

**SRN: PES2UG20CS815**



**EC -Campus**

**Subject: Big Data**

**Subject Code: UE20CS322**

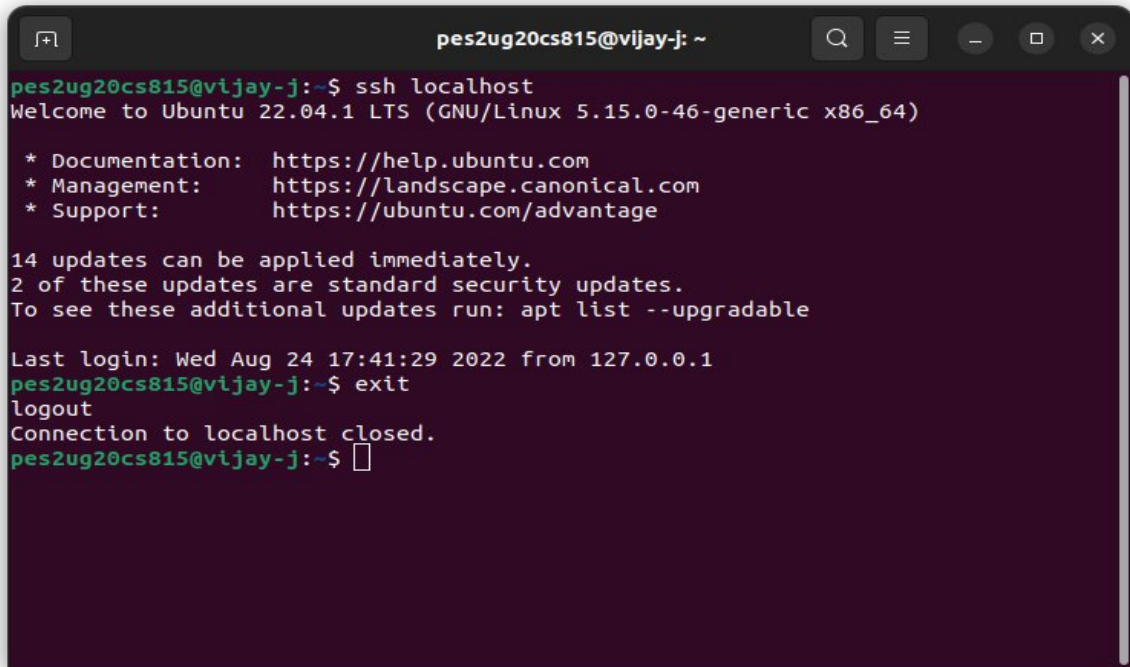
**Name: Vijay.J**

**SRN: PES2UG20CS815**

**Section: J**

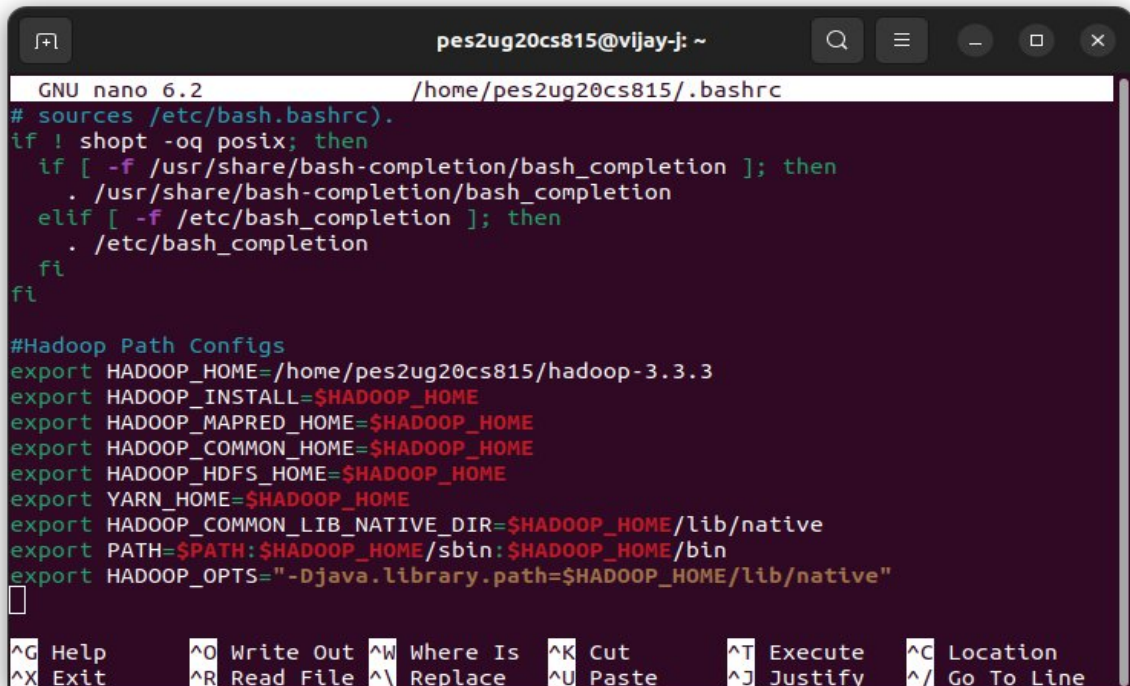
SRN: PES2UG20CS815

1a)



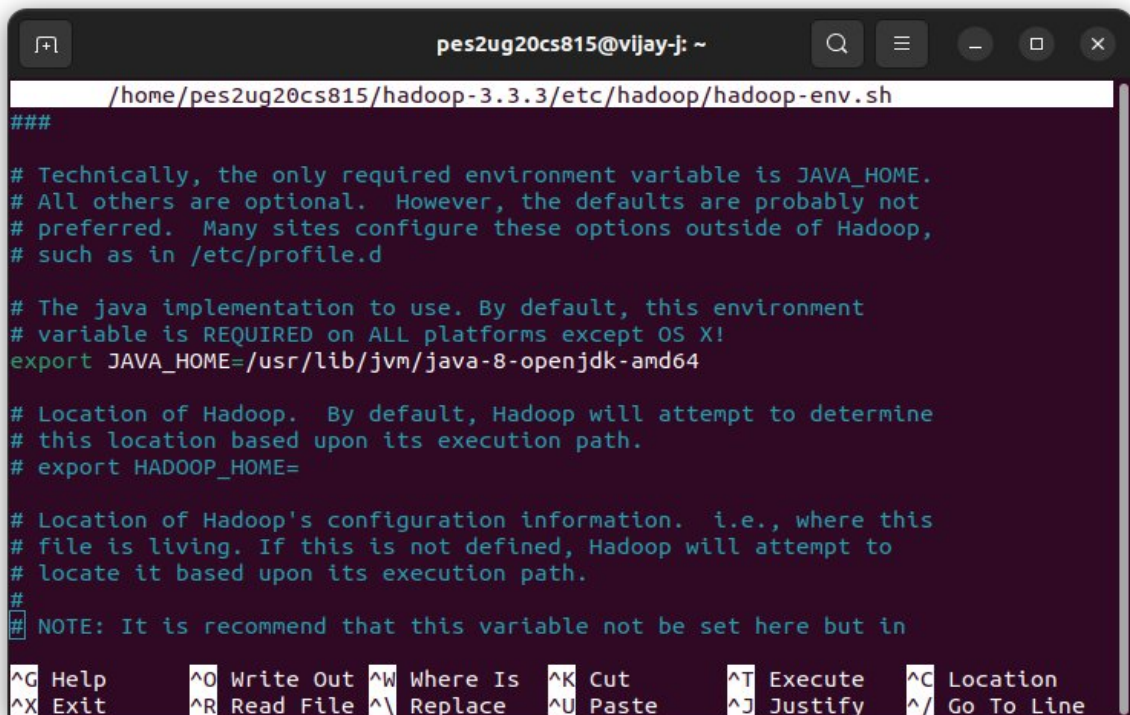
```
pes2ug20cs815@vijay-j: ~  
pes2ug20cs815@vijay-j:~$ ssh localhost  
Welcome to Ubuntu 22.04.1 LTS (GNU/Linux 5.15.0-46-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/advantage  
  
14 updates can be applied immediately.  
2 of these updates are standard security updates.  
To see these additional updates run: apt list --upgradable  
  
Last login: Wed Aug 24 17:41:29 2022 from 127.0.0.1  
pes2ug20cs815@vijay-j:~$ exit  
logout  
Connection to localhost closed.  
pes2ug20cs815@vijay-j:~$
```

