Exp.No: 1

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"
exportHADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoopstreaming-3.2.3.jar
export HADOOP_LOG_DIR=$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.h

sudo nano hadoop-env.h

JAVA_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 4: 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 5: 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>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADO
        OP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
        </property>
</configuration>
```

Step 6: Add this file in yarn-site.xml

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

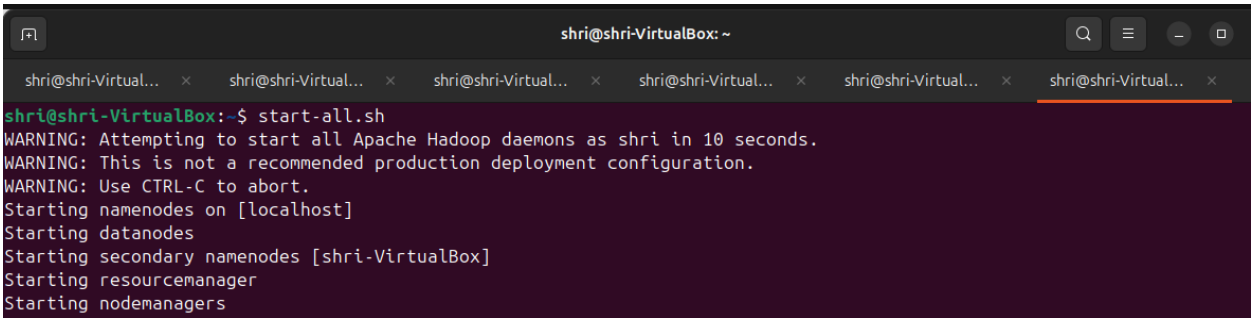
yarn-site.xml

```
<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_PREPEND_DISTCACHE,HADOOP_YARN_
        HOME,HADOOP_MAPRED_HOME</value>
        </property>
</configuration>
```

