210701174

Exp.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 Development Kit

The default Ubuntu repositories contain Java 8 and Java 11 both. But, Install Java 8 because hive only works on this version. Use the following command to install it.

\$sudo apt update&&sudo apt install openjdk-8-jdk

Step 2: Verify the Java version

Once installed, verify the installed version of Java with the following

command: \$ java -version Output:

```
naveen@ubuntu:=$ java -version
openjdk version "11.0.24" 2024-07-16
OpenJDK Runtime Environment (build 11.0.24+8-post-Ubuntu-1ubuntu322.04)
OpenJDK 64-Bit Server VM (build 11.0.24+8-post-Ubuntu-1ubuntu322.04, mixed mode, sharing)
naveen@ubuntu:=$
```

Step 3: Install SSH

SSH (Secure Shell) installation is vital for Hadoop as it enables secure communication between nodes in the Hadoop cluster. This ensures data integrity, confidentiality, and allows for efficient distributed processing of data across the cluster. **\$sudo apt install ssh**

Step 4 : Create the hadoop user :

All the Hadoop components will run as the user that you create for Apache Hadoop, and the user will also be used for logging in to Hadoop's web interface. Run the command to create user and set password:

\$ sudo adduser hadoop Output:

```
sanjay@sanjay-VirtualBox:-$ sudo adduser hadoop
Adding user 'hadoop' ...
Adding new group 'hadoop' (1001) ...
Adding new user 'hadoop' (1001) with group 'hadoop (1001)' ...
adduser: The home directory '/home/hadoop' already exists. Not copying from '/etc/skel'.
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
password updated successfully
Changing the user information for hadoop
Enter the new value, or press ENTER for the default
Full Name []:
    Room Number []:
    Work Phone []:
    Home Phone []:
    Other []:
    Is the information correct? [Y/n] y
Adding new user 'hadoop' to supplemental / extra groups 'users' ...
sanjay@sanjay-VirtualBox:-$ su - hadoop
Password:
hadoop@sanjay-VirtualBox:-$ []
```

Step 5: Switch user

Switch to the newly created hadoop user:

\$ su - hadoop

Step 6 : Configure SSH

Now configure password-less SSH access for the newly created hadoop user, so didn't enter the key to save file and passphrase. Generate an SSH keypair (generate Public and Private Key Pairs)first

\$ssh-keygen-t rsa

```
hadoop@sanjay-VirtualBox: S 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:yu8Hsie3mbQ7UifnfH6iam4kFLRRbEb9zVYGUtbaYyg hadoop@sanjay-VirtualBox
The key's randomart image is:
 ----[RSA 3072]----+
      ,0+0. ..+0 |
.0+ . 0 +|
            E..=0.
       0*.B
       +0*++ 0 .
        X@*.+.0
    -[SHA256]-----
hadoop@sanjay-VirtualBox:-$
```

Step 7 : Set permissions :

Next, append the generated public keys from id_rsa.pub to authorized_keys and set proper permission:

\$ cat ~/.ssh/id rsa.pub >> ~/.ssh/authorized keys

\$ chmod 640 ~/.ssh/authorized keys

Step 8 : SSH to the localhost

Next, verify the password less SSH authentication with the following command:

\$ ssh localhost

You will be asked to authenticate hosts by adding RSA keys to known hosts. Type yes and hit Enter to authenticate the localhost:

```
hadoop@ubuntu:/home/naveen$ ssh localhost
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.8.0-40-generic x86_64)

* Documentation: https://help.ubuntu.com

* Management: https://landscape.canonical.com

* Support: https://ubuntu.com/pro

Expanded Security Maintenance for Applications is not enabled.

7 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The list of available updates is more than a week old.
To check for new updates run: sudo apt update
Last login: Mon Aug 19 13:36:03 2024 from 127.0.0.1
hadoop@ubuntu:-$
```

Step 9: Switch user

Again switch to hadoop. So, First, change the user to hadoop with the following command: \$ su-hadoop

Step 10: Install hadoop

Next, download the latest version of Hadoop using the wget command:

\$ wgethttps://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz

Once downloaded, extract the downloaded file:

\$ tar -xvzf hadoop-3.3.6.tar.gz

Next, rename the extracted directory to hadoop:

\$ mv hadoop-3.3.6 hadoop

```
hadoop@sanjay-Virtual8ox:-$ mv hadoop-3.3.6 hadoop
hadoop@sanjay-Virtual8ox:-$ ls
hadoop@sanjay-Virtual8ox:-$ []
hadoop@sanjay-Virtual8ox:-$ []
```

Next, you will need to configure Hadoop and Java Environment Variables on your system. Open the ~/.bashrc file in your favorite text editor. Use nano editior, to pasting the code we use ctrl+shift+v for saving the file ctrl+x and ctrl+y, then hit enter:

Next, you will need to configure Hadoop and Java Environment Variables on your system.