2a)



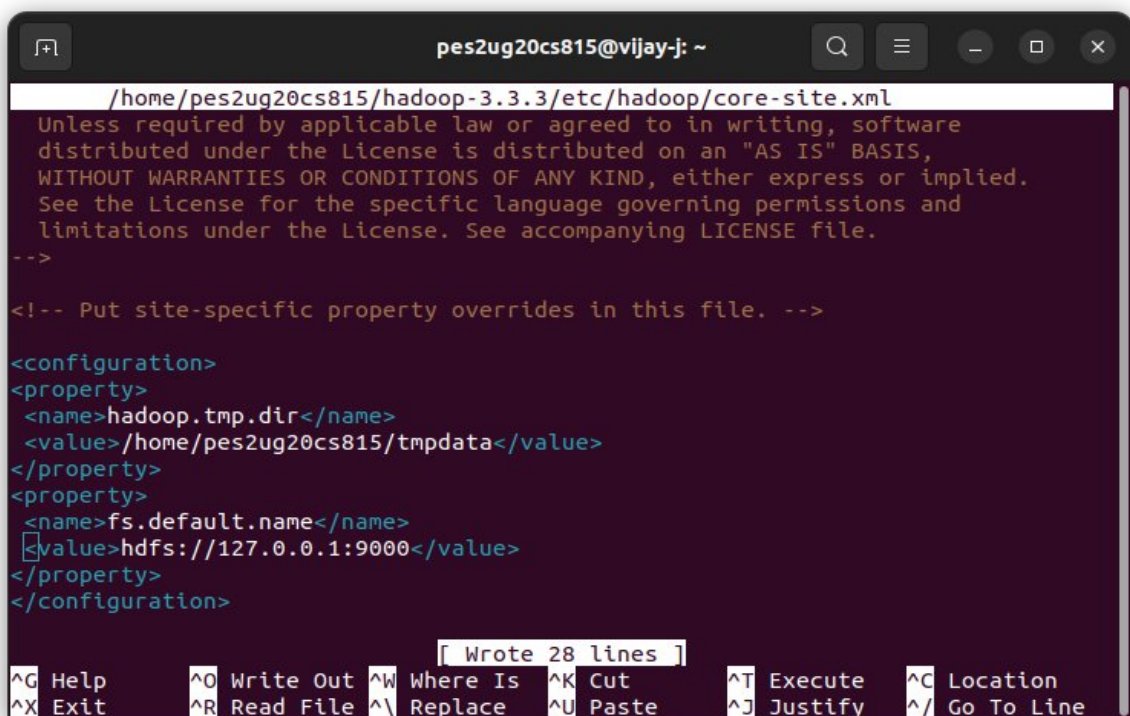
```
GNU nano 6.2 /home/pes2ug20cs815/.bashrc  
# sources /etc/bash.bashrc).  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
  
#Hadoop Path Configs  
export HADOOP_HOME=/home/pes2ug20cs815/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   ^C Location  
^X Exit      ^R Read File ^\ Replace    ^U Paste     ^J Justify   ^_ Go To Line
```

2b)



