Exp-1

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

Aim:

To successfully install, configure, and run Hadoop on a local system using a single-node setup.

Procedure:

1. Install Java and SSH:

o Update your package lists and install OpenJDK 8 and SSH.

```
sudo apt update
sudo apt install openjdk-8-jdk java -
version # Verify Java installation sudo
apt install ssh
```

2. Create Hadoop User:

Add a dedicated user for Hadoop and generate SSH keys for passwordless SSH.

```
sudo adduser hadoop
su - hadoop # Switch to Hadoop user ssh-keygen
-t rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 640 ~/.ssh/authorized_keys ssh localhost #
Test SSH connection to localhost
```

3. Download and Install Hadoop:

o Download the latest Hadoop version (3.3.6), extract the tarball, and move it to the desired location.

```
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz tar -xvzf hadoop-3.3.6.tar.gz mv hadoop-3.3.6 hadoop
```

4. Configure Environment Variables:

o Update. bashrc to include Hadoop and Java paths.

```
nano ~/.bashrc

# Add the following lines at the end
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=$HOME/hadoop

export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
source ~/.bashrc # Apply changes
```

5. Edit Hadoop Configuration Files:

- o Modify configuration files to set up the necessary Hadoop directories and services.
 - o core-site.xml:

```
nano $HADOOP_HOME/etc/hadoop/core-site.xml #
Add between <configuration></configuration>:
```

o hdfs-site.xml:

```
nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml Add:
```

o **mapred-site.xml**:

cp \$HADOOP_HOME/etc/hadoop/mapred-site.xml.template \$HADOOP_HOME/etc/hadoop/mapred-site.xml nano \$HADOOP_HOME/etc/hadoop/mapred-site.xml

Add:

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

o yarn-site.xml:

nano \$HADOOP_HOME/etc/hadoop/yarn-site.xml Add:

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

6. Format the NameNode:

Format the HDFS NameNode.

hdfs namenode -format

7. Start Hadoop:

 Start Hadoop services (NameNode, DataNode, ResourceManager, and NodeManager). start-all.sh

jps # Verify running services

8. Access Web Interfaces:

• Verify that Hadoop is running by accessing the following URLs:

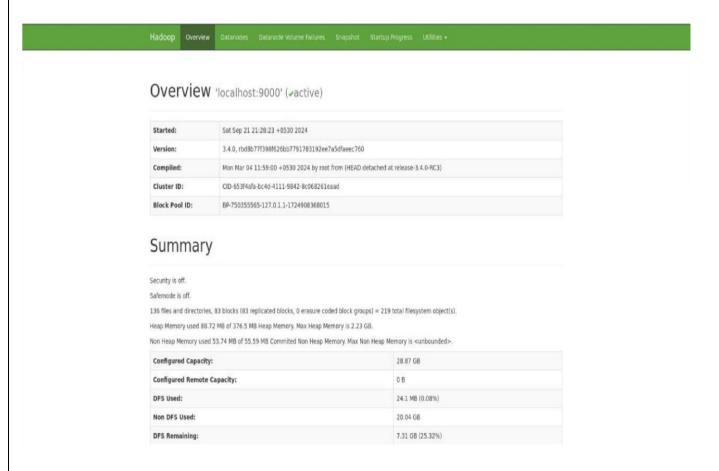
■ NameNode: http://localhost:9870 ■

Resource Manager: http://localhost:8088

9. Stop Hadoop Cluster:

Stop all Hadoop services.

stop-all.sh



RESULT:

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