Đại học Khoa học Tự nhiên – ĐHQG TP.HCM

----------□□----------

Course: DỮ LIỆU LỚN

Lab 1: A Gentle Introduction to Hadoop

Class: 21KHMT1

Teacher:

Mrs. Nguyễn Ngọc Thảo

Mr. Bùi Huỳnh Trung Nam

Mr. Đỗ Trọng Lễ

Student information:

21127456 – Võ Cao Trí

21127608 – Trần Trung Hiếu

21127668 – Đinh Quang Phong

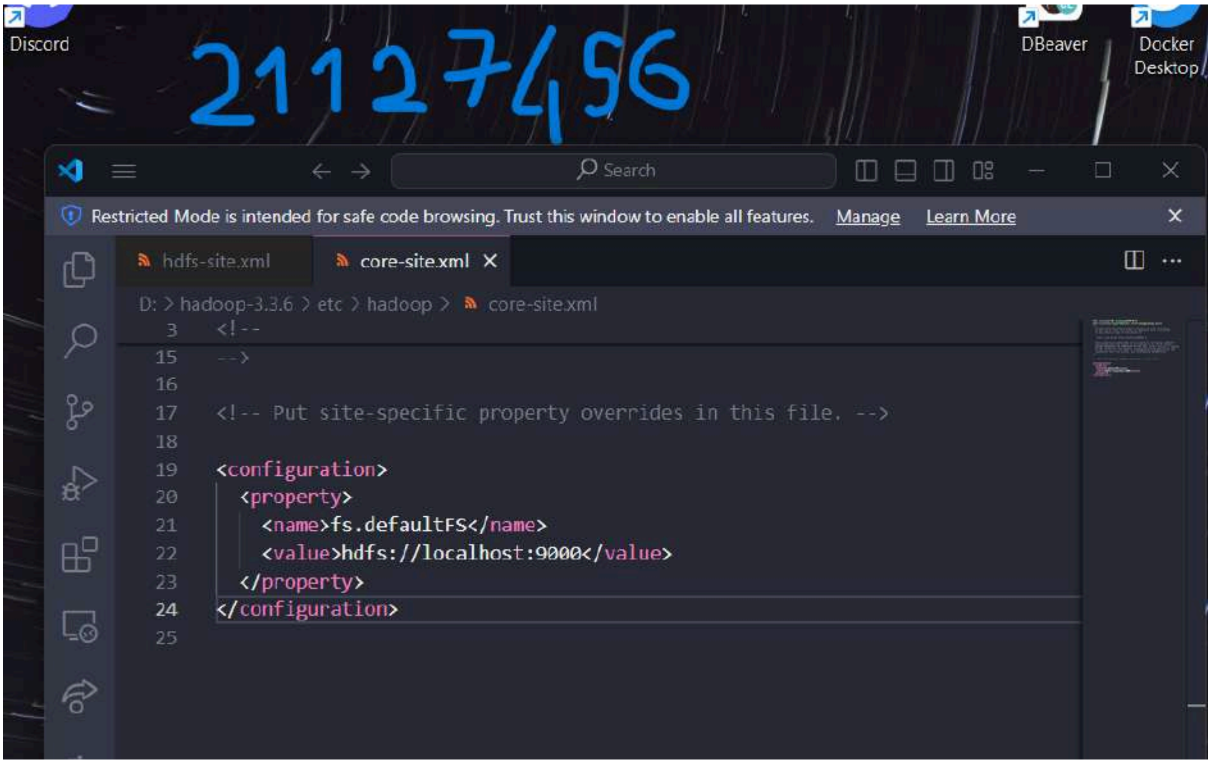
Assign task:

NO	Name	Task	Completion level
1	Tri, Hieu, Phong	Setting up Single-node Hadoop Cluster	100%
2	Hieu, Tri	Paper Reading	100%
3	Phong	Running a warm-up problem: Word Count	100%
4	Hieu, Tri, Phong	Write report	100%

1. Setting up SNC - Single Node Cluster

a. The result of Setting up SNC:

