Ex no: 8

Hadoop - Map Reduce

Setup a Single Node Hadoop cluster and show all the process through WEB UI.

Aim:

To setup a Single Node Hadoop cluster and show all the process through WEB UI.

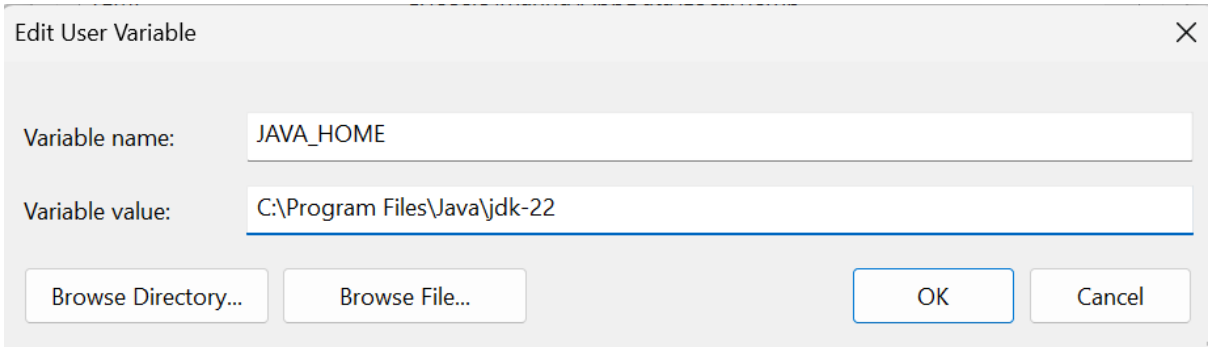
Procedure:

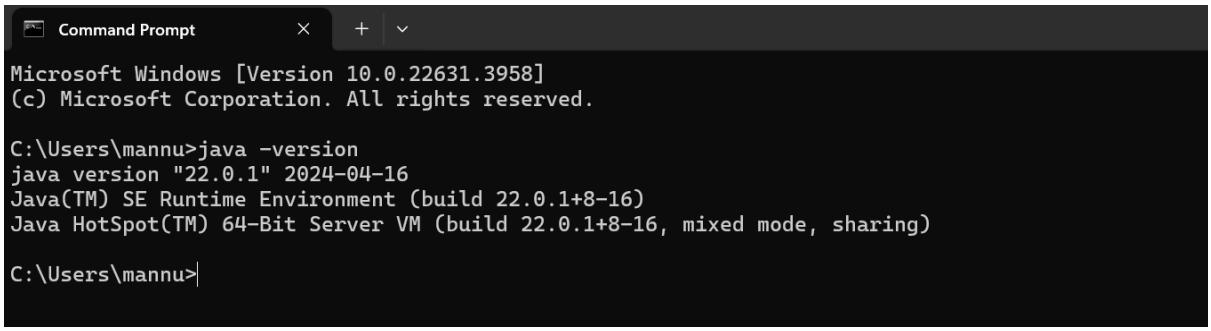
1. Install java version 8 or higher.
2. In the environment variables create a variable in the user var
3. iables and name it as “JAVA_HOME” and specify the JDK file path in the variable value.
4. To check whether Java’s path is set properly open command prompt and type “java -version”.
5. Java’s version will be displayed if the java configuration is done properly.
6. Now to install Hadoop goto the official Apache Hadoop website and download the latest version.
7. Extract the tar file using WinRar into the C drive.
8. After extracting the Hadoop file, open the environment variables and create a new variable “HADOOP_HOME” in the user variable with the value of the variable as the file path of the Hadoop’s bin folder
9. In the Hadoop file create a new folder called “data” inside which you should create two more folders one is “namenode” and the other is “datanode” folders.
10. To configure Hadoop edit the following files which are located in the etc folder :
    - core-site.xml
    - hdfs-site.xml
    - mapred-site.xml
    - yarn-site.xml
11. After editing the above files open hadoop-env file in notepad and set the java home location to it.
12. Now goto command prompt and type the command “hdfs namenode -format”. This command is used to format the HDFS (Hadoop Distributed File System) namenode. This command initializes the directory structure of the namenode by creating the necessary file system metadata.
13. NameNode works on the Master System. The primary purpose of Namenode is to manage all the MetaData. Metadata is the list of files stored in HDFS(Hadoop Distributed File System). As we know the data is stored in the form of blocks in a Hadoop cluster. So the DataNode on which or the location at which that block of the file is stored is mentioned in MetaData. All information regarding the logs of

