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

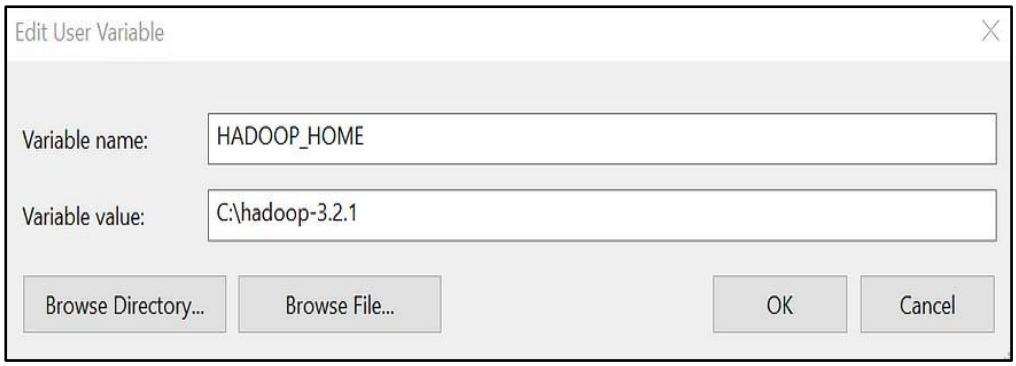
4. **Description:**

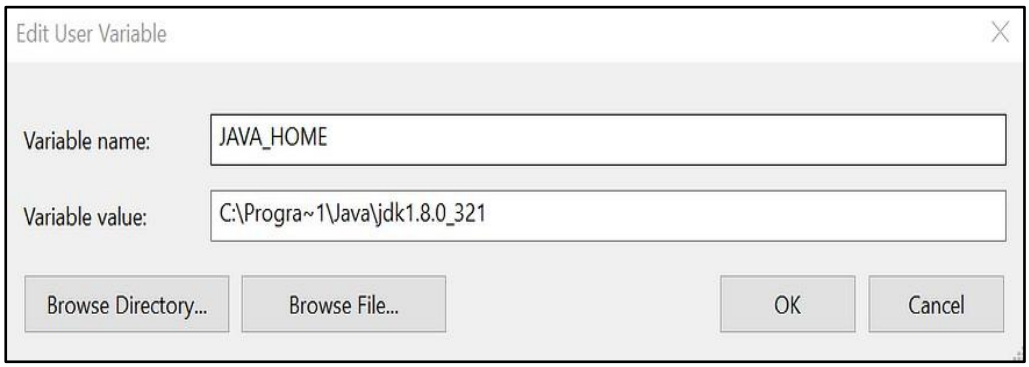
   Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model.

   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.