Open the ~/.bashrc file in your favorite text editor:

\$ nano ~/.bashrc

Append the below lines to file.

```
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=$HADOOP_HOME
export HADOOP_HOFS_HOME=$HADOOP_HOME
export HADOOP_YARN_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"
```

Save and close the file. Then, activate the environment variables with the following command:

s\$ source ~/.bashrc

Next, open the Hadoop environment variable file:

\$ nano \$HADOOP_HOME/etc/hadoop/hadoop-

env.sh

Search for the "export JAVA HOME" and configure it.

JAVA HOME=/usr/lib/jvm/java-8-openjdk-amd64

Save and close the file when you are finished.

Step 11: Configuring Hadoop:

First, you will need to create the namenode and datanode directories inside the Hadoop user home directory. Run the following command to create both directories:

\$ cd hadoop/

\$mkdir -p ~/hadoopdata/hdfs/{namenode,datanode}

• Next, edit the core-site.xml file and

```
The list of available updates is more than a week old.
To check for new updates run: sudo apt update
Last login: Mon Aug 19 13:36:03 2024 from 127.0.0.1
hadoop@ubuntu:~$ mkdir -p ~/hadoop/hdfs/(namenode,datanode)
```

update with your system hostname:

\$nano \$HADOOP HOME/etc/hadoop/core-site.xml

Change the following name as per your system hostname:

Save and close the file.

Then, edit the hdfs-site.xml file:

\$nano \$HADOOP HOME/etc/hadoop/hdfs-site.xml

• Change the NameNode and DataNode directory paths as shown below:

• Then, edit the mapred-site.xml file:

\$nano \$HADOOP HOME/etc/hadoop/mapred-site.xml

Make the following changes:

• Then, edit the yarn-site.xml file:

\$nano \$HADOOP_HOME/etc/hadoop/yarnsite.xml

• Make the following changes:

Save the file and close it.

Step 12 – Start Hadoop Cluster

Before starting the Hadoop cluster. You will need to format the Namenode as a hadoop user.

Run the following command to format the Hadoop Namenode:

\$hdfs namenode -format

Once the namenode directory is successfully formatted with hdfs file system, you will see the message "Storage directory /home/hadoop/hadoopdata/hdfs/namenode has been successfully formatted "

Then start the Hadoop cluster with the following command.

\$ start-all.sh

```
hadoop@ubuntu:~$ 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]

Starting datanodes

Starting secondary namenodes [ubuntu]

SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".

SLF4J: Defaulting to no-operation (NOP) logger implementation

SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further detail

s.

Starting resourcemanager

Starting nodemanagers
```

You can now check the status of all Hadoop services using the jps command:

\$ jps

```
hadoop@ubuntu:~$ jps
4624 ResourceManager
5077 Jps
4745 NodeManager
4427 SecondaryNameNode
4204 DataNode
4079 NameNode
```

Step 13 – Access Hadoop Namenode and Resource Manager

 First we need to know our ipaddress, In Ubuntu we need to install net-tools to run ipconfig command,

If you installing net-tools for the first time switch to default user:

