

BIG DATA

HADOOP HANDS ON

1a.png

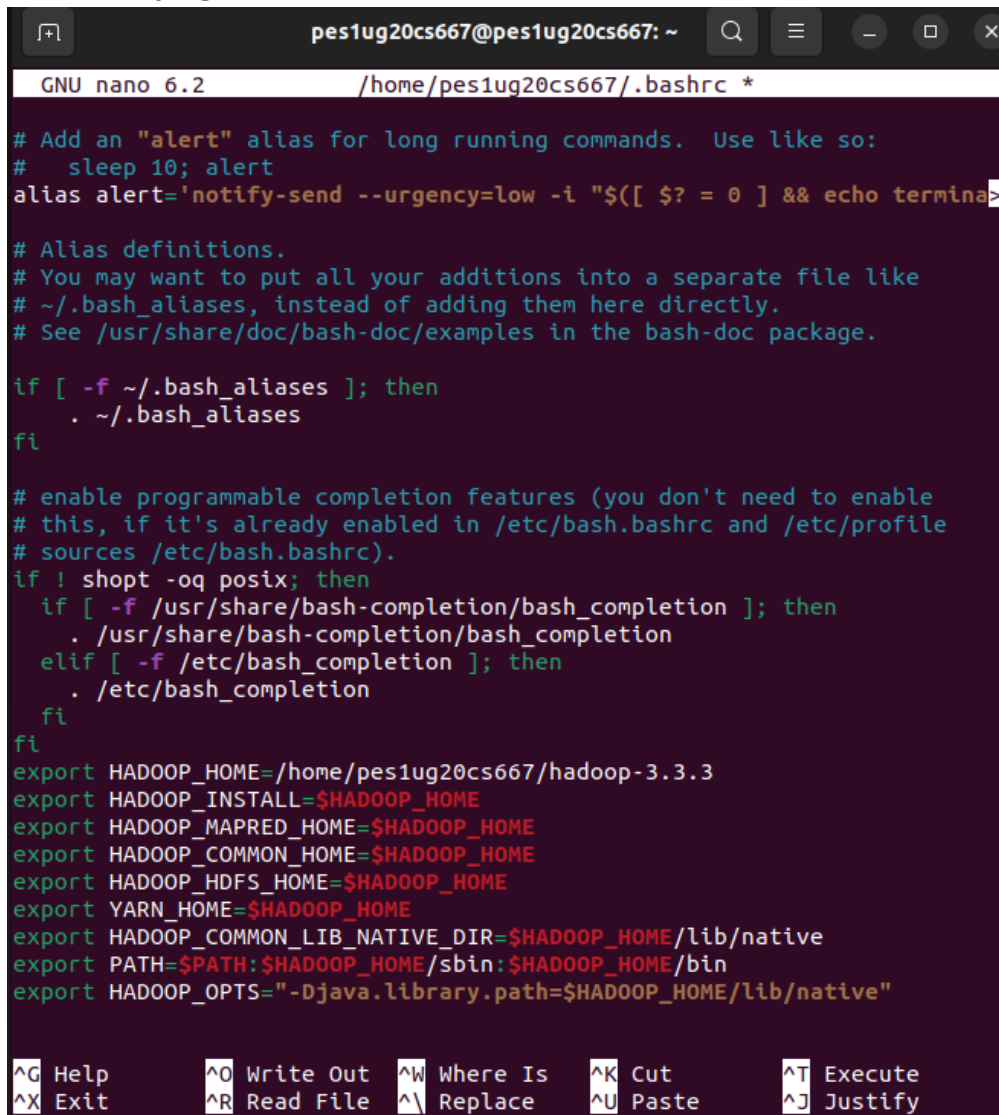
```
pes1ug20cs667@pes1ug20cs667:~$ ssh localhost
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.15.0-46-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

0 updates can be applied immediately.

Last login: Fri Aug 26 14:42:24 2022 from 127.0.0.1
pes1ug20cs667@pes1ug20cs667:~$ exit
logout
Connection to localhost closed.
pes1ug20cs667@pes1ug20cs667:~$
```

