→ create single node cluster
Refer hadoop 2.6.0 installation file

```
sdevsinx@bdt0:~/Desktop$ cd ~

# Update the source list
sdevsinx@bdt0:~$ sudo apt update

# The OpenJDK project is the default version of Java
# that is provided from a supported Ubuntu repository.
sdevsinx@bdt0:~$ sudo apt install default-jdk
sdevsinx@bdt0:~$ sudo apt install openjdk-8-jdk

sdevsinx@bdt0:~$ java -version
```

```
sdevsinx@bdt0:~$ java -version
openjdk version "11.0.19" 2023-04-18
OpenJDK Runtime Environment (build 11.0.19+7-post-Ubuntu-0ubuntu122.04.1)
OpenJDK 64-Bit Server VM (build 11.0.19+7-post-Ubuntu-0ubuntu122.04.1, mixed mod
e, sharing)
sdevsinx@bdt0:~$
```

```
sdevsinx@bdt0:~$ sudo update-alternatives  --config java
```

```
hduser@bdt0:~$ sudo update-alternatives --config java
There are 2 choices for the alternative java (providing /usr/bin/java).

  Selection    Path                                            Priority   Status
------------------------------------------------------------
* 0            /usr/lib/jvm/java-11-openjdk-amd64/bin/java      1111      auto m
ode
  1            /usr/lib/jvm/java-11-openjdk-amd64/bin/java      1111      manual
 mode
  2            /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java   1081      manual
 mode

Press <enter> to keep the current choice[*], or type selection number: 2
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java to pro
vide /usr/bin/java (java) in manual mode
hduser@bdt0:~$ sudo update-alternatives --config java
There are 2 choices for the alternative java (providing /usr/bin/java).

  Selection    Path                                            Priority   Status
------------------------------------------------------------
  0            /usr/lib/jvm/java-11-openjdk-amd64/bin/java      1111      auto m
ode
  1            /usr/lib/jvm/java-11-openjdk-amd64/bin/java      1111      manual
 mode
* 2            /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java   1081      manual
 mode

Press <enter> to keep the current choice[*], or type selection number:
hduser@bdt0:~$
```

```
sdevsinx@bdt0:~$ sudo update-alternatives  --config javac
```

```
sdevsinx@bdt0:~$ sudo update-alternatives --config javac
There are 2 choices for the alternative javac (providing /usr/bin/javac).

  Selection    Path                                            Priority   Status
------------------------------------------------------------
* 0            /usr/lib/jvm/java-11-openjdk-amd64/bin/javac    1111      auto mod
e
  1            /usr/lib/jvm/java-11-openjdk-amd64/bin/javac    1111      manual m
ode
  2            /usr/lib/jvm/java-8-openjdk-amd64/bin/javac     1081      manual m
ode

Press <enter> to keep the current choice[*], or type selection number: 2
update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/bin/javac to provid
e /usr/bin/javac (javac) in manual mode
sdevsinx@bdt0:~$ sudo update-alternatives --config javac
There are 2 choices for the alternative javac (providing /usr/bin/javac).

  Selection    Path                                            Priority   Status
------------------------------------------------------------
  0            /usr/lib/jvm/java-11-openjdk-amd64/bin/javac    1111      auto mod
e
  1            /usr/lib/jvm/java-11-openjdk-amd64/bin/javac    1111      manual m
ode
* 2            /usr/lib/jvm/java-8-openjdk-amd64/bin/javac     1081      manual m
ode

Press <enter> to keep the current choice[*], or type selection number:
sdevsinx@bdt0:~$
```

```
sdevsinx@bdt0:~$ java -version
```

```
hduser@bdt0:~$ java -version
openjdk version "1.8.0_362"
OpenJDK Runtime Environment (build 1.8.0_362-8u372-ga~us1-0ubuntu1~22.04-b09)
OpenJDK 64-Bit Server VM (build 25.362-b09, mixed mode)
hduser@bdt0:~$
```

```
sdevsinx@bdt0:~$ sudo addgroup hadoop
```

```
sdevsinx@bdt0:~$ sudo addgroup hadoop
Adding group `hadoop' (GID 1001) ...
Done.
```

```
sdevsinx@bdt0:~$ sudo adduser --ingroup hadoop hduser
```

```
sdevsinx@bdt0:~$ sudo adduser --ingroup hadoop hduser
Adding user `hduser' ...
Adding new user `hduser' (1001) with group `hadoop' ...
Creating home directory `/home/hduser' ...
Copying files from `/etc/skel' ...
New password:
BAD PASSWORD: The password is shorter than 8 characters
Retype new password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
        Full Name []:
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] Y
```

```
sdevsinx@bdt0:~$ sudo apt install ssh
```

```
sdevsinx@bdt0:~$ which ssh
```

```
sdevsinx@bdt0:~$ which ssh
/usr/bin/ssh
sdevsinx@bdt0:~$
```

```
sdevsinx@bdt0:~$ which sshd
```

```
sdevsinx@bdt0:~$ which sshd
/usr/sbin/sshd
sdevsinx@bdt0:~$
```

```
sdevsinx@bdt0:~$ su hduser
```

```
sdevsinx@bdt0:~$ su hduser
Password:
hduser@bdt0:/home/sdevsinx$ cd
```

```
hduser@bdt0:/home/sdevsinx$ cd
```

```
hduser@bdt0:~$ ssh-keygen -t rsa -P ""
```

```
hduser@bdt0:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):
Created directory '/home/hduser/.ssh'.
Your identification has been saved in /home/hduser/.ssh/id_rsa
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:Syc+3LTZKrVxvoRs9SxHPzFDhlZKga7kL7V+25Sgn50 hduser@bdt0
The key's randomart image is:
+---[RSA 3072]----+
|            .o.. |
|           .. +  |
|          .  + o |
|         . .. o  |
|        S + ...+ |
|       + XoBo+..=|
|        =.X*= ++.|
|        .+.=o+= o|
|         .+.o=.E |
+----[SHA256]-----+
```

```
hduser@bdt0:~$ cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

```
hduser@bdt0:~$ ssh localhost
```

```
hduser@bdt0:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ED25519 key fingerprint is SHA256:2jT3f3S4nTuql8EM+43hELWVJ+Jgqq1wQM/ioNa0D+4.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Welcome to Ubuntu 22.04.2 LTS (GNU/Linux 5.19.0-42-generic x86_64)

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

Expanded Security Maintenance for Applications is not enabled.

0 updates can be applied immediately.

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status


The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

hduser@bdt0:~$
```

```
hduser@bdt0:~$ wget
https://archive.apache.org/dist/hadoop/common/hadoop-2.6.0/hadoop-2.6.0.tar.gz
```

```
hduser@bdt0:~$ tar xvzf hadoop-2.6.0.tar.gz
```

```
hduser@bdt0:~$ mv hadoop-2.6.0 hadoop
```

```
hduser@bdt0:~$ mv hadoop /usr/local/
```

```
hduser@bdt0:~$ mv hadoop /usr/local/
mv: cannot move 'hadoop' to '/usr/local/hadoop': Permission denied
```

```
hduser@bdt0:~$ sudo hadoop /usr/local/
```

```
hduser@bdt0:~$ sudo hadoop /usr/local/
[sudo] password for hduser:
hduser is not in the sudoers file.  This incident will be reported.
hduser@bdt0:~$
```

```
hduser@bdt0:~$ su sdevsinx
```

```
hduser@bdt0:~$ su sdevsinx
Password:
sdevsinx@bdt0:/home/hduser$
```

```
sdevsinx@bdt0:/home/hduser$ sudo adduser hduser sudo
```

```
sdevsinx@bdt0:/home/hduser$ sudo adduser hduser sudo
[sudo] password for sdevsinx:
Adding user `hduser' to group `sudo' ...
Adding user hduser to group sudo
Done.
sdevsinx@bdt0:/home/hduser$
```

```
sdevsinx@bdt0:/home/hduser$ sudo su hduser
```

```
hduser@bdt0:~$ sudo mv hadoop /usr/local/
```

```
hduser@bdt0:~$ sudo chown hduser:hadoop /usr/local/hadoop/
```

Configuring files

```
hduser@bdt0:~$ update-alternatives --config java
```

```
hduser@bdt0:~$ update-alternatives --config java
There is only one alternative in link group java (providing /usr/bin/java): /usr
/lib/jvm/java-11-openjdk-amd64/bin/java
Nothing to configure.
hduser@bdt0:~$
```

```
hduser@bdt0:~$ vim .bashrc
```

```
hduser@bdt0:~$ vim .bashrc
```

```
119 #HADOOP VARIABLES START
120 #export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
120 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
121 export HADOOP_INSTALL=/usr/local/hadoop
122 export PATH=$PATH:$HADOOP_INSTALL/bin
123 export PATH=$PATH:$HADOOP_INSTALL/sbin
124 export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
125 export HADOOP_COMMON_HOME=$HADOOP_INSTALL
126 export HADOOP_HDFS_HOME=$HADOOP_INSTALL
127 export YARN_HOME=$HADOOP_INSTALL
128 export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
129 export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
130 #HADOOP VARIABLES END
```

