Assignment No 6

Q) Create single node hadoop cluster having the machine name as your name and upload the project data in HDFS. Upload all the screenshots that would describe the steps taken in completing this assignment.

Complete Process of Hadoop single node cluster Setup

1) First install vmware in local system.

2) After installation of vmware step up any operating system in vmware workstation(for ex: linux , microsoft).

3) Java Setup in Linux virtual operating system:

Steps for Java Setup:

Step 1: start a virtual operating system in VMware.

Step 2: start ip address of virtual operating system.

commands for start ip address:

nmcli conn show

nmcli conn up device_name

for ip check:

ip addr show

Step 3: Transfer Java tar file to virtual operating system using any file transfer tool.

Step 4: Extract the tar file using below command.

tar -xvf file_name.

Step 5: Move and rename the extracted file /use/local location using the below command.

mv file_name /use/local/new_file_name

Step 6: create an environment variable for Java in **/etc/profile** location using command **vi /etc/profile.**

Step 7: Refresh the desktop with the help of command **source /etc/profile.**

Step 8: Type **Jps** and **Java verison** commands for checking setup of Java.

Step 9: Go to **/etc/hostname** location and change machine host name.

Step 10: Then go to **/etc/hosts** location and change ip address and machine name.

4) Process of Hadoop Setup in Linux:

This process is start after complete setup of java in linux.

Steps for setup of Hadoop:

Step 1: Transfer hadoop tar file to virtual operating system using any file transfer tool.

Step 2: Extract tar file using **tar -xvf file_name** command

Step 3: Move and rename the extracted file to **/usr/local** location using command mv **file_name /usr/local/new_file_name.**

Step 4: Create an environment variable for hadoop setup in Linux in **/etc/profile** directory.

Step 5: Set the path of Java file in the hadoop environment file. Location of the hadoop environment file: **/usr/local/hadoop/etc/hadoop**

File name: **hadoop-env.sh**

Step 6: rewrite XML files in hadoop.

Files location: **usr/local/hadoop/etc/hadoop**

Files: **core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml.**

Step 7: Refresh the desktop using the source command.

source /etc/profile

Step 8: Before setup of hadoop namenode change machine name in **/usr/local/hadoop/etc/hadoop/slaves** file location.

Step 9: Format the name node of hadoop using **hdfs namenode -format** command.

Step 10: Start all Storage and Processing tools of hadoop using the below commands.

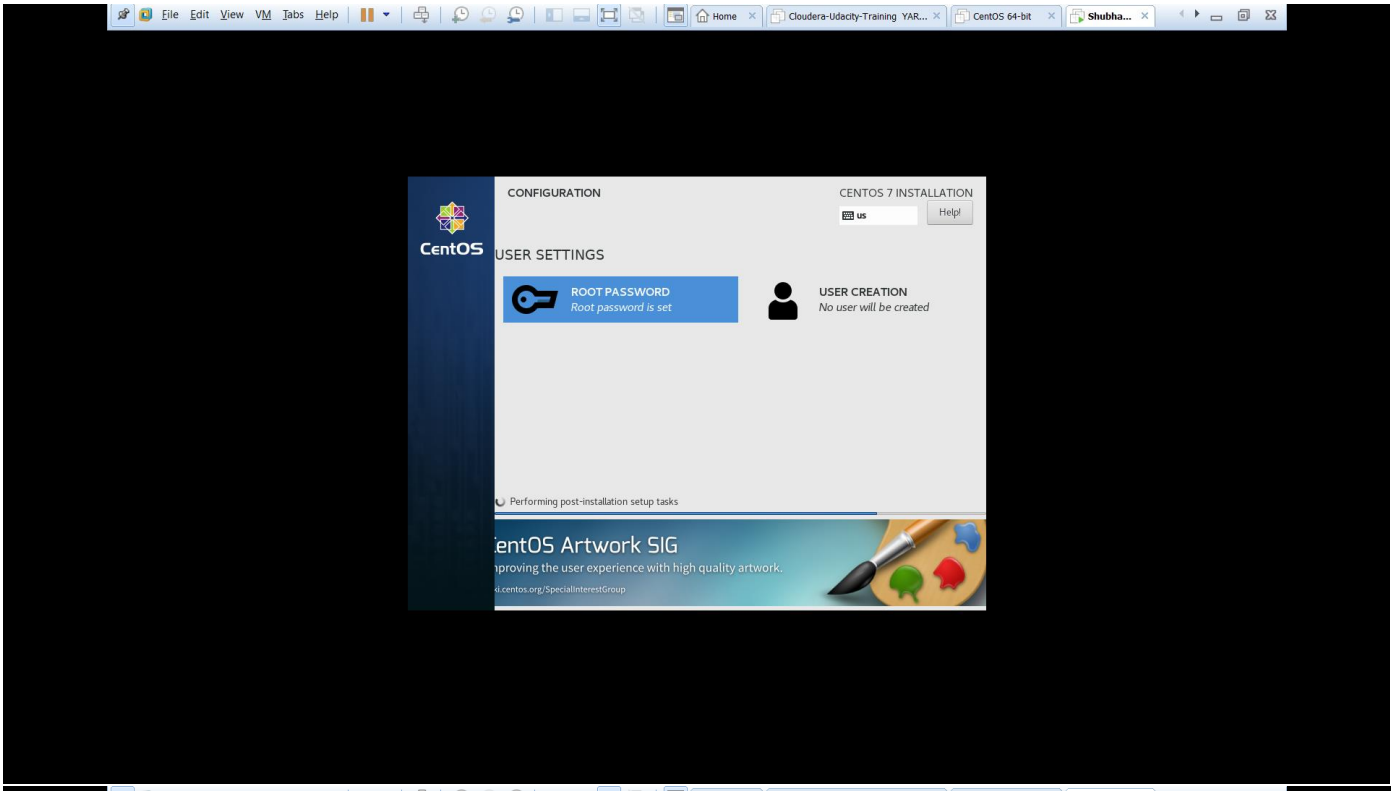
commands:

