

PROCEDURE:

Step 1

First we need to install the prerequisites for hadoop. To do that type the following commands on the linux terminal

```
$sudo apt update  
$sudo apt-get upgrade
```

Step 2

Now install adopt openjdk 11

i) First add the adopt openjdk repo to update list

```
$wget -qO- https://adoptopenjdk.jfrog.io/adoptopenjdk/api/gpg/key/public | sudo apt-key add -
```

ii) Now add the software properties

```
$apt-get install -y software-properties-common
```

iii) Install java now

```
$apt-get install adoptopenjdk-11-hotspot
```

iv) Check if java is installed using these commands

```
$java -version
```

```
$javac -version
```

Step 3

Now install openssh client

```
$sudo apt install openssh-server openssh-client -y
```

Step 4

Now create a hadoop user

```
$sudo adduser hadoop
```

i) Create a password for your hadoop user and remember it

ii) Also add the necessary details it asks

Step 5

Now switch hadoop user

\$su – hadoop

NOTE: ENTER THE PASSWORD YOU CREATED WHEN PROMPTED.

Step 6

Generate a key

ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa

Step 7

Now store it as a public key using cat command

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

Step 8

Set permissions using chmod command

\$chmod 0600 ~/.ssh/authorized_keys

Step 9

Now start ssh

\$ssh localhost

Step 10

Hadoop tar.gz file will be there on your pc. Locate it and unzip it using

\$tar xzf <location of your file>/hadoop-3.3.0.tar.gz

Step 11

Now after hadoop installs, you are supposed to modify 6 files.

i).bashrc

To modify this file you need root permissions so first switch to microstack user

\$su – microstack

password for this is 'microstack'

Now switch to root using

\$sudo su

Enter the same password if prompted

Now modify the file and add the following properties at the end of the file

To modify

\$sudo nano .bashrc

Add the following

```
export HADOOP_HOME=/home/hadoop/hadoop/hadoop-3.3.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$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 PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS"-Djava.library.path=$HADOOP_HOME/lib/nativ"
```

SAVE AND EXIT NANO USING Ctrl+O and Ctrl+X

hadoop-env.sh

Now switch to hadoop user using

\$su – hadoop

Now go to hadoop-env.sh location using

\$cd /home/hadoop/hadoop/hadoop-3.3.0/hadoop/etc

Now find java location using

\$update-alternatives --list java

And note the location we need that.

Now add the properties

\$sudo nano hadoop-env.sh

export JAVA_HOME=<your java location>

SAVE AND EXIT NANO USING Ctrl+O and Ctrl+X

core-site.xml

\$sudo nano core-site.xml

ADD THE PROPERTIES AT THE END OF THE FILE

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/home/hadoop/tmpdata</value>

</property>

```
<property>

  <name>fs.default.name</name>

  <value>hdfs://127.0.0.1:9000</value>

</property>
```

```
</configuration>
```

SAVE AND EXIT NANO USING Ctrl+O and Ctrl+X

hdfs-site.xml

\$sudo nano hdfs-site.xml

\$sudo nano hdfs-site.xml

Add the properties

```
<configuration>

<property>

  <name>dfs.data.dir</name>

  <value>/home/hadoop/dfsdata/namenode</value>

</property>

<property>

  <name>dfs.data.dir</name>

  <value>/home/hadoop/dfsdata/datanode</value>

</property>

<property>

  <name>dfs.replication</name>

  <value>1</value>

</property>

</configuration>
```

SAVE AND EXIT NANO USING Ctrl+O and Ctrl+X

mapred-site.xml

\$sudo nano mapred-site.xml

Add the properties

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

SAVE AND EXIT NANO USING Ctrl+O and Ctrl+X

yarn-site.xml

\$sudo nano yarn-site.xml

Add the properties

```
<configuration>

<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.resourcemanager.hostname</name>

  <value>127.0.0.1</value>

</property>

<property>

  <name>yarn.acl.enable</name>

  <value>0</value>

</property>

<property>

  <name>yarn.nodemanager.env-whitelist</name>
```

```
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
```

```
</property>
```

```
</configuration>
```

SAVE AND EXIT NANO USING Ctrl+O and Ctrl+X

Step 12

Now start the namenode

```
$hdfs namenode -format
```

Step 13

Navigate back to hadoop-3.3.0/bin folder

```
$cd ~
```

```
$cd ~
```

```
$cd bin
```

Step 14

Now start hadoop

```
$/start-dfs.sh
```

```
$/start-yarn.sh
```

Step 15

Now run jps

```
$jps
```

Step 16

Open browser and type

<http://localhost:9870>

Step 17

Come back to the same terminal

```
$hdfs dfs -mkdir /inputdir
```

Step 18

Now create a folder called sample

\$mkdir sample

Step 19

Go inside sample folder and create a text file and input some data

\$cd sample

\$sudo nano sample.txt

TYPE YOUR DATA AND SAVE AND EXIT USING CTRL+O AND CTRL+X

Now go back to bin using

\$ cd ~

Step 20

Put the sample folder inside inputdir we created using

\$hdfs dfs -put /home/hadoop/hadoop/hadoop-3.3.0/bin/sample /inputdir

Step 21

Now run the jar file of mapreduce

\$hadoop jar /home/hadoop/hadoop/hadoop-3.3.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.0.jar grep /inputtext/sample/sample.txt /output '(am)'

Step 22

Now open your browser, Navigate to utilities-> Browse the filesystem -> Now open output folder and download the **part-r-00000** file

Step 23

Open the file and check the output.