```
119 #HADOOP VARIABLES START
120 export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
121 export HADOOP_INSTALL=/usr/local/hadoop
122 export PATH=$PATH:$HADOOP_INSTALL/bin
123 export PATH=$PATH:$HADOOP_INSTALL/sbin
124 export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
125 export HADOOP_COMMON_HOME=$HADOOP_INSTALL
126 export HADOOP_HDFS_HOME=$HADOOP_INSTALL
127 export YARN_HOME=$HADOOP_INSTALL
128 export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
129 export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
130 #HADOOP VARIABLES END
~
                                                    130,1         Bot
```

```
hduser@bdt0:~$ source ~/.bashrc
```

```
hduser@bdt0:~$ javac --version
```

```
hduser@bdt0:~$ javac --version
javac 11.0.19
hduser@bdt0:~$
```

```
hduser@bdt0:~$ readlink -f /usr/bin/javac
```

```
hduser@bdt0:~$ readlink -f /usr/bin/javac
/usr/lib/jvm/java-11-openjdk-amd64/bin/javac
hduser@bdt0:~$
```

```
hduser@bdt0:~$ readlink -f /usr/bin/javac
/usr/lib/jvm/java-8-openjdk-amd64/bin/javac
hduser@bdt0:~$
```

```
hduser@bdt0:~$ nano /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

```
# A string representing this instance of hadoop. $USER by default.
export HADOOP_IDENT_STRING=$USER
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64


^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
```

```
# A string representing this instance of hadoop. $USER by default.
export HADOOP_IDENT_STRING=$USER
#export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64


^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
```

```
hduser@bdt0:~$ sudo mkdir -p /app/hadoop/tmp
hduser@bdt0:~$ sudo chown hduser:hadoop /app/hadoop/tmp/
```

```
hduser@bdt0:~$ nano /usr/local/hadoop/etc/hadoop/core-site.xml
 19 <configuration>
 20          <property>
 21                  <name>hadoop.tmp.dir</name>
 22                  <value>/app/hadoop/tmp</value>
 23                  <description>A base for other temporary directories.</
description>
 24          </property>
 25          <property>
 26                  <name>fs.default.name</name>
 27                  <value>hdfs://localhost:54310</value>
 28                  <description>The name of the default file system. A URI
whose scheme and authority determine the FileSystem implementation. The       uri's
scheme determines the config property (fs.SCHEME.impl) naming the       FileSystem
implementation class. The uri's authority is used to determine    the host, port,
etc. for a filesystem.</description>
 29          </property>
 30 </configuration>
```

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

<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/app/hadoop/tmp</value>
                <description>A base for other temporary directories.</descripti>
        </property>
        <property>
                <name>fs.default.name</name>
                <value>hdfs://localhost:54310</value>
                <description>The name of the default file system. A URI whose s>
        </property>
</configuration>


