Assignment No 6 – Mayur Borse

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.

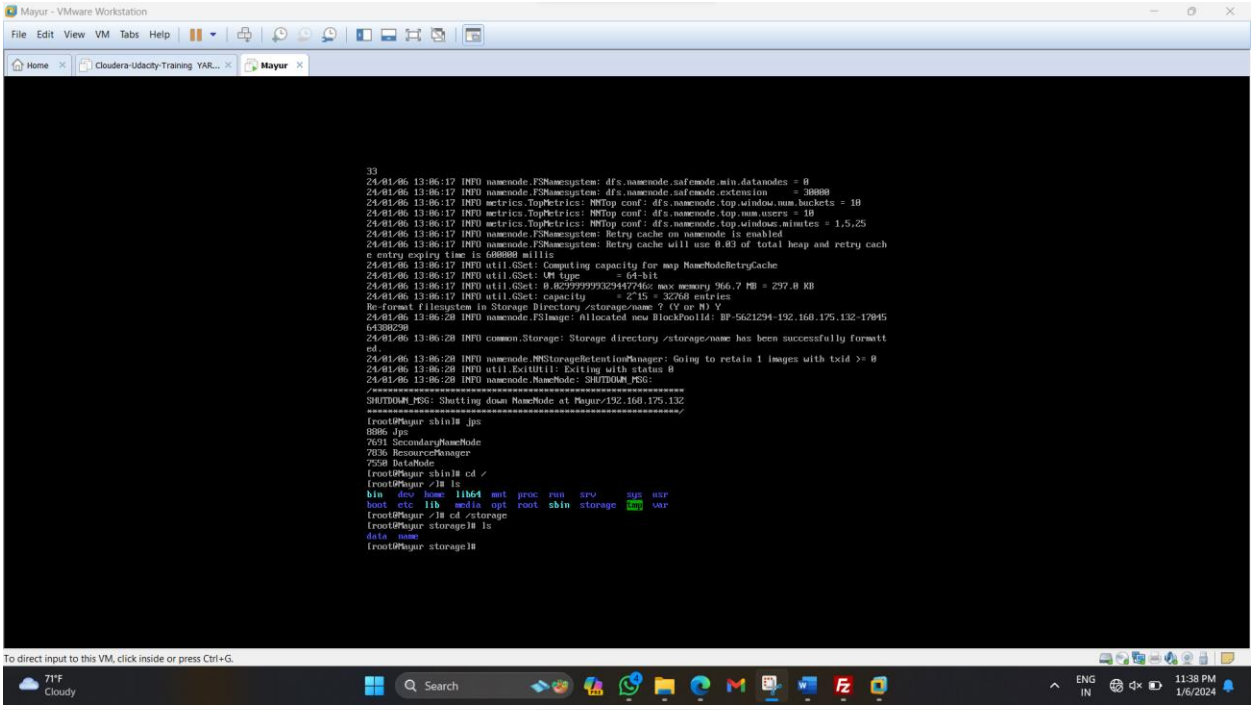
commands:

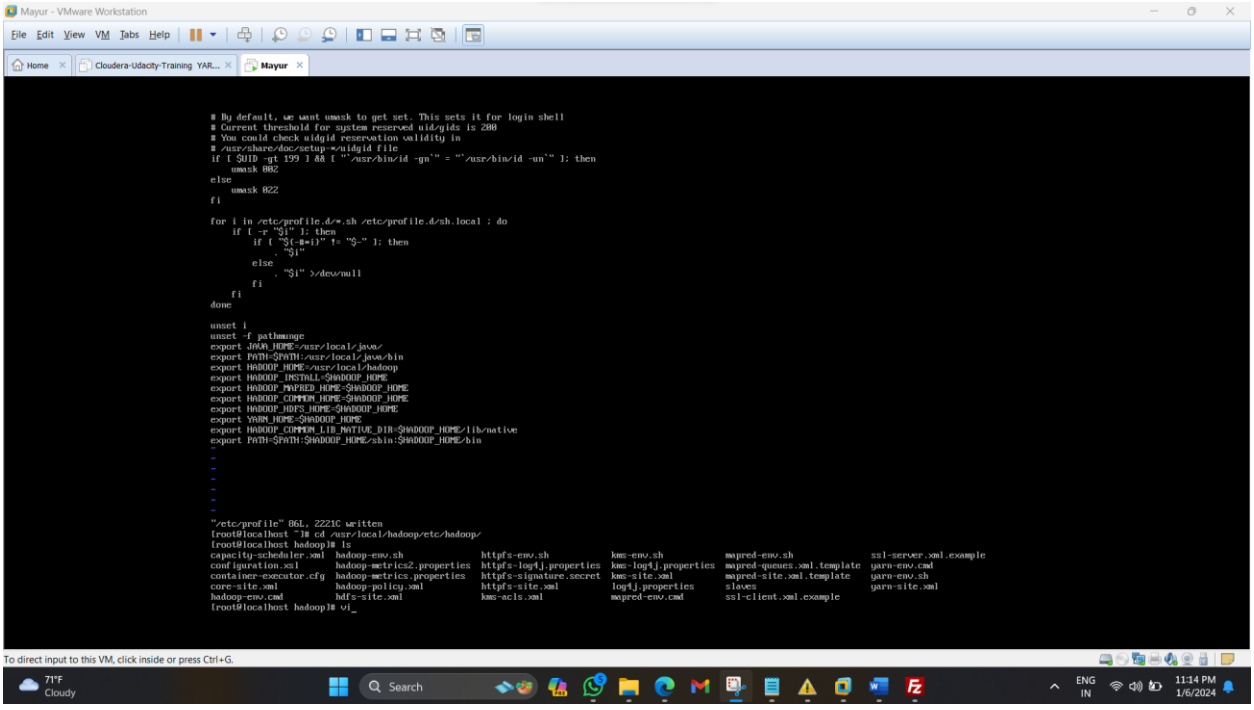
sh start-dfs.sh

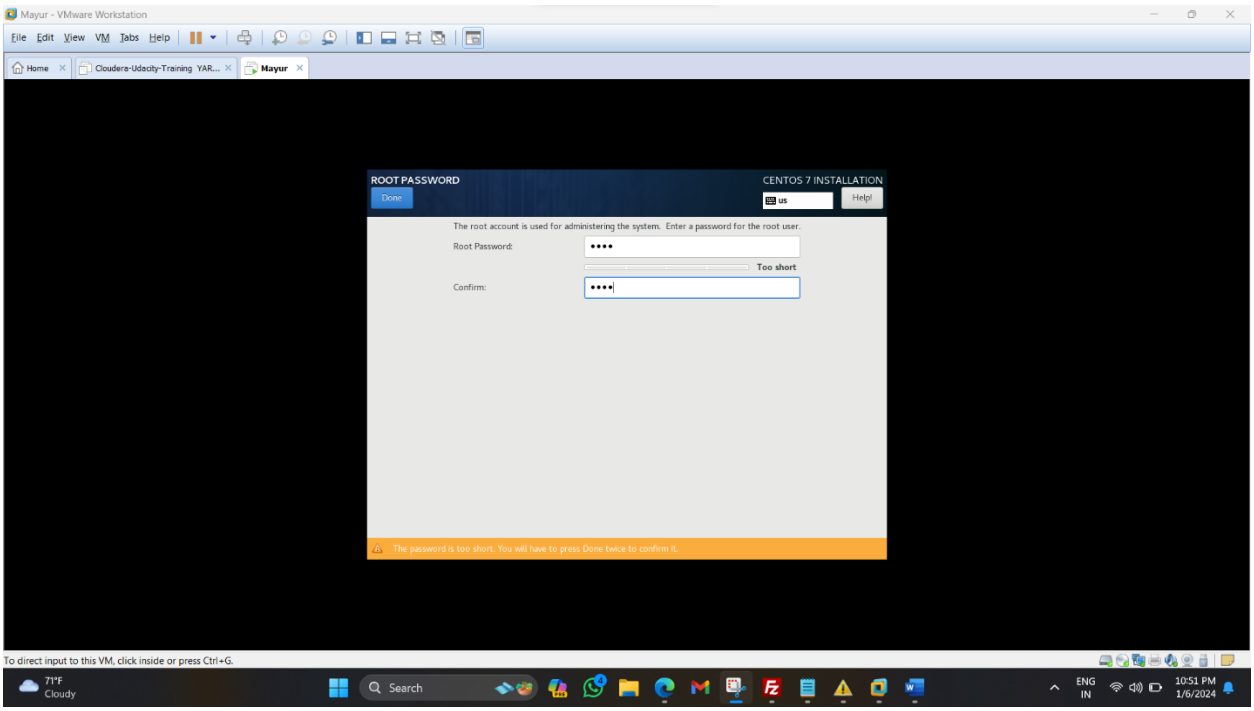
sh start-yarn.sh

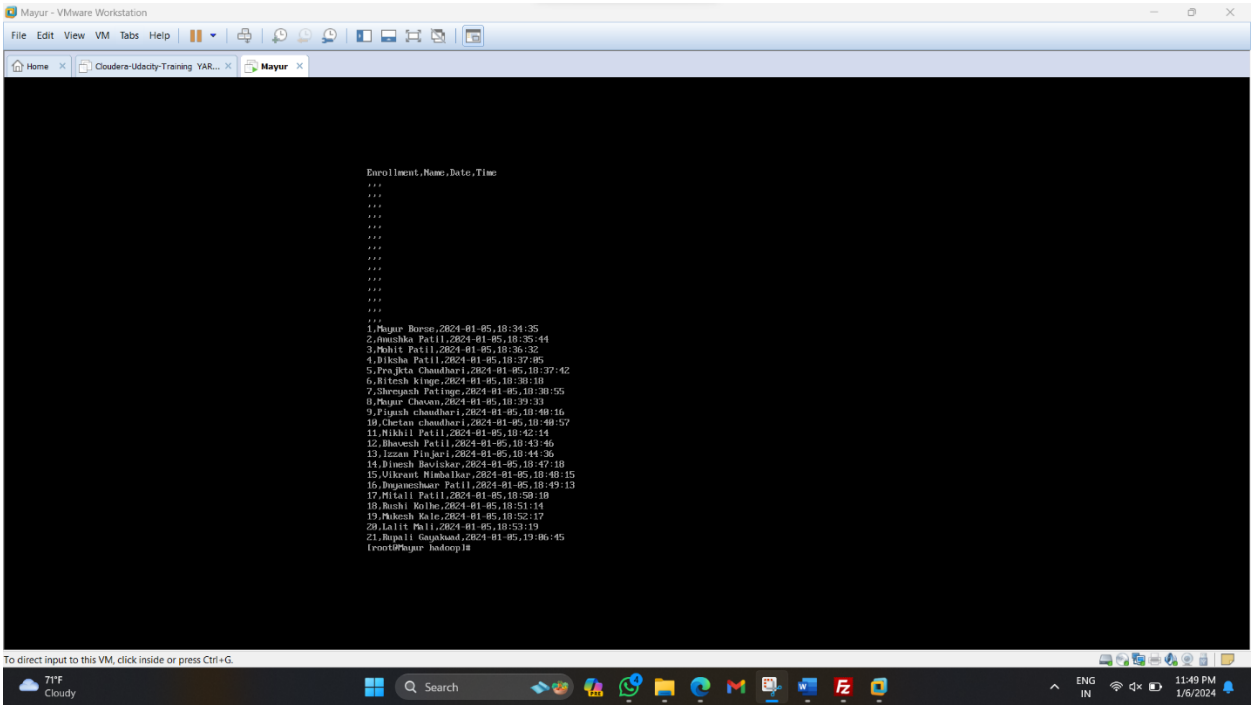
Step 11: Check all tools of Hadoop are working or not using Jps command.The tools are working mains hadoop setup in linux is completed. In that point hadoop single node cluster setup is completed.

Outcomes


Mayur - VMware Workstation
File Edit View VM Tabs Help
Home
Cloudera-Udacity-Training YAR...
Mayur
33
24/01/06 13:06:17 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
24/01/06 13:06:17 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
24/01/06 13:06:17 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
24/01/06 13:06:17 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
24/01/06 13:06:17 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
24/01/06 13:06:17 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
24/01/06 13:06:17 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cach
e entry expiry time is 600000 millis
24/01/06 13:06:17 INFO util.GSet: Computing capacity for map NameNodeRetryCache
