UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

NAAC GRADE A+ ACCREDITED UNIVERSITY

Experiment No. 1.1

Student Name: Rishav Kumar
UID: 22MCC20039
Branch: MCA - CCD
Section/Group: MCD-1/ Grp A
Semester: IV
Date of Performance: 18th Jan 24
Subject Name: Big Data Analytics
Subject Code: 22CAP-782

1. Aim/Overview of the practical:

Hadoop Installation: Ubuntu Operating System in stand-alone mode.

2. Code for practical:

- Open terminal in Ubuntu and use following command to install JDK 11

sudo apt install default-jdk default-jre -y

- First, create a new user named hadoop:

sudo adduser Hadoop

- To enable superuser privileges to the new user, add it to the sudo group:

sudo usermod -aG sudo Hadoop

- Switch to the user hadoop:

sudo su - hadoop

- Next, install the OpenSSH server and client:

sudo apt install openssh-server openssh-client -y

- Now, use the following command to generate private and public keys:

ssh-keygen -t rsa

- Now, add the public key to authorized_keys:

cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

- Use the chmod command to change the file permissions of authorized_keys:

sudo chmod 640 ~/.ssh/authorized_keys

- Finally, verify the SSH configuration:

ssh localhost

- Download Hadoop using following command

wget https://downloads.apache.org/hadoop/common/stable/hadoop-3.3.6.tar.gz

- Extract the file using the following command:

tar -xvzf hadoop-3.3.4.tar.gz

- Next, move the extracted file to the /usr/local/hadoop using the following command:

sudo mv hadoop-3.3.4 /usr/local/hadoop

- Now, create a directory using mkdir command to store logs:

sudo mkdir /usr/local/hadoop/logs

- Finally, change the ownership of the /usr/local/hadoop to the user hadoop:

sudo chown -R hadoop:hadoop /usr/local/Hadoop

- Open the .bashrc file using the following command:

sudo nano ~/.bashrc

- Paste following lines in file:

```
export HADOOP_HOME=/usr/local/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 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"
```

- To enable the changes, source the .bashrc file:

source ~/.bashrc

- First, open the hadoop-env.sh file:

sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh

- Paste the following lines in the file to add the path of the Java:

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64

export HADOOP_CLASSPATH+=" $HADOOP_HOME/lib/*.jar"
```

- Next, change your current working directory to /usr/local/hadoop/lib:

cd /usr/local/hadoop/lib

- Here, download the javax activation file:

sudo wget https://jcenter.bintray.com/javax/activation/javax.activation-api/1.2.0/javax.activation-api-1.2.0.jar

- Once done, check the Hadoop version in Ubuntu:

hadoop version

- First, open the core-site.xml file using the following command:

sudo nano $HADOOP_HOME/etc/hadoop/core-site.xml

- And add the following lines in between <configuration> </configuration>:

```
<property>
    <name>fs.default.name</name>
    <value>hdfs://0.0.0.0:9000</value>
    <description>The default file system URI</description>
  </property>
```

- Create a directory to store node metadata using the following command:

sudo mkdir -p /home/hadoop/hdfs/{namenode,datanode}

- And change the ownership of the created directory to the hadoop user:

sudo chown -R hadoop:hadoop /home/hadoop/hdfs

- So first open the configuration file:

sudo nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml

- And paste the following line in between <configuration> ... </configuration>:

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

- To do that, first, open the configuration file using the following command:

sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml

- And paste the following line in between <configuration> ... </configuration>:

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

- First, open the configuration file:

sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

- Paste the following in between <configuration> ... </configuration>:

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

