Exp No: 8 **Date: 03.09.2024**

SET UP A SINGLE HADOOP CLUSTER AND SHOW THE PROCESS USING WEB UI

AIM:

To set up a single hadoop cluster and show the process using web UI.

PROCEDURE:

Installation steps:

Step 1: Download and install Java

Hadoop is built on Java, so you must have Java installed on your PC. You can get the most recent version of Java from the official website. After downloading, follow the installation wizard to install Java on your system.

JDK: https://www.oracle.com/java/technologies/javase-downloads.html

Step 2: Download Hadoop

Hadoop can be downloaded from the Apache Hadoop website. Make sure to have the latest stable release of Hadoop. Once downloaded, extract the contents to a convenient location.

Hadoop: https://hadoop.apache.org/releases.html

Step 3: Set Environment Variables

You must configure environment variables after downloading and unpacking Hadoop. Launch the Start menu, type “Edit the system environment variables,” and select the result. This will launch the System Properties dialogue box. Click on “Environment Variables” button to open.

Click “New” under System Variables to add a new variable. Enter the variable name “HADOOP_HOME” and the path to the Hadoop folder as the variable value. Then press “OK.”

Then, under System Variables, locate the “Path” variable and click “Edit.” Click “New” in the Edit Environment Variable window and enter “%HADOOP_HOME%bin” as the variable value. To close all the windows, use the “OK” button.

Step 4: Setup Hadoop

You must configure Hadoop in this phase by modifying several configuration files. Navigate to the "etc/hadoop" folder in the Hadoop folder. You must make changes to three files:

core-site.xml

hdfs-site.xml

mapred-site.xml

Open each file in a text editor and edit the following properties:

In core-site.xml

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

In hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/hadoop-3.3.1/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
```

```
    <value>file:/hadoop-3.3.1/data/datanode</value>
  </property>
</configuration>
```

In mapred-site.xml

