Name-Manas Kumar Manna
24MCS105 (M.tech AI 1st sem)

AI-Lab[22-Oct-2024]

Setup single node Hadoop cluster and Apply HDFS Commands on Single Node Hadoop Cluster.

- **Prequisites**

1. Java 8 runtime environment (JRE)
2. Apache Hadoop 3.3.6

1 - Download Hadoop binary package

The first step is to download Hadoop binaries from the official website(https://hadoop.apache.org/releases.html). The binary packagesize is about 696 MB.

2 - Unpack the package

After finishing the file download, we should unpack the package using 7zip or WinRar.

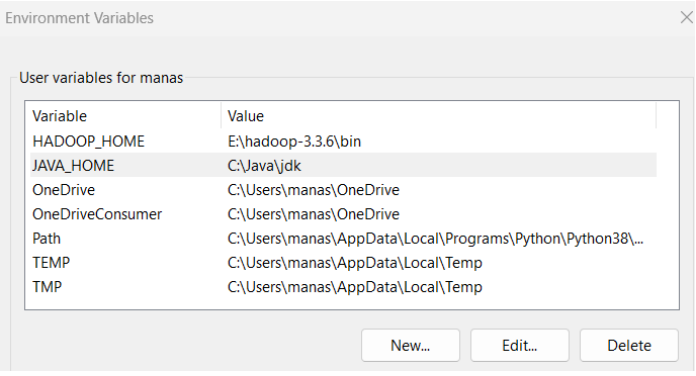
3 -Java installation

Java is required to run Hadoop. If java is not installed , We have to install it. After finishing the file download we open a new command prompt, we should unpack the package.

4 - Configure environment variables

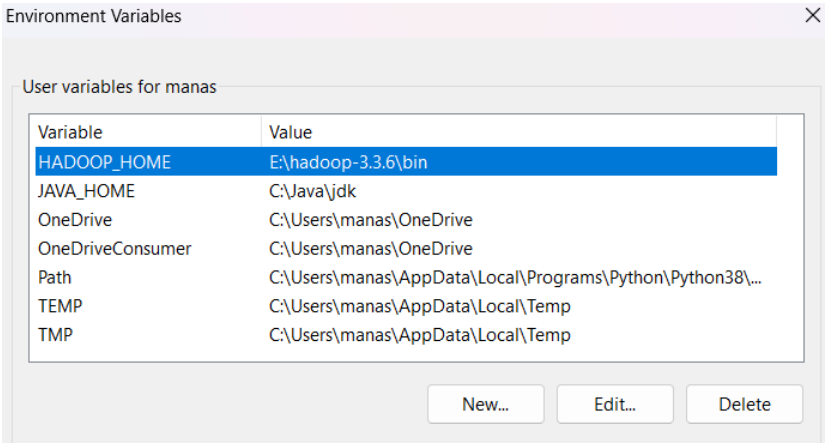
Now we've downloaded and unpacked all the files we need to configure two important environment variables.

We configure **JAVA_HOME** environment variable by adding new environment variable.
Variable name : JAVA_HOME Variable value: Java path



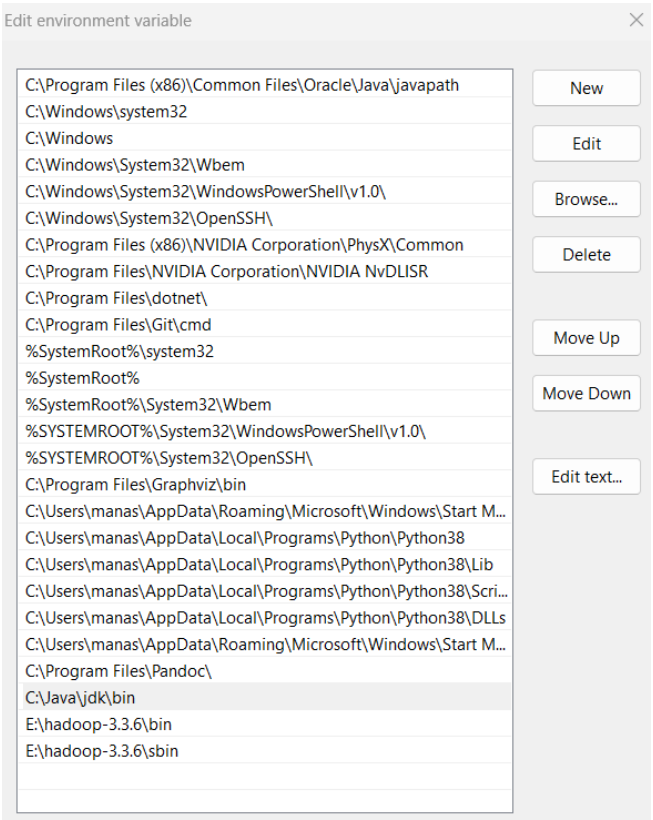
The same with **HADOOP_HOME** environment variable-