```
<value>mapreduce_shuffle</value>
</property>
```

- Finally, use the following command to validate the Hadoop configuration and to format the HDFS NameNode:

hdfs namenode -format

- Start with starting the NameNode and DataNode:

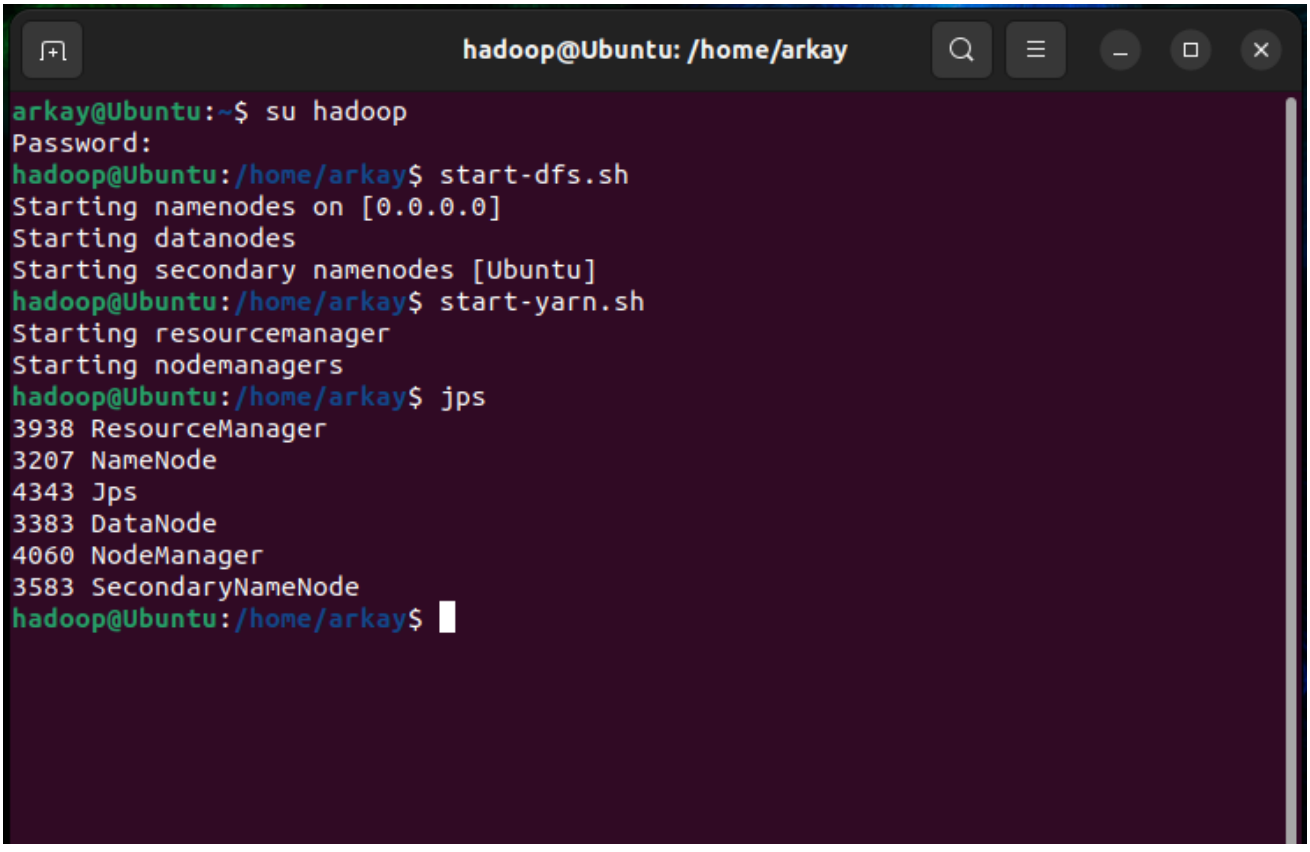
start-dfs.sh

- Start the node manager and resource manager:

start-yarn.sh

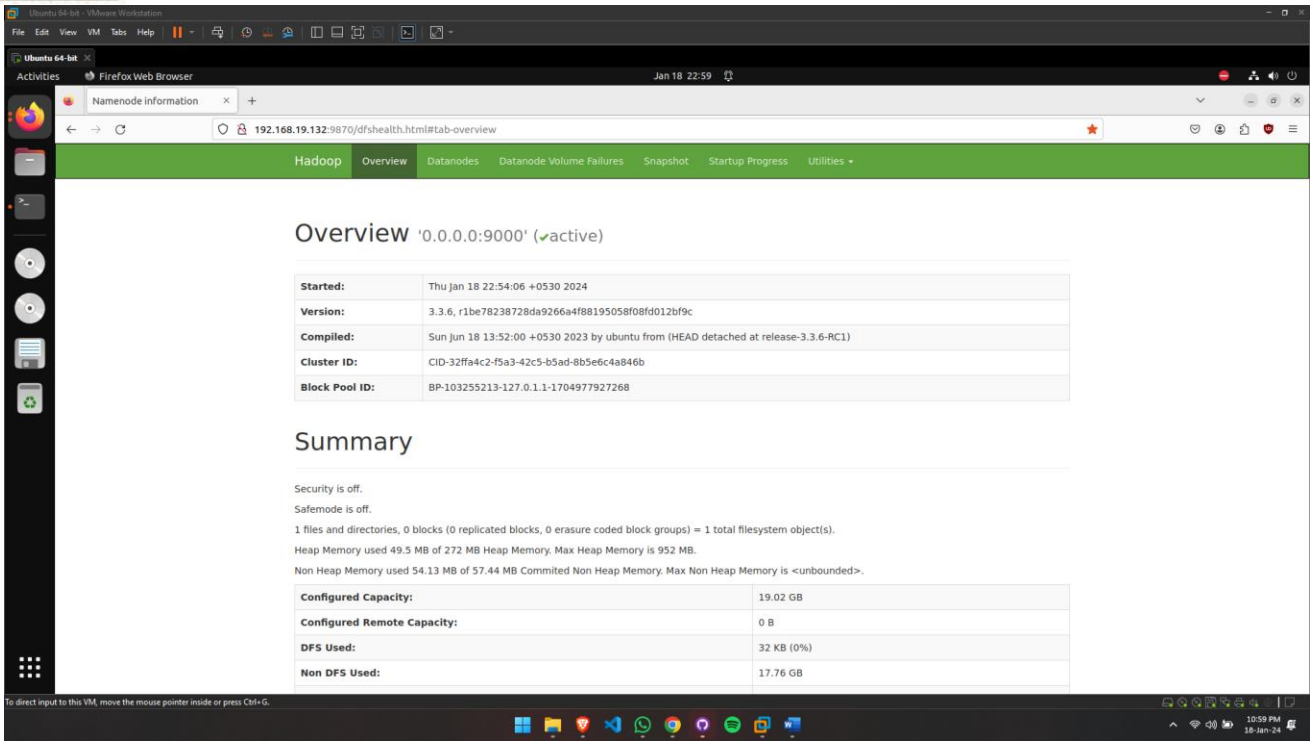
- To verify whether the services are running as intended, use the following command:

jps



- To access the Hadoop web interface, you will have to know your IP and append the port no 9870 in your address bar:

http://server-IP:9870


Activities
Firefox Web Browser
Jan 18 22:59
Namenode information
192.168.19.132:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview '0.0.0.0:9000' (active)
Started: Thu Jan 18 22:54:06 +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-32ffa4c2-f5a3-42c5-b5ad-8b5e6c4a846b
Block Pool ID: BP-103255213-127.0.1.1-1704977927268
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 49.5 MB of 272 MB Heap Memory. Max Heap Memory is 952 MB.
Non Heap Memory used 54.13 MB of 57.44 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.
Configured Capacity: 19.02 GB
Configured Remote Capacity: 0 B
DFS Used: 32 KB (0%)
Non DFS Used: 17.76 GB

********** THE END **********