```
<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>localhost:54311</value>
  </property>
</configuration>
```

Save the changes in each file.

Step 5: Format Hadoop NameNode

You must format the NameNode before you can start Hadoop. Navigate to the Hadoop bin folder using a command prompt. Execute this command:

hdfs namenode -format

Step 6: Start Hadoop

To start Hadoop, open a command prompt and navigate to the Hadoop bin folder. Run the following command:

start-dfs.cmd

start-yarn.cmd

This command will start all the required Hadoop services, including the NameNode, DataNode, and JobTracker. Wait for a few minutes until all the services are started.

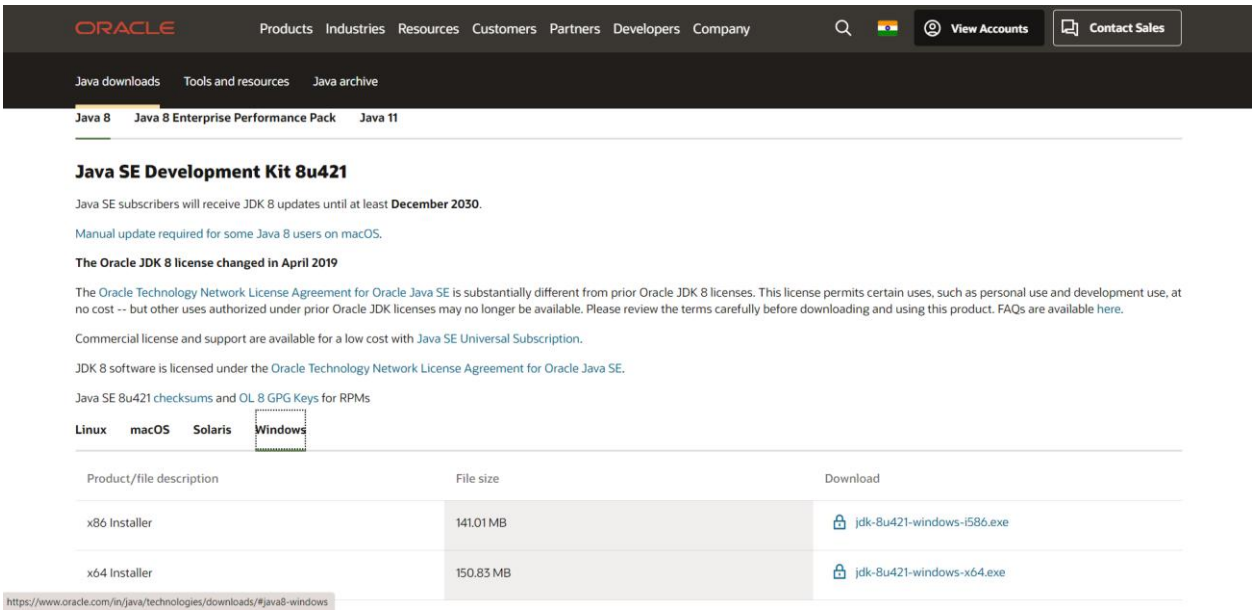
Step 7: Verify Hadoop Installation

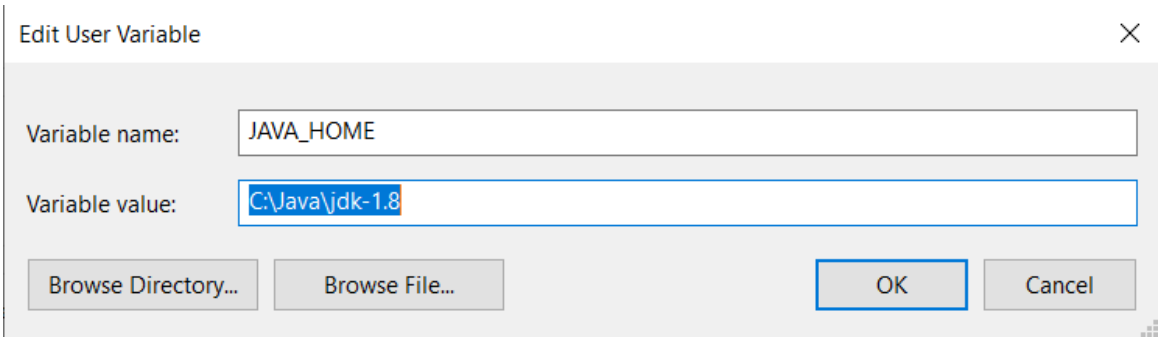
To ensure that Hadoop is properly installed, open a web browser and go to http://localhost:9870. This will launch the web interface for the Hadoop NameNode. You should see a page with Hadoop cluster information.

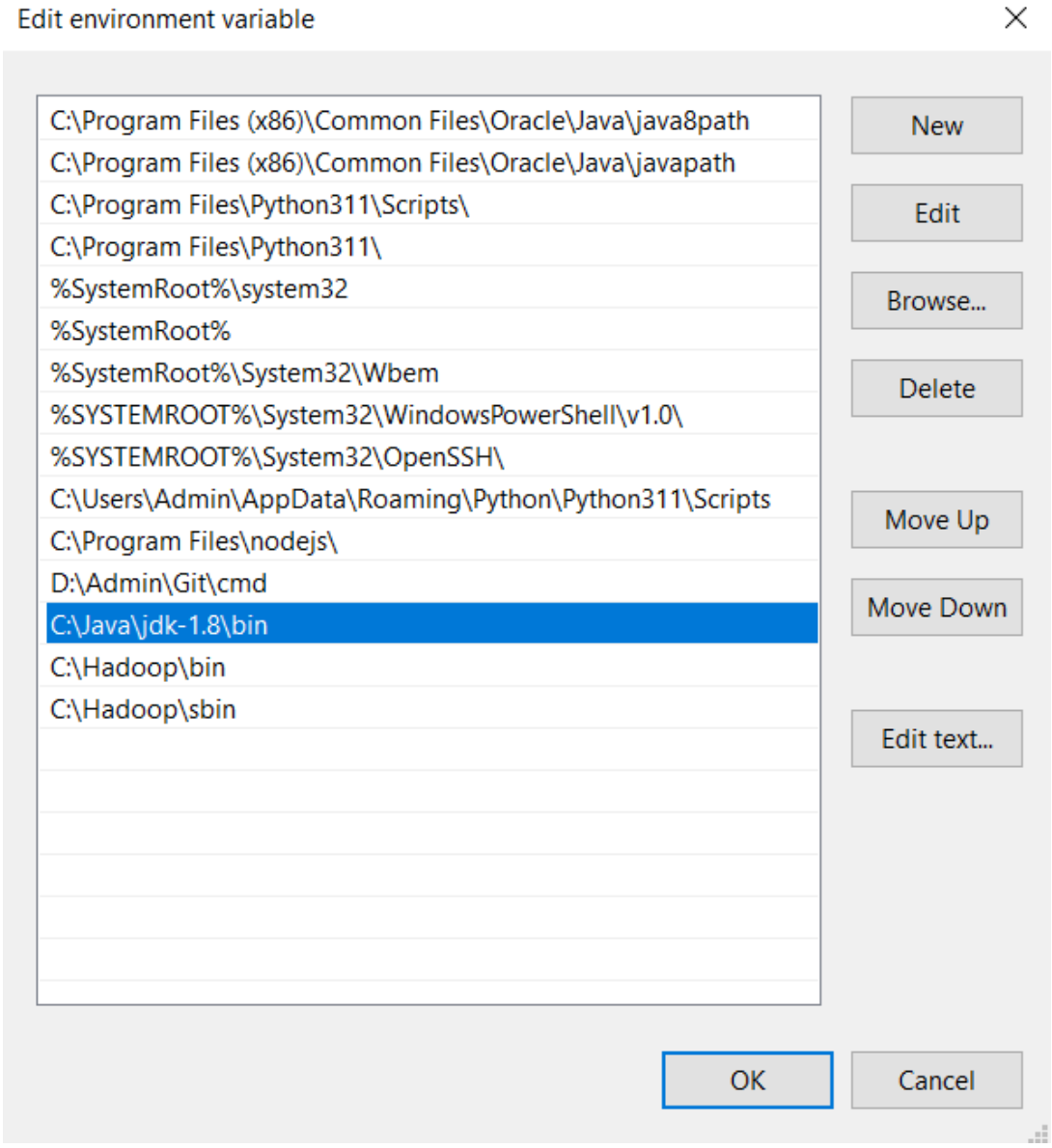
OUTPUT:

Installation Screenshots:

Step 1: Download and install Java

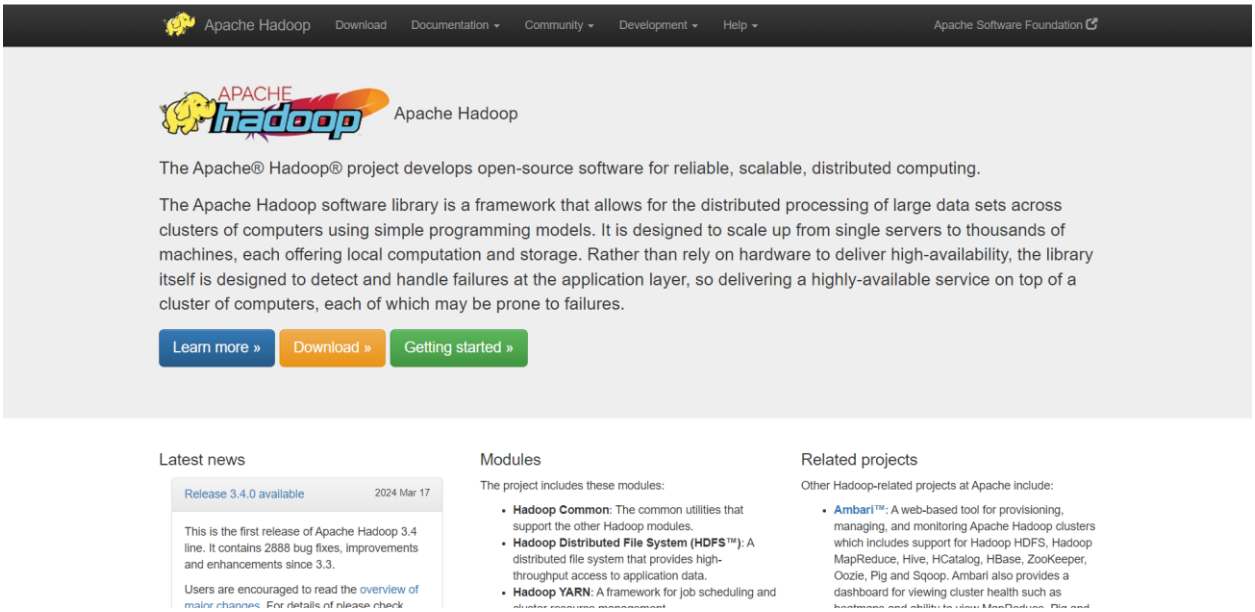




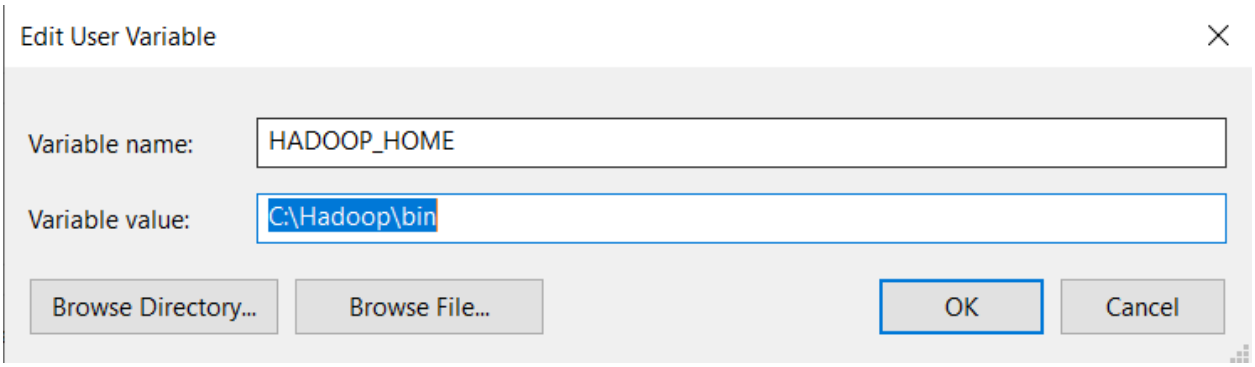


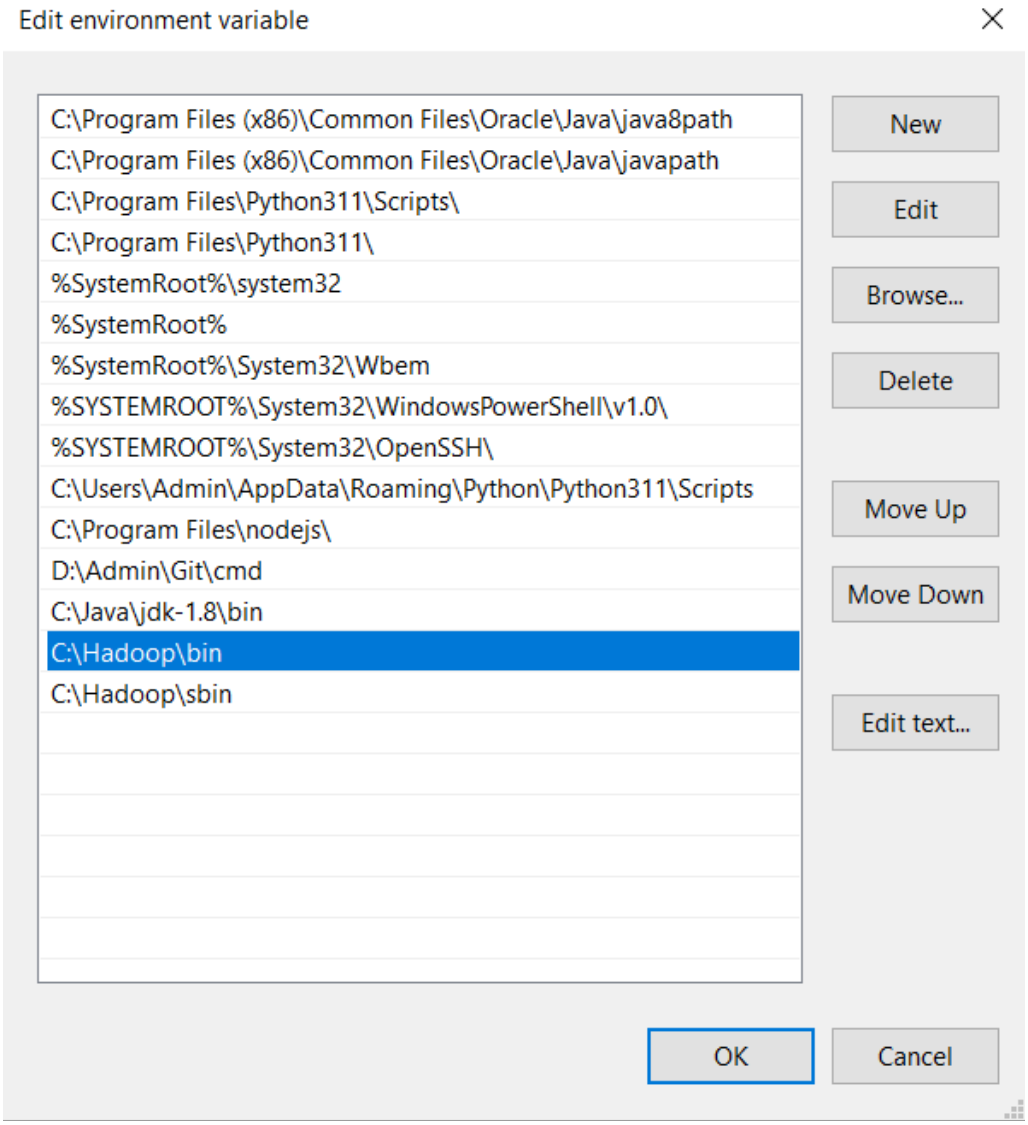