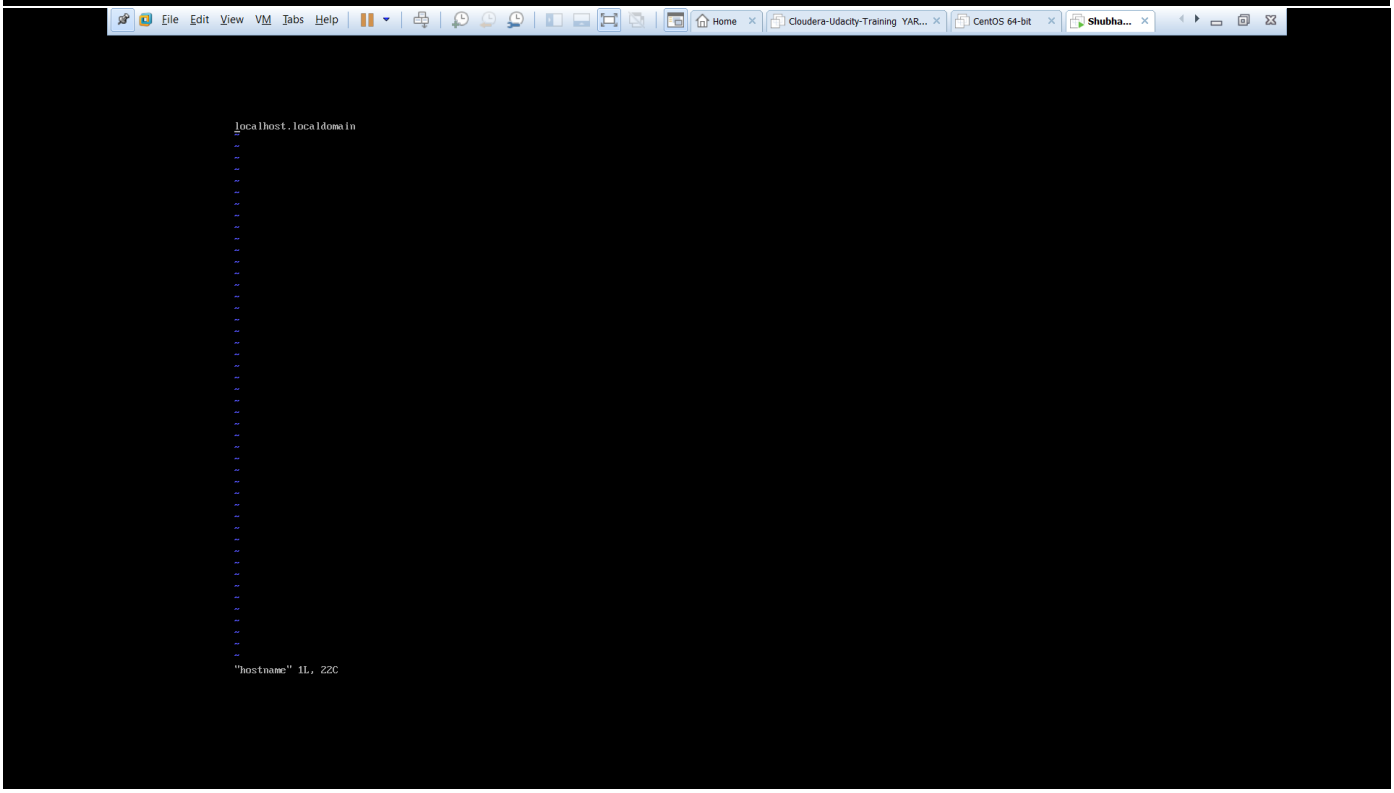
sh start-dfs.sh

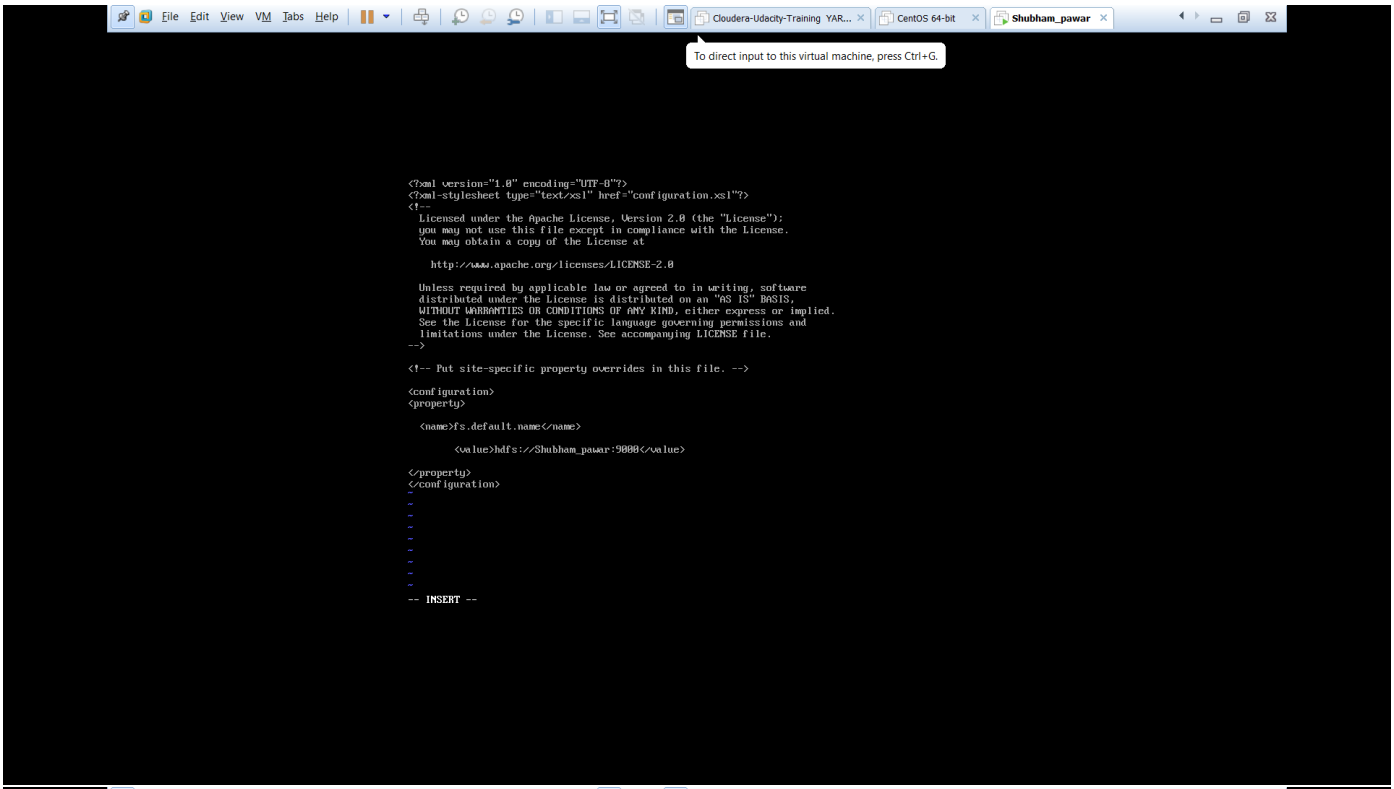
sh start-yarn.sh