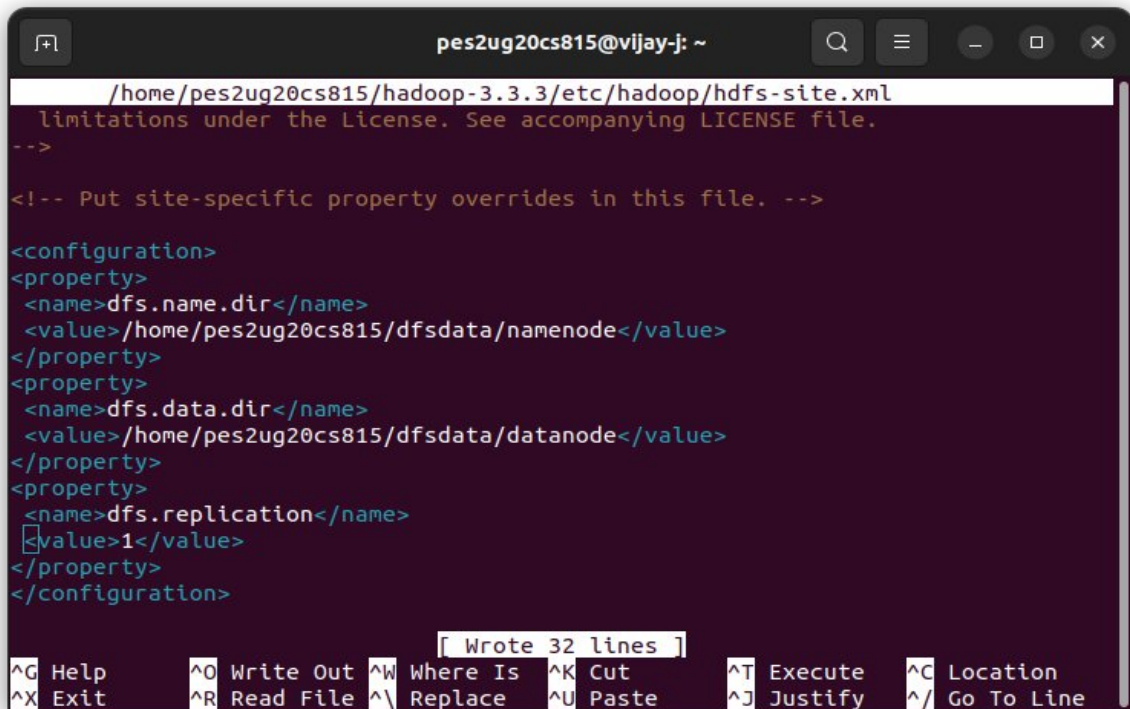
```
pes2ug20cs815@vijay-j: ~  
/home/pes2ug20cs815/hadoop-3.3.3/etc/hadoop/hadoop-env.sh  
###  
# Technically, the only required environment variable is JAVA_HOME.  
# All others are optional.  However, the defaults are probably not  
# preferred.  Many sites configure these options outside of Hadoop,  
# such as in /etc/profile.d  
  
# The java implementation to use.  By default, this environment  
# variable is REQUIRED on ALL platforms except OS X!  
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64  
  
# Location of Hadoop.  By default, Hadoop will attempt to determine  
# this location based upon its execution path.  
# export HADOOP_HOME=  
  
# Location of Hadoop's configuration information.  i.e., where this  
# file is living.  If this is not defined, Hadoop will attempt to  
# locate it based upon its execution path.  
#  
## NOTE: It is recommend that this variable not be set here but in  
  
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location  
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line
```

2c)