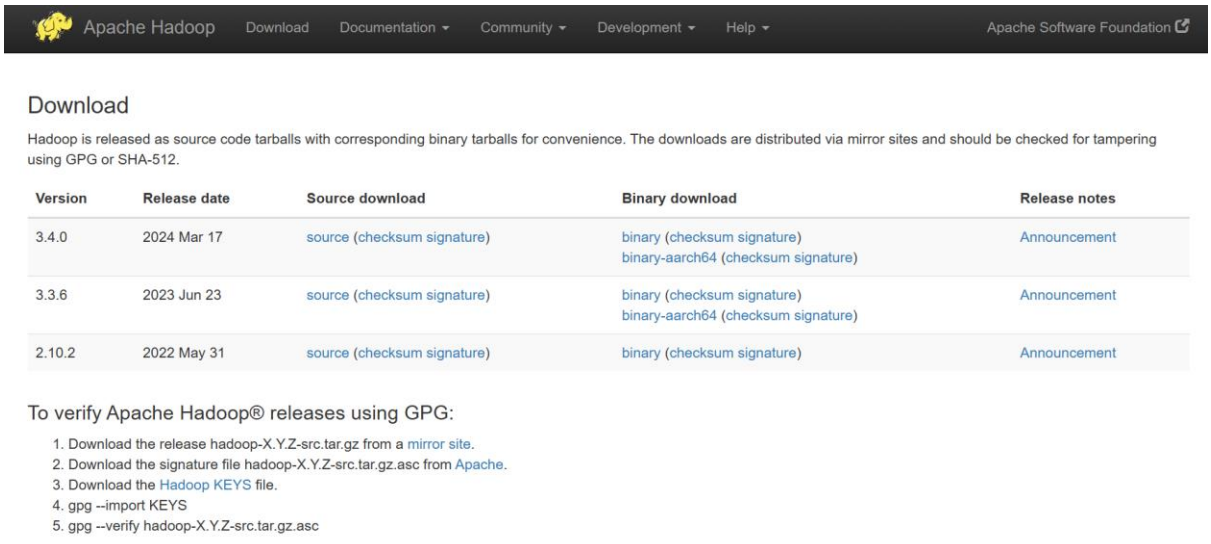
the transactions happening in a Hadoop cluster (when or who read/wrote the data) will be stored in MetaData. MetaData is stored in the memory.

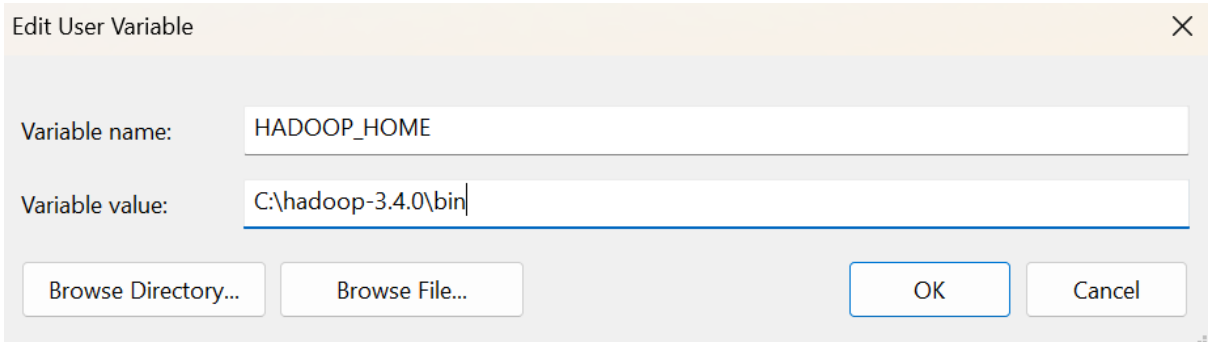
14. Now type the commands “start-dfs.cmd” and “start-yarn.cmd”.This command will start all the daemons at once.
15. Daemons mean Process. Hadoop Daemons are a set of processes that run on Hadoop.
16. Goto the browser and type “localhost:9870” this will open Hadoop in the browser.