24/01/06 13:06:17 INFO util.GSet: VM type = 64-bit
24/01/06 13:06:17 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 297.0 KB
24/01/06 13:06:17 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory /storage/name ? (Y or N) Y
24/01/06 13:06:28 INFO namenode.FSImage: Allocated new BlockPoolId: BP-5621294-192.168.175.132-17045
64308290
24/01/06 13:06:28 INFO common.Storage: Storage directory /storage/name has been successfully formatt
ed.
24/01/06 13:06:28 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
24/01/06 13:06:28 INFO util.ExitUtil: Exiting with status 0
24/01/06 13:06:28 INFO namenode.NameNode: SHUTDOWN_MSG:
/**
SHUTDOWN_MSG: Shutting down NameNode at Mayur/192.168.175.132
**/
[root@Mayur sbin]# jps
8886 Jps
7691 SecondaryNameNode
7836 ResourceManager
7558 DataNode
[root@Mayur sbin]# cd /
[root@Mayur /]# ls
bin dev home lib64 mnt proc run srv sys usr
boot etc lib media opt root sbin storage tmp var
[root@Mayur /]# cd /storage
[root@Mayur storage]# ls
data name
[root@Mayur storage]#
To direct input to this VM, click inside or press Ctrl+G.
71°F Cloudy
Search
ENG IN
11:38 PM 1/6/2024


