HADOOP 3.4.1 STEP BY STEP INSTALLATION ON UBUNTU 24.04.1 LTS

Hardware specification:
OS: Ubuntu 24.04.1 LTS
Memory: 4 GB
CPU: 2 cores
Storage: 25 GB

Prerequisites:
Java Development Kit (JDK)
SSH key pair
Hadoop version 3.4.1

STEP 1: Install JDK

 # apt install openjdk-8-jdk

```
root@kk:~# apt install openjdk-8-jdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  adwaita-icon-theme alsa-topology-conf alsa-ucm-conf at-spi2-common
  fonts-dejavu-extra gsettings-desktop-schemas gtk-update-icon-cache
  libatk-bridge2.0-0t64 libatk-wrapper-java libatk-wrapper-java-jni
```

```
root@kk:~# java -version
openjdk version "1.8.0_432"
OpenJDK Runtime Environment (build 1.8.0_432-8u432-ga~us1-0ubuntu2~24.04-ga)
OpenJDK 64-Bit Server VM (build 25.432-bga, mixed mode)
root@kk:~#
```

STEP 2: Generate SSH key pair from "hadoop" user

```
root@kk:~# su hadoop
hadoop@kk:/root$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa):
Created directory '/home/hadoop/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:ezHyiSXL1bj4Ikf5/pvZYpCpWYh/UMN6yEA91I/GVP8 hadoop@kk
The key's randomart image is:
+---[RSA 3072]----+
|       o.. ..    |
|      . o o  .   |
|     .   = o  .  |
|      .   Bo.  . |
|       +SB*+.   E|
|      ..O@*=     |
|       o*B=.     |
|      . *oo o+   |
|       o +oo=o.  |
+----[SHA256]-----+
hadoop@kk:/root$
```

```
hadoop@kk:/$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hadoop@kk:/$ chmod 640 ~/.ssh/authorized_keys
hadoop@kk:/$
```

STEP 3: Install Hadoop

 # wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz

```
hadoop@kk:~$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz
--2025-01-21 11:09:23--  https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 2a04:4e42::644, 151.101.2.132
Connecting to dlcdn.apache.org (dlcdn.apache.org)|2a04:4e42::644|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 974002355 (929M) [application/x-gzip]
Saving to: 'hadoop-3.4.1.tar.gz'

hadoop-3.4.1.tar.gz                 100%[=======================================================================================>] 928.88M  1.15MB/s    in 7m 46s

2025-01-21 11:17:10 (1.99 MB/s) - 'hadoop-3.4.1.tar.gz' saved [974002355/974002355]

hadoop@kk:~$
```

Extract folder and rename

 # tar -xvzf hadoop-3.4.1.tar.gz

 # mv hadoop-3.4.1 hadoop

Setup Hadoop and Java Environment variables

 # nano ~/.bashrc

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop export
HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

```
##############################

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
##############################
hadoop@kk:~$
```

Load the configuration in the current environment

 # source ~/.bashrc

```
hadoop@kk:~$ source ~/.bashrc
hadoop@kk:~$
```

STEP 3: Configure Hadoop environment variable file

Edit the **$HADOOP_HOME/etc/hadoop/hadoop-env.sh** file to configure **JAVA_HOME**

```
hadoop@kk:~$ nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

Add **export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64**

```
  GNU nano 7.2                                  /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh *

# Technically, the only required environment variable is JAVA_HOME.
# All others are optional.  However, the defaults are probably not
# preferred.  Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

# The language environment in which Hadoop runs. Use the English
# environment to ensure that logs are printed as expected.
```

Create directory for **namenode** and **datanode**

```
hadoop@kk:~/hadoop$ mkdir -p ~/hadoopdata/hdfs/{namenode,datanode}
```

Edit the **$HADOOP_HOME/etc/hadoop/core-site.xml** file and update with your system hostname

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

Add below line

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

```
  GNU nano 7.2                                  /home/hadoop/hadoop/etc/hadoop/core-site.xml *
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

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

Edit the **$HADOOP_HOME/etc/hadoop/hdfs-site.xml** file and change the namenode and datanode directory paths