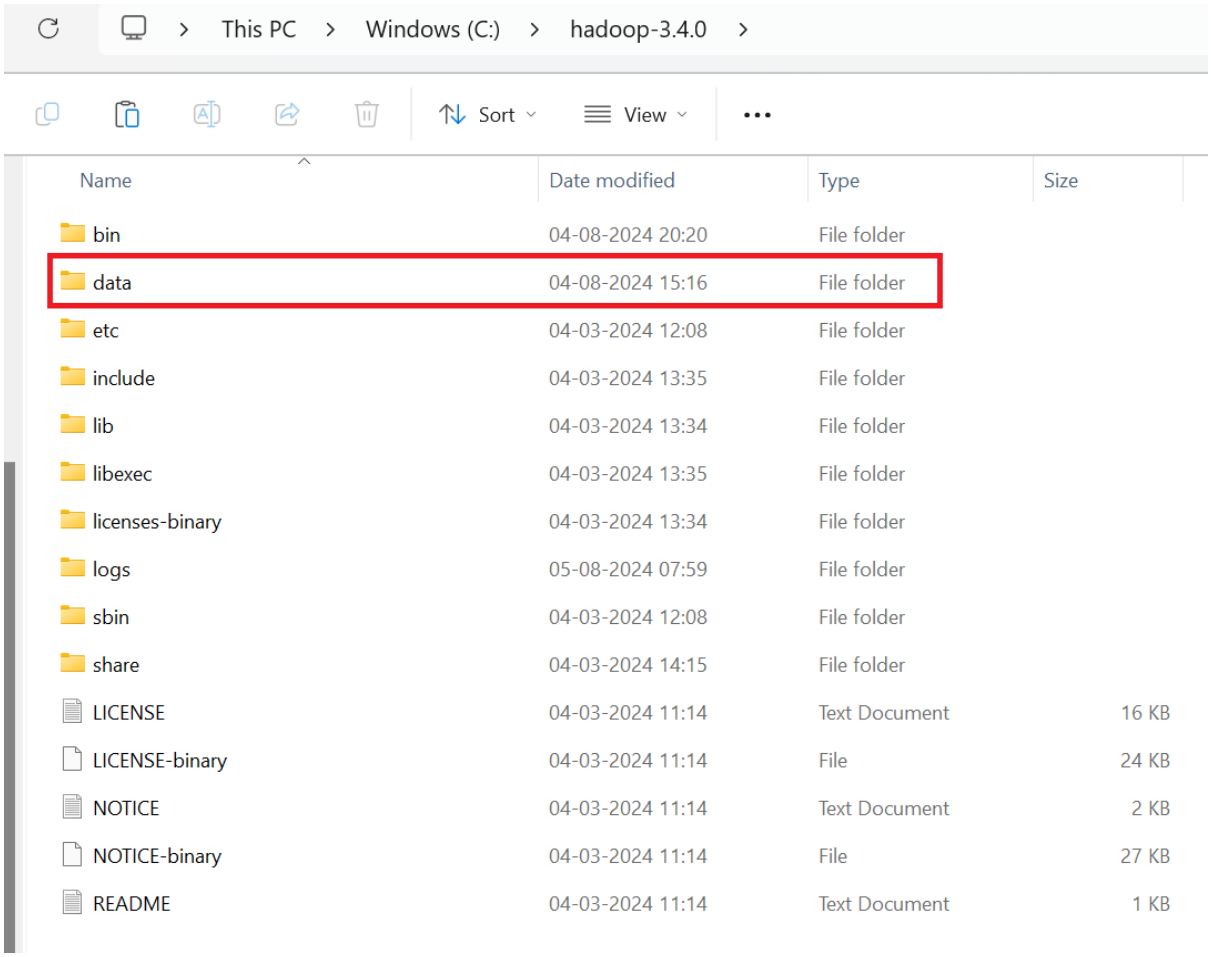
Output:

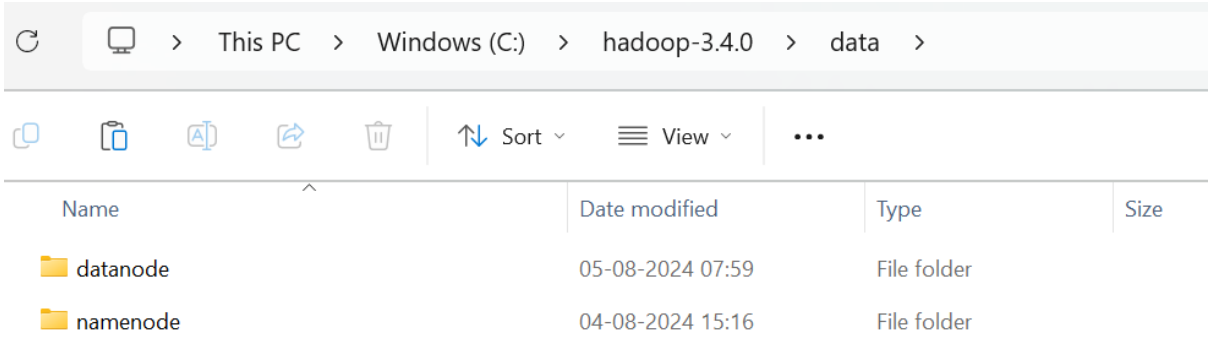







Edit User Variable
Variable name:
HADOOP_HOME
Variable value:
C:\hadoop-3.4.0\bin
Browse Directory...
Browse File...
OK
Cancel


This PC
Windows (C:)
hadoop-3.4.0
Sort
View
Name
Date modified
Type
Size
bin 04-08-2024 20:20 File folder
data 04-08-2024 15:16 File folder
etc 04-03-2024 12:08 File folder
include 04-03-2024 13:35 File folder
lib 04-03-2024 13:34 File folder
libexec 04-03-2024 13:35 File folder
licenses-binary 04-03-2024 13:34 File folder
logs 05-08-2024 07:59 File folder
sbin 04-03-2024 12:08 File folder
share 04-03-2024 14:15 File folder
LICENSE 04-03-2024 11:14 Text Document 16 KB
LICENSE-binary 04-03-2024 11:14 File 24 KB
NOTICE 04-03-2024 11:14 Text Document 2 KB
NOTICE-binary 04-03-2024 11:14 File 27 KB
README 04-03-2024 11:14 Text Document 1 KB