Mayur - VMware Workstation
File Edit View VM Tabs Help
Home
Cloudera-Udacity-Training YAR...
Mayur
# By default, we want umask to get set. This sets it for login shell
# Current threshold for system reserved uid/gids is 200
# You could check uidgid reservation validity in
# /usr/share/doc/setup-*/uidgid file
if [$UID -gt 199] && ["`/usr/bin/id -gn`" = "`/usr/bin/id -un`"]; then
umask 002
else
umask 022
fi
for i in /etc/profile.d/*.sh /etc/profile.d/sh.local ; do
if [-r "$i"]; then
if ["${-#*i}" != "$-"]; then
. "$i"
else
. "$i" >/dev/null
fi
fi
done
unset i
unset -f pathmunge
export JAVA_HOME=/usr/local/java/
export PATH=$PATH:/usr/local/java/bin
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=$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
"/etc/profile" 86L, 2221C written
[root@localhost ~]# cd /usr/local/hadoop/etc/hadoop/
[root@localhost hadoop]# ls
capacity-scheduler.xml hadoop-env.sh httpfs-env.sh kms-env.sh mapred-env.sh ssl-server.xml.example
configuration.xsl hadoop-metrics2.properties httpfs-log4j.properties kms-log4j.properties mapred-queues.xml.template yarn-env.cmd
container-executor.cfg hadoop-metrics.properties httpfs-signature.secret kms-site.xml mapred-site.xml.template yarn-env.sh
core-site.xml hadoop-policy.xml httpfs-site.xml log4j.properties slaves yarn-site.xml
hadoop-env.cmd hdfs-site.xml kms-acls.xml mapred-env.cmd ssl-client.xml.example
[root@localhost hadoop]# vi
To direct input to this VM, click inside or press Ctrl+G.
71°F Cloudy
Search
ENG IN
11:14 PM 1/6/2024


Mayur - VMware Workstation
File Edit View VM Tabs Help
Home
Cloudera-Udacity-Training YAR...
Mayur
ROOT PASSWORD
Done
CENTOS 7 INSTALLATION
us
Help!
The root account is used for administering the system. Enter a password for the root user.
Root Password:
Too short
Confirm:
The password is too short. You will have to press Done twice to confirm it.
To direct input to this VM, click inside or press Ctrl+G.
71°F
Cloudy
Search
ENG
IN
10:51 PM
1/6/2024


Mayur - VMware Workstation
File Edit View VM Tabs Help
Home
Cloudera-Udacity-Training YAR...
Mayur
Enrollment,Name,Date,Time
1,Mayur Borse,2024-01-05,18:34:35
2,Anushka Patil,2024-01-05,18:35:44
3,Mohit Patil,2024-01-05,18:36:32
4,Diksha Patil,2024-01-05,18:37:05
5,Prajkta Chaudhari,2024-01-05,18:37:42
6,Ritesh kinge,2024-01-05,18:38:18
7,Shreyash Patinge,2024-01-05,18:38:55
8,Mayur Chavan,2024-01-05,18:39:33
9,Piyush chaudhari,2024-01-05,18:40:16
10,Chetan chaudhari,2024-01-05,18:40:57
11,Nikhil Patil,2024-01-05,18:42:14
12,Bhavesh Patil,2024-01-05,18:43:46
13,Izzan Pinjari,2024-01-05,18:44:36
14,Dinesh Baviskar,2024-01-05,18:47:18
15,Vikrant Nimbalkar,2024-01-05,18:48:15
16,Dnyaneshwar Patil,2024-01-05,18:49:13
17,Mitali Patil,2024-01-05,18:50:18
18,Rushi Kolhe,2024-01-05,18:51:14
19,Mahesh Kale,2024-01-05,18:52:17
20,Lalit Mali,2024-01-05,18:53:19
21,Rupali Gayakwad,2024-01-05,19:06:45
[root@Mayur hadoop]#
To direct input to this VM, click inside or press Ctrl+G.
71°F
Cloudy
Search
ENG
IN
11:49 PM
1/6/2024