Exp.1 210701224

Downloading and installing Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

Aim:

To Download and install Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

Procedure:

Step 1 : Install java jdk 8

First of all you must install Java JDK 8 on your system. You can just type this command to install java jdk on your system.

sudo apt install openjdk-8-jdk

To check it's there cd /usr/lib/jvm

Step 2 : Add this configuration on you bash file

Now just open .bashrc file and paste these commands.

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin
export HADOOP_HOME=~/hadoop-3.2.3/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh

(ssh — secure shell — protocol used to securely connect to remote server/system — transfers data in encrypted form)

sudo apt-get install ssh

now go to hadoop.apache.org website download the tar file
(hadoop.apache.org — download tar file of hadoop.)

tar -zxvf ~/Downloads/hadoop-3.2.3.tar.gz

(Extract the tar file)
cd hadoop-3.2.3/etc/hadoop

now open hadoop-env.hsudo nano hadoop-env.hJAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64 (set the path for JAVA_HOME)

Step 3 : Add this file in core-site.xml

Now add this configuration in core-site.xml file.

core-site.xml

```
<configuration>
 <property>
 <name>fs.defaultFS</name>
 <value>hdfs://localhost:9000</value>  </property>
 <property>
<name>hadoop.proxyuser.dataflair.groups</name> <value>*</value>
 </property>
 <property>
<name>hadoop.proxyuser.dataflair.hosts</name> <value>*</value>
 </property>
 <property>
<name>hadoop.proxyuser.server.hosts</name> <value>*</value>
 </property>
 <property>
<name>hadoop.proxyuser.server.groups</name> <value>*</value>
 </property>
</configuration>
```

Step 3 : Add this file in hdfs-site.xml

Now add this configuration in hdfs-site.xml file.

hdfs-site.xml

```
<configuration>
 <property>
 <name>dfs.replication</name>
 <value>1</value>
 </property>
</configuration>
```

Step 4: Add this file in mapred-site.xml

Now add this configuration in mapred-site.xml file.

mapred-site.xml

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

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPR
ED_HOME/share/hadoop/mapreduce/lib/*</value>
 </property>
</configuration>
```

Step 4: Add this file in yarn-site.xml

Now add this configuration in yarn-site.xml file.

yarn-site.xml

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

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_
CONF_DIR,CLASSPATH_PREP
END_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
 </property>
</configuration>
```

ssh

```
ssh localhost
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
hadoop-3.2.3/bin/hdfs namenode -format
```