2a.png

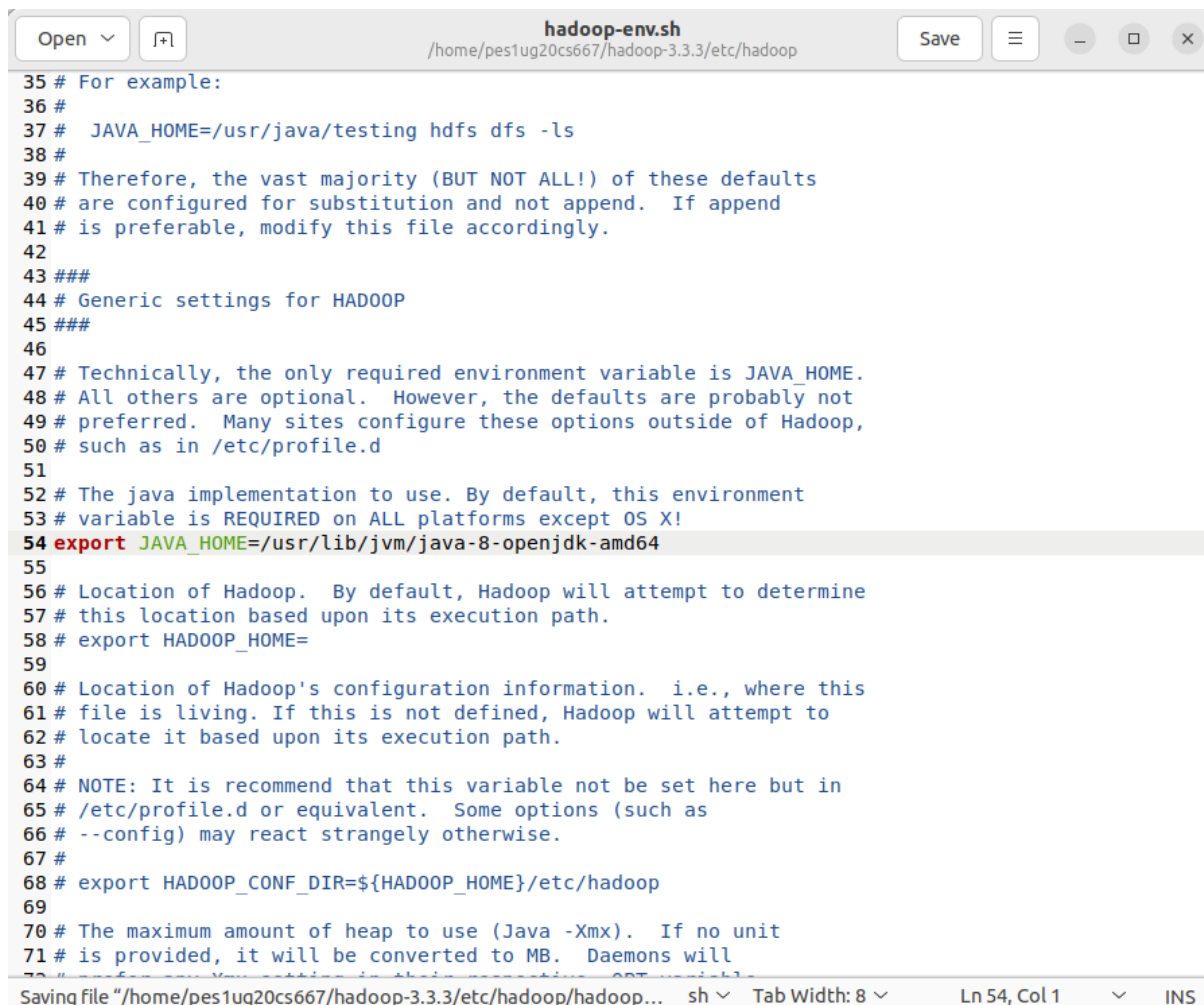


```
GNU nano 6.2 /home/pes1ug20cs667/.bashrc *  
  
# Add an "alert" alias for long running commands. Use like so:  
#   sleep 10; alert  
alias alert='notify-send --urgency=low -i "${?} = 0" && echo termina>  
  
# Alias definitions.  
# You may want to put all your additions into a separate file like  
# ~/.bash_aliases, instead of adding them here directly.  
# See /usr/share/doc/bash-doc/examples in the bash-doc package.  
  
if [ -f ~/.bash_aliases ]; then  
    . ~/.bash_aliases  
fi  
  
# enable programmable completion features (you don't need to enable  
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile  
# sources /etc/bash.bashrc).  
if ! shopt -oq posix; then  
    if [ -f /usr/share/bash-completion/bash_completion ]; then  
        . /usr/share/bash-completion/bash_completion  
    elif [ -f /etc/bash_completion ]; then  
        . /etc/bash_completion  
    fi  
fi  
  
export HADOOP_HOME=/home/pes1ug20cs667/hadoop-3.3.3  
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/native"  
  
^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute  
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify
```