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

core-site.xml

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.data.dir</name>
<value>file:///C:/hadoop-3.4.0/data/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>file:///C:/hadoop-3.4.0/data/datanode</value>
</property>
</configuration>
```

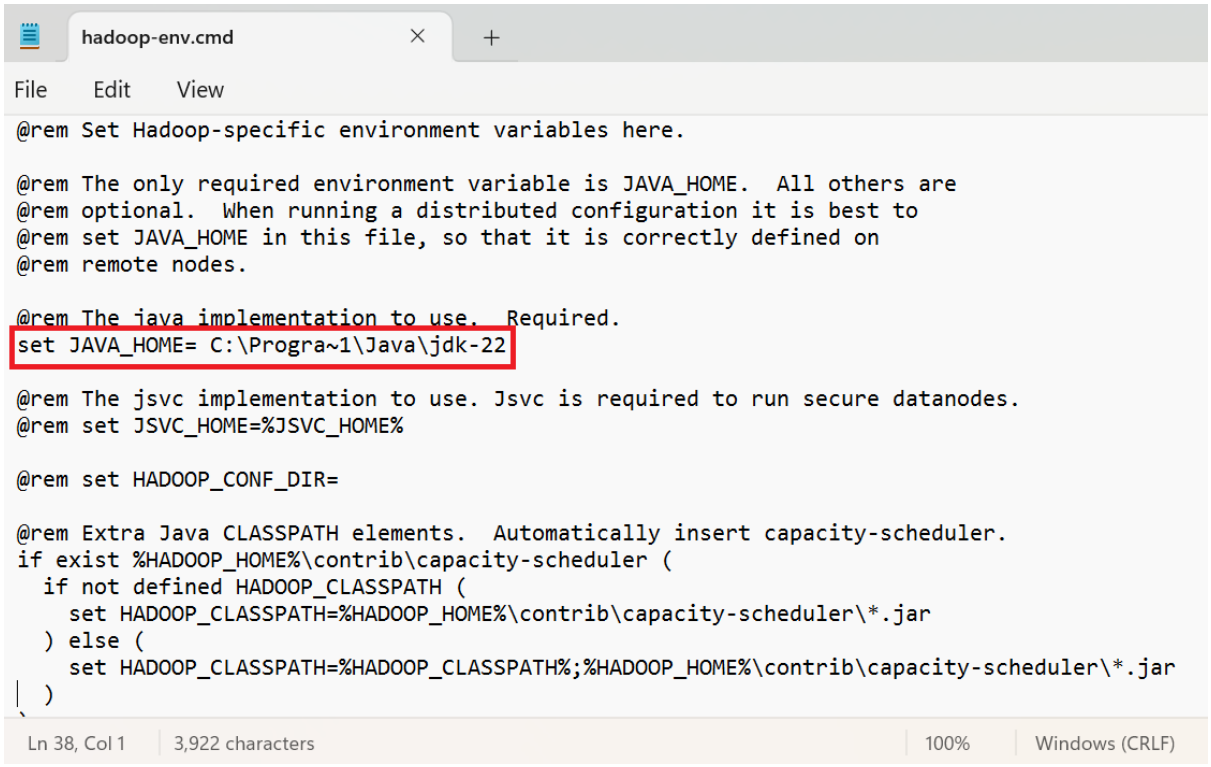
hdfs-site.xml

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

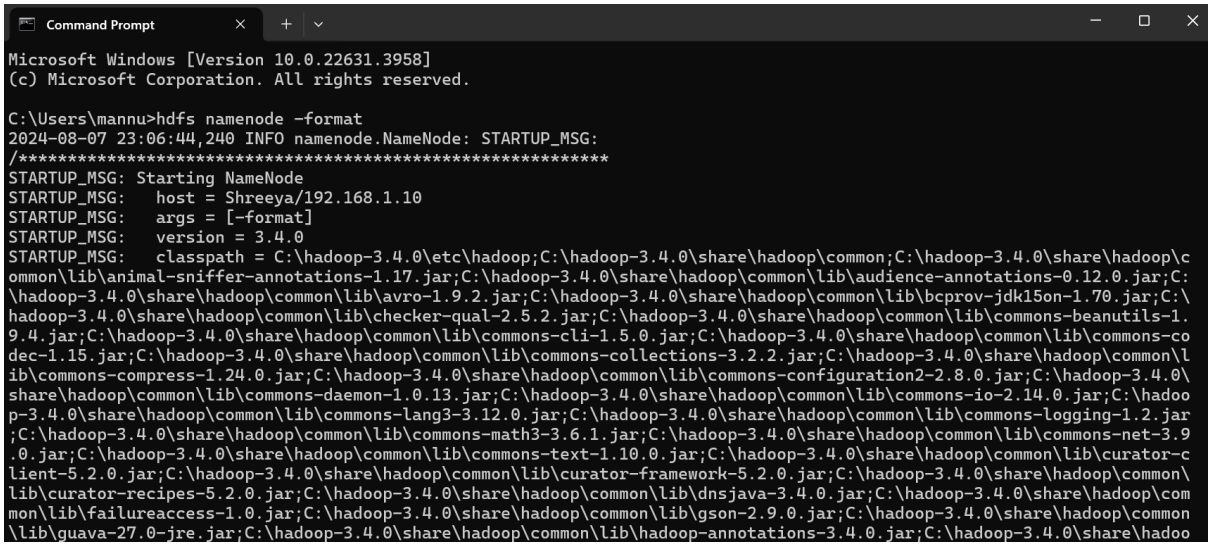
mapred-site.xml

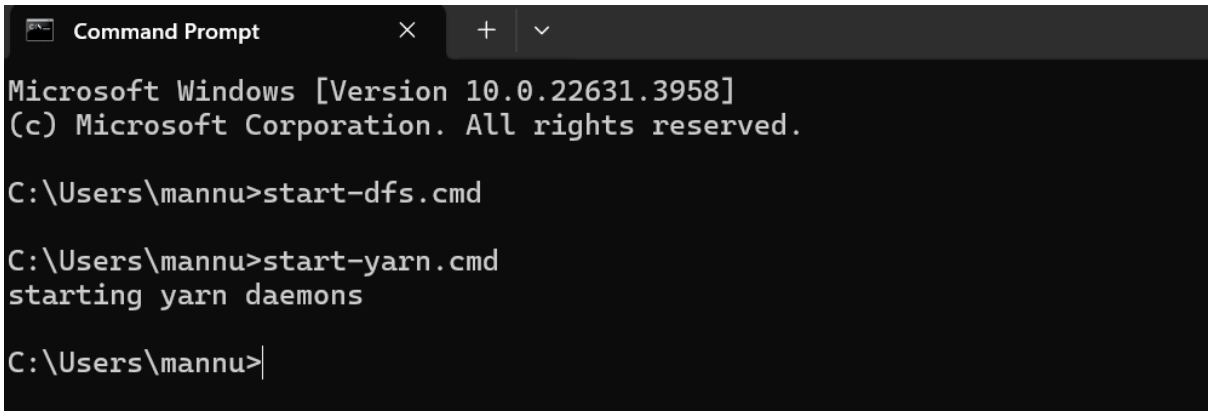
```
<configuration>
<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
```

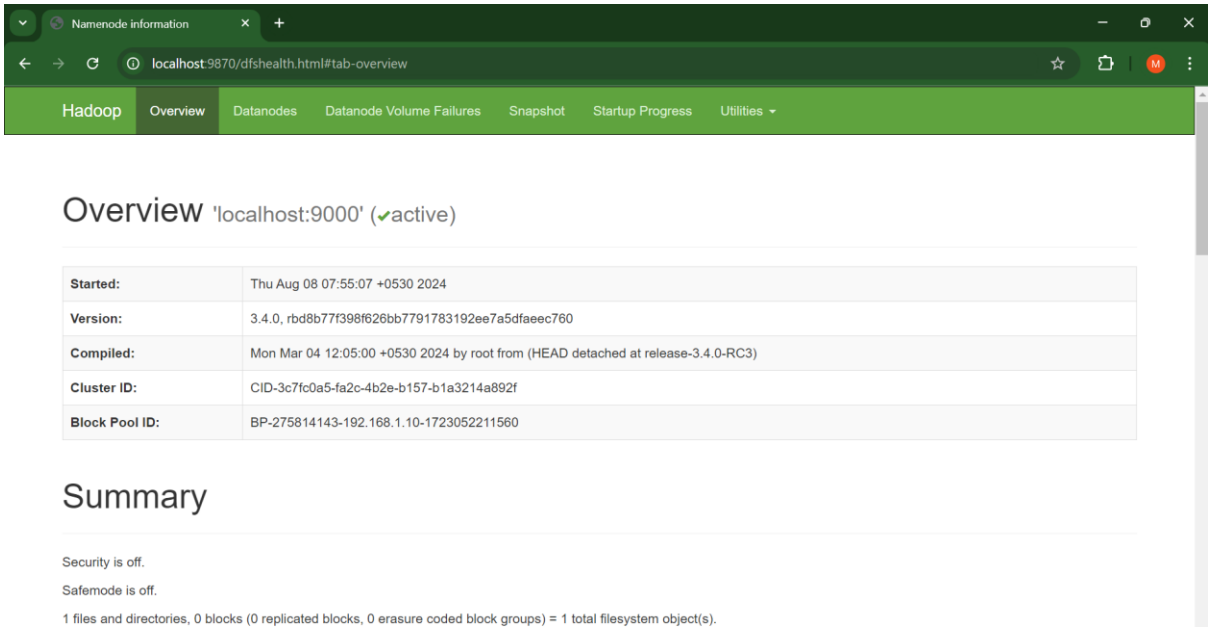
yarn-site.xml



hadoop-env file


Command Prompt
Microsoft Windows [Version 10.0.22631.3958]
(c) Microsoft Corporation. All rights reserved.
C:\Users\mannu>hdfs namenode -format
2024-08-07 23:06:44,240 INFO namenode.NameNode: STARTUP_MSG:
/**
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = Shreeya/192.168.1.10
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.4.0
STARTUP_MSG: classpath = C:\hadoop-3.4.0\etc\hadoop;C:\hadoop-3.4.0\share\hadoop\common;C:\hadoop-3.4.0\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\audience-annotations-0.12.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\avro-1.9.2.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\bcprov-jdk15on-1.70.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\checker-qual-2.5.2.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-beanutils-1.9.4.