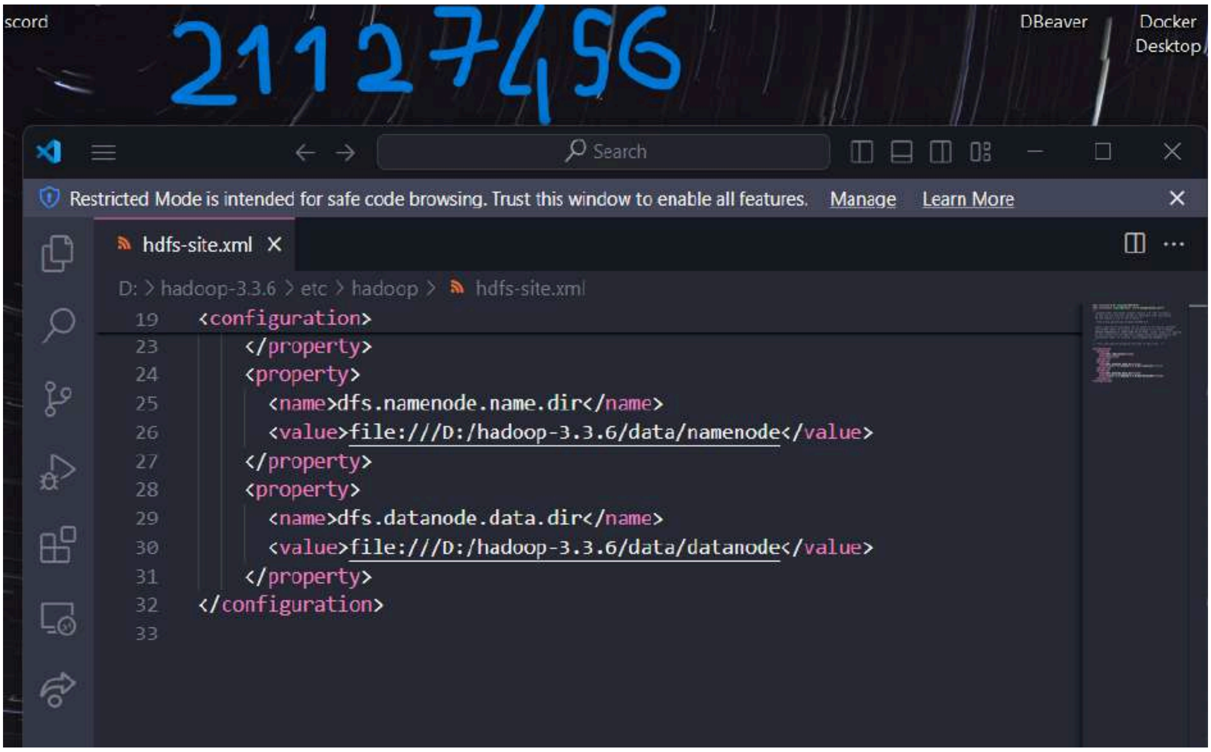
- 21127456
  - Set up core-site.xmr



Restricted Mode is intended for safe code browsing. Trust this window to enable all features.

```
<!--
-->

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

<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

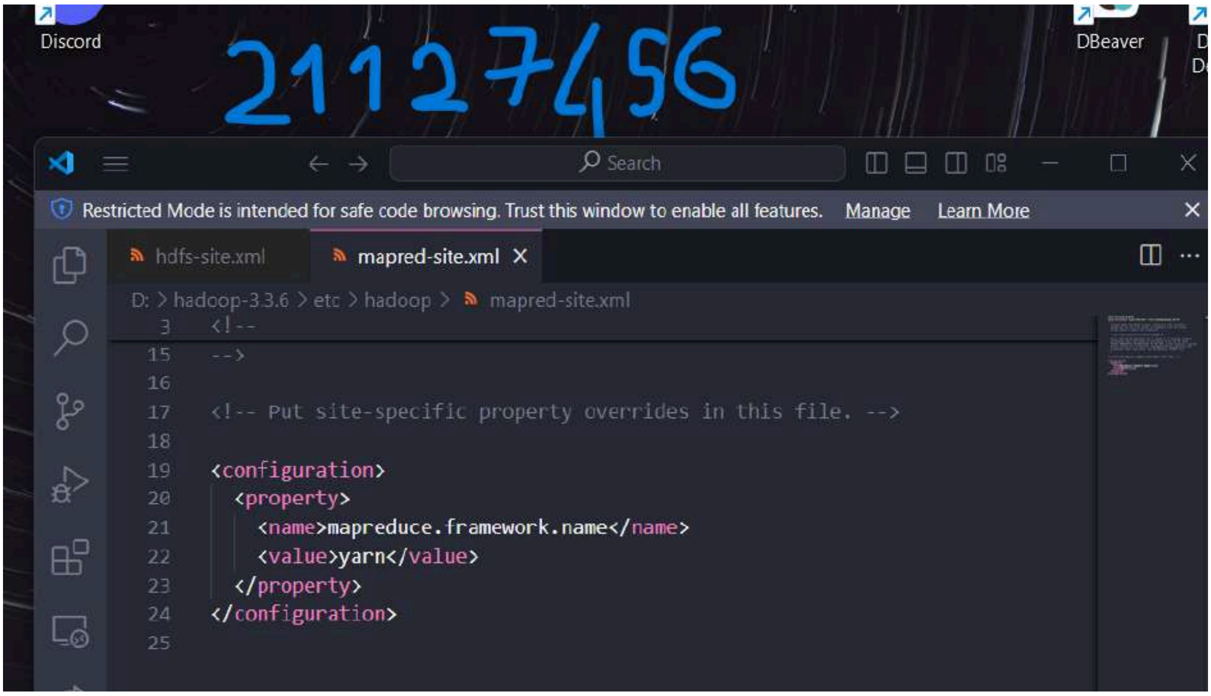
  - Set up hdfs-site.xml



Restricted Mode is intended for safe code browsing. Trust this window to enable all features.

```
<configuration>
    </property>
    <property>
      <name>dfs.namenode.name.dir</name>
      <value>file:///D:/hadoop-3.3.6/data/namenode</value>
    </property>
    <property>
      <name>dfs.datanode.data.dir</name>
      <value>file:///D:/hadoop-3.3.6/data/datanode</value>
    </property>
</configuration>
```

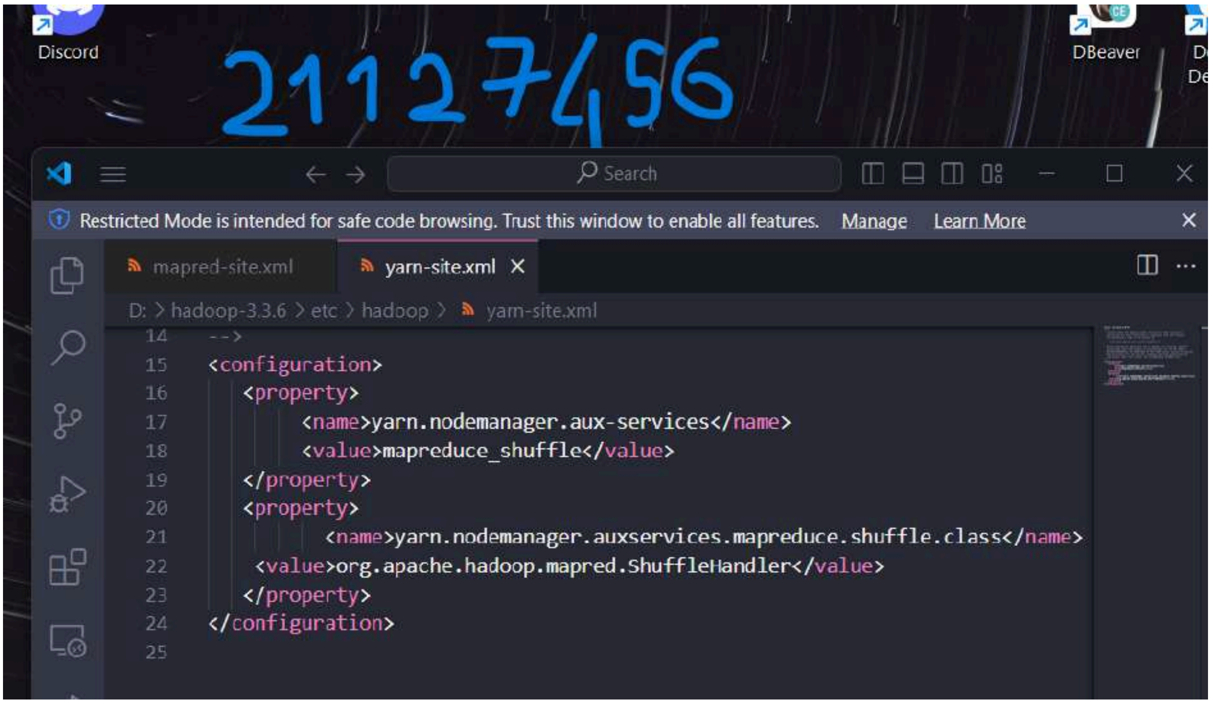
- Set up mapred-site.xml



Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More