2b.png

NAME : NIKHIL RAJU MOHITE
SRN : PES1UG20CS667

SEC : K



```
35 # For example:
36 #
37 # JAVA_HOME=/usr/java/testing hdfs dfs -ls
38 #
39 # Therefore, the vast majority (BUT NOT ALL!) of these defaults
40 # are configured for substitution and not append. If append
41 # is preferable, modify this file accordingly.
42
43 ###
44 # Generic settings for HADOOP
45 ###
46
47 # Technically, the only required environment variable is JAVA_HOME.
48 # All others are optional. However, the defaults are probably not
49 # preferred. Many sites configure these options outside of Hadoop,
50 # such as in /etc/profile.d
51
52 # The java implementation to use. By default, this environment
53 # variable is REQUIRED on ALL platforms except OS X!
54 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
55
56 # Location of Hadoop. By default, Hadoop will attempt to determine
57 # this location based upon its execution path.
58 # export HADOOP_HOME=
59
60 # Location of Hadoop's configuration information. i.e., where this
61 # file is living. If this is not defined, Hadoop will attempt to
62 # locate it based upon its execution path.
63 #
64 # NOTE: It is recommend that this variable not be set here but in
65 # /etc/profile.d or equivalent. Some options (such as
66 # --config) may react strangely otherwise.
67 #
68 # export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop
69
70 # The maximum amount of heap to use (Java -Xmx). If no unit
71 # is provided, it will be converted to MB. Daemons will
72 # use the same value as the master.
```

2c.png



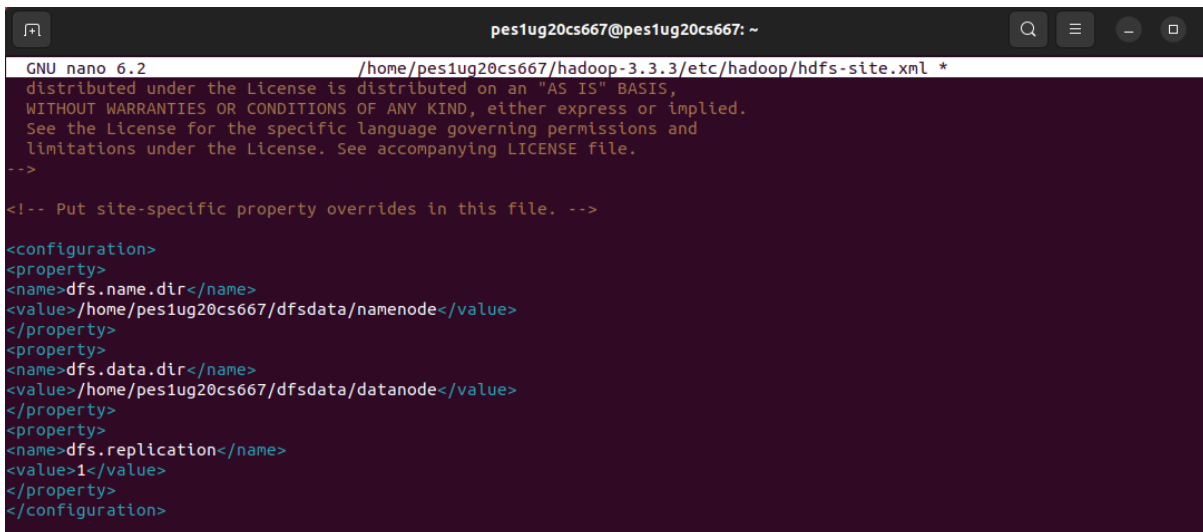
```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>hadoop.tmp.dir</name>
<value>/home/pes1ug20cs667/tmpdata</value>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>
```