^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
```

```
hduser@bdt0:~$ nano /usr/local/hadoop/etc/hadoop/mapred-site.xml
 19 <configuration>
 20          <property>
 21                  <name>mapred.job.tracker</name>
 22                  <value>localhost:54311</value>
 23                  <description>The host and port that the MapReduce job
tracker runs at. If "local", then jobs are run in-process as a single map      and
reduce task.</description>
 24          </property>
 25 </configuration>
```

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

<configuration>
        <property>
                <name>mapred.job.tracker</name>
                <value>localhost:54311</value>
                <description>The host and port that the MapReduce job tracker r>
        </property>
</configuration>

^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
```

```
hduser@bdt0:~$ sudo mkdir -p /usr/local/hadoop_store/hdfs/namenode

hduser@bdt0:~$ sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode

hduser@bdt0:~$ sudo chown -R hduser:hadoop /usr/local/hadoop_store

hduser@bdt0:~$ nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
 19 <configuration>
 20           <property>
 21                   <name>dfs.replication</name>
 22                   <value>1</value>
 23                   <description>Default block replication. The actual number
of replications can be specified when the file is created. The default is      used
if replication is not specified in create time.</description>
 24           </property>
 25 <!--      <property>
 26                   <name>dfs.blocksize</name>
 27                   <value>27M</value>
 28           </property>
 29 -->
 30           <property>
 31                   <name>dfs.namenode.name.dir</name>
 32                   <value>file:/usr/local/hadoop_store/hdfs/namenode</value>
 33           </property>
 34           <property>
 35                   <name>dfs.datanode.data.dir</name>
 36                   <value>file:/usr/local/hadoop_store/hdfs/datanode</value>
 37           </property>
 38 </configuration>
```

```
<configuration>
        <property>
                <name>dfs.replication</name>
                <value>1</value>
                <description>Default block replication. The actual number of re>
        </property>
<!--    <property>
                <name>dfs.blocksize</name>
                <value>27M</value>
        </property>
-->
        <property>
                <name>dfs.namenode.name.dir</name>
                <value>file:/usr/local/hadoop_store/hdfs/namenode</value>
        </property>
        <property>
                <name>dfs.datanode.data.dir</name>
                <value>file:/usr/local/hadoop_store/hdfs/datanode</value>
        </property>
</configuration>


^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
```

```
hduser@bdt0:~$ hadoop namenode -format
```

```
hduser@bdt0:~$ hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

23/05/26 23:52:09 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = bdt0/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.6.0
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/
hadoop/common/lib/jasper-compiler-5.5.23.jar:/usr/local/hadoop/share/hadoop/comm
on/lib/commons-digester-1.8.jar:/usr/local/hadoop/share/hadoop/common/lib/jsp-ap
i-2.1.jar:/usr/local/hadoop/share/hadoop/common/lib/stax-api-1.0-2.jar:/usr/loca
l/hadoop/share/hadoop/common/lib/commons-configuration-1.6.jar:/usr/local/hadoop
/share/hadoop/common/lib/jersey-core-1.9.jar:/usr/local/hadoop/share/hadoop/comm
on/lib/jasper-runtime-5.5.23.jar:/usr/local/hadoop/share/hadoop/common/lib/curat
or-recipes-2.6.0.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-core-asl-
1.9.13.jar:/usr/local/hadoop/share/hadoop/common/lib/snappy-java-1.0.4.1.jar:/us
r/local/hadoop/share/hadoop/common/lib/avro-1.7.4.jar:/usr/local/hadoop/share/ha
```

```
/mapreduce/hadoop-mapreduce-examples-2.6.0.jar:/usr/local/hadoop/share/hadoop/ma
preduce/hadoop-mapreduce-client-app-2.6.0.jar:/contrib/capacity-scheduler/*.jar:
/contrib/capacity-scheduler/*.jar
STARTUP_MSG:   build = https://git-wip-us.apache.org/repos/asf/hadoop.git -r e34
96499ecb8d220fba99dc5ed4c99c8f9e33bb1; compiled by 'jenkins' on 2014-11-13T21:10
Z
STARTUP_MSG:   java = 11.0.19
************************************************************/
23/05/26 23:52:09 INFO namenode.NameNode: registered UNIX signal handlers for [T
ERM, HUP, INT]
23/05/26 23:52:09 INFO namenode.NameNode: createNameNode [-format]
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/26 23:52:10 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Formatting using clusterid: CID-53fae62b-4e8f-4dd6-bb2d-2b25c9cb0540
23/05/26 23:52:10 INFO namenode.FSNamesystem: No KeyProvider found.
23/05/26 23:52:10 INFO namenode.FSNamesystem: fsLock is fair:true
23/05/26 23:52:10 INFO blockmanagement.DatanodeManager: dfs.block.invalidate.lim
it=1000
23/05/26 23:52:10 INFO blockmanagement.DatanodeManager: dfs.namenode.datanode.re
gistration.ip-hostname-check=true
23/05/26 23:52:10 INFO blockmanagement.BlockManager: dfs.namenode.startup.delay.
block.deletion.sec is set to 000:00:00:00.000
23/05/26 23:52:10 INFO blockmanagement.BlockManager: The block deletion will sta
rt around 2023 May 26 23:52:10
23/05/26 23:52:10 INFO util.GSet: Computing capacity for map BlocksMap
23/05/26 23:52:10 INFO util.GSet: VM type       = 64-bit
23/05/26 23:52:10 INFO util.GSet: 2.0% max memory 1000 MB = 20 MB
23/05/26 23:52:10 INFO util.GSet: capacity      = 2^21 = 2097152 entries
23/05/26 23:52:10 INFO blockmanagement.BlockManager: dfs.block.access.token.enab
le=false
23/05/26 23:52:10 INFO blockmanagement.BlockManager: defaultReplication
= 1
```

```
23/05/26 23:52:10 INFO blockmanagement.BlockManager: maxReplication
= 512
23/05/26 23:52:10 INFO blockmanagement.BlockManager: minReplication
= 1
23/05/26 23:52:10 INFO blockmanagement.BlockManager: maxReplicationStreams
= 2
23/05/26 23:52:10 INFO blockmanagement.BlockManager: shouldCheckForEnoughRacks
= false
23/05/26 23:52:10 INFO blockmanagement.BlockManager: replicationRecheckInterval
= 3000
23/05/26 23:52:10 INFO blockmanagement.BlockManager: encryptDataTransfer
= false
23/05/26 23:52:10 INFO blockmanagement.BlockManager: maxNumBlocksToLog
= 1000
23/05/26 23:52:11 INFO namenode.FSNamesystem: fsOwner             = hduser (auth
:SIMPLE)
23/05/26 23:52:11 INFO namenode.FSNamesystem: supergroup          = supergroup
23/05/26 23:52:11 INFO namenode.FSNamesystem: isPermissionEnabled = true
23/05/26 23:52:11 INFO namenode.FSNamesystem: HA Enabled: false
23/05/26 23:52:11 INFO namenode.FSNamesystem: Append Enabled: true
23/05/26 23:52:11 INFO util.GSet: Computing capacity for map INodeMap
23/05/26 23:52:11 INFO util.GSet: VM type       = 64-bit
23/05/26 23:52:11 INFO util.GSet: 1.0% max memory 1000 MB = 10 MB
23/05/26 23:52:11 INFO util.GSet: capacity      = 2^20 = 1048576 entries
23/05/26 23:52:11 INFO namenode.NameNode: Caching file names occuring more than
10 times
23/05/26 23:52:11 INFO util.GSet: Computing capacity for map cachedBlocks
23/05/26 23:52:11 INFO util.GSet: VM type       = 64-bit
23/05/26 23:52:11 INFO util.GSet: 0.25% max memory 1000 MB = 2.5 MB
23/05/26 23:52:11 INFO util.GSet: capacity      = 2^18 = 262144 entries
23/05/26 23:52:11 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pc
t = 0.9990000128746033
23/05/26 23:52:11 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanode
s = 0
23/05/26 23:52:11 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension
  = 30000
23/05/26 23:52:11 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
23/05/26 23:52:11 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total
 heap and retry cache entry expiry time is 600000 millis
23/05/26 23:52:11 INFO util.GSet: Computing capacity for map NameNodeRetryCache
```

```
23/05/26 23:52:11 INFO util.GSet: VM type       = 64-bit
23/05/26 23:52:11 INFO util.GSet: 0.029999999329447746% max memory 1000 MB = 307
.2 KB
23/05/26 23:52:11 INFO util.GSet: capacity      = 2^15 = 32768 entries
23/05/26 23:52:11 INFO namenode.NNConf: ACLs enabled? false
23/05/26 23:52:11 INFO namenode.NNConf: XAttrs enabled? true
23/05/26 23:52:11 INFO namenode.NNConf: Maximum size of an xattr: 16384
Re-format filesystem in Storage Directory /usr/local/hadoop_store/hdfs/namenode
? (Y or N) Y
23/05/26 23:54:52 INFO namenode.FSImage: Allocated new BlockPoolId: BP-188099665
9-127.0.1.1-1685125492455
23/05/26 23:54:52 INFO common.Storage: Storage directory /usr/local/hadoop_store
/hdfs/namenode has been successfully formatted.
23/05/26 23:54:52 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
23/05/26 23:54:52 INFO util.ExitUtil: Exiting with status 0
23/05/26 23:54:52 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at bdt0/127.0.1.1
************************************************************/
hduser@bdt0:~$
```

```
hduser@bdt0:~$ start-all.sh
```

```
hduser@bdt0:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/26 23:57:11 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-na
menode-bdt0.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-da
tanode-bdt0.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hd
user-secondarynamenode-bdt0.out
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/26 23:57:34 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resource
manager-bdt0.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-n
odemanager-bdt0.out
hduser@bdt0:~$
```

```
hduser@bdt0:~$ jps
```

```
hduser@bdt0:~$ jps
18113 Jps
17242 DataNode
17468 SecondaryNameNode
17613 ResourceManager
17103 NameNode
17727 NodeManager
hduser@bdt0:~$
```

```
hduser@bdt0:~$ netstat -plten | grep java
hduser@bdt0:~$ netstat -plten | grep java
(Not all processes could be identified, non-owned process info
 will not be shown, you would have to be root to see it all.)
tcp        0      0 0.0.0.0:50090           0.0.0.0:*               LISTEN
1001       96806      17468/java
tcp        0      0 0.0.0.0:50075           0.0.0.0:*               LISTEN
1001       97904      17242/java
tcp        0      0 0.0.0.0:50070           0.0.0.0:*               LISTEN
1001       95008      17103/java
tcp        0      0 0.0.0.0:50020           0.0.0.0:*               LISTEN
1001       96416      17242/java
tcp        0      0 0.0.0.0:50010           0.0.0.0:*               LISTEN
1001       97898      17242/java
tcp        0      0 127.0.0.1:54310         0.0.0.0:*               LISTEN
1001       95220      17103/java
tcp6       0      0 :::43715                :::*                    LISTEN
1001       102989     17727/java
tcp6       0      0 :::8088                 :::*                    LISTEN
1001       101182     17613/java
tcp6       0      0 :::8042                 :::*                    LISTEN
1001       103000     17727/java
tcp6       0      0 :::8040                 :::*                    LISTEN
1001       102996     17727/java
tcp6       0      0 :::8033                 :::*                    LISTEN
1001       103774     17613/java
tcp6       0      0 :::8032                 :::*                    LISTEN
1001       100571     17613/java
tcp6       0      0 :::8031                 :::*                    LISTEN
1001       100563     17613/java
tcp6       0      0 :::8030                 :::*                    LISTEN
1001       99746      17613/java
hduser@bdt0:~$
```

```
hduser@bdt0:~$ stop-all.sh
hduser@bdt0:~$ stop-all.sh
This script is Deprecated. Instead use stop-dfs.sh and stop-yarn.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/27 00:06:20 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/27 00:06:43 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
no proxyserver to stop
hduser@bdt0:~$
```

→ Alternate to start hadoop server

```
hduser@bdt0:~$ start-dfs.sh
```

```
hduser@bdt0:~$ start-dfs.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/27 00:10:49 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-na
menode-bdt0.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-da
tanode-bdt0.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hd
user-secondarynamenode-bdt0.out
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.
util.KerberosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.
6.0.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.
security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflect
ive access operations
WARNING: All illegal access operations will be denied in a future release
23/05/27 00:11:09 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
```

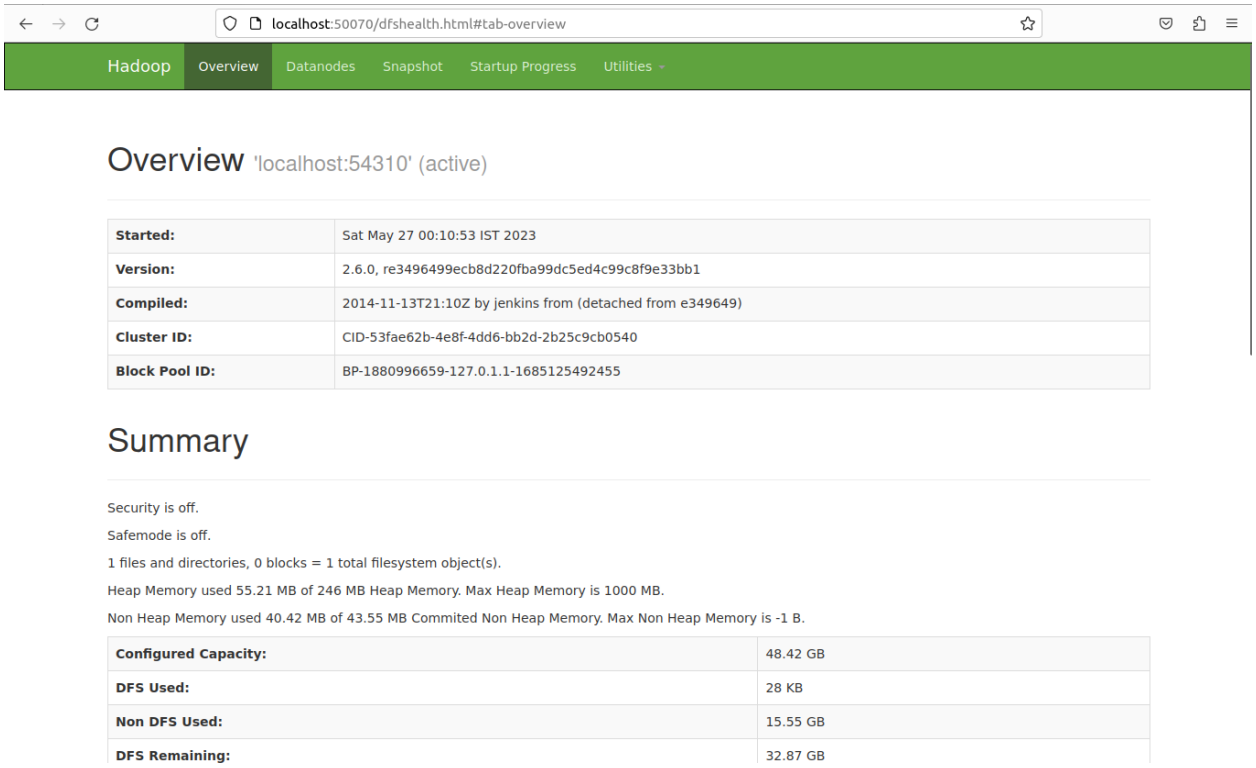
```
hduser@bdt0:~$ start-yarn.sh
```

```
hduser@bdt0:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resource
manager-bdt0.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-n
odemanager-bdt0.out
hduser@bdt0:~$
```