```
ssh
ssh localhost
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
```

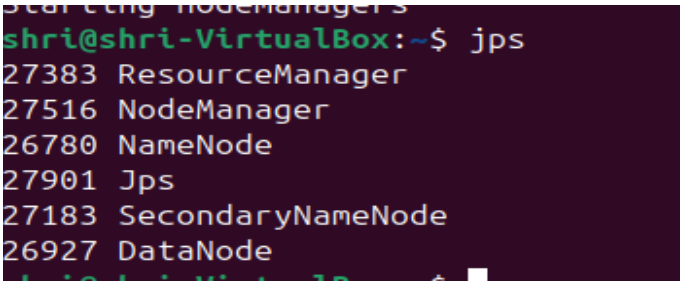
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
hadoop-3.2.3/bin/hdfs namenode -format
Format the file system
export PDSH_RCMD_TYPE=ssh

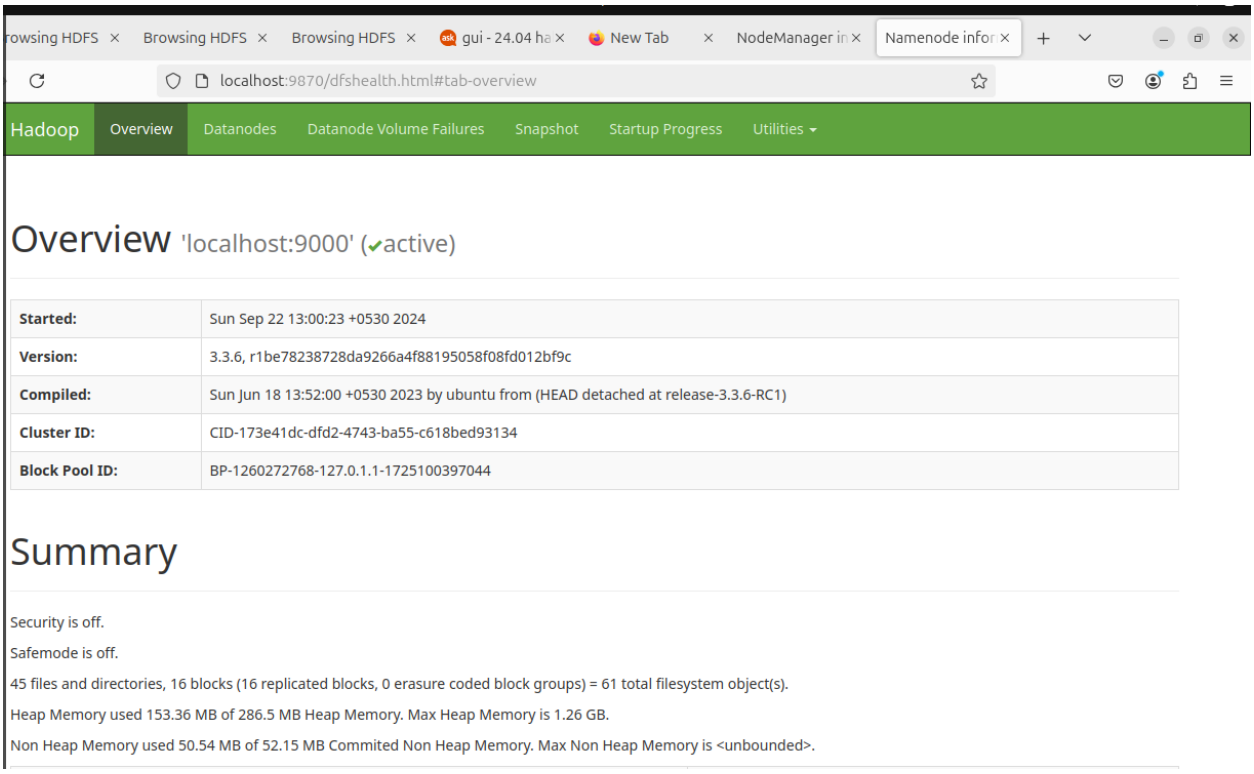
Step 7: Start hadoop
To start, type the command below:
start-all.sh (Start NameNode daemon and DataNode daemon)

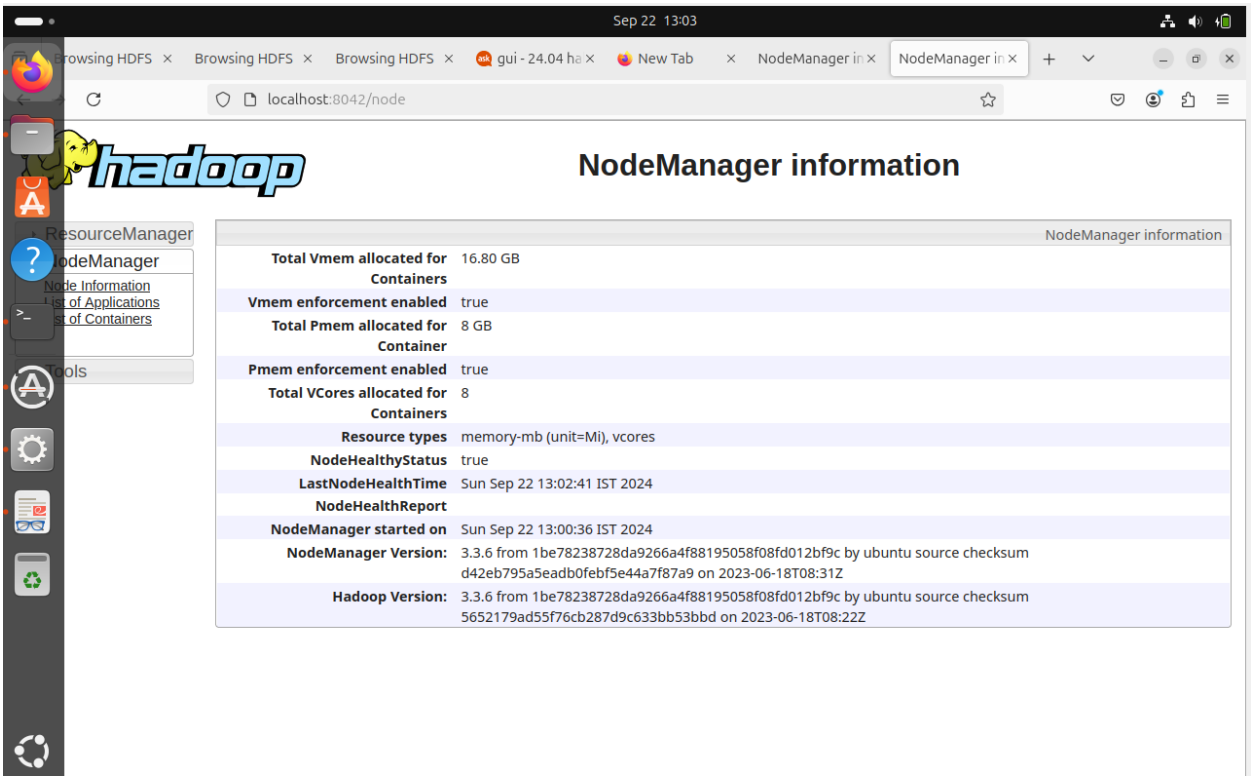


This is how you can install hadoop on your ubuntu operating system and start using on your system.

Step 8: Check the status using jps
jps




Browsing HDFS
Browsing HDFS
Browsing HDFS
gui - 24.04 ha
New Tab
NodeManager in
Namenode infor
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:9000' (active)
Started: Sun Sep 22 13:00:23 +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-173e41dc-dfd2-4743-ba55-c618bed93134
Block Pool ID: BP-1260272768-127.0.1.1-1725100397044
Summary
Security is off.
Safemode is off.
45 files and directories, 16 blocks (16 replicated blocks, 0 erasure coded block groups) = 61 total filesystem object(s).
Heap Memory used 153.36 MB of 286.5 MB Heap Memory. Max Heap Memory is 1.26 GB.
Non Heap Memory used 50.54 MB of 52.15 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.


Sep 22 13:03
Browsing HDFS
Browsing HDFS
Browsing HDFS
gui - 24.04 ha
New Tab
NodeManager in
NodeManager in
localhost:8042/node
hadoop
NodeManager information
ResourceManager
NodeManager
Node Information
List of Applications
List of Containers
Tools
NodeManager information
Total Vmem allocated for Containers 16.80 GB
Vmem enforcement enabled true
Total Pmem allocated for Container 8 GB
Pmem enforcement enabled true
Total VCores allocated for Containers 8
Resource types memory-mb (unit=Mi), vcores
NodeHealthyStatus true
LastNodeHealthTime Sun Sep 22 13:02:41 IST 2024
NodeHealthReport
NodeManager started on Sun Sep 22 13:00:36 IST 2024
NodeManager Version: 3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum d42eb795a5eadb0febf5e44a7f87a9 on 2023-06-18T08:31Z
Hadoop Version: 3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum 5652179ad55f76cb287d9c633bb53bbd on 2023-06-18T08:22Z

Step 9: Stop Hadoop Cluster

To stop the Hadoop all services, run the following command:

stop-all.sh

```
shri@shri-VirtualBox:~$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as shri in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [shri-VirtualBox]
Stopping nodemanagers
Stopping resourcemanager
```

RESULT:

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