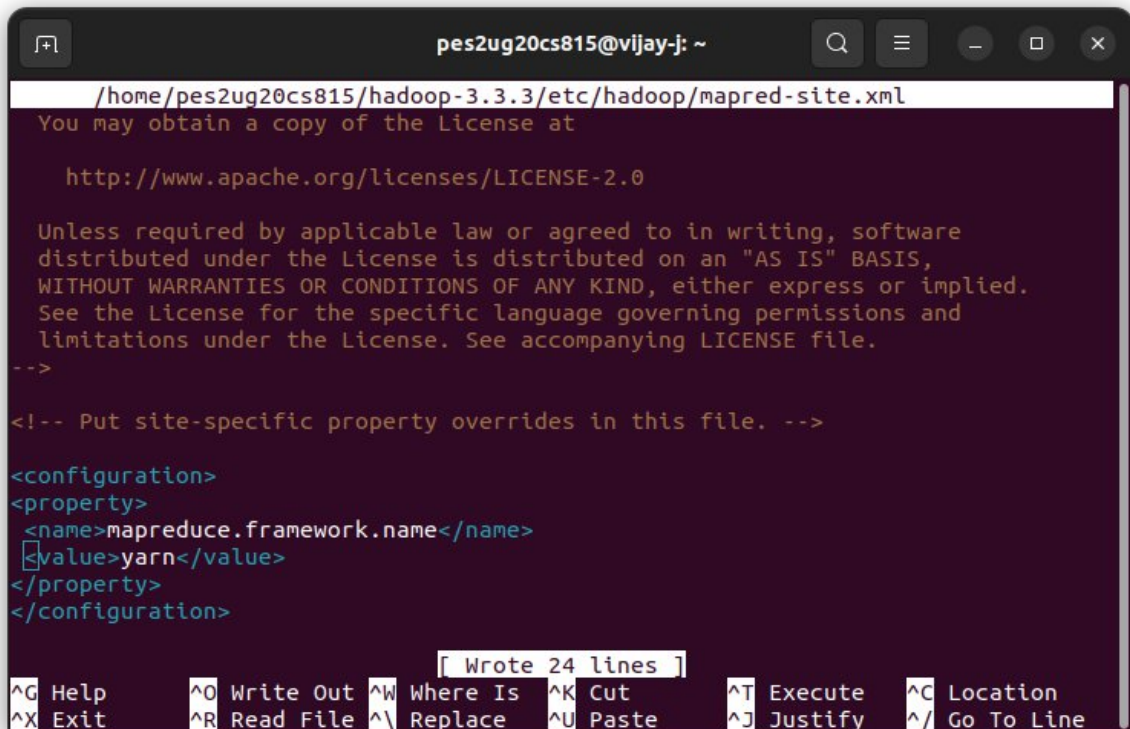
```
pes2ug20cs815@vijay-j: ~  
/home/pes2ug20cs815/hadoop-3.3.3/etc/hadoop/core-site.xml  
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
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>hadoop.tmp.dir</name>  
  <value>/home/pes2ug20cs815/tmpdata</value>  
</property>  
<property>  
  <name>fs.default.name</name>  
  <value>hdfs://127.0.0.1:9000</value>  
</property>  
</configuration>  
  
[ Wrote 28 lines ]  
  
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location  
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line
```

2d)



```
pes2ug20cs815@vijay-j: ~  
/home/pes2ug20cs815/hadoop-3.3.3/etc/hadoop/hdfs-site.xml  
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/pes2ug20cs815/dfsdata/namenode</value>  
</property>  
<property>  
  <name>dfs.data.dir</name>  
  <value>/home/pes2ug20cs815/dfsdata/datanode</value>  
</property>  
<property>  
  <name>dfs.replication</name>  
  <value>1</value>  
</property>  
</configuration>  
  
[ Wrote 32 lines ]  
^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location  
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/ Go To Line
```

2e)



```
pes2ug20cs815@vijay-j: ~  
/home/pes2ug20cs815/hadoop-3.3.3/etc/hadoop/mapred-site.xml  
You may obtain a copy of the License at  
  
http://www.apache.org/licenses/LICENSE-2.0  
  
Unless required by applicable law or agreed to in writing, software  
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>mapreduce.framework.name</name>  
  <value>yarn</value>  
</property>  
</configuration>  
  
[ Wrote 24 lines ]  
^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location  
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/ Go To Line
```



2f)

```

pes2ug20cs815@vijay-j: ~
GNU nano 6.2 /home/pes2ug20cs815/hadoop-3.3.3/etc/hadoop/yarn-site.xml
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
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.
-->
<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.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>