Step 24

STOP

```
hadoop@microstack-Veriton-M200-H110: ~/hadoop/hadoop-3...
bash: export: `HADOOP_OPTS-Djava.library.path=/home/hadoop/hadoop/hadoop-3.3.0/lib/nativ': not a valid identifier
microstack@microstack-Veriton-M200-H110:/home/hadoop/hadoop/etc/hadoop$ sudo nano core-site.xml
[sudo] password for microstack:
microstack@microstack-Veriton-M200-H110:/home/hadoop/hadoop/etc/hadoop$ sudo nano hdfs-site.xml
microstack@microstack-Veriton-M200-H110:/home/hadoop/hadoop/etc/hadoop$ sudo nano mapred-site.xml
microstack@microstack-Veriton-M200-H110:/home/hadoop/hadoop/etc/hadoop$ sudo nano yarn-site.xml
microstack@microstack-Veriton-M200-H110:/home/hadoop/hadoop/etc/hadoop$ su - hadoop
Password:
hadoop@microstack-Veriton-M200-H110:~$ hdfs namenode -format
2021-09-30 20:16:23,584 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = microstack-Veriton-M200-H110/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.0
STARTUP_MSG:   classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/sh
```

```
hadoop@microstack-Veriton-M200-H110: ~/hadoop/hadoop-3...
2021-09-30 20:16:26,326 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at microstack-Veriton-M200-H110/127.0.1.1
*****/
hadoop@microstack-Veriton-M200-H110:~$ ./start-dfs.sh
-bash: ./start-dfs.sh: No such file or directory
hadoop@microstack-Veriton-M200-H110:~$ ./start-dfs.sh
-bash: ./start-dfs.sh: No such file or directory
hadoop@microstack-Veriton-M200-H110:~$ l
hadoop/  hadoop-3.3.0.tar.gz  hadoop-3.3.0.tar.gz.1  hadoopdata/
hadoop@microstack-Veriton-M200-H110:~$ cd hadoop/hadoop
-bash: cd: hadoop/hadoop: No such file or directory
hadoop@microstack-Veriton-M200-H110:~$ cd hadoop
hadoop@microstack-Veriton-M200-H110:~/hadoop$ l
bin/      hadoop-3.3.0/  libexec/  LICENSE.txt  NOTICE.txt  share/
dfsdata/  include/      LICENSE-binary  logs/      README.txt  tmpdata/
etc/      lib/          licenses-binary/  NOTICE-binary  sbin/
hadoop@microstack-Veriton-M200-H110:~/hadoop$ cd hadoop-3.3.0
hadoop@microstack-Veriton-M200-H110:~/hadoop/hadoop-3.3.0$ cd bin
hadoop@microstack-Veriton-M200-H110:~/hadoop/hadoop-3.3.0/bin$ hdfs dfs -mkdir /inputdir
hadoop@microstack-Veriton-M200-H110:~/hadoop/hadoop-3.3.0/bin$ l
container-executor*  hdfs*  mapred.cmd*  yarn*
hadoop*  hdfs.cmd*  oom-listener*  yarn.cmd*
Starting NameNode
```



```
hadoop@microstack-Veriton-M200-H110:~/hadoop/hadoop-3.3.0/bin$ hdfs dfs -put /home/hadoop/hadoop/hadoop-3.3.0/bin/sample /inputtext
```

```
hadoop@microstack-Veriton-M200-H110: ~/hadoop/hadoop-3.3.0/bin
hadoop@microstack-Veriton-M200-H110:~/hadoop/hadoop-3.3.0/bin$ hadoop jar /home/hadoop/hadoop/hadoop-3.3.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.0.jar grep /inputtext/sample/sample.txt /output '(am)'
2021-09-30 11:00:53,477 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-09-30 11:00:53,754 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1633013328334_0017
2021-09-30 11:00:53,952 INFO input.FileInputFormat: Total input files to process : 1
2021-09-30 11:00:54,070 INFO mapreduce.JobSubmitter: number of splits:1
2021-09-30 11:00:54,303 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1633013328334_0017
2021-09-30 11:00:54,303 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-09-30 11:00:54,431 INFO conf.Configuration: resource-types.xml not found
2021-09-30 11:00:54,431 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-09-30 11:00:54,473 INFO impl.YarnClientImpl: Submitted application application_1633013328334_0017
2021-09-30 11:00:54,498 INFO mapreduce.Job: The url to track the job: http://microstack-Veriton-M200-H110:8088/proxy/application_1633013328334_0017/
2021-09-30 11:00:54,498 INFO mapreduce.Job: Running job: job_1633013328334_0017
2021-09-30 11:00:59,563 INFO mapreduce.Job: Job job_1633013328334_0017 running in uber mode : false
2021-09-30 11:00:59,565 INFO mapreduce.Job: map 0% reduce 0%
2021-09-30 11:01:02,595 INFO mapreduce.Job: map 100% reduce 0%
2021-09-30 11:01:06,632 INFO mapreduce.Job: map 100% reduce 100%
```

```
Total megabyte-milliseconds taken by all reduce tasks=1871872
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Map output bytes=11
  Map output materialized bytes=19
  Input split bytes=131
  Combine input records=0
  Combine output records=0
```

Open part-r-00000 ~/Downloads Save

1 8 hello

```
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=105
File Output Format Counters
  Bytes Written=5
hadoop@microstack-Veriton-M200-H110:~/hadoop/hadoop-3.3.0/bin$
```

localhost:9870/explorer.html#/output2

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Sep 30 21:06	1	128 MB	._SUCCESS	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	5 B	Sep 30 21:06	1	128 MB	part-r-00000	<input type="checkbox"/>

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2020.