→ Hadoop Web Interface

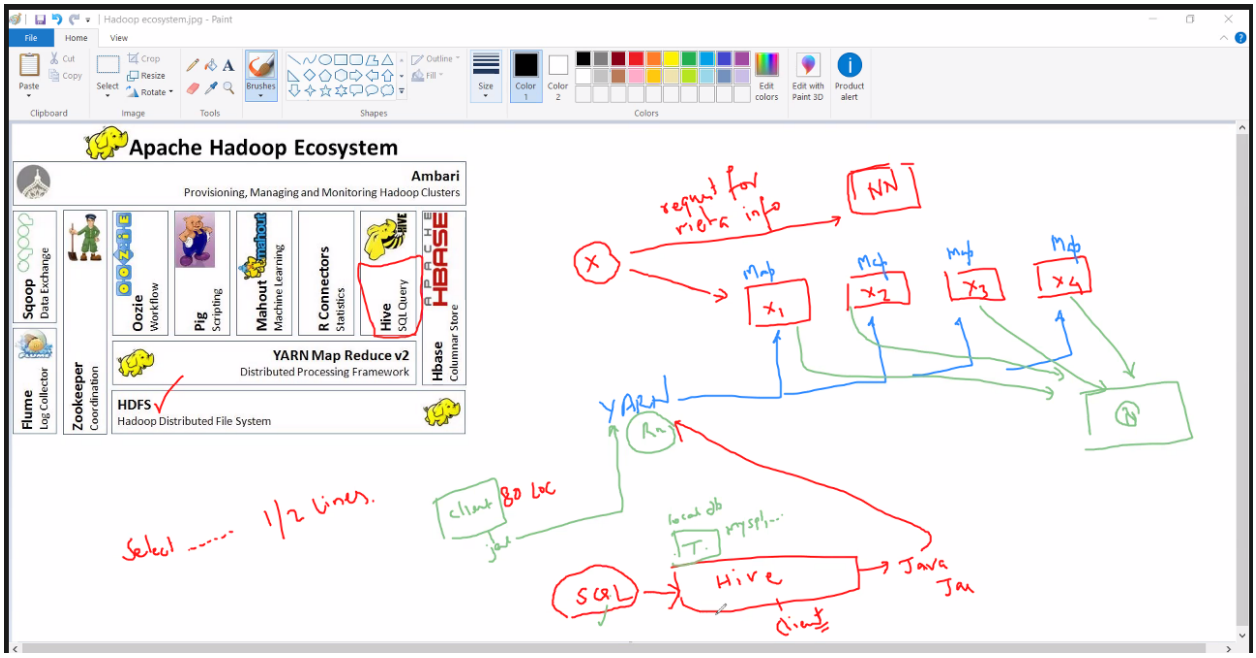
To launch Web UI of the NameNode Daemon, open below web link

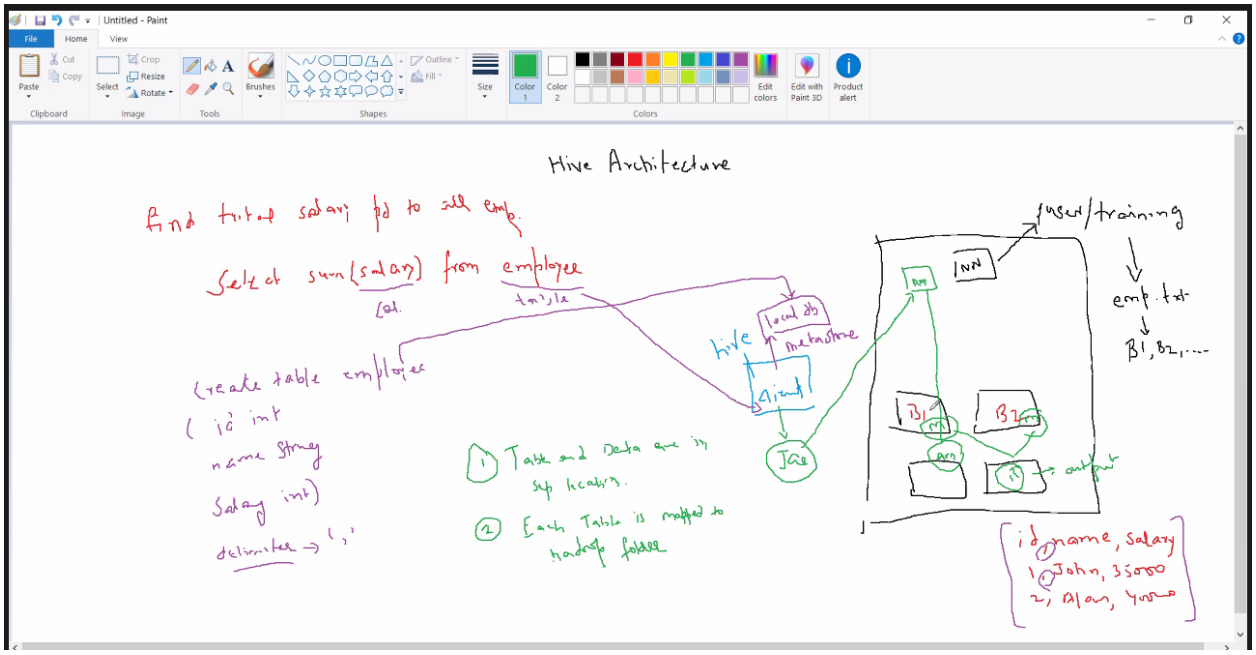
http://localhost:50070/



→ Hive

a. Created by facebook
b. Used to submit SQL queries
c. Compiles SQL query and convert it into jar file and then jar file will be run by YARN
d. It needs a local database, so we first create a local database to store the data from the splits/blocks stored on DataNodes
e. The local database should be created considering the structure of data/file being read from DataNodes
f. Table and data are in separate folder/location
g. Table should be mapped to hadoop folder
h. Table is created in meta-store
i. Table is just for schema reference
j.





→ hive CLI

start hive cli by using command ‘hive’

```
[bigdatalab456422@ip-10-1-1-204 ~]$ hive
```

```
[bigdatalab456422@ip-10-1-1-204 ~]$ hive
WARNING: Use "yarn jar" to launch YARN applications.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/log4j-slf4j-impl-2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2023-06-08 06:02:50,539 main WARN JNDI lookup class is not available because this JRE does not support JNDI. JNDI string lookups will not be available, continuing confi
guration. Ignoring java.lang.ClassNotFoundException: org.apache.logging.log4j.core.lookup.JndiLookup

Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/jars/hive-common-2.1.1-cdh6.2.1.jar!/hive-log4j2.properties As
ync: false

WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>
```

set hive.cli.print.current.db = true // sets console to print the current database in use

```
hive> set hive.cli.print.current.db = true ;
```

```
hive> set hive.cli.print.current.db = true ;
hive (default)>
```

SHOW DATABASES

```
hive (default)> SHOW DATABASES;
```

```
hive (default)> SHOW DATABASES ;
OK
01_piyali
01piyali
03jan2023
03march2023
03march2023demo
yuvrajyadav
zaq
zeus_pract
zone
Time taken: 1.537 seconds, Fetched: 1028 row(s)
hive (default)>
```

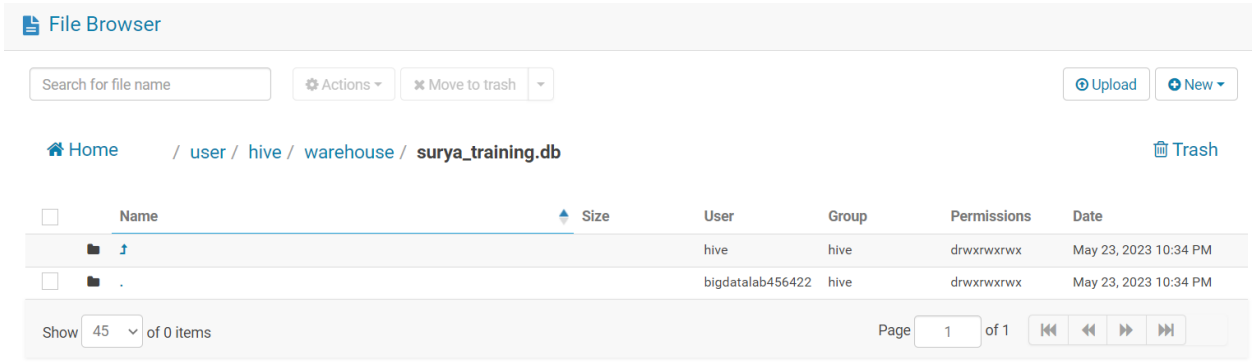
CREATE DATABASE

```
hive (default)> CREATE DATABASE sandeep_training;
```