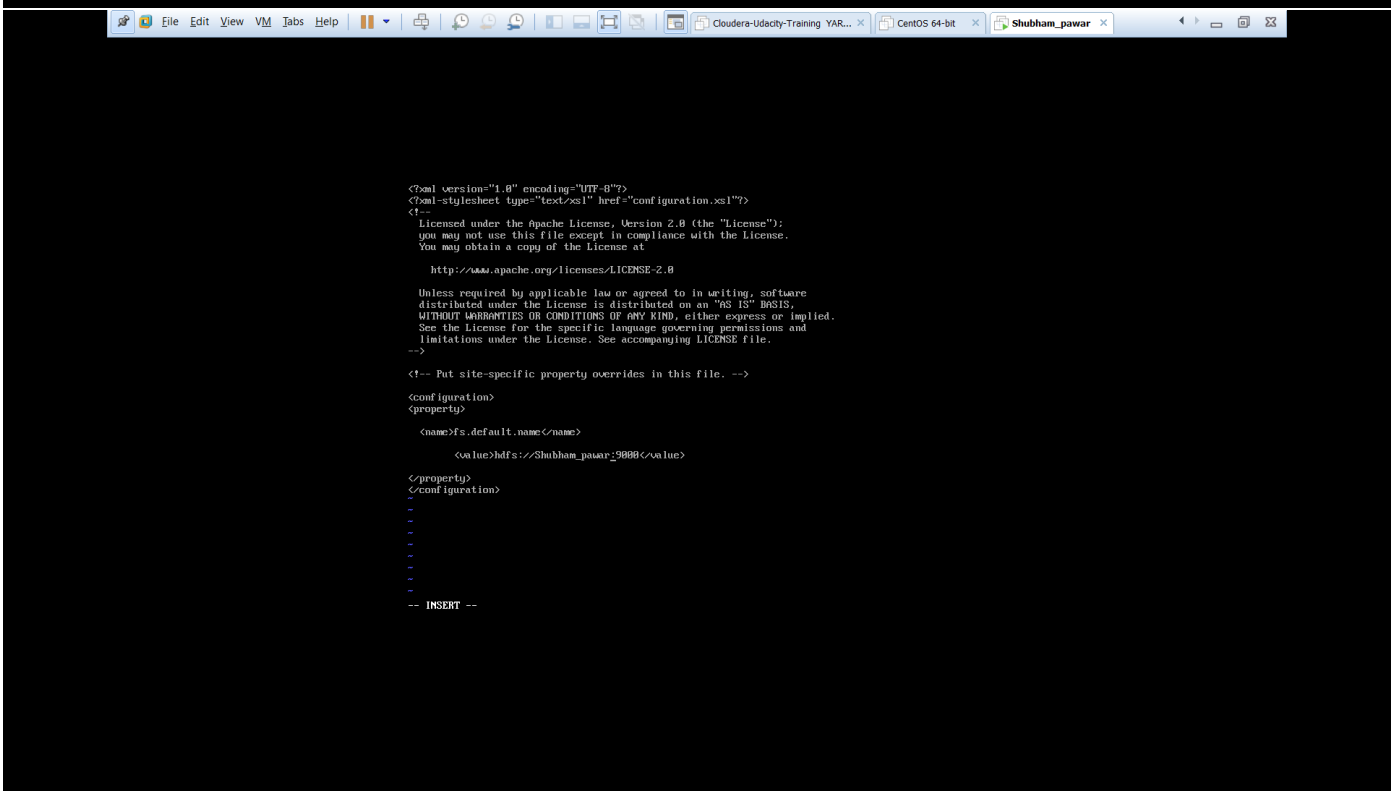
Step 11: Check all tools of Hadoop are working or not using **Jps command**.The tools are working mains hadoop setup in linux is completed. In that point hadoop single node cluster setup is completed.