NAME : NIKHIL RAJU MOHITE
SRN : PES1UG20CS667

SEC : K

2d.png

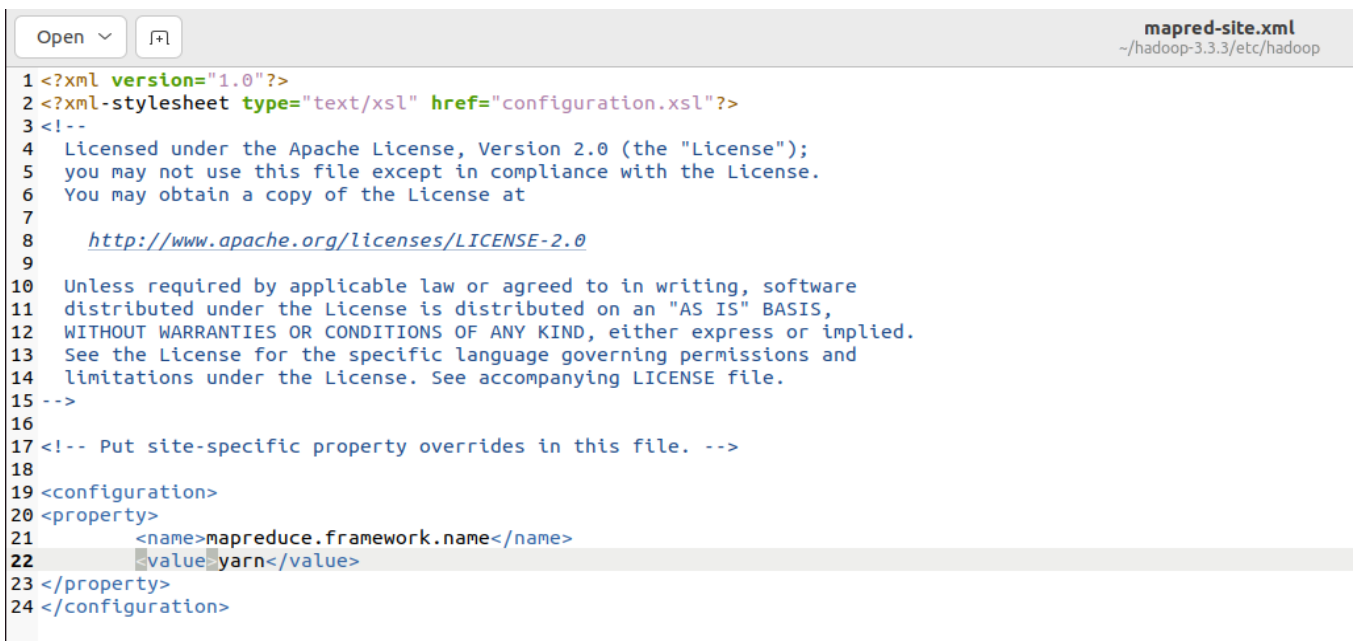


```
GNU nano 6.2 /home/pes1ug20cs667/hadoop-3.3.3/etc/hadoop/hdfs-site.xml *
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.name.dir</name>
<value>/home/pes1ug20cs667/dfsdata/namenode</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/home/pes1ug20cs667/dfsdata/datanode</value>
</property>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>
```