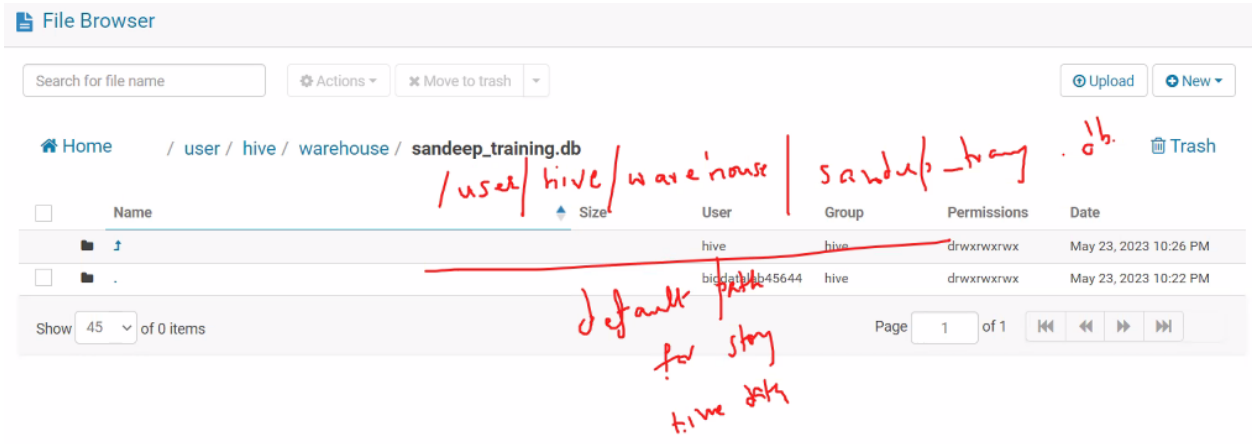
```
hive (default)> CREATE DATABASE sandeep_training ;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. Database sandeep_training already exists
hive (default)>
```

```
hive (default)> CREATE DATABASE surya_training ;
OK
Time taken: 0.2 seconds
hive (default)>
```

visit this path in hue tool to check contents of database
/user/hive/warehouse/surya_training.db



sandeep_training in hue



change current database

```
hive (default)> USE surya_training ;
```

show tables in current database

```
hive (surya_training)> SHOW TABLES ;
OK
Time taken: 0.067 seconds
hive (surya_training)>
```

→ while creating table, remember:

a. No. of cols = no of fields
b. Proper data types to be used
c. Proper delimiter to be used

Create table nyse

```
hive (surya_training)> create table nyse (exchange_name string,
stock_id string, stk_date date,
                    > open double, high double, low double,
                    > close double, volume bigint, adj_close double
                    >
                    > ) row format delimited
                    >
                    > fields terminated by ','
                    >
                    > stored as textfile;
OK
Time taken: 0.132 seconds

hive (surya_training)> SHOW TABLES ;
OK
nyse
Time taken: 0.054 seconds, Fetched: 1 row(s)
hive (surya_training)>
```

DESC

```
hive (surya_training)> DESC nyse ;
```

```
hive (surya_training)> DESC nyse ;
OK
exchange_name           string
stock_id                string
stk_date                date
open                    double
high                    double
low                     double
close                   double
volume                  bigint
adj_close               double
Time taken: 0.069 seconds, Fetched: 9 row(s)
hive (surya_training)>
```

DESC FORMATTED

```
hive (surya_training)> DESC FORMATTED  nyse ;
```

```
hive (surya_training)> DESC FORMATTED nyse ;
OK
# col_name              data_type               comment

exchange_name           string
stock_id                string
stk_date                date
open                    double
high                    double
low                     double
close                   double
volume                  bigint
adj_close               double

# Detailed Table Information
Database:               surya_training
OwnerType:              USER
Owner:                  bigdatalab456422
CreateTime:             Wed May 24 06:13:46 UTC 2023
LastAccessTime:         UNKNOWN
Retention:              0
Location:               hdfs://nameservice1/user/hive/warehouse/surya_training.db/nyse
Table Type:             MANAGED_TABLE
Table Parameters:
        COLUMN_STATS_ACCURATE   {\"BASIC_STATS\":\"true\"}
        numFiles                0
        numRows                 0
        rawDataSize             0
        totalSize               0
        transient_lastDdlTime   1684908826

# Storage Information
SerDe Library:          org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:            org.apache.hadoop.mapred.TextInputFormat
OutputFormat:           org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed:             No
Num Buckets:            -1
Bucket Columns:         []
Sort Columns:           []
Storage Desc Params:
        field.delim             ,
        serialization.format    ,
Time taken: 0.077 seconds, Fetched: 40 row(s)
hive (surya_training)>
```

try to print record from nyse table, but it'll not show any record, since table is empty and we've not moved the data to warehouse/database directory

```
hive (surya_training)> SELECT * FROM nyse ;
OK
Time taken: 0.366 seconds
hive (surya_training)>
```

move NYSE.csv to database location as seen in DESC FORMATTED command

```
[bigdatalab456422@ip-10-1-1-204 ~]$ hadoop fs -mv
/user/bigdatalab456422/training/NYSE.csv
/user/hive/warehouse/surya_training.db/nyse
[bigdatalab456422@ip-10-1-1-204 ~]$
```

print record from nyse table

```
hive (surya_training)> SELECT * FROM nyse LIMIT 10 ;
```

```
hive (surya_training)> SELECT * FROM nyse LIMIT 10 ;
OK
NYSE    AEA     2010-02-08      4.42    4.42    4.21    4.24    205500  4.24
NYSE    AEA     2010-02-05      4.42    4.54    4.22    4.41    194300  4.41
NYSE    AEA     2010-02-04      4.55    4.69    4.39    4.42    233800  4.42
NYSE    AEA     2010-02-03      4.65    4.69    4.5     4.55    182100  4.55
NYSE    AEA     2010-02-02      4.74    5.0     4.62    4.66    222700  4.66
NYSE    AEA     2010-02-01      4.84    4.92    4.68    4.75    194800  4.75
NYSE    AEA     2010-01-29      4.97    5.05    4.76    4.83    222900  4.83
NYSE    AEA     2010-01-28      5.12    5.22    4.81    4.98    283100  4.98
NYSE    AEA     2010-01-27      4.82    5.16    4.79    5.09    243500  5.09
NYSE    AEA     2010-01-26      5.18    5.18    4.81    4.84    554800  4.84
Time taken: 0.397 seconds, Fetched: 10 row(s)
hive (surya_training)>
```

→ # find the total volume for each stock

SELECT stock_id, sum (volume) FROM nyse GROUP BY stock_id ;

```
hive (surya_training)> SELECT stock_id, sum (volume) FROM nyse GROUP
BY stock_id ;
```