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

Add below line

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```

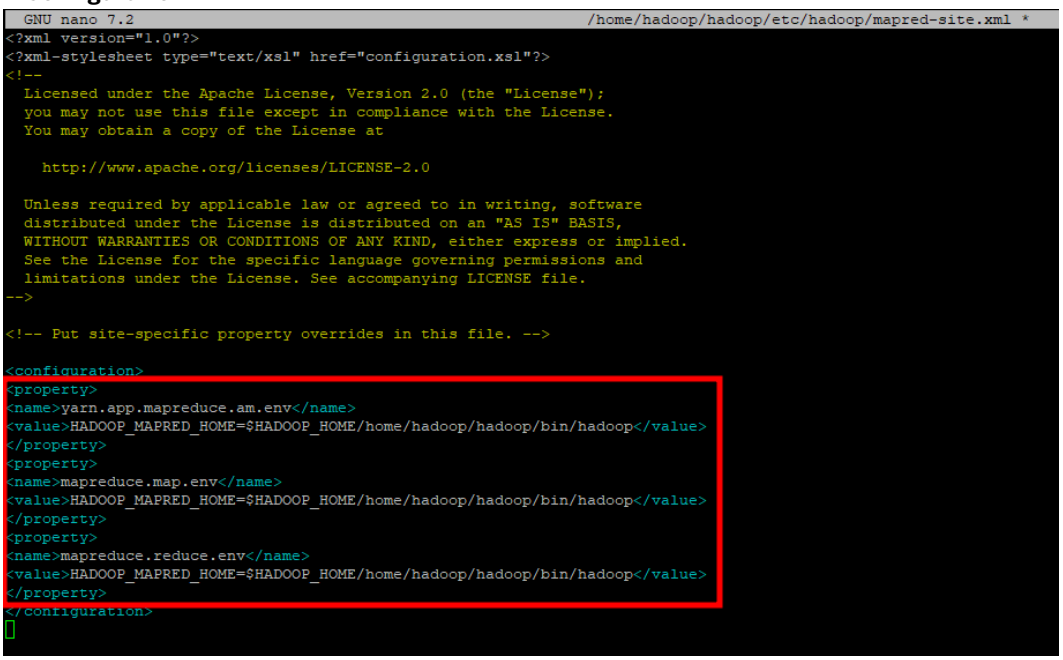
```
  GNU nano 7.2                                  /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml *
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
    </property>
</configuration>
```

Edit the **$HADOOP_HOME/etc/hadoop/mapred-site.xml** file

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

Add below line

```
<configuration>
 <property>
  <name>yarn.app.mapreduce.am.env</name>
  <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
 </property>
 <property>
  <name>mapreduce.map.env</name>
  <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
 </property>
 <property>
  <name>mapreduce.reduce.env</name>
  <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
 </property>
</configuration>
```



Edit the **$HADOOP_HOME/etc/hadoop/yarn-site.xml** file

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

Add below line

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

```
  GNU nano 7.2                                  /home/hadoop/hadoop/etc/hadoop/yarn-site.xml *
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
```

STEP 4: Start Hadoop Cluster

Format the Hadoop Namenode

```
hadoop@kk:~/hadoop$ hdfs namenode -format
WARNING: /home/hadoop/hadoop/logs does not exist. Creating.
2025-01-21 12:19:55,118 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = kk/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.4.1
STARTUP_MSG:   classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/
```

Initiate the Hadoop cluster with **start-all.sh**

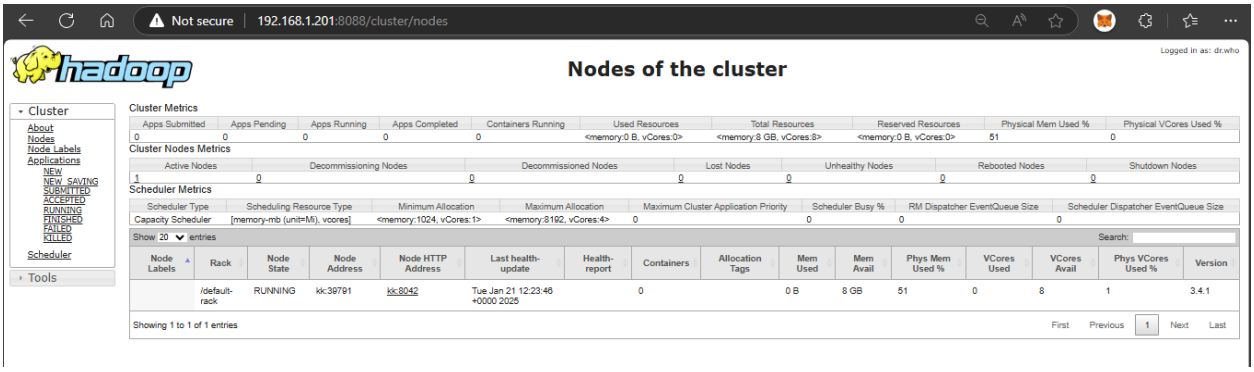
```
hadoop@kk:~/hadoop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Starting datanodes
Starting secondary namenodes [kk]
kk: Warning: Permanently added 'kk' (ED25519) to the list of known hosts.
Starting resourcemanager
Starting nodemanagers
hadoop@kk:~/hadoop$
```

