Installing Hadoop on Ubuntu 20.04

- Install java
  1. sudo apt-get update
  2. sudo apt-get install default-jdk

  Check java is installed or not

- Create Hadoop Group and user
  1. sudo addgroup hadoop
  2. sudo adduser --ingroup hadoop hduser
  3. groups hduser

- Install SSH
  1. sudo apt-get install ssh

  check ssh and sshd

- Create and setud SSH
  1. su hduser
  2. ssh-keygen
  3. cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
  4. ssh localhost

- Install Hadoop

  Download Hadoop
  1. tar xvzf hadoop-3.3.4.tar.gz
  2. sudo -v
  3. su mainuser
  4. sudo adduser hduser sudo
  5. su hduser
  6. sudo mkdir -p /usr/local/hadoop
  7. cd hadoop-3.3.4
  8. sudo mv * /usr/local/hadoop
  9. sudo chown -R hduser:hadoop /usr/local/hadoop

- Setup Configuration Files:-

  Open New Terminal Window:- update-alternatives --config java

1. sudo nano ~/.bashrc

```
#HADOOP VARIABLES START
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-i386
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
#HADOOP VARIABLES END
```

2. source ~/.bashrc

3. sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh

```
export JAVA_HOME=”JAVA PATH”
```

4. sudo mkdir -p /app/hadoop/tmp
5. sudo chown hduser:hadoop /app/hadoop/tmp
6. sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml

```
<configuration>
<property>
<name>hadoop.tmp.dir</name>
<value>/app/hadoop/tmp</value>
<description>A base for other temporary
directories.</description>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:54310</value>
<description>The name of the default file system.
A URI whose
scheme and authority determine the FileSystem implementation.
The
uri's scheme determines the config property (fs.SCHEME.impl)
naming
the FileSystem implementation class.
used to
The uri's authority is
determine the host, port, etc. for a filesystem.</description>
</property>
</configuration>
```

7. sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:54311</value>
<description>The host and port that the MapReduce job tracker
runs
at.
If "local", then jobs are run in-process as a single map
and reduce task.
</description>
</property>
</configuration>
```

8. sudo mkdir -p /usr/local/hadoop_store/hdfs/namenode
9. sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode
10. sudo chown -R hduser:hadoop /usr/local/hadoop_store
11. sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
<description>Default block replication.
The actual number of replications can be specified when the
file is created.
The default is used if replication is not specified in create
time.
</description>
</property>
<property>
<name>dfs.block.size</name>
<value>1048576</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>file:/usr/local/hadoop_store/hdfs/namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>file:/usr/local/hadoop_store/hdfs/datanode</value>
</property>
</configuration>
```

12. hadoop namenode -format
13. su mainuser
14. cd /usr/local/hadoop/sbin
15. ls
16. sudo su hduser
17. start-dfs.sh
18. start-yarn.sh
19. jps

20 stop-yarn.sh
21. stop-dfs.sh

- Setting up Trash:-
  1. touch example.desktop
     hadoop fs -put example.desktop /
  2. Go to web UI
     - http://localhost:9870/
  3. cd /*usr*/local/hadoop/etc/hadoop/
  4. sudo nano core-site.xml

```
<property>
<name>fs.trash.interval</name>
<value>3</value>
</property>
<property>
<name>fs.trash.checkpoint.interval</name>
<value>1</value>
</property>
```

22. hadoop fs –chmod –R 755 /user

Permissions:-

hdfs-site.xml:-
<property>
<name>dfs.permissions.enabled</name>
<value>true</value>
</property>

core-site.xml:-
<property>
<name>hadoop.http.staticuser.user</name>
<value>hduser</value>
</property>