```
<!--
-->

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

<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

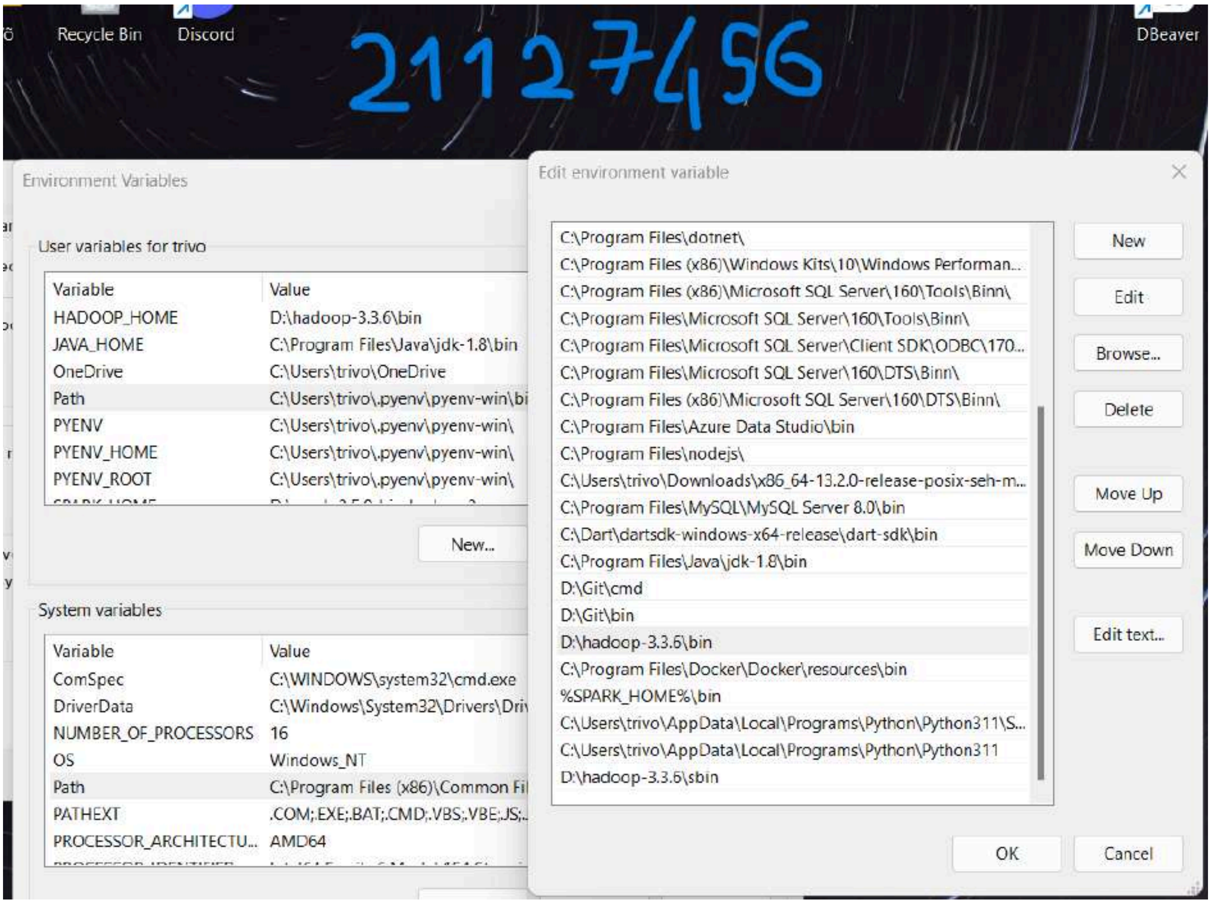
- Set up yarn-site.xml



Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More

```
-->
<configuration>
   <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
   </property>
   <property>
          <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
   </property>
</configuration>
```

