HADOOP INSTALL

COMMANDS

HADOOP VERSION – 2.9.0

JAVA – 8 (1.8.0)

WORKS ON ANY UBUNTU VERSION NO HDUSER REQUIRED

Installing and moving (make sure the java version is same as given below)

sudo apt install openjdk-8-jdk -y

java -version

sudo apt install ssh -y

ssh-keygen -t rsa -P ""

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

chmod 600 ~/.ssh/authorized_keys

ssh localhost

tar -xzf hadoop-2.9.0.tar.gz

sudo mv hadoop-2.9.0 /usr/local/hadoop

Set Environment Variables

nano ~/.bashrc **### go to end and paste it**

 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

 export HADOOP_HOME=/usr/local/hadoop

 export PATH=$PATH:$HADOOP_HOME/bin

 export PATH=$PATH:$HADOOP_HOME/sbin

 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 HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"

export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
```

source ~/.bashrc

Step 3: Configure Hadoop

cd $HADOOP_HOME/etc/hadoop

nano core-site.xml

```
<configuration>
 <property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
 </property>
</configuration>
```

nano hdfs-site.xml

```
<configuration>
 <property>
  <name>dfs.replication</name>
  <value>1</value>
 </property>
 <property>
```

```xml
        <name>dfs.namenode.name.dir</name>
        <value>file:///usr/local/hadoop/hdfs/namenode</value>
      </property>
      <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:///usr/local/hadoop/hdfs/datanode</value>
      </property>
      <property>
        <name>dfs.namenode.rpc-address</name>
        <value>localhost:9000</value>
      </property>
      <property>
        <name>dfs.namenode.http-address</name>
        <value>localhost:9870</value>
      </property>
    </configuration>
```

cp mapred-site.xml.template mapred-site.xml

nano mapred-site.xml

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

nano yarn-site.xml

```
<configuration>
 <property>
  <name>yarn.resourcemanager.address</name>
  <value>localhost:8032</value>
 </property>
 <property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
 </property>
</configuration>
```

nano hadoop-env.sh

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

hdfs namenode -format

start-dfs.sh

start-yarn.sh

go to http://localhost:9870 for Hadoop dashboard and http://localhost:8088 for manager

MAP REDUCE

COMMANDS

Create a folder called analyzelogs in the home and paste following files in it:

1) **Link CLASSPATH & MANIFEST**
2) **Link log file**
3) **Mapper**
4) **Reducer**
5) **Driver**

These files can be downloaded from the url -
https://www.youtube.com/watch?v=gXQT7vHB59Q

cd ~/analyzelogs

convert the accesslog.txt to csv by opening it in libre calc and setting rows 2, separated by other –

####################################

ls

export CLASSPATH="$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.9.0.jar:$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.9.0.jar:$HADOOP_HOME/share/hadoop/common/hadoop-common-2.9.0.jar:~/analyzelogs/SalesCountry/*:$HADOOP_HOME/lib/*"

javac -d . SalesMapper.java SalesCountryReducer.java SalesCountryDriver.java

nano Manifest.txt

######### paste following in the Manifest.txt

Main-Class: SalesCountry.SalesCountryDriver

####################################

jar -cfm analyzelogs.jar Manifest.txt SalesCountry/*.class

start-dfs.sh

start-yarn.sh

```
mkdir ~/input2000

cp access_log_short.csv ~/input2000/

$HADOOP_HOME/bin/hdfs dfs -put ~/input2000 /

$HADOOP_HOME/bin/hadoop jar analyzelogs.jar /input2000 /output2000

$HADOOP_HOME/bin/hdfs dfs -cat /output2000/part-00000

jar -cfm analyzelogs.jar Manifest.txt SalesCountry/*.class

cp access_log_short.csv ~/input2000/

$HADOOP_HOME/bin/hdfs dfs -put ~/input2000 /

$HADOOP_HOME/bin/hadoop jar analyzelogs.jar /input2000 /output2000
```

HIVE INSTALL

COMMANDS

HIVE VERSION – 1.2.2

JAVA – 8 (1.8.0)

sudo cp apache-hive-1.2.2-bin.tar.gz /usr/local/

nano ~/.bashrc **### go to end and paste it**

      export HIVE_HOME=/usr/local/hive

source ~/.bashrc

cd /usr/local

sudo tar -xvzf apache-hive-1.2.2-bin.tar.gz

sudo mv apache-hive-1.2.2-bin hive

start-dfs.sh

start-yarn.sh

cd hadoop/bin/

hdfs dfs -mkdir -p /user/hive/warehouse

hdfs dfs -mkdir -p /tmp/hive

hdfs dfs -chmod 777 /tmp

hdfs dfs -chmod 777 /user/hive/warehouse

hdfs dfs -chmod 777 /tmp/hive

cd /usr/local/hive/bin/

sudo chmod -R 755 /usr/local/hive

sudo chown -R omkar:omkar /usr/local/hive

./schematool -initSchema -dbType derby

./hive