\Temp

System variables

Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
NUMBER_OF_PROCESSORS	4
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\Wi...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64
PROCESSOR_IDENTIFIER	Intel64 Family 6 Model 158 Stepping 9 GenuineIntel

4. Once inside the path variable configuration window, click “new” button and add the following variables.

```
%JAVA_HOME%\bin
```

```
%HADOOP_HOME%\bin
```

5. Hadoop Configuration :

For Hadoop Configuration we need to modify Six files that are listed below-

1. Core-site.xml

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

2. Mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

3. Hdfs-site.xml

```
<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>
```

4. Yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
```

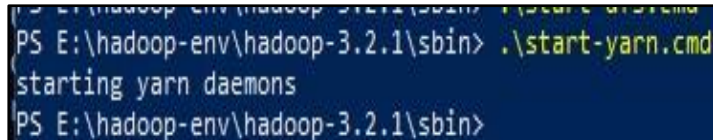
<property>

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

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

</property>

</configuration>



```
PS E:\hadoop-env\hadoop-3.2.1\sbin> .\start-yarn.cmd
starting yarn daemons
PS E:\hadoop-env\hadoop-3.2.1\sbin>
```

5. Hadoop-env.cmd

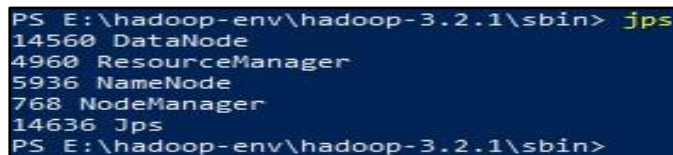
Set "JAVA_HOME=C:\Java" (On C:\java this is path to file jdk.18.0)

6. Create two folders datanode and namenode

1. Create folder "data" under "C:\Hadoop-2.8.0"
2. Create folder "datanode" under "C:\Hadoop-2.8.0\data"
3. Create folder "namenode" under "C:\Hadoop-2.8.0\data"

6. To make sure that all services started successfully, we can run the following command:

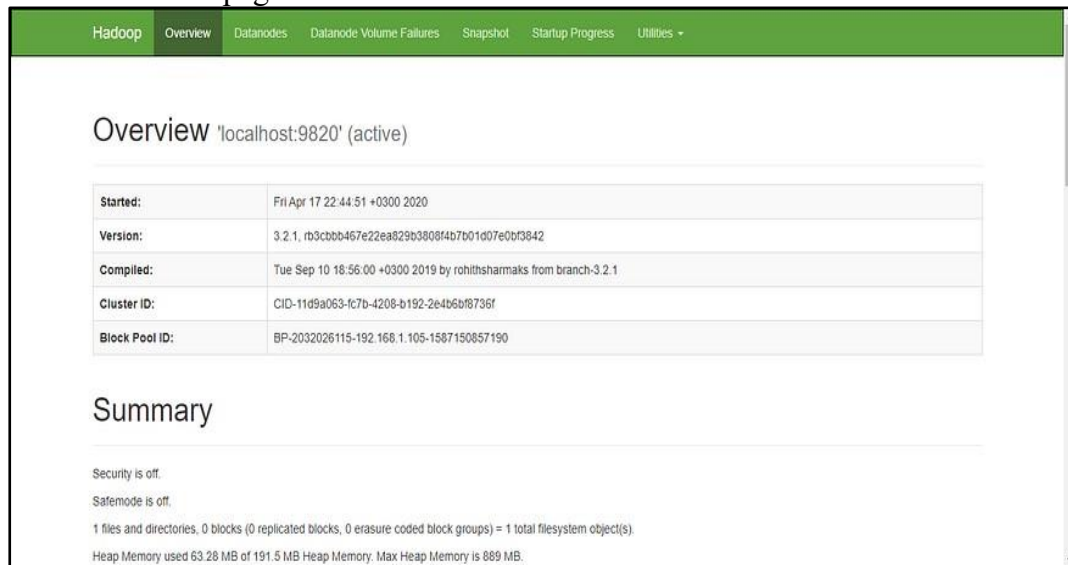
jps



```
PS E:\hadoop-env\hadoop-3.2.1\sbin> jps
14560 DataNode
4960 ResourceManager
5936 NameNode
768 NodeManager
14636 Jps
PS E:\hadoop-env\hadoop-3.2.1\sbin>
```

7. Hadoop Web UI

Name node web page



The screenshot shows the Hadoop Name Node Web Page for 'localhost:9820' (active). The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The 'Overview' tab is selected.

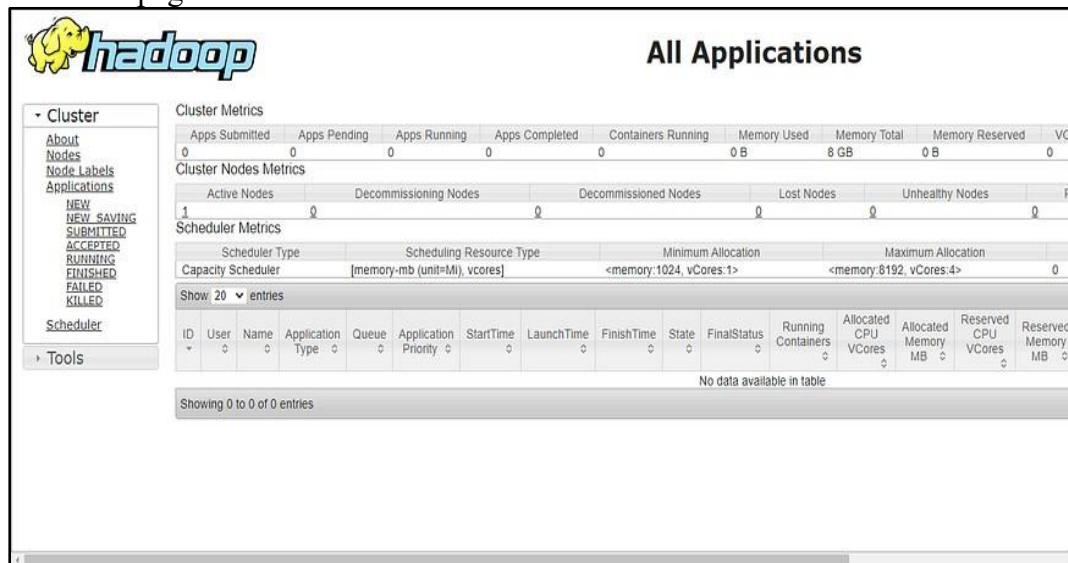
Overview 'localhost:9820' (active)

Started:	Fri Apr 17 22:44:51 +0300 2020
Version:	3.2.1, rd3cbb467e22ea829b3808f4b7b01d07e0b3842
Compiled:	Tue Sep 10 18:56:00 +0300 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-11d9a063-fc7b-4208-b192-2e4b6b78736f
Block Pool ID:	BP-2032026115-192.168.1.105-1587150857190

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 63.28 MB of 191.5 MB Heap Memory. Max Heap Memory is 889 MB.

Yarn web page



The screenshot shows the Hadoop Yarn web page titled 'All Applications'. It features the Hadoop logo and a sidebar with navigation links: Cluster, About, Nodes, Node Labels, Applications, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, and Scheduler. The 'Cluster' link is selected.

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores
0	0	0	0	0	0 B	8 GB	0 B	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Retired Nodes
1	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Showing 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Reserved CPU Vcores	Reserved Memory MB
No data available in table															

Showing 0 to 0 of 0 entries

6. Learning Outcomes:

- Learnt how to install Hadoop single one cluster.
- Learnt how to execute simple applications using Hadoop successfully.