- Set up in Environment Variable



User variables for trivo

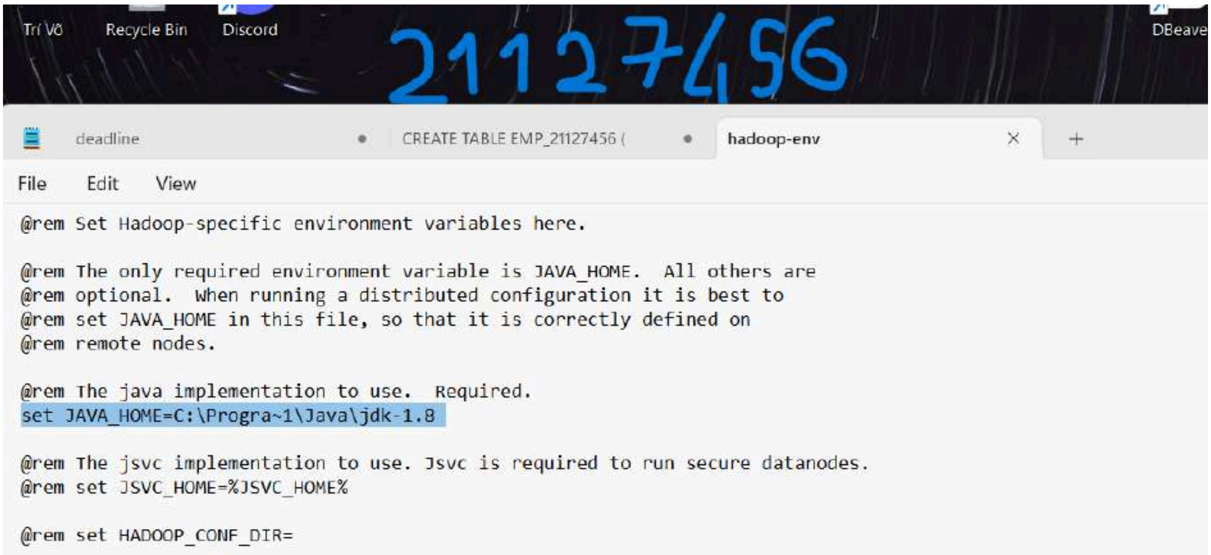
Variable	Value
HADOOP_HOME	D:\hadoop-3.3.6\bin
JAVA_HOME	C:\Program Files\Java\jdk-1.8\bin
OneDrive	C:\Users\trivo\OneDrive
Path	C:\Users\trivo\.pyenv\pyenv-win\bi
PYENV	C:\Users\trivo\.pyenv\pyenv-win\
PYENV_HOME	C:\Users\trivo\.pyenv\pyenv-win\
PYENV_ROOT	C:\Users\trivo\.pyenv\pyenv-win\

System variables

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\Dri
NUMBER_OF_PROCESSORS	16
OS	Windows_NT
Path	C:\Program Files (x86)\Common Fi
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;...
PROCESSOR_ARCHITECTU...	AMD64

```
C:\Program Files\dotnet\
C:\Program Files (x86)\Windows Kits\10\Windows Performan...
C:\Program Files (x86)\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\170...
C:\Program Files\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files\Azure Data Studio\bin
C:\Program Files\nodejs\
C:\Users\trivo\Downloads\x86_64-13.2.0-release-posix-seh-m...
C:\Program Files\MySQL\MySQL Server 8.0\bin
C:\Dart\dartsdk-windows-x64-release\dart-sdk\bin
C:\Program Files\Java\jdk-1.8\bin
D:\Git\cmd
D:\Git\bin
D:\hadoop-3.3.6\bin
C:\Program Files\Docker\Docker\resources\bin
%SPARK_HOME%\bin
C:\Users\trivo\AppData\Local\Programs\Python\Python311\S...
C:\Users\trivo\AppData\Local\Programs\Python\Python311
D:\hadoop-3.3.6\sbin
```