```
C:\Users\Admin>java -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.421-b09, mixed mode)
```

Step 2: Download Hadoop



Step 3: Set Environment Variables



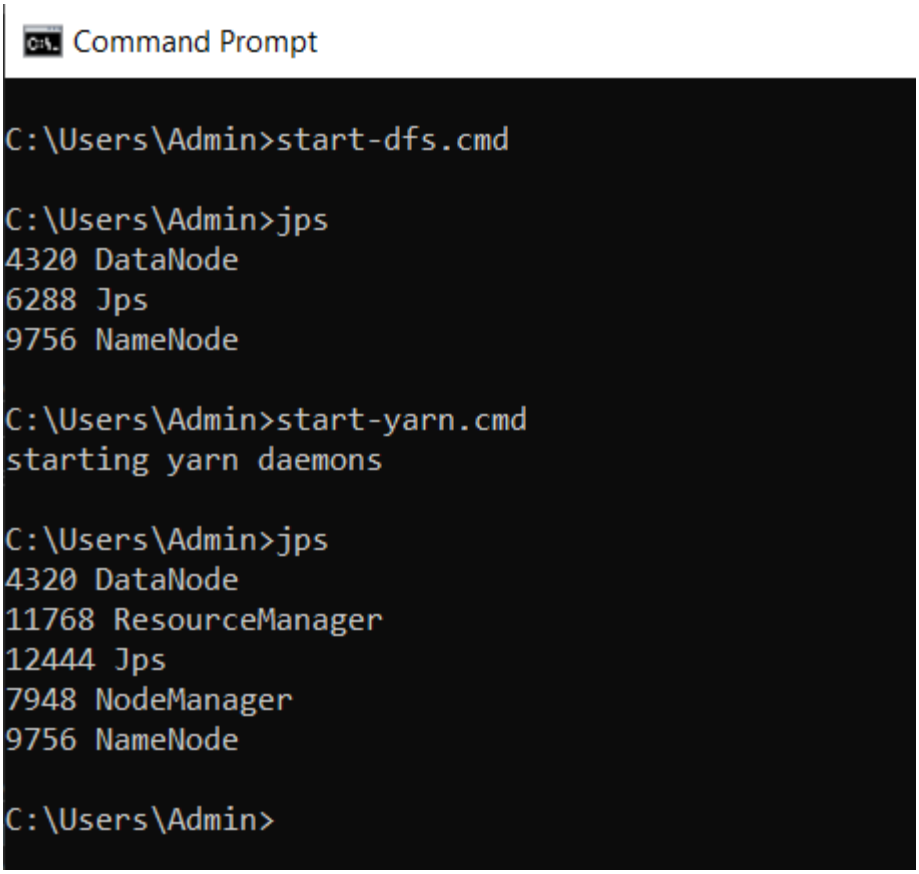