Variable name : HADOOP_HOME Variable value: C:\Hadoop\hadoop-3.3.6



- **Configure PATH environment variable**

Once we finish setting up the above two environment variables, we need to add the bin folders to the PATH environment variable.



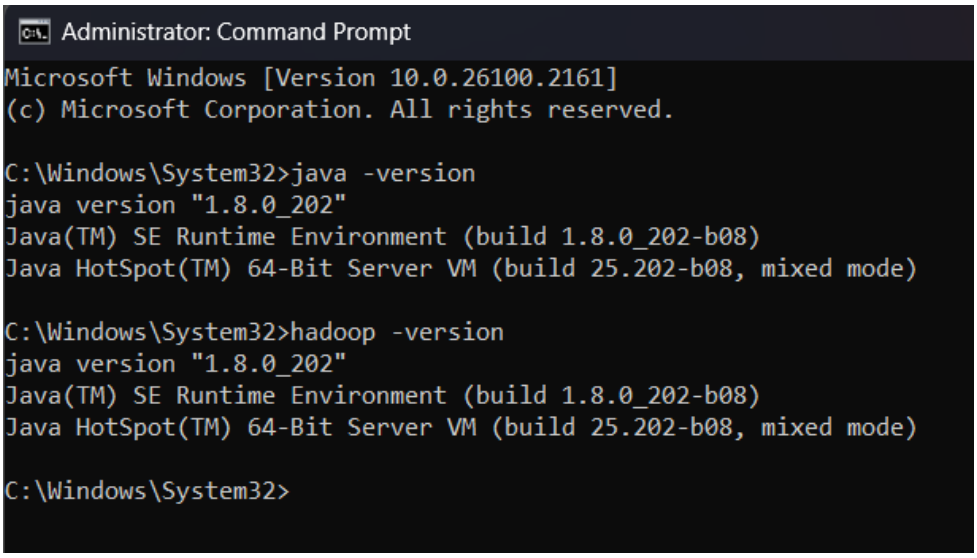
5 - Verification of Installation

Once We complete the installation, Close terminal window and open a new one and run the following command to verify:

java -version

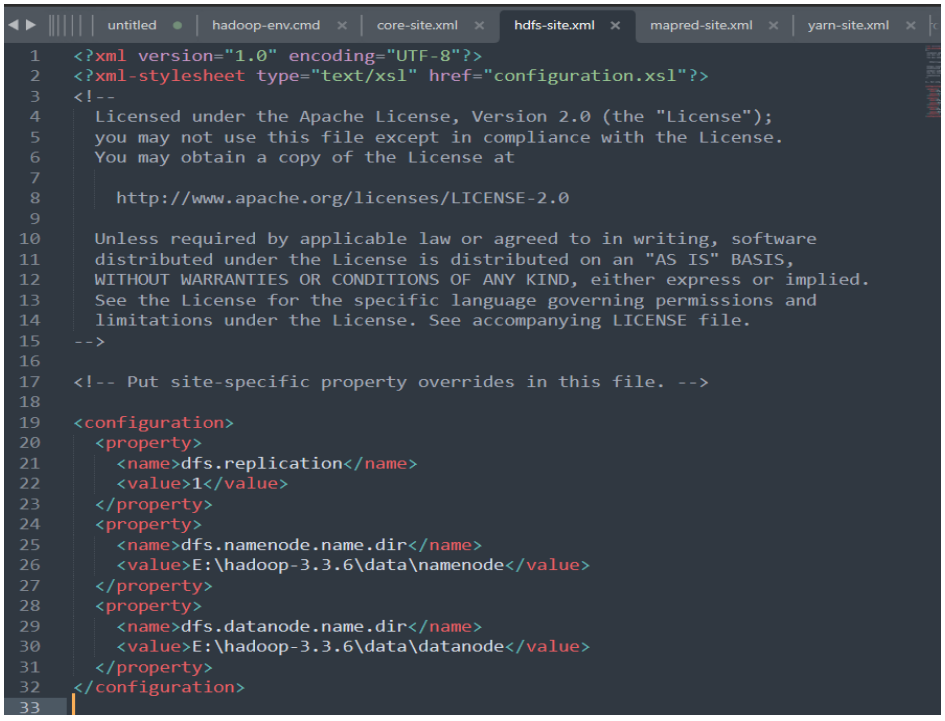
We can also be able to run the following command:

hadoop –version



6 - Configure Hadoop

Now we are ready to configure the most important part - Hadoop configurations which involves Core, YARN, MapReduce, HDFS configuration



7 – Initialize HDFS

Run the following command in Command Prompt

hdfs namenode –format

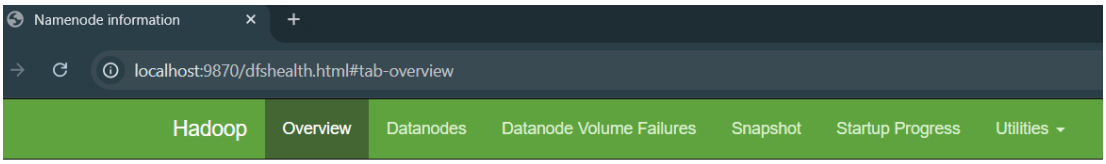
```
C:\Windows\System32>hdfs namenode -format
2024-10-29 10:14:15,312 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = Manas/192.168.1.196
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = E:\hadoop-3.3.6\etc\hadoop;E:\hadoop-3.3.6\share\hadoop\common;E:\hadoop-3.3.6\share\hadoop\c
ommon\lib\animal-sniffer-annotations-1.17.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\audience-annotations-0.5.0.jar;E:\
hadoop-3.3.6\share\hadoop\common\lib\avro-1.7.7.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\checker-qual-2.5.2.jar;E:\ha
doop-3.3.6\share\hadoop\common\lib\commons-beanutils-1.9.4.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-cli-1.2.j
ar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-codec-1.15.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-collec
tions-3.2.2.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-compress-1.21.jar;E:\hadoop-3.3.6\share\hadoop\common\li
b\commons-configuration2-2.8.0.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-daemon-1.0.13.jar;E:\hadoop-3.3.6\sha
re\hadoop\common\lib\commons-io-2.8.0.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-lang3-3.12.0.jar;E:\hadoop-3.3
.6\share\hadoop\common\lib\commons-logging-1.1.3.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-math3-3.1.1.jar;E:\
hadoop-3.3.6\share\hadoop\common\lib\commons-net-3.9.0.jar;E:\hadoop-3.3.6\share\hadoop\common\lib\commons-text-1.10.0.j
```

8 - Start HDFS daemons

Run the following command to start HDFS daemons in Command Prompt:

start-all.cmd

Verify HDFS web portal UI through this link http://localhost:9870/dfshealth.html#tab-overview.



Overview 'localhost:9000' (active)