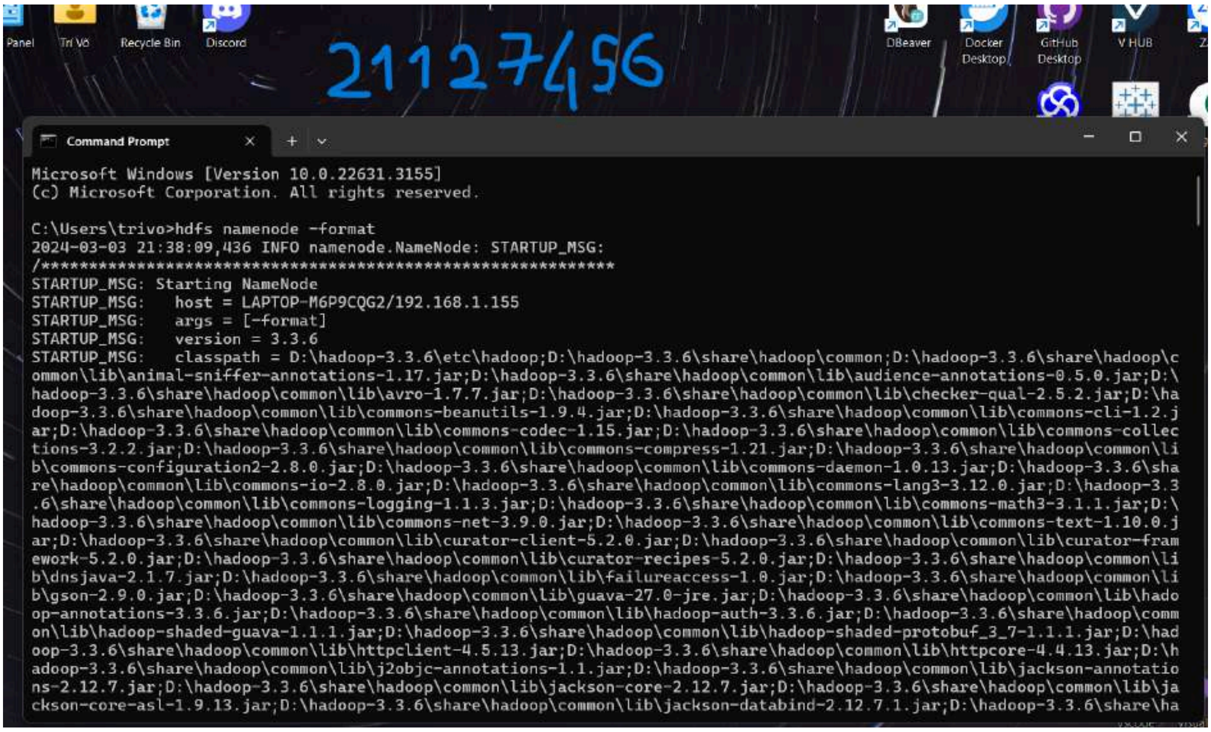
- Set up JAVA_HOME of hadoop-env.cmd



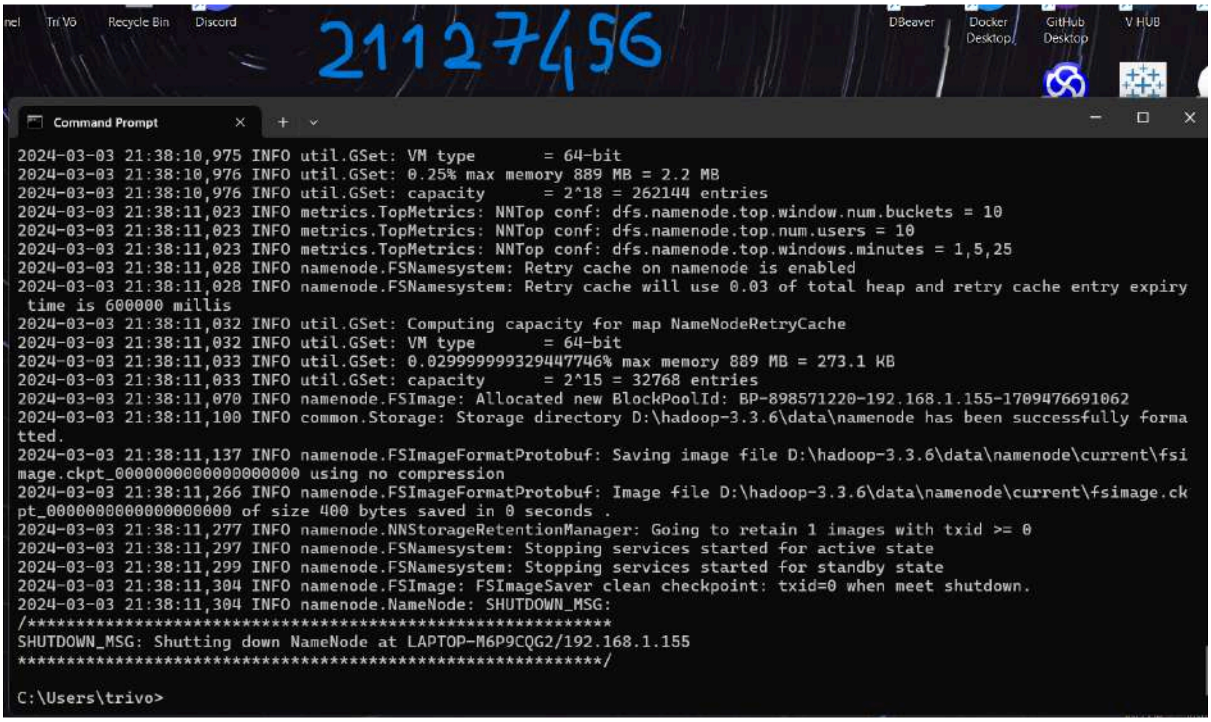
```
@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA_HOME.  All others are
@rem optional.  When running a distributed configuration it is best to
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.

@rem The java implementation to use.  Required.
set JAVA_HOME=C:\Progra~1\Java\jdk-1.8

@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%

@rem set HADOOP_CONF_DIR=
```

Result:

- HDFS namenode format screen

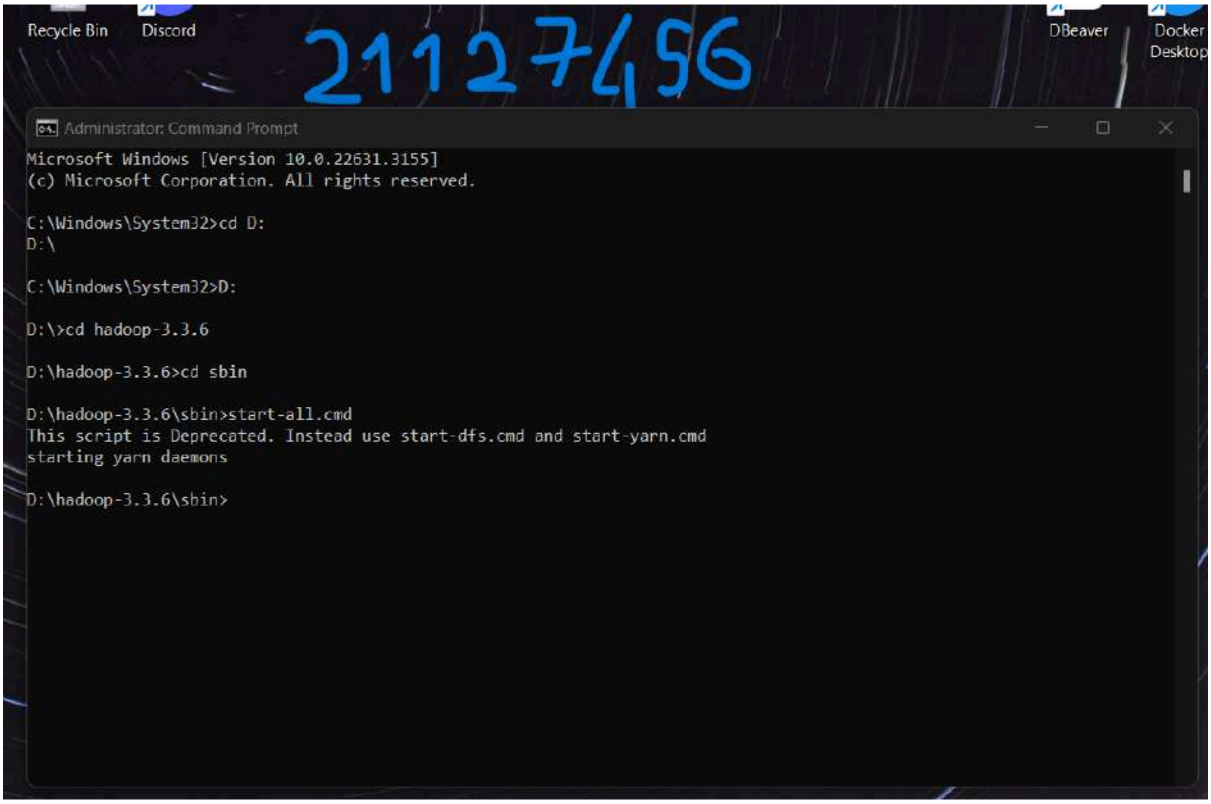


```
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Users\trivo>hdfs namenode -format
2024-03-03 21:38:09,436 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = LAPTOP-M6P9CQG2/192.168.1.155
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = D:\hadoop-3.3.6\etc\hadoop;D:\hadoop-3.3.6\share\hadoop\common;D:\hadoop-3.3.6\share\hadoop\c
ommon\lib\animal-sniffer-annotations-1.17.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\audience-annotations-0.5.0.jar;D:\
hadoop-3.3.6\share\hadoop\common\lib\avro-1.7.7.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\checker-qual-2.5.2.jar;D:\ha
doop-3.3.6\share\hadoop\common\lib\commons-beanutils-1.9.4.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-cli-1.2.j
ar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-codec-1.15.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-collec
tions-3.2.2.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-compress-1.21.jar;D:\hadoop-3.3.6\share\hadoop\common\li
b\commons-configuration2-2.8.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-daemon-1.0.13.jar;D:\hadoop-3.3.6\sha
re\hadoop\common\lib\commons-io-2.8.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-lang3-3.12.0.jar;D:\hadoop-3.3
.6\share\hadoop\common\lib\commons-logging-1.1.3.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-math3-3.1.1.jar;D:\
hadoop-3.3.6\share\hadoop\common\lib\commons-net-3.9.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-text-1.10.0.j
ar;D:\hadoop-3.3.6\share\hadoop\common\lib\curator-client-5.2.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\curator-fram
ework-5.2.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\curator-recipes-5.2.0.jar;D:\hadoop-3.3.6\share\hadoop\common\li
b\dnsjava-2.1.7.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\failureaccess-1.0.jar;D:\hadoop-3.3.6\share\hadoop\common\li
b\gson-2.9.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\guava-27.0-jre.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\hado
op-annotations-3.3.6.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-auth-3.3.6.jar;D:\hadoop-3.3.6\share\hadoop\comm
on\lib\hadoop-shaded-guava-1.1.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-shaded-protobuf_3_7-1.1.1.jar;D:\had
oop-3.3.6\share\hadoop\common\lib\httpclient-4.5.13.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\httpcore-4.4.13.jar;D:\h
adoop-3.3.6\share\hadoop\common\lib\j2objc-annotations-1.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jackson-annotatio
ns-2.12.7.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jackson-core-2.12.7.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\ja
ckson-core-asl-1.9.13.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jackson-databind-2.12.7.1.jar;D:\hadoop-3.3.6\share\ha
```