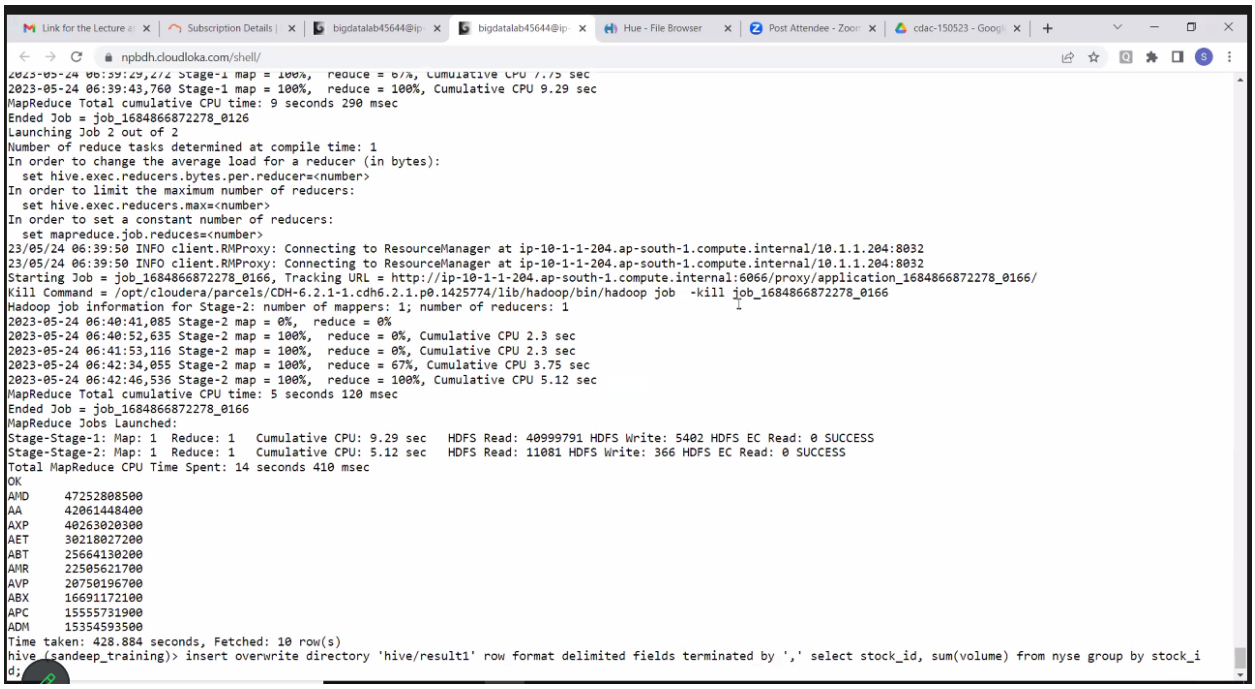
```
hive (surya_training)> SELECT stock_id, sum(volume) FROM nyse GROUP BY stock_id ;
Query ID = bigdatalab456422_20230608064000_5694e1d7-32ee-4c8c-aad7-f43c129cf95c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 06:40:00 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 06:40:00 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1624, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1624/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1624
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 06:40:09,263 Stage-1 map = 0%,  reduce = 0%
2023-06-08 06:40:16,435 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.08 sec
2023-06-08 06:40:22,571 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.07 sec
MapReduce Total cumulative CPU time: 8 seconds 70 msec
Ended Job = job_1685754149182_1624
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.07 sec   HDFS Read: 41000657 HDFS Write: 5441 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 70 msec
OK
AA      42061448400
AAI     5246821400
AAN     817567400
AAP     2802701500
AAR     49882000
AAV     834246600
AYR     472431800
AZN     3418077300
AZO     3366821200
AZZ     323604700
Time taken: 23.324 seconds, Fetched: 203 row(s)
hive (surya_training)>
```

→ # highest vol - top 10 stocks

SELECT stock_id, sum(volume) AS total_vol FROM nyse GROUP BY stock_id ORDER BY total_vol DESC LIMIT 10;

```
hive (surya_training)> SELECT stock_id, sum(volume) AS total_vol FROM
nyse GROUP BY stock_id ORDER BY  total_vol DESC LIMIT 10;
```

```
hive> SELECT stock_id, sum(volume) AS total_vol FROM nyse GROUP BY stock_id ORDER BY  total_vol DESC LIMIT 10;
Query ID = bigdatalab456422_20230608070630_2edc1c11-771d-48ea-9066-6146beed2cae
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 07:06:33 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 07:06:33 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1638, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1638/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1638
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 07:06:42,332 Stage-1 map = 0%,  reduce = 0%
2023-06-08 07:06:50,554 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.2 sec
2023-06-08 07:06:58,757 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.46 sec
MapReduce Total cumulative CPU time: 7 seconds 460 msec
Ended Job = job_1685754149182_1638
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 07:06:59 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 07:06:59 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1639, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1639/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1639
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-06-08 07:07:10,974 Stage-2 map = 0%,  reduce = 0%
2023-06-08 07:07:18,127 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.74 sec
2023-06-08 07:07:25,300 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.16 sec
MapReduce Total cumulative CPU time: 4 seconds 160 msec
Ended Job = job_1685754149182_1639

MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.46 sec   HDFS Read: 40999570 HDFS Write: 5402 HDFS EC Read: 0 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.16 sec   HDFS Read: 11096 HDFS Write: 366 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 620 msec
OK
AMD     47252808500
AA      42061448400
AXP     40263020300
AET     30218027200
ABT     25664130200
AMR     22505621700
AVP     20750196700
ABX     16691172100
APC     15555731900
ADM     15354593500
Time taken: 55.688 seconds, Fetched: 10 row(s)
hive>
```



```
```

→ # find all time low price for each stock

SELECT stock_id, min(low) FROM nyse GROUP BY stock_id;

```
hive (surya_training)> SELECT stock_id, min(low) FROM nyse GROUP BY
stock_id;
```

```
hive> SELECT stock_id, min(low) FROM nyse GROUP BY stock_id;
Query ID = bigdatalab456422_20230608073534_bcef23c9-b3a3-47ee-b111-9bf9835023b5
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 07:35:34 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 07:35:34 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1660, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1660/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1660
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 07:35:44,568 Stage-1 map = 0%,  reduce = 0%
2023-06-08 07:35:53,827 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.55 sec
2023-06-08 07:36:02,005 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.39 sec
MapReduce Total cumulative CPU time: 8 seconds 390 msec
Ended Job = job_1685754149182_1660
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.39 sec   HDFS Read: 41000523 HDFS Write: 4363 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 390 msec
OK
AA      4.97
AAI     1.28
AAN     6.94
AAP     24.03

AYE     2.95
AYI     10.7
AYN     6.93
AYR     2.54
AZN     22.29
AZO     19.5
AZZ     2.06
Time taken: 29.661 seconds, Fetched: 203 row(s)
hive>
```

→ # find all time high price for each stock

SELECT stock_id, max(high) FROM nyse GROUP BY stock_id;

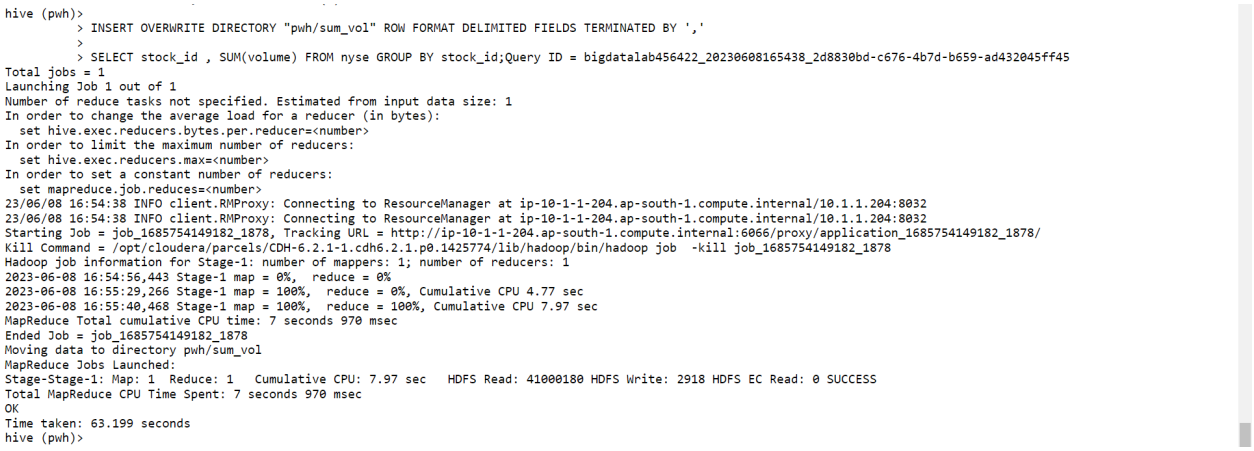
```
hive (surya_training)> SELECT stock_id, max(high) FROM nyse GROUP BY
stock_id;
```

```
hive> SELECT stock_id, max(high) FROM nyse GROUP BY stock_id;
Query ID = bigdatalab456422_20230608074158_9cf6cba4-2023-4704-9b7f-8682fc3bdad7
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 07:41:58 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 07:41:58 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1662, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1662/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1662
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 07:42:10,090 Stage-1 map = 0%,  reduce = 0%
2023-06-08 07:42:19,356 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.61 sec
2023-06-08 07:42:27,562 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.64 sec
MapReduce Total cumulative CPU time: 9 seconds 640 msec
Ended Job = job_1685754149182_1662
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.64 sec   HDFS Read: 41000525 HDFS Write: 4521 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 640 msec
OK
AA      94.62
AAI     57.88
AAN     35.21
AAP     83.65

ADY     44.0
AEA     23.94
AEB     26.5

AYE     65.48
AYI     66.89
AYN     15.42
AYR     41.31
AZN     145.41
AZO     169.99
AZZ     59.2
Time taken: 30.089 seconds, Fetched: 203 row(s)
hive>
```

→ # save total volume for each stock in file

INSERT OVERWRITE DIRECTORY ‘hive/result1’ ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ SELECT stock_id, sum(volume) FROM nyse GROUP BY stock_id;