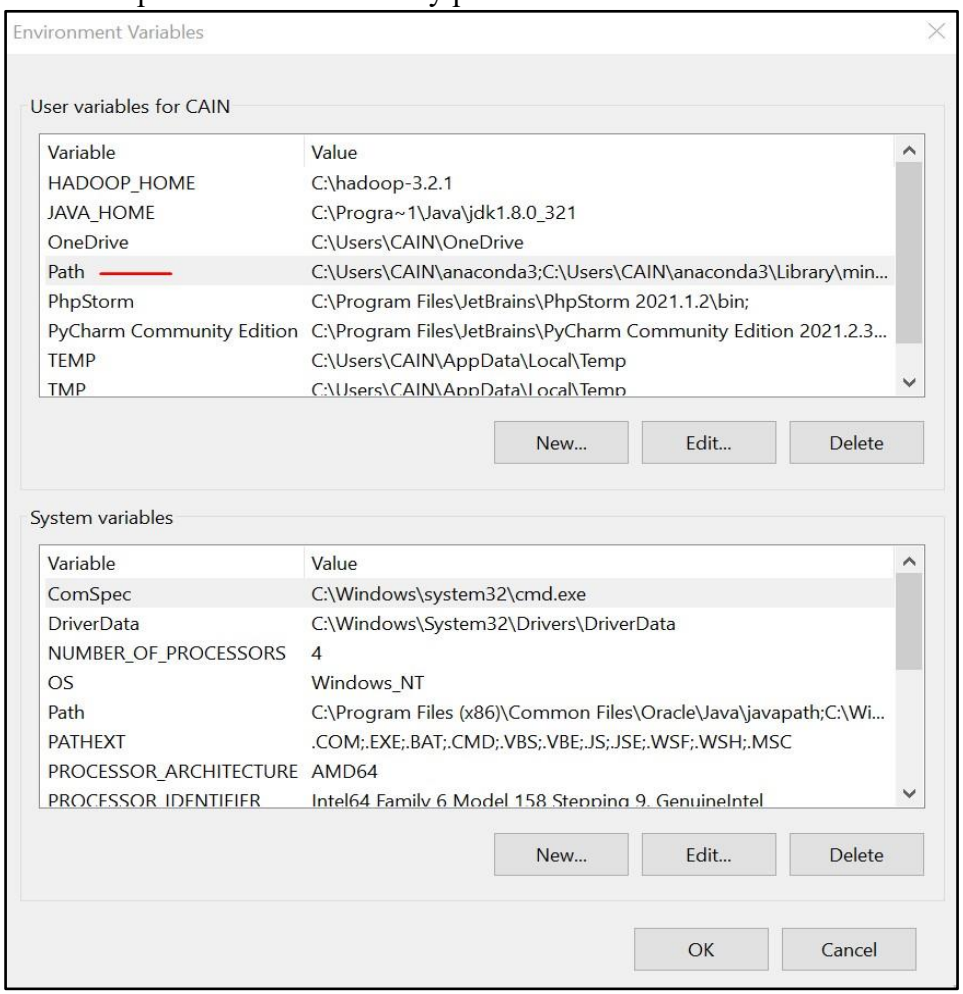
5. **Procedure:**

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.





3. Set Hadoop and Java bin directory path



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>
```

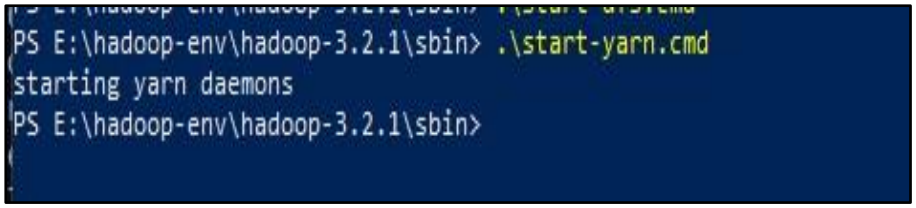
   3. Hdfs-site.xml

```
<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>
```

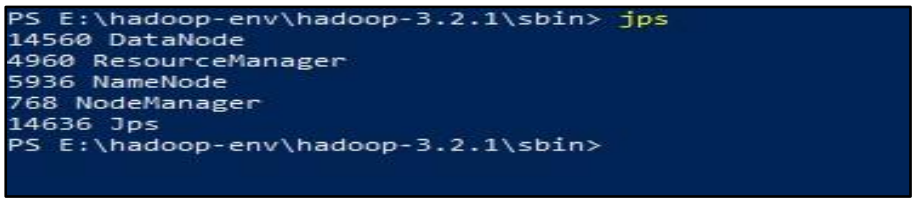
```
  <property>

<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
   <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```



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"