format the file system

```
export PDSH_RCMD_TYPE=ssh
```

Step 5 : Start hadoop

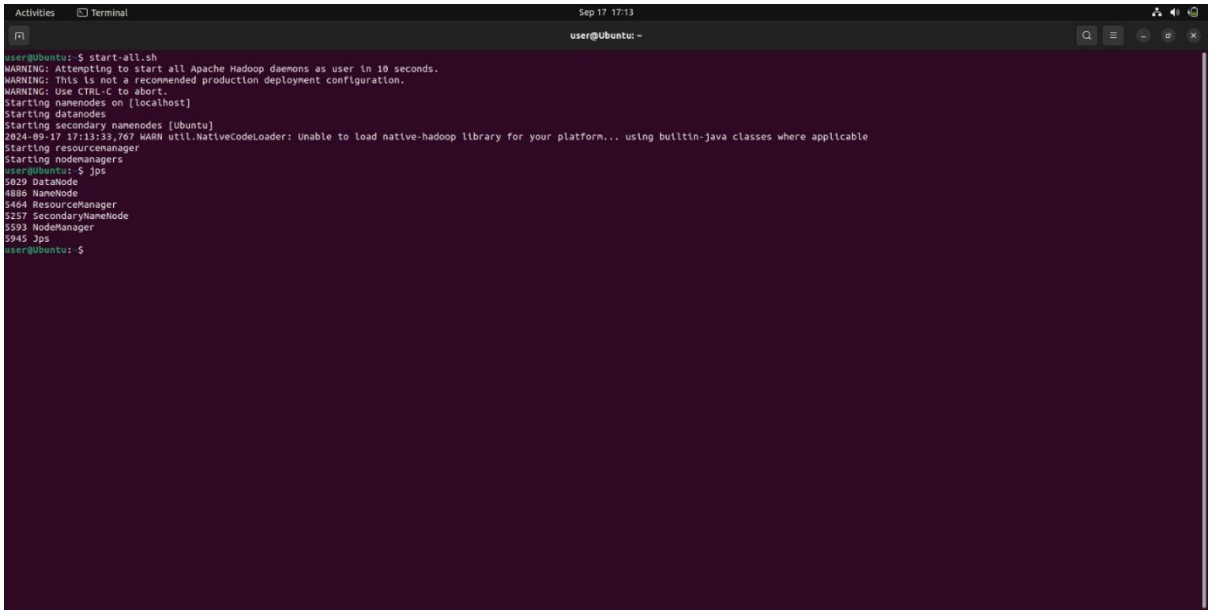
To start

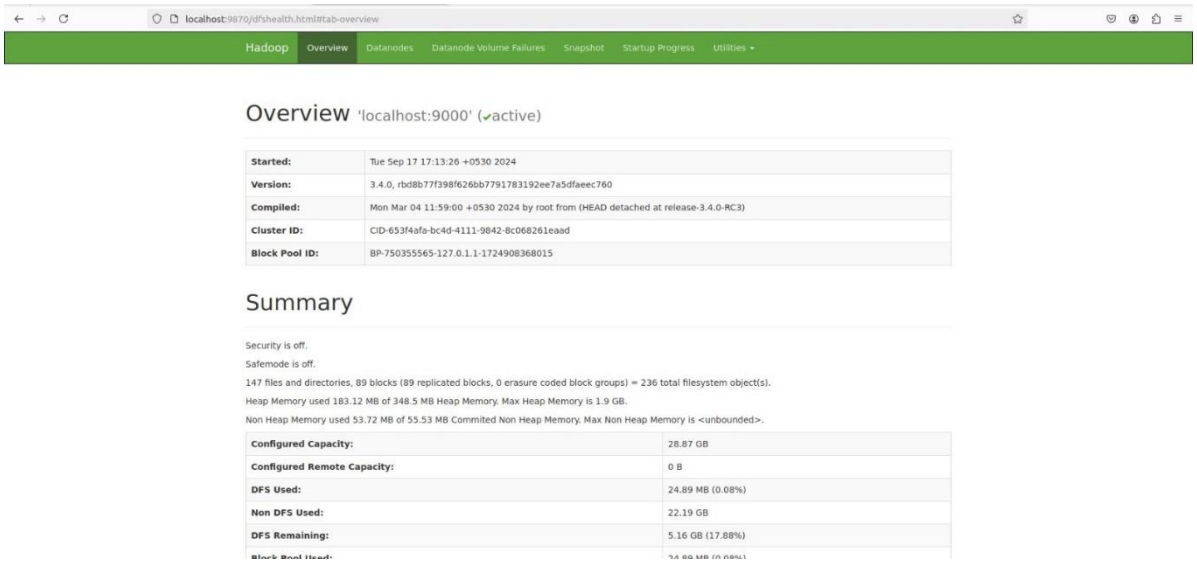
start-all.sh(Start NameNode daemon and DataNode daemon)

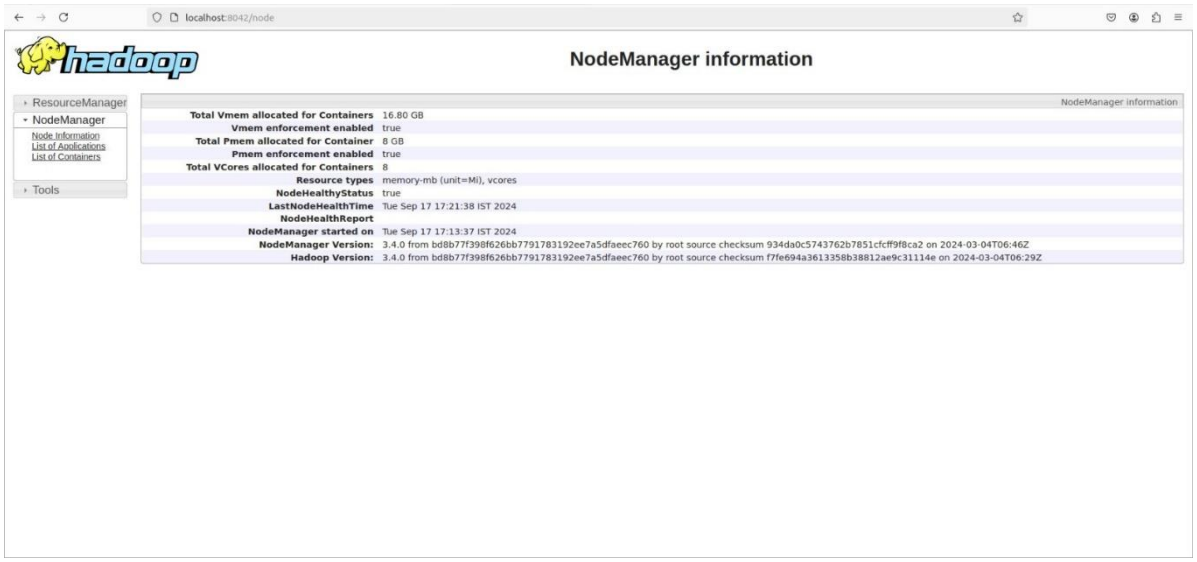
This is how you can install hadoop on your ubuntu operating system and start using on your system.

Step 6 : Check the status using jps

Jps







Result:

The step-by-step installation and configuration of Hadoop on Ubutu linux system have been successfully completed.