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-cli-1.5.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-codec-1.15.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-compress-1.24.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-configuration2-2.8.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-daemon-1.0.13.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-io-2.14.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-lang3-3.12.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-logging-1.2.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-math3-3.6.1.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-net-3.9.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\commons-text-1.10.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\curator-client-5.2.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\curator-framework-5.2.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\curator-recipes-5.2.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\dnsjava-3.4.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\failureaccess-1.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\gson-2.9.0.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\guava-27.0-jre.jar;C:\hadoop-3.4.0\share\hadoop\common\lib\hadoop-annotations-3.4.0.jar;C:\hadoop-3.4.0\share\hadoo


Command Prompt
Microsoft Windows [Version 10.0.22631.3958]
(c) Microsoft Corporation. All rights reserved.
C:\Users\mannu>start-dfs.cmd
C:\Users\mannu>start-yarn.cmd
starting yarn daemons
C:\Users\mannu>


Namenode information
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:9000' (active)
Started: Thu Aug 08 07:55:07 +0530 2024
Version: 3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled: Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID: CID-3c7fc0a5-fa2c-4b2e-b157-b1a3214a892f
Block Pool ID: BP-275814143-192.168.1.10-1723052211560
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 62.7 MB of 80 MB Heap Memory. Max Heap Memory is 1000 MB.

Non Heap Memory used 52.45 MB of 54.88 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	352.98 GB
Configured Remote Capacity:	0 B
DFS Used:	321 B (0%)
Non DFS Used:	154.21 GB
DFS Remaining:	198.77 GB (56.31%)
Block Pool Used:	321 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	Thu Aug 08 07:55:07 +0530 2024
Last Checkpoint Time	Thu Aug 08 07:53:35 +0530 2024
Last HA Transition Time	Never
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 2

Journal Manager	State
FileJournalManager(root=\tmp\hadoop-mannu\dfs\name)	EditLogFileOutputStream(\tmp\hadoop-mannu\dfs\name\current\edits_inprogress_0000000000000000002)

NameNode Storage

Storage Directory	Type	State
\tmp\hadoop-mannu\dfs\name	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	352.98 GB	321 B (0%)	198.77 GB (56.31%)	321 B	1

Hadoop, 2024.

Result:

Thus to setup a Single Node Hadoop cluster and show all the process through WEB UI has been executed successfully.