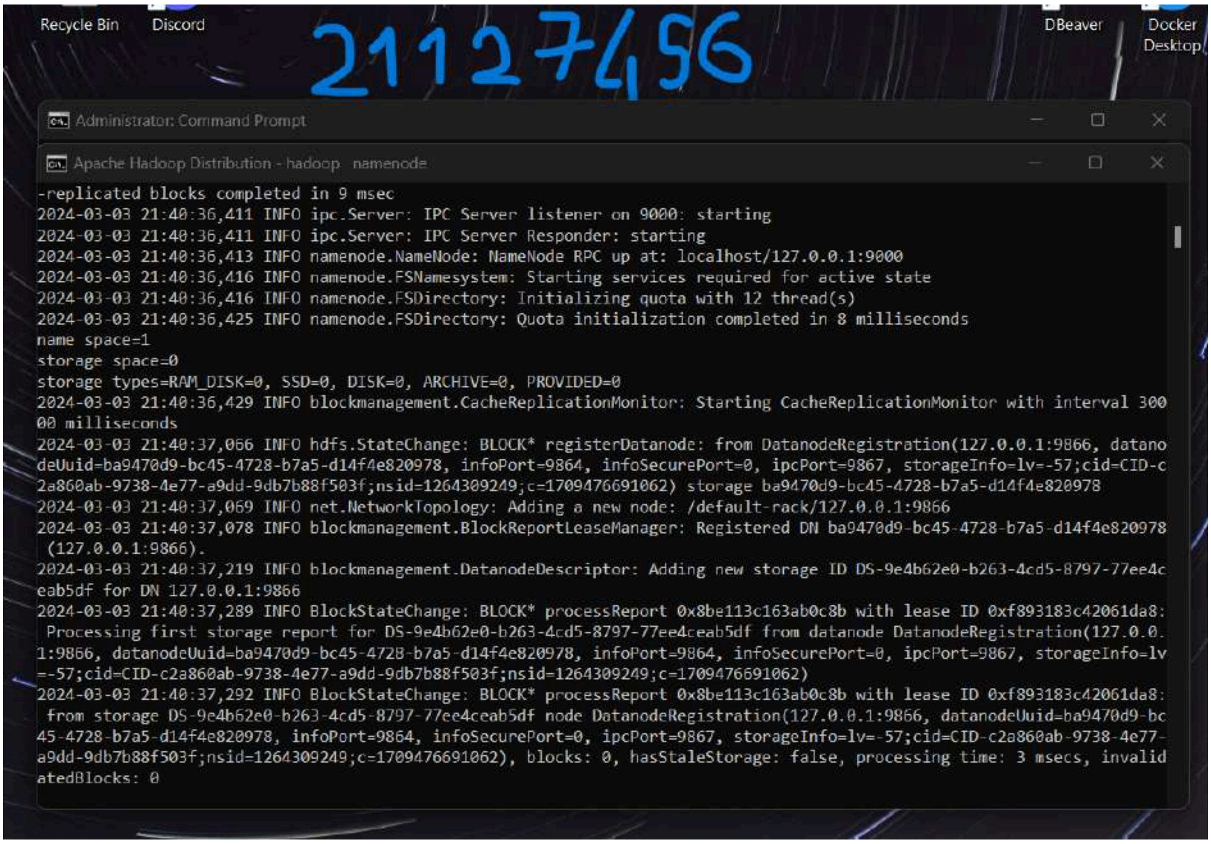
```
2024-03-03 21:38:10,975 INFO util.GSet: VM type       = 64-bit
2024-03-03 21:38:10,976 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
2024-03-03 21:38:10,976 INFO util.GSet: capacity      = 2^18 = 262144 entries
2024-03-03 21:38:11,023 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2024-03-03 21:38:11,023 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-03-03 21:38:11,023 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-03-03 21:38:11,028 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-03-03 21:38:11,028 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry
 time is 600000 millis
2024-03-03 21:38:11,032 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-03-03 21:38:11,032 INFO util.GSet: VM type       = 64-bit
2024-03-03 21:38:11,033 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
2024-03-03 21:38:11,033 INFO util.GSet: capacity      = 2^15 = 32768 entries
2024-03-03 21:38:11,070 INFO namenode.FSImage: Allocated new BlockPoolId: BP-898571220-192.168.1.155-1709476691062
2024-03-03 21:38:11,100 INFO common.Storage: Storage directory D:\hadoop-3.3.6\data\namenode has been successfully forma
tted.
2024-03-03 21:38:11,137 INFO namenode.FSImageFormatProtobuf: Saving image file D:\hadoop-3.3.6\data\namenode\current\fsi
mage.ckpt_0000000000000000000 using no compression
2024-03-03 21:38:11,266 INFO namenode.FSImageFormatProtobuf: Image file D:\hadoop-3.3.6\data\namenode\current\fsimage.ck
pt_0000000000000000000 of size 400 bytes saved in 0 seconds .
2024-03-03 21:38:11,277 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-03-03 21:38:11,297 INFO namenode.FSNamesystem: Stopping services started for active state
2024-03-03 21:38:11,299 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-03-03 21:38:11,304 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-03-03 21:38:11,304 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at LAPTOP-M6P9CQG2/192.168.1.155
************************************************************/

C:\Users\trivo>
```

- start-all.cmd screen



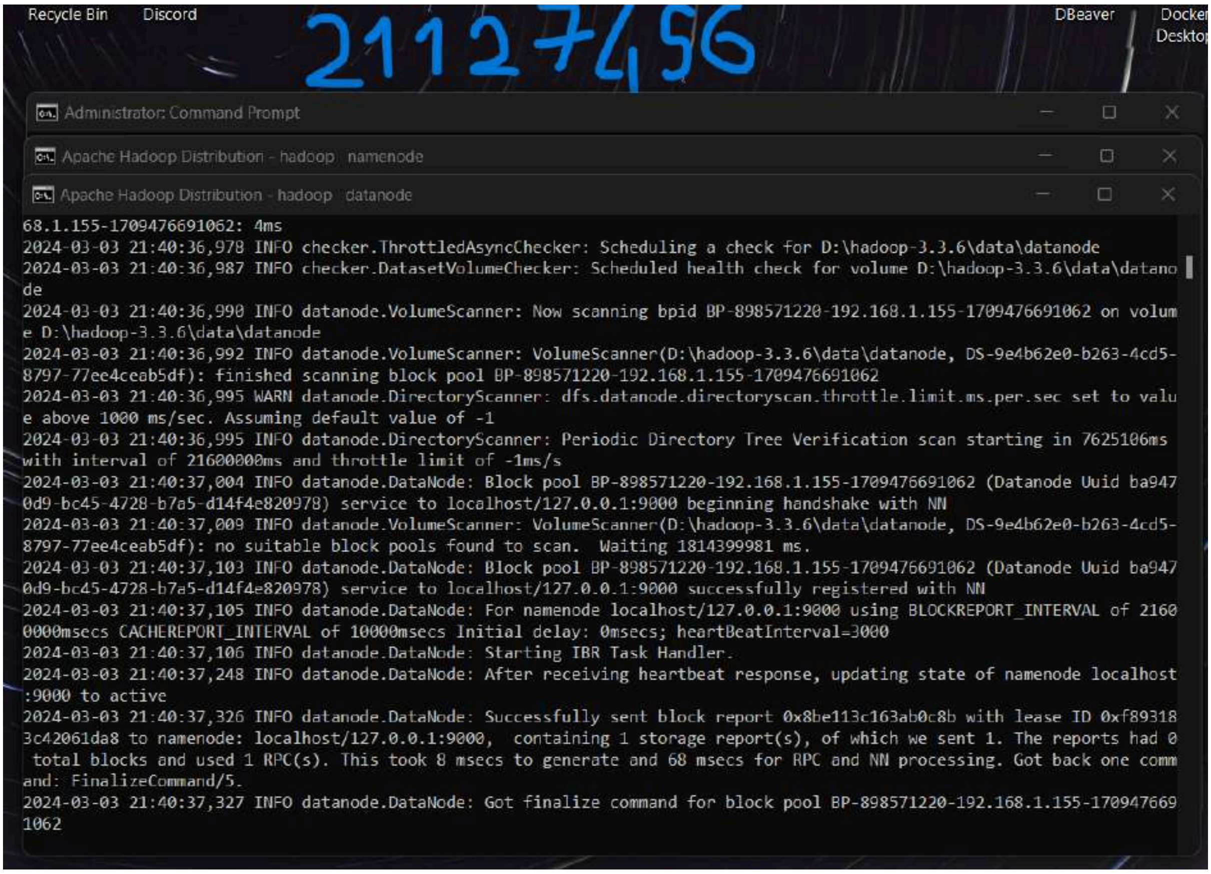
```
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd D:
D:\

C:\Windows\System32>D:

D:\>cd hadoop-3.3.6

D:\hadoop-3.3.6>cd sbin

D:\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop-3.3.6\sbin>
```