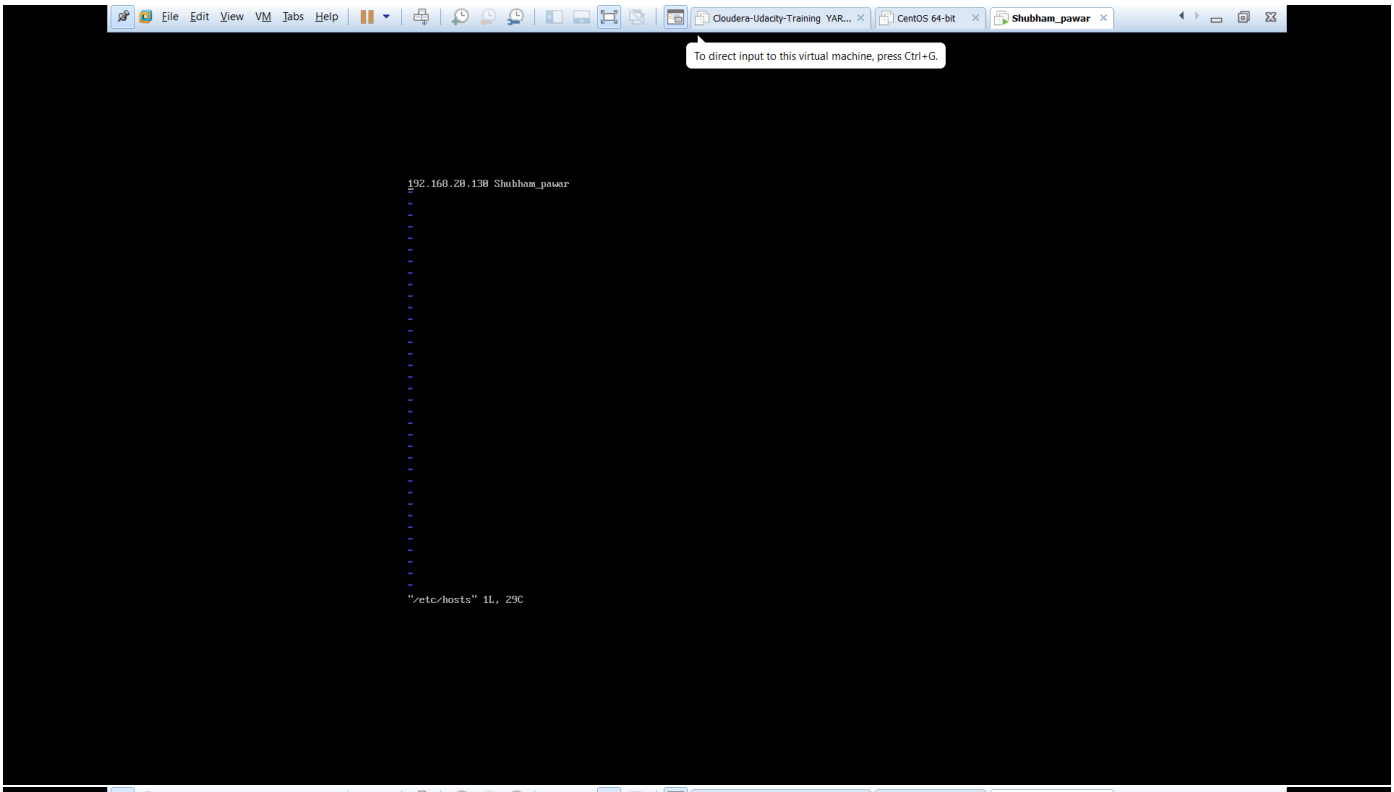
Screenshots

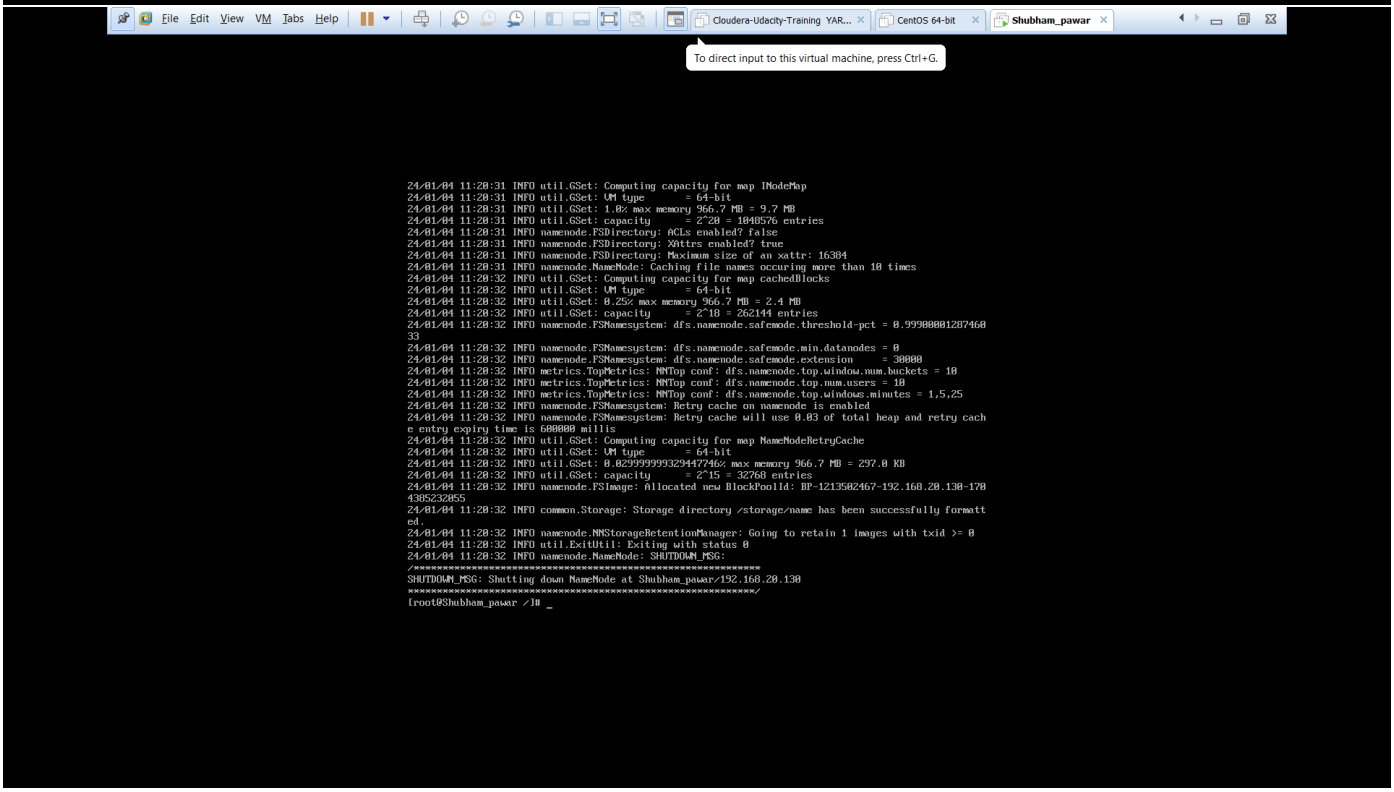


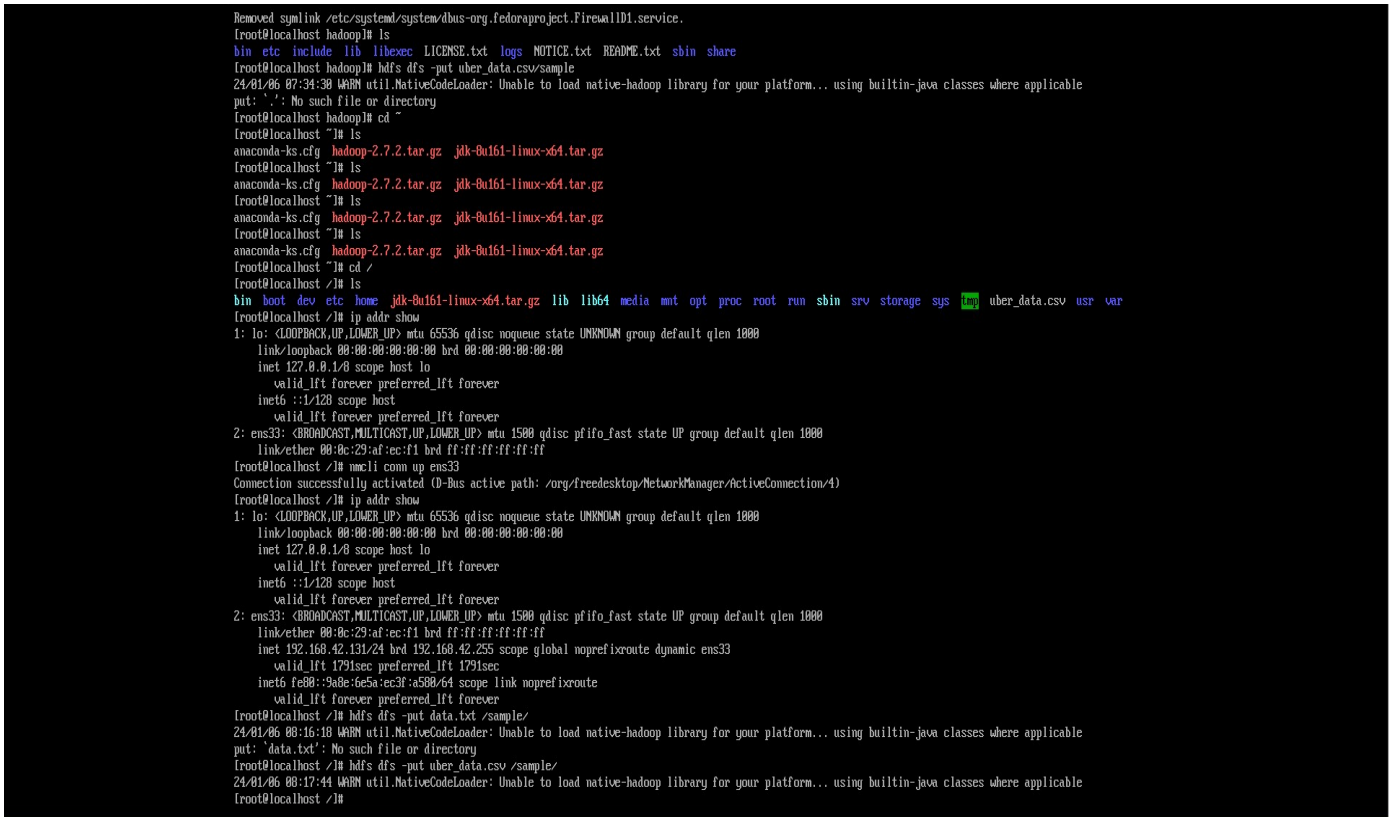



File Edit View VM Tabs Help
Cloudera-Udacity-Training YAR...
CentOS 64-bit
Shubham_pawar
To direct input to this virtual machine, press Ctrl+G.
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
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>fs.default.name</name>
<value>hdfs://Shubham_pawar:9000</value>
</property>
</configuration>
~
~
~
~
~
~
~
~
~
-- INSERT --


File Edit View VM Tabs Help
Cloudera-Udacity-Training YAR...
CentOS 64-bit
Shubham_pawar
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
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>fs.default.name</name>
<value>hdfs://Shubham_pawar:9000</value>
</property>
</configuration>
~
~
~
~
~
~
~
~
~
-- INSERT --


File Edit View VM Tabs Help
Cloudera-Udacity-Training YAR...
CentOS 64-bit
Shubham_pawar
To direct input to this virtual machine, press Ctrl+G.
192.168.20.130 Shubham_pawar
"/etc/hosts" 1L, 29C