\$sudo apt install net-tools

Then run if config command to know our ip address: **if config**

```
hadoop@sanjay-VirtualBox:-/hadoop$ ifconfig
enp8s3: flags=4163-UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.1.6 netmask 255.255.255.0 broadcast 192.168.1.255
    inet6 2401:4900:1c28:46c4:ef76c:b206:abe3:2d45 prefixlen 64 scopeid 0x0<global>
    inet6 2401:4900:1c28:46c4:ed13:53f4:5c05:50c6 prefixlen 64 scopeid 0x0<global>
    inet6 fe80::112b:300a:9242:51f3 prefixlen 64 scopeid 0x20linet6 fe80::12b:300a:9242:51f3 prefixlen 64 scopeid 0x20RX packets 645228 bytes 934388358 (934.3 MB)
    RX errors 0 dropped 0 overruns 0 frame 0
    IX packets 93618 bytes 8998032 (8.9 MB)
    IX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

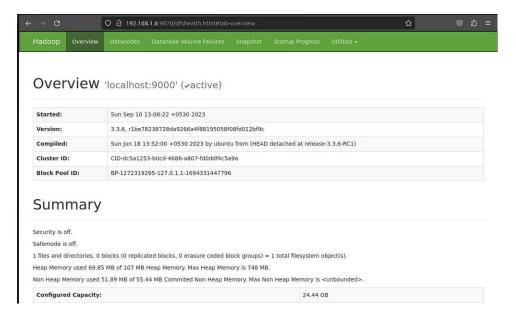
lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10host>
    loop txqueuelen 1000 (Local Loopback)
    RX packets 3331 bytes 491873 (491.8 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    IX packets 3331 bytes 491873 (491.8 KB)
    IX prors 0 dropped 0 overruns 0 carrier 0 collisions 0

hadoop@sanjay-VirtualBox:-/hadoop$ []
```

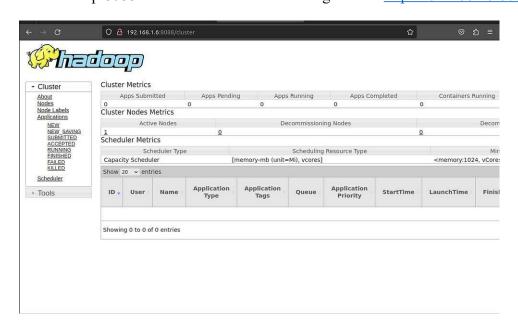
Here my ip address is 192.168.1.6.

- To access the Namenode, open your web browser and visit the URL http://your-serverip:9870.
- You should see the following screen:

http://192.168.1.6:9870



To access Resource Manage, open your web browser and visit the URL http://your-serverip:8088. You should see the following screen: http://192.168.16:8088



Step 14 – Verify the Hadoop Cluster

At this point, the Hadoop cluster is installed and configured. Next, we will create some directories in the HDFS filesystem to test the Hadoop.

Let's create some directories in the HDFS filesystem using the following command:

\$ hdfsdfs -mkdir /test1 \$ hdfsdfs -mkdir /logs Next, run the following command to list the above directory:

\$ hdfs dfs -ls /

You should get the following output:

```
hadoop@sanjay-VirtualBox:-/hadoop$ hdfs dfs ·ls /
Found 2 items
drwxr-xr-x - hadoop supergroup 0 2023-09-10 13:16 /logs
drwxr-xr-x - hadoop supergroup 0 2023-09-10 13:16 /test1
hadoop@sanjay-VirtualBox:-/hadoop$ 

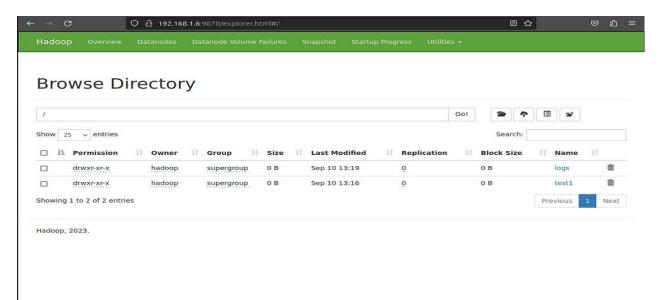
| Adoop@sanjay-VirtualBox:-/hadoop$ |
```

Also, put some files to hadoop file system. For the example, putting log files from host machine to hadoop file system.

\$ hdfs dfs -put /var/log/* /logs/

You can also verify the above files and directory in the Hadoop Namenode web interface.

Go to the web interface, click on the Utilities => Browse the file system. You should see your directories which you have created earlier in the following screen:



Step 15 – Stop Hadoop Cluster

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

\$ stop-all.sh

```
hadoop@ubuntu:~/hadoop$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
Stopping nodemanagers
Stopping resourcemanager
```

Result:

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