HADOOP INSTALLATION

1. sudo apt-get update
2. sudo adduser hduser
3. sudo usermod -aG sudo hduser
4. groups hduser
5. *--LOGIN TO HDUSER--*

6. cd Downloads
7. sudo tar xfz jdk-8u45-linux-x64.tar.gz
8. sudo mv jdk1.8.0_45 /usr/local/java
9. gedit ~/.bashrc

Add Contents:

 # JAVA PATH VARIABLES

export JAVA_HOME=/usr/local/java

export PATH=$JAVA_HOME/bin:{$PATH}

export CLASSPATH=$JAVA_HOME/lib

 # END OF JAVA PATH VARIABLES

10. source ~/.bashrc
11. java -version
12. sudo apt-get install ssh
13. sudo apt-get install rsync
14. sudo apt-get install openssh-server
15. sudo apt-get install openssh-client
16. ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
17. sudo cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
18. ssh localhost
19. exit
20. sudo tar xfz hadoop-3.1.4.tar.gz
21. sudo mv hadoop-3.1.4 /usr/local/hadoop

22. sudo nano ~/.bashrc

Add Contents:

#HADOOP VARIABLES START

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"

export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar

#HADOOP VARIABLES END

23. source ~/.bashrc
24. sudo nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh

Content change:

   export JAVA_HOME=/usr/local/java

25. sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml

Add contents:

<configuration>

<property>

```xml
        <name>fs.default.name</name>

        <value>hdfs://localhost:9000</value>

        </property>

        </configuration>
```

26.sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

```xml
<configuration>


<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
<name>yarn.nodemanager.vmem-check-enabled</name>
<value>false</value>
</property>
<property>
<name>yarn.nodemanager.vmem-pmem-ratio</name>
<value>4</value>
</property>
```

</configuration>

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

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP_MAPRED_HOME=/usr/local/hadoop</value>

</property>

<property>

<name>mapreduce.map.env</name>

<value>HADOOP_MAPRED_HOME=/usr/local/hadoop</value>

</property>

<property>

<name>mapreduce.reduce.env</name>

<value>HADOOP_MAPRED_HOME=/usr/local/hadoop</value>

</property>

<property>

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

<value>/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoo

p/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/share/hadoop/yarn:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr/local/java/lib/tools.jar</value>

</property>

</configuration>

28.sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1</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>

29.sudo mkdir -p /usr/local/hadoop_store/hdfs

30.sudo chown hduser:hduser -R /usr/local/hadoop

31.sudo chown hduser:hduser -R /usr/local/hadoop_store

32.sudo chmod -R 777 /usr/local/hadoop

33.sudo chmod -R 777 /usr/local/hadoop_store

34.hdfs namenode -format

35.start-all.sh

36.jps

37.gedit wcinput.txt

38.hdfs dfs -mkdir -p /user/hduser/wordcount/input

39.hdfs dfs -put wcinput.txt /user/hduser/wordcount/input

40.hdfs dfs -ls /user/hduser/wordcount/input

41.hdfs dfs -cat /user/hduser/wordcount/input/wcinput.txt

42.hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-
   mapreduce-examples-3.1.4.jar wordcount
   /user/hduser/wordcount/input /user/hduser/wordcount/output

43.hdfs dfs -ls /user/hduser/wordcount/output

44.hdfs dfs -cat /user/hduser/wordcount/output/part-r-00000

45.stop-all.sh