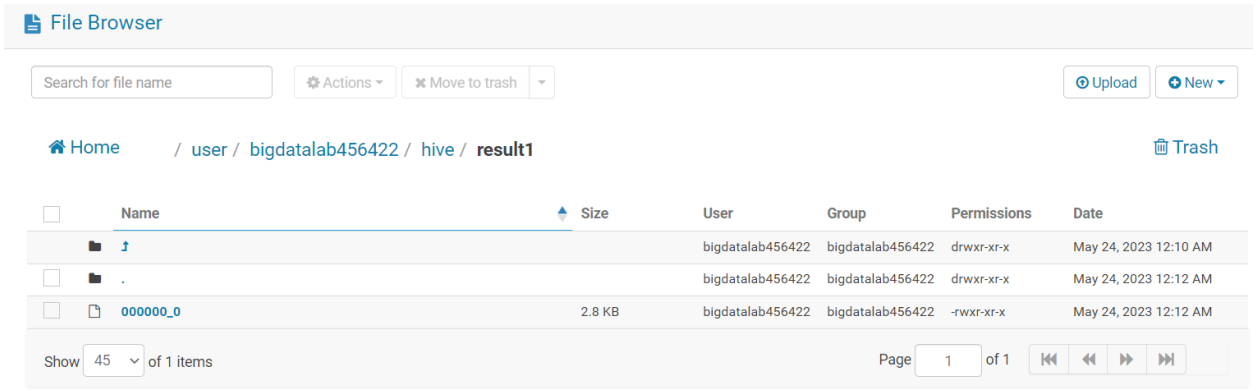
```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/result1' ROW
FORMAT DELIMITED FIELDS TERMINATED BY ',' SELECT stock_id,
sum(volume) FROM nyse GROUP BY stock_id;
```

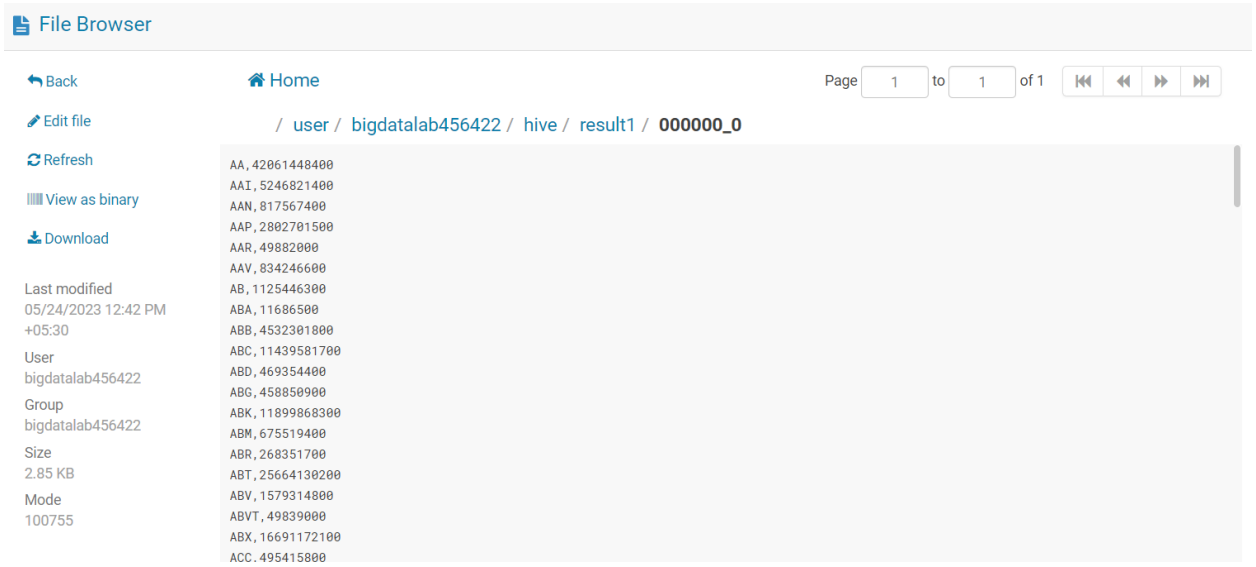


```
hive (pwh)>
         > INSERT OVERWRITE DIRECTORY "pwh/sum_vol" ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
         >
         > SELECT stock_id , SUM(volume) FROM nyse GROUP BY stock_id;Query ID = bigdatalab456422_20230608165438_2d8830bd-c676-4b7d-b659-ad432045ff45
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 16:54:38 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 16:54:38 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1878, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1878/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1878
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 16:54:56,443 Stage-1 map = 0%,  reduce = 0%
2023-06-08 16:55:29,266 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.77 sec
2023-06-08 16:55:40,468 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.97 sec
MapReduce Total cumulative CPU time: 7 seconds 970 msec
Ended Job = job_1685754149182_1878
Moving data to directory pwh/sum_vol
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.97 sec   HDFS Read: 41000180 HDFS Write: 2918 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 970 msec
OK
Time taken: 63.199 seconds
hive (pwh)>
```

```
Query ID =
bigdatalab456422_20230524071018_0a7286c1-718f-4a18-897d-7c8567317622
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size:
1
In order to change the average load for a reducer (in bytes):
 set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
 set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
 set mapreduce.job.reduces=<number>
23/05/24 07:10:19 INFO client.RMProxy: Connecting to ResourceManager
at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/24 07:10:19 INFO client.RMProxy: Connecting to ResourceManager
at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0374, Tracking URL =
http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/applicati
on_1684866872278_0374/
Kill Command =
/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/
hadoop job  -kill job_1684866872278_0374
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2023-05-24 07:11:26,556 Stage-1 map = 0%,  reduce = 0%
2023-05-24 07:11:54,284 Stage-1 map = 100%,  reduce = 0%, Cumulative
CPU 6.1 sec
```

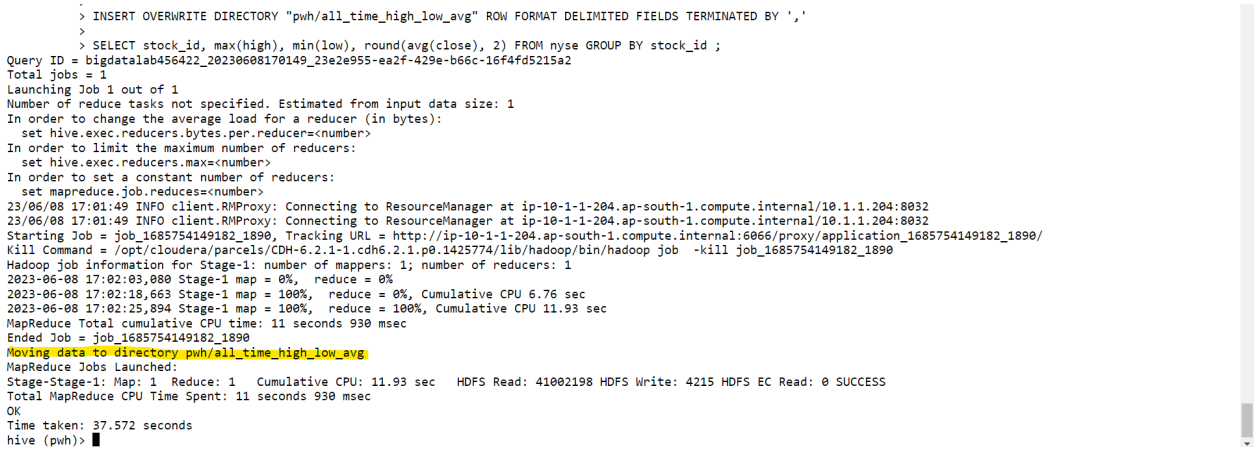
```
2023-05-24 07:12:14,119 Stage-1 map = 100%,  reduce = 100%,
Cumulative CPU 8.52 sec
MapReduce Total cumulative CPU time: 8 seconds 520 msec
Ended Job = job_1684866872278_0374
Moving data to directory hive/result1
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.52 sec   HDFS
Read: 41000225 HDFS Write: 2918 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 520 msec
OK
Time taken: 118.308 seconds
hive (surya_training)>
```





→ # save AllTimeHigh, AllTimeLow, AverageClose for each stock in file
INSERT OVERWRITE DIRECTORY ‘hive/result2’ ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ SELECT stock_id, max(high), min(low), round(avg(close), 2) FROM nyse GROUP BY stock_id;