File Edit View VM Tabs Help
Cloudera-Udacity-Training YAR...
CentOS 64-bit
Shubham_pawar
To direct input to this virtual machine, press Ctrl+G.
24/01/04 11:20:31 INFO util.GSet: Computing capacity for map INodeMap
24/01/04 11:20:31 INFO util.GSet: VM type = 64-bit
24/01/04 11:20:31 INFO util.GSet: 1.0% max memory 966.7 MB = 9.7 MB
24/01/04 11:20:31 INFO util.GSet: capacity = 2^20 = 1048576 entries
24/01/04 11:20:31 INFO namenode.FSDirectory: ACLs enabled? false
24/01/04 11:20:31 INFO namenode.FSDirectory: XAttrs enabled? true
24/01/04 11:20:31 INFO namenode.FSDirectory: Maximum size of an xattr: 16384
24/01/04 11:20:31 INFO namenode.NameNode: Caching file names occuring more than 10 times
24/01/04 11:20:32 INFO util.GSet: Computing capacity for map cachedBlocks
24/01/04 11:20:32 INFO util.GSet: VM type = 64-bit
24/01/04 11:20:32 INFO util.GSet: 0.25% max memory 966.7 MB = 2.4 MB
24/01/04 11:20:32 INFO util.GSet: capacity = 2^18 = 262144 entries
24/01/04 11:20:32 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
24/01/04 11:20:32 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
24/01/04 11:20:32 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
24/01/04 11:20:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
24/01/04 11:20:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
24/01/04 11:20:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
24/01/04 11:20:32 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
24/01/04 11:20:32 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
24/01/04 11:20:32 INFO util.GSet: Computing capacity for map NameNodeRetryCache
24/01/04 11:20:32 INFO util.GSet: VM type = 64-bit
24/01/04 11:20:32 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 297.0 KB
24/01/04 11:20:32 INFO util.GSet: capacity = 2^15 = 32768 entries
24/01/04 11:20:32 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1213582467-192.168.20.130-1704385232855
24/01/04 11:20:32 INFO common.Storage: Storage directory /storage/name has been successfully formatted.
24/01/04 11:20:32 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
24/01/04 11:20:32 INFO util.ExitUtil: Exiting with status 0
24/01/04 11:20:32 INFO namenode.NameNode: SHUTDOWN_MSG:
/**
SHUTDOWN_MSG: Shutting down NameNode at Shubham_pawar/192.168.20.130
**/
[root@Shubham_pawar /]#


Removed symlink /etc/systemd/system/dbus-org.fedoraproject.FirewallD1.service.
[root@localhost hadoop]# ls
bin etc include lib libexec LICENSE.txt logs NOTICE.txt README.txt sbin share
[root@localhost hadoop]# hdfs dfs -put uber_data.csv/sample
24/01/06 07:34:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: `.': No such file or directory
[root@localhost hadoop]# cd ~
[root@localhost ~]# ls
anaconda-ks.cfg hadoop-2.7.2.tar.gz jdk-8u161-linux-x64.tar.gz
[root@localhost ~]# ls
anaconda-ks.cfg hadoop-2.7.2.tar.gz jdk-8u161-linux-x64.tar.gz
[root@localhost ~]# ls
anaconda-ks.cfg hadoop-2.7.2.tar.gz jdk-8u161-linux-x64.tar.gz
[root@localhost ~]# ls
anaconda-ks.cfg hadoop-2.7.2.tar.gz jdk-8u161-linux-x64.tar.gz
[root@localhost ~]# cd /
[root@localhost /]# ls
bin boot dev etc home jdk-8u161-linux-x64.tar.gz lib lib64 media mnt opt proc root run sbin srv storage sys tmp uber_data.csv usr var
[root@localhost /]# ip addr show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: ens33: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:af:ec:f1 brd ff:ff:ff:ff:ff:ff
[root@localhost /]# nmcli conn up ens33
Connection successfully activated (D-Bus active path: /org/freedesktop/NetworkManager/ActiveConnection/4)
[root@localhost /]# ip addr show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: ens33: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:af:ec:f1 brd ff:ff:ff:ff:ff:ff
    inet 192.168.42.131/24 brd 192.168.42.255 scope global noprefixroute dynamic ens33
       valid_lft 1791sec preferred_lft 1791sec
    inet6 fe80::9a8e:6e5a:ec3f:a580/64 scope link noprefixroute
       valid_lft forever preferred_lft forever
[root@localhost /]# hdfs dfs -put data.txt /sample/
24/01/06 08:16:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: `data.txt': No such file or directory
[root@localhost /]# hdfs dfs -put uber_data.csv /sample/
24/01/06 08:17:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@localhost /]#