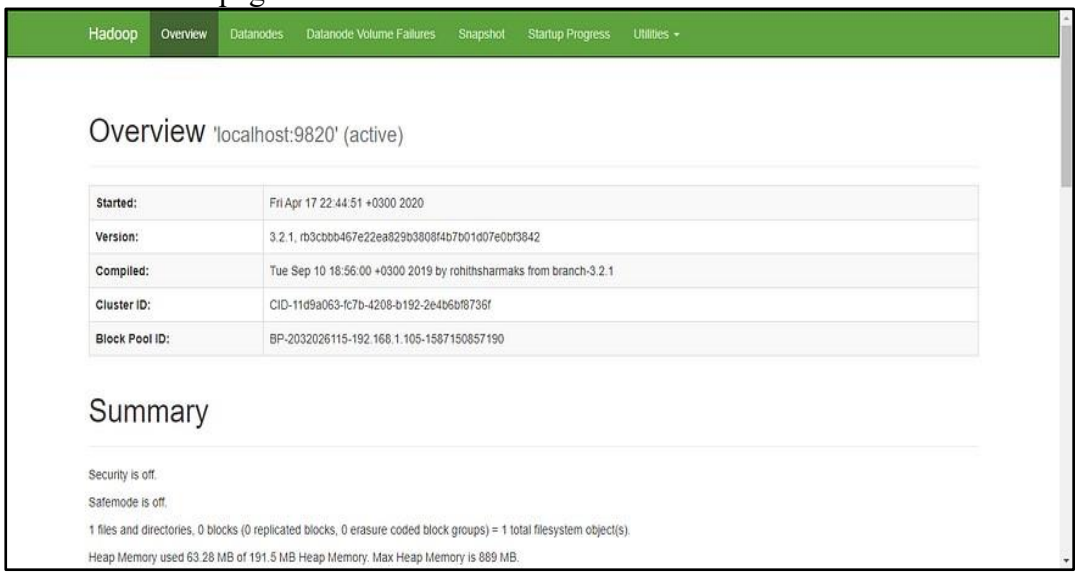
6. To make sure that all services started successfully, we can run the following command:

jps

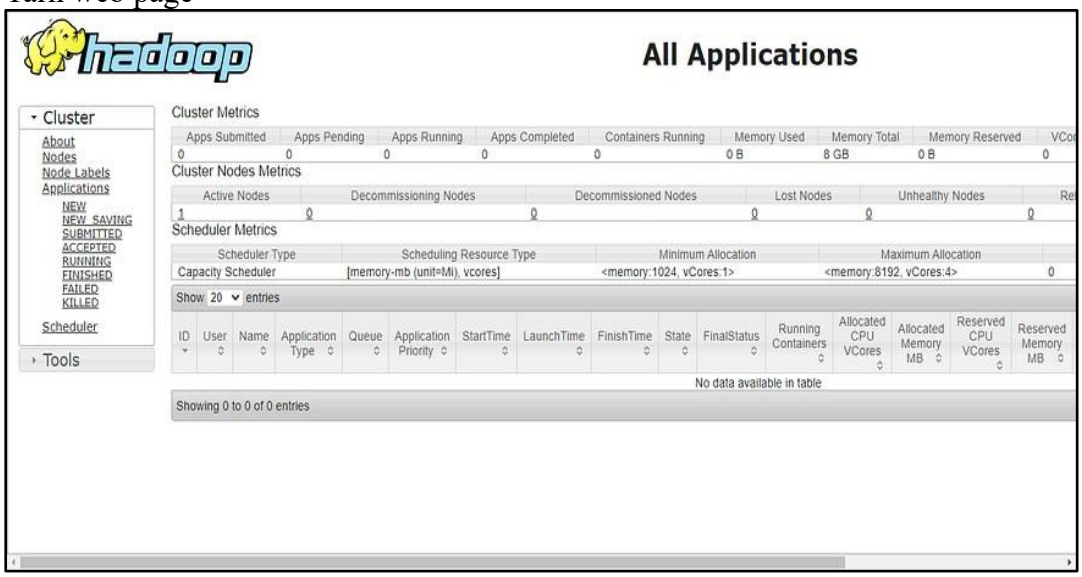


7. **Hadoop Web UI**

Name node web page



Yarn web page



6. Learning Outcomes:

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