```
C:\Users\Admin>hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /C:/Hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
```

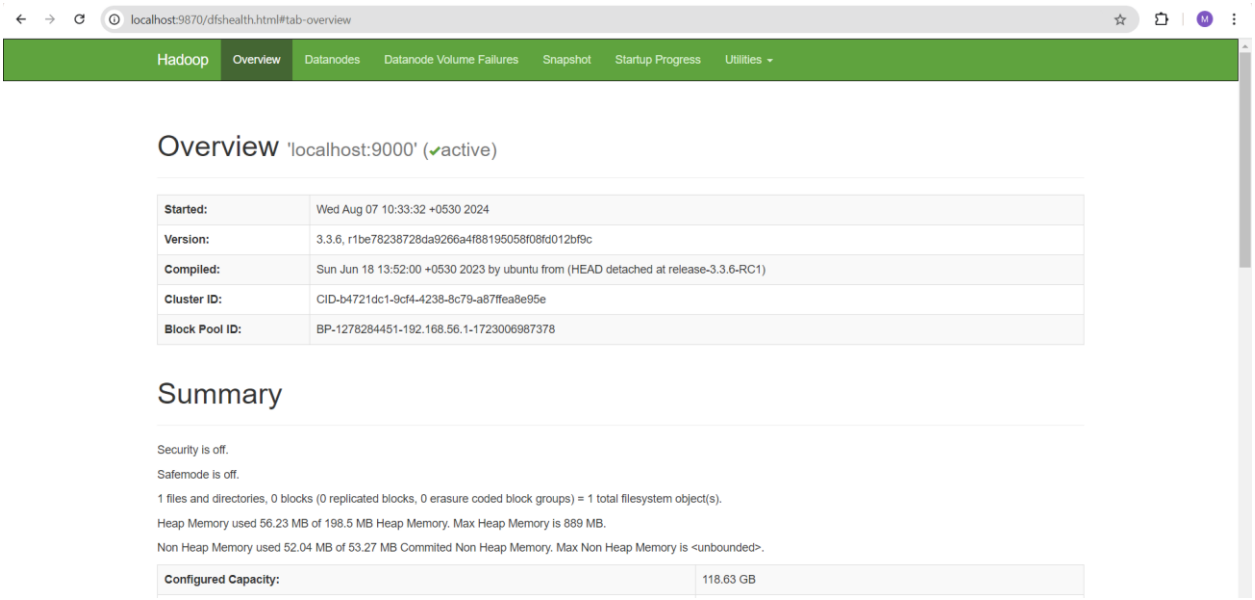
Step 4: Format Hadoop NameNode