2e.png



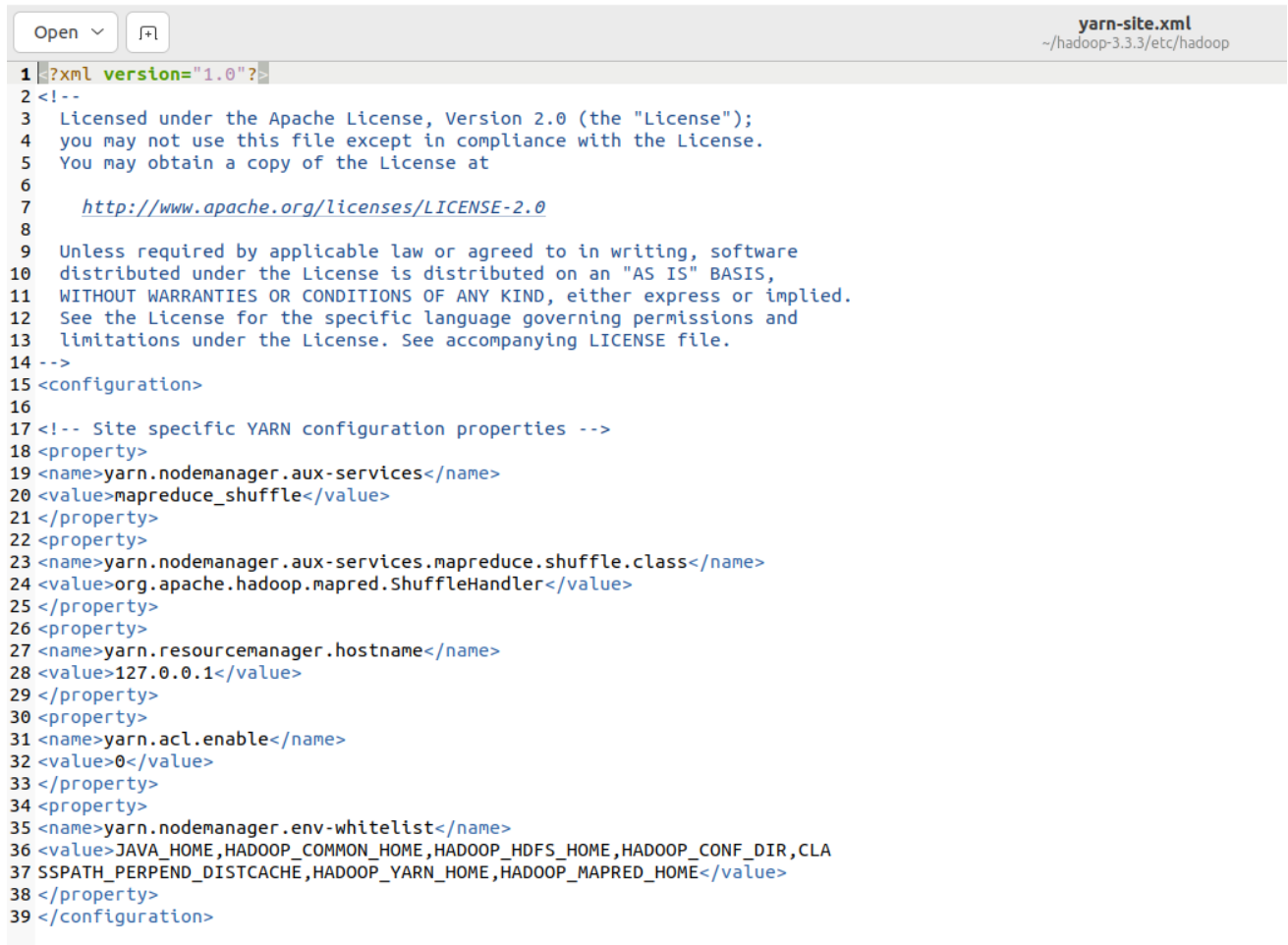
```
Open  [icon] mapred-site.xml
~/hadoop-3.3.3/etc/hadoop

1 <?xml version="1.0"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20 <property>
21 <name>mapreduce.framework.name</name>
22 <value>yarn</value>
23 </property>
24 </configuration>
```