```

3a)

```

pes2ug20cs815@vijay-j: ~
2022-08-24 20:28:05,270 INFO util.GSet: capacity = 2^14 = 16384 entries
2022-08-24 20:28:05,446 INFO namenode.FSImage: Allocated new BlockPoolId: BP-114
7643129-127.0.1.1-1661353085401
2022-08-24 20:28:06,139 INFO common.Storage: Storage directory /home/pes2ug20cs8
15/dfsdata/namenode has been successfully formatted.
2022-08-24 20:28:06,367 INFO namenode.FSImageFormatProtobuf: Saving image file /
home/pes2ug20cs815/dfsdata/namenode/current/fsimage.ckpt_00000000000000000000 usi
ng no compression
2022-08-24 20:28:06,804 INFO namenode.FSImageFormatProtobuf: Image file /home/pe
s2ug20cs815/dfsdata/namenode/current/fsimage.ckpt_00000000000000000000 of size 40
8 bytes saved in 0 seconds .
2022-08-24 20:28:07,039 INFO namenode.NNStorageRetentionManager: Going to retain
1 images with txid >= 0
2022-08-24 20:28:07,134 INFO namenode.FSNamesystem: Stopping services started fo
r active state
2022-08-24 20:28:07,134 INFO namenode.FSNamesystem: Stopping services started fo
r standby state
2022-08-24 20:28:07,160 INFO namenode.FSImage: FSImageSaver clean checkpoint: tx
id=0 when meet shutdown.
2022-08-24 20:28:07,161 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at vijay-j/127.0.1.1
*****/
pes2ug20cs815@vijay-j: ~$

```

SRN: PES2UG2OCS815

4a)

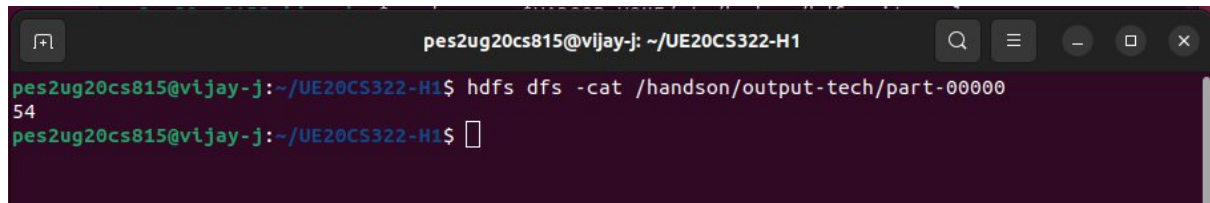
```
pes2ug20cs815@vijay-j:~$ cd hadoop-3.3.3/sbin/
./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as pes2ug20cs815 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [vijay-j]
Starting resourcemanager
Starting nodemanagers
pes2ug20cs815@vijay-j:~/hadoop-3.3.3/sbin$ jps
6120 NameNode
5755 SecondaryNameNode
5483 NameNode
5996 ResourceManager
5597 DataNode
6269 Jps
pes2ug20cs815@vijay-j:~/hadoop-3.3.3/sbin$
```

6a)

```
pes2ug20cs815@vijay-j:~$ hdfs dfs -cat /example/output/part-r-00000
Bytes Written=272
.
Criminals 1
I 2
It's 1
So 1
That's 1
a 5
a... 1
able 1
an 1
are 1
bat! 2
be 2
become 1
black, 1
cowardly 1
creature 1
disguise 1
hearts. 1
into 1
it! 1
lot. 1
must 2
my 1
night, 1
of 1
omen. 1
shall 1
strike 1
superstitious 1
terrible... 1
terror 1
the 1
their 1
to 1
pes2ug20cs815@vijay-j:~$
```

**SRN: PES2UG2OCS815**

**7a)**



```
pes2ug20cs815@vijay-j: ~/UE20CS322-H1
pes2ug20cs815@vijay-j:~/UE20CS322-H1$ hdfs dfs -cat /handson/output-tech/part-00000
54
pes2ug20cs815@vijay-j:~/UE20CS322-H1$
```

The image shows a terminal window with a dark background. The title bar at the top reads 'pes2ug20cs815@vijay-j: ~/UE20CS322-H1'. The terminal content shows a user prompt 'pes2ug20cs815@vijay-j:~/UE20CS322-H1\$' followed by the command 'hdfs dfs -cat /handson/output-tech/part-00000'. The output of the command is '54'. Below the output, the prompt 'pes2ug20cs815@vijay-j:~/UE20CS322-H1\$' is shown again with a cursor, indicating the command has completed.