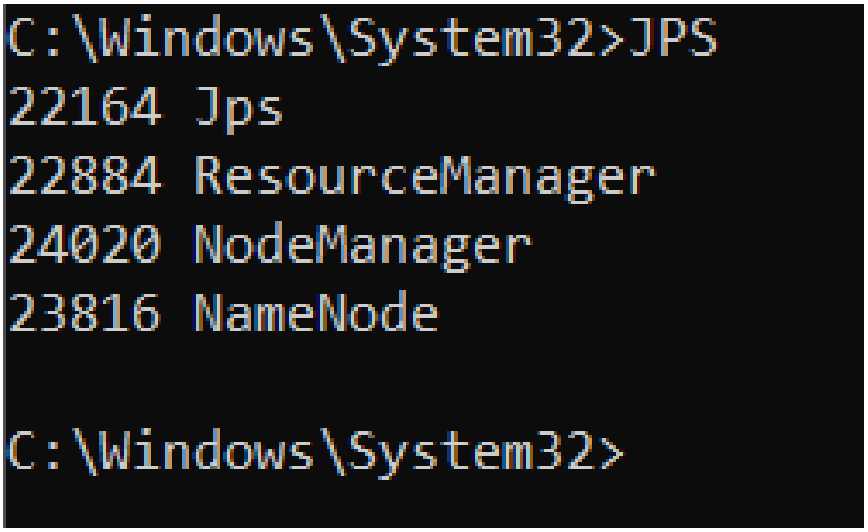
Started:	Tue Oct 29 10:27:26 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-891bd78f-c90c-4979-b240-f46069696303
Block Pool ID:	BP-932695391-192.168.1.196-1730177794675

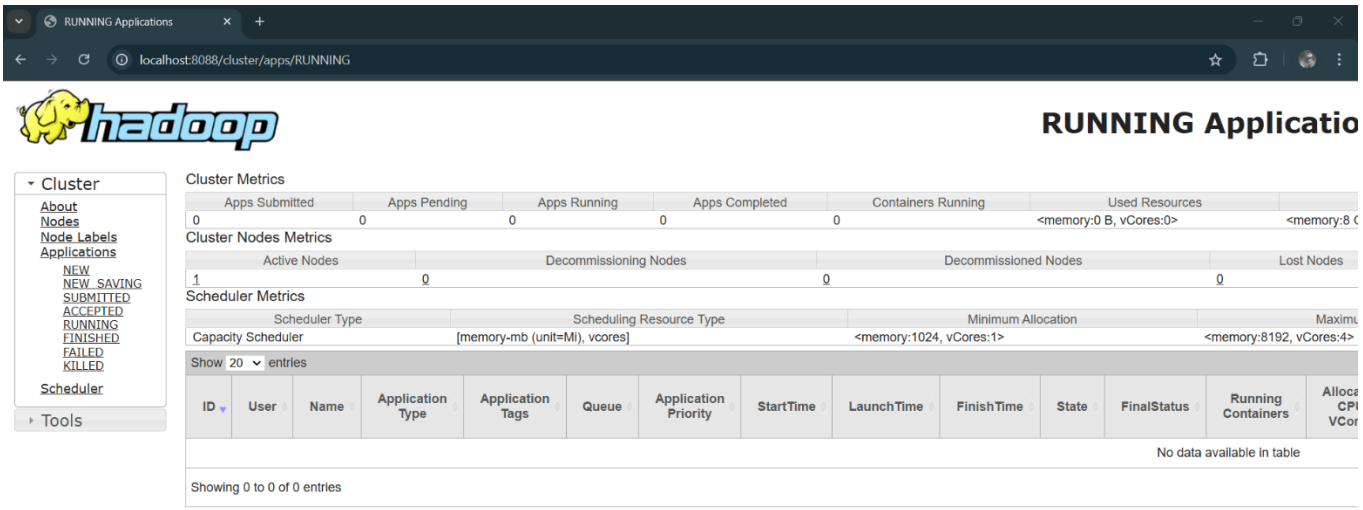
9 - Start YARN daemons

Run the following command in an elevated Command Prompt window (Run as administrator) to start YARN daemons:

%HADOOP_HOME%\sbin\start-yarn.cmd
You can verify YARN resource manager UI when all services are started successfully.

http://localhost:8088




RUNNING Applications
localhost:8088/cluster/apps/RUNNING
hadoop
RUNNING Applicatio
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Used Resources
0
0
0
0
0
<memory:0 B, vCores:0>
<memory:8
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
1
0
0
0
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximu
Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>
Show 20 entries
ID
User
Name
Application Type
Application Tags
Queue
Application Priority
StartTime
LaunchTime
FinishTime
State
FinalStatus
Running Containers
No data available in table
Showing 0 to 0 of 0 entries