2f.png

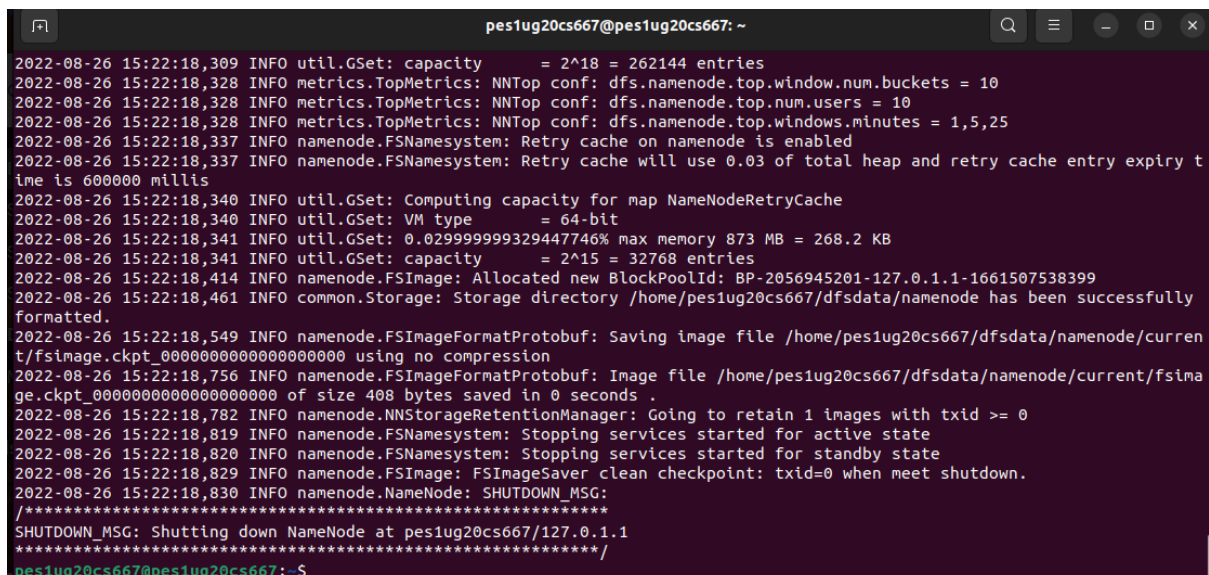
NAME : NIKHIL RAJU MOHITE
SRN : PES1UG20CS667

SEC : K



```
1 <?xml version="1.0"?>
2 <!--
3 Licensed under the Apache License, Version 2.0 (the "License");
4 you may not use this file except in compliance with the License.
5 You may obtain a copy of the License at
6
7 http://www.apache.org/licenses/LICENSE-2.0
8
9 Unless required by applicable law or agreed to in writing, software
10 distributed under the License is distributed on an "AS IS" BASIS,
11 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
12 See the License for the specific language governing permissions and
13 limitations under the License. See accompanying LICENSE file.
14 -->
15 <configuration>
16
17 <!-- Site specific YARN configuration properties -->
18 <property>
19 <name>yarn.nodemanager.aux-services</name>
20 <value>mapreduce_shuffle</value>
21 </property>
22 <property>
23 <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
24 <value>org.apache.hadoop.mapred.ShuffleHandler</value>
25 </property>
26 <property>
27 <name>yarn.resourcemanager.hostname</name>
28 <value>127.0.0.1</value>
29 </property>
30 <property>
31 <name>yarn.acl.enable</name>
32 <value>0</value>
33 </property>
34 <property>
35 <name>yarn.nodemanager.env-whitelist</name>
36 <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLA
37 SSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
38 </property>
39 </configuration>
```