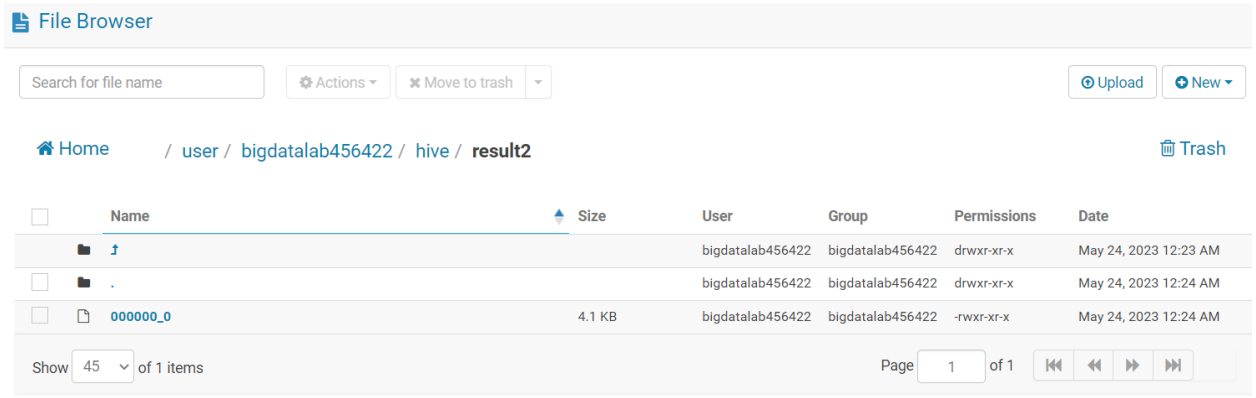
```
hive (surya_training)> INSERT OVERWRITE DIRECTORY 'hive/result2' ROW
FORMAT DELIMITED FIELDS TERMINATED BY ',' SELECT stock_id, max(high),
min(low), round(avg(close), 2
) FROM nyse GROUP BY stock_id;
```

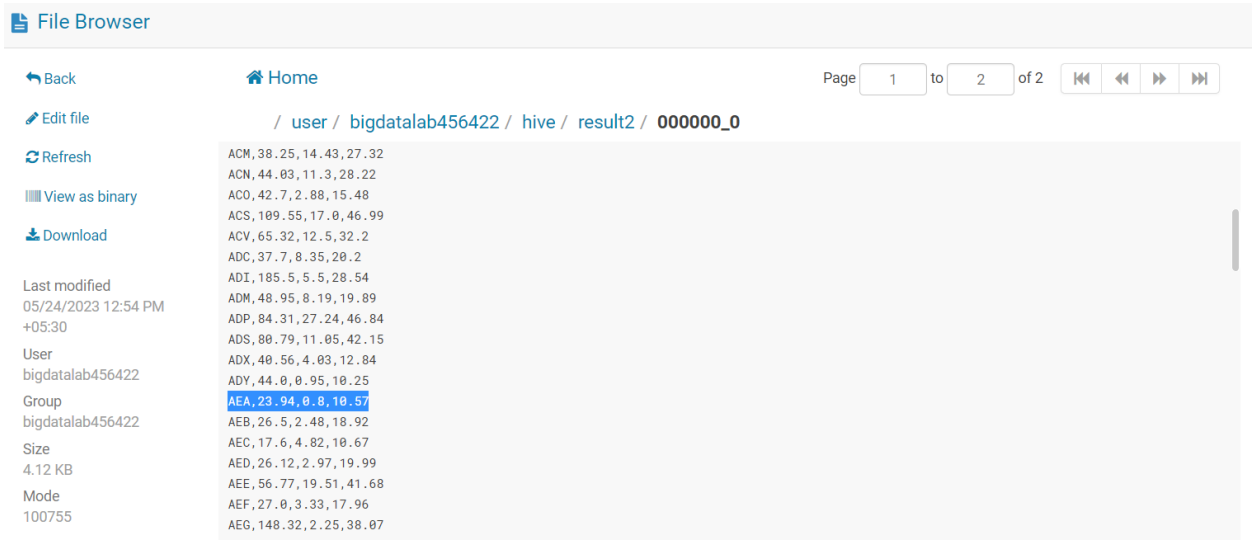


```
        > INSERT OVERWRITE DIRECTORY "pwh/all_time_high_low_avg" ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
        >
        > SELECT stock_id, max(high), min(low), round(avg(close), 2) FROM nyse GROUP BY stock_id ;
Query ID = bigdatalab456422_20230608170149_23e2e955-ea2f-429e-b66c-16f4fd5215a2
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 17:01:49 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 17:01:49 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1890, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1890/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1890
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 17:02:03,080 Stage-1 map = 0%,  reduce = 0%
2023-06-08 17:02:18,663 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.76 sec
2023-06-08 17:02:25,894 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.93 sec
MapReduce Total cumulative CPU time: 11 seconds 930 msec
Ended Job = job_1685754149182_1890
Moving data to directory pwh/all_time_high_low_avg
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.93 sec   HDFS Read: 41002198 HDFS Write: 4215 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 930 msec
OK
Time taken: 37.572 seconds
hive (pwh)>
```

```
Query ID =
bigdatalab456422_20230524072336_7a5d4b20-1939-4828-8699-8a49ed1fef3f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size:
1
In order to change the average load for a reducer (in bytes):
 set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
 set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
 set mapreduce.job.reduces=<number>
23/05/24 07:23:36 INFO client.RMProxy: Connecting to ResourceManager
at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/24 07:23:36 INFO client.RMProxy: Connecting to ResourceManager
at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0396, Tracking URL =
http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/applicati
on_1684866872278_0396/
Kill Command =
/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/
hadoop job  -kill job_1684866872278_0396
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
```

```
2023-05-24 07:23:51,461 Stage-1 map = 0%,  reduce = 0%
2023-05-24 07:24:07,120 Stage-1 map = 100%,  reduce = 0%, Cumulative
CPU 8.59 sec
2023-05-24 07:24:16,561 Stage-1 map = 100%,  reduce = 100%,
Cumulative CPU 13.89 sec
MapReduce Total cumulative CPU time: 13 seconds 890 msec
Ended Job = job_1684866872278_0396
Moving data to directory hive/result2
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 13.89 sec   HDFS
Read: 41002190 HDFS Write: 4215 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 890 msec
OK
Time taken: 41.592 seconds
hive (surya_training)>
```





→ # save total volume of each stock in another table

CREATE TABLE stkvol AS SELECT stock_id, sum(volume) AS total_vol FROM nyse GROUP BY stock_id;

```
hive (surya_training)> SHOW TABLES ;
OK
nyse
Time taken: 0.034 seconds, Fetched: 1 row(s)

hive (surya_training)> CREATE TABLE stkvol AS SELECT stock_id,
sum(volume) AS total_vol FROM nyse GROUP BY stock_id;
```

```
        > CREATE TABLE stkvol AS
        >
        > SELECT stock_id, SUM(volume) FROM nyse GROUP BY stock_id;
Query ID = bigdatalab456422_20230608171134_45a85c52-728f-43f6-a663-c75f79e95d3b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/06/08 17:11:34 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/08 17:11:34 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_1901, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_1901/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_1901
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-06-08 17:11:43,708 Stage-1 map = 0%,  reduce = 0%
2023-06-08 17:11:52,892 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.66 sec
2023-06-08 17:12:00,032 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.13 sec
MapReduce Total cumulative CPU time: 6 seconds 130 msec
Ended Job = job_1685754149182_1901
Moving data to directory hdfs://nameservice1/user/hive/warehouse/pwh.db/stkvol
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.13 sec   HDFS Read: 41000278 HDFS Write: 2987 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 130 msec
OK
Time taken: 28.299 seconds
hive (pwh)>
```

```
Query ID =
bigdatalab456422_20230524073028_d3a92865-6298-4b0b-8208-c8a9bd75ac9d
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size:
1
In order to change the average load for a reducer (in bytes):
 set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
 set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
 set mapreduce.job.reduces=<number>
23/05/24 07:30:28 INFO client.RMProxy: Connecting to ResourceManager
at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/05/24 07:30:28 INFO client.RMProxy: Connecting to ResourceManager
at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1684866872278_0397, Tracking URL =
http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/applicati
on_1684866872278_0397/
Kill Command =
/opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/
hadoop job  -kill job_1684866872278_0397
```

```
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2023-05-24 07:30:37,678 Stage-1 map = 0%,  reduce = 0%
2023-05-24 07:30:47,000 Stage-1 map = 100%,  reduce = 0%, Cumulative
CPU 4.11 sec
2023-05-24 07:30:54,291 Stage-1 map = 100%,  reduce = 100%,
Cumulative CPU 7.21 sec
MapReduce Total cumulative CPU time: 7 seconds 210 msec
Ended Job = job_1684866872278_0397
Moving data to directory
hdfs://nameservice1/user/hive/warehouse/surya_training.db/stkvol
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.21 sec   HDFS
Read: 41000385 HDFS Write: 2998 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 210 msec
OK
Time taken: 28.605 seconds

hive (surya_training)> SHOW TABLES ;
OK
nyse
stkvol
Time taken: 0.037 seconds, Fetched: 2 row(s)

hive (surya_training)> SELECT * FROM stkvol LIMIT 10;
OK
AA      42061448400
AAI     5246821400
AAN     817567400
AAP     2802701500
AAR     49882000
AAV     834246600
AB      1125446300
ABA     11686500
ABB     4532301800
ABC     11439581700
Time taken: 0.081 seconds, Fetched: 10 row(s)
hive (surya_training)>
```