```
C:\Users\Admin>hdfs namenode -format
2024-08-07 09:22:38,053 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = DESKTOP-TF65P79/192.168.56.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = C:\Hadoop\etc\hadoop;C:\Hadoop\share\hadoop\common;C:\Hadoop\share\hadoop\common\lib\animal-s
niffer-annotations-1.17.jar;C:\Hadoop\share\hadoop\common\lib\audience-annotations-0.5.0.jar;C:\Hadoop\share\hadoop\comm
on\lib\avro-1.7.7.jar;C:\Hadoop\share\hadoop\common\lib\checker-qual-2.5.2.jar;C:\Hadoop\share\hadoop\common\lib\commons
-beanutils-1.9.4.jar;C:\Hadoop\share\hadoop\common\lib\commons-cli-1.2.jar;C:\Hadoop\share\hadoop\common\lib\commons-cod
ec-1.15.jar;C:\Hadoop\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\Hadoop\share\hadoop\common\lib\commons-co
mpress-1.21.jar;C:\Hadoop\share\hadoop\common\lib\commons-configuration2-2.8.0.jar;C:\Hadoop\share\hadoop\common\lib\com
mons-daemon-1.0.13.jar;C:\Hadoop\share\hadoop\common\lib\commons-io-2.8.0.jar;C:\Hadoop\share\hadoop\common\lib\commons-
lang3-3.12.0.jar;C:\Hadoop\share\hadoop\common\lib\commons-logging-1.1.3.jar;C:\Hadoop\share\hadoop\common\lib\commons-m
ath3-3.1.1.jar;C:\Hadoop\share\hadoop\common\lib\commons-net-3.9.0.jar;C:\Hadoop\share\hadoop\common\lib\commons-text-1.
10.0.jar;C:\Hadoop\share\hadoop\common\lib\curator-client-5.2.0.jar;C:\Hadoop\share\hadoop\common\lib\curator-framework-
5.2.0.jar;C:\Hadoop\share\hadoop\common\lib\curator-recipes-5.2.0.jar;C:\Hadoop\share\hadoop\common\lib\dnsjava-2.1.7.ja
```

Step 5: Start Hadoop



Step 6: Verify Hadoop Installation



Configured Capacity:	118.63 GB
Configured Remote Capacity:	0 B
DFS Used:	148 B (0%)
Non DFS Used:	97.34 GB
DFS Remaining:	21.29 GB (17.95%)
Block Pool Used:	148 B (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Wed Aug 07 10:33:32 +0530 2024
Last Checkpoint Time	Wed Aug 07 10:33:07 +0530 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 1

Journal Manager	State
FileJournalManager(root=C:\Hadoop\data\namenode)	EditLogFileOutputStream(C:\Hadoop\data\namenode\current\edits_inprogress_0000000000000000001)

NameNode Storage

Storage Directory	Type	State
C:\Hadoop\data\namenode	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	118.63 GB	148 B (0%)	21.29 GB (17.95%)	148 B	1

Hadoop, 2023.

RESULT:

Thus, to set up a single hadoop cluster and show the process using web UI was completed successfully.