3a.png



```
pes1ug20cs667@pes1ug20cs667: ~
2022-08-26 15:22:18,309 INFO util.GSet: capacity = 2^18 = 262144 entries
2022-08-26 15:22:18,328 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-08-26 15:22:18,328 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-08-26 15:22:18,328 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-08-26 15:22:18,337 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-08-26 15:22:18,337 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry t
ime is 600000 millis
2022-08-26 15:22:18,340 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-08-26 15:22:18,340 INFO util.GSet: VM type = 64-bit
2022-08-26 15:22:18,341 INFO util.GSet: 0.02999999999329447746% max memory 873 MB = 268.2 KB
2022-08-26 15:22:18,341 INFO util.GSet: capacity = 2^15 = 32768 entries
2022-08-26 15:22:18,414 INFO namenode.FSImage: Allocated new BlockPoolId: BP-2056945201-127.0.1.1-1661507538399
2022-08-26 15:22:18,461 INFO common.Storage: Storage directory /home/pes1ug20cs667/dfsdata/namenode has been successfully
formatted.
2022-08-26 15:22:18,549 INFO FSImageFormatProtobuf: Saving image file /home/pes1ug20cs667/dfsdata/namenode/current
t/fsimage.ckpt_00000000000000000000 using no compression
2022-08-26 15:22:18,756 INFO namenode.FSImageFormatProtobuf: Image file /home/pes1ug20cs667/dfsdata/namenode/current/fsima
ge.ckpt_00000000000000000000 of size 408 bytes saved in 0 seconds .
2022-08-26 15:22:18,782 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-08-26 15:22:18,819 INFO namenode.FSNamesystem: Stopping services started for active state
2022-08-26 15:22:18,820 INFO namenode.FSNamesystem: Stopping services started for standby state
2022-08-26 15:22:18,829 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2022-08-26 15:22:18,830 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at pes1ug20cs667/127.0.1.1
*****/
pes1ug20cs667@pes1ug20cs667: ~$
```

4a.png

NAME : NIKHIL RAJU MOHITE
SRN : PES1UG20CS667

SEC : K

```
pes1ug20cs667@pes1ug20cs667: ~/hadoop-3.3.3/sbin
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as pes1ug20cs667 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [pes1ug20cs667]
Starting resourcemanager
Starting nodemanagers
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$ jps
5362 Jps
4213 NameNode
4762 ResourceManager
4476 SecondaryNameNode
4332 DataNode
4911 NodeManager
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin$
```

6a.png

```
pes1ug20cs667@pes1ug20cs667:~$ hdfs dfs -cat /example/output/part-r-00000
bye      1
hello    1
is       1
nikhil   1
this     1
pes1ug20cs667@pes1ug20cs667:~$
```

7a.png

```
pes1ug20cs667@pes1ug20cs667: ~/hadoop-3.3.3/sbin/UE20CS322-H1
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin/UE20CS322-H1$ hdfs dfs -cat /handson/output-tech/part-00000
54
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin/UE20CS322-H1$
```

NAME : NIKHIL RAJU MOHITE
SRN : PES1UG20CS667

SEC : K

```
pes1ug20cs667@pes1ug20cs667: ~/hadoop-3.3.3/sbin/UE20CS322-H1
pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin/UE20CS322-H1$ python3 eval-rr.pyc
Starting Hadoop Installation verification...
Running clean up...
Setting up HDFS directories...
[sudo] password for pes1ug20cs667:
Sorry, try again.
[sudo] password for pes1ug20cs667:
Creating mapper and reducer scripts...
Running Hadoop MR job...
Verifying output...
Sending result to server...
Server response: Success. Your marks are 5 / 5.0
Running clean up...

Auto-evaluation concluded.

pes1ug20cs667@pes1ug20cs667:~/hadoop-3.3.3/sbin/UE20CS322-H1$
```