Verify running status of Hadoop services

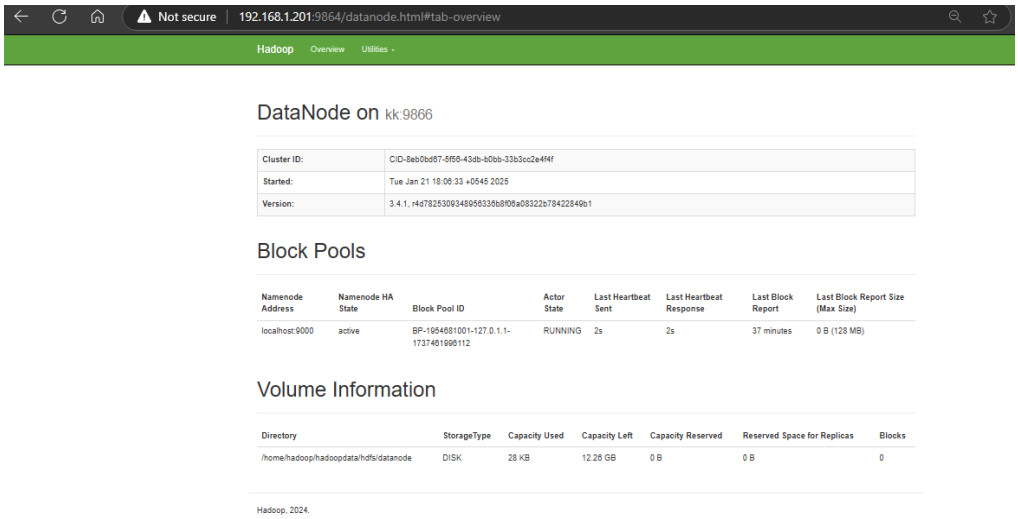
 # jps

```
hadoop@kk:~/hadoop$ jps
5605 ResourceManager
5081 NameNode
5434 SecondaryNameNode
5214 DataNode
6063 Jps
5743 NodeManager
hadoop@kk:~/hadoop$
```

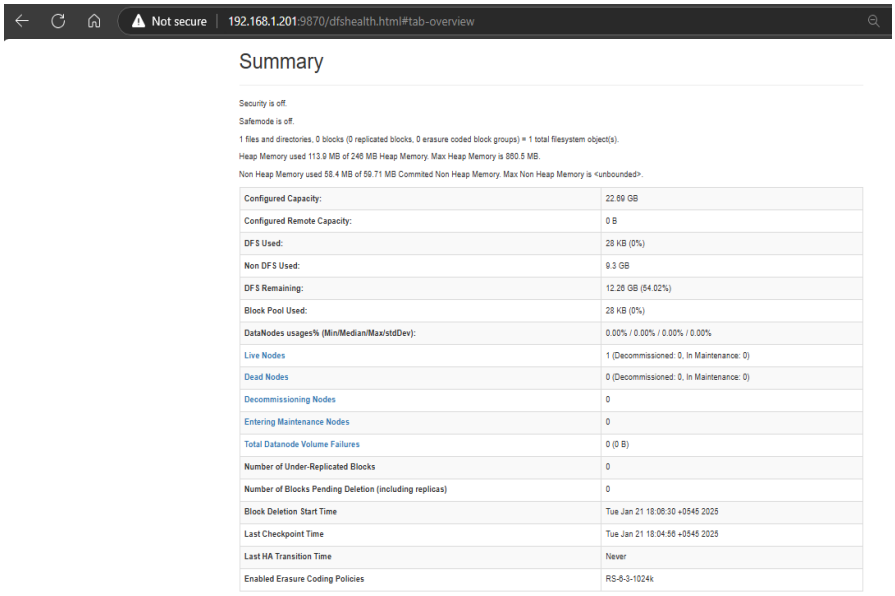
STEP 5: Access Dashboard



DATANODE



NAMENODE



STEP 6: Create directories in HDFS filesystem

 # hdfs dfs -mkdir /kaushal

```
hadoop@kk:/$ hdfs dfs -mkdir /kaushal
```

```
hadoop@kk:~/hadoop$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2025-01-21 13:03 /kaushal
hadoop@kk:~/hadoop$ bin/hdfs dfs -touchz /kaushal/hello.txt
hadoop@kk:~/hadoop$
```

Browse file