- ○ namenode screen


Recycle Bin
Discord
21127456
DBeaver
Docker Desktop
Administrator: Command Prompt
Apache Hadoop Distribution - hadoop namenode

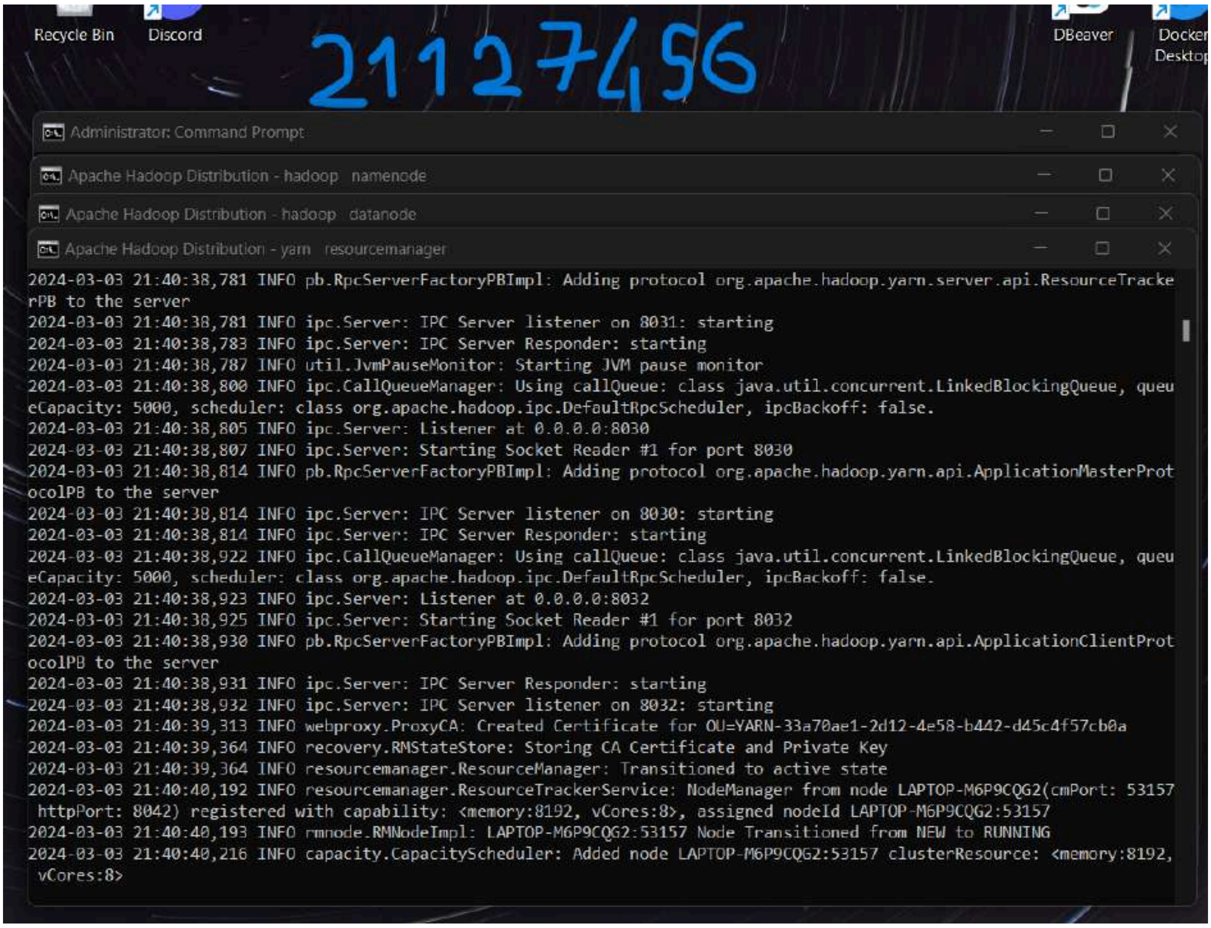
```
-replicated blocks completed in 9 msec
2024-03-03 21:40:36,411 INFO ipc.Server: IPC Server listener on 9000: starting
2024-03-03 21:40:36,411 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:40:36,413 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000
2024-03-03 21:40:36,416 INFO namenode.FSNamesystem: Starting services required for active state
2024-03-03 21:40:36,416 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)
2024-03-03 21:40:36,425 INFO namenode.FSDirectory: Quota initialization completed in 8 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2024-03-03 21:40:36,429 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 300
00 milliseconds
2024-03-03 21:40:37,066 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datano
deUuid=ba9470d9-bc45-4728-b7a5-d14f4e820978, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-c
2a860ab-9738-4e77-a9dd-9db7b88f503f;nsid=1264309249;c=1709476691062) storage ba9470d9-bc45-4728-b7a5-d14f4e820978
2024-03-03 21:40:37,069 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2024-03-03 21:40:37,078 INFO blockmanagement.BlockReportLeaseManager: Registered DN ba9470d9-bc45-4728-b7a5-d14f4e820978
 (127.0.0.1:9866).
2024-03-03 21:40:37,219 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-9e4b62e0-b263-4cd5-8797-77ee4c
eab5df for DN 127.0.0.1:9866
2024-03-03 21:40:37,289 INFO BlockStateChange: BLOCK* processReport 0x8be113c163ab0c8b with lease ