HADOOP-3.2.1 Insatallation with commands

By Dipanshu Modi

```
sudo apt install openjdk-8-jdk
cd /usr/lib/jvm
make sure its there

open bashrc file
sudo nano ~/.bashrc
paste these
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin

export HADOOP_HOME=~/Downloads/hadoop-3.2.1/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export
HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.2.1.jar
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh
export HIVE_HOME=~/Downloads/apache-hive-3.1.2-bin
export PATH=$PATH:~/Downloads/apache-hive-3.1.2-bin/bin

come out of the file

sudo apt-get install ssh
( ssh - secure shell - protocol used to securely connect to remote server/system -
transfers data in encrypted form)

(parallel shell tool to run commands accross multiple nodes in cluster)

now go to hadoop.apache.org website
download the tar file
(hadoop.apache.org - download tar file of hadoop.)
tar -zxvf ~/Downloads/hadoop-3.2.1.tar.gz
(Extract the tar file)
cd hadoop-3.2.1/etc/hadoop

now open hadoop-env.h
```

```
sudo nano hadoop-env.h
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
(set the path for JAVA_HOME)
edit all the files-


core-site.xml
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
    <property>
        <name>hadoop.proxyuser.dataflair.groups</name>
        <value>*</value>
    </property>
    <property>
        <name>hadoop.proxyuser.dataflair.hosts</name>
        <value>*</value>
     </property>
     <property>
        <name>hadoop.proxyuser.server.hosts</name>
        <value>*</value>
     </property>
     <property>
        <name>hadoop.proxyuser.server.groups</name>
        <value>*</value>
     </property>
</configuration>



hdfs-site.xml
<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
</configuration>




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

```
        <name>mapreduce.application.classpath</name>

<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoo
p/mapreduce/lib/*</value>
    </property>
</configuration>



yarn-site.xml
<configuration>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREP
END_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
    </property>
</configuration>

ssh localhost
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys

~/Downloads/hadoop-3.2.1/bin/hdfs namenode -format
(Format the filesystem)
export PDSH_RCMD_TYPE=ssh

~/Downloads/hadoop-3.2.1/sbin/start-dfs.sh
(Start NameNode daemon and DataNode daemon)
localhost:9870
/// Hadoop Insatlled :)
Make the HDFS directories required to execute MapReduce jobs:
~/Downloads/hadoop-3.2.1/bin/hdfs dfs -mkdir /user
~/Downloads/hadoop-3.2.1/bin/hdfs dfs -mkdir /user/<username>

Copy the input files into the distributed filesystem:
~/Downloads/hadoop-3.2.1/bin/hdfs dfs -mkdir input
~/Downloads/hadoop-3.2.1/bin/hdfs dfs -put
~/Downloads/hadoop-3.2.1/etc/hadoop/*.xml input
~/Downloads/hadoop-3.2.1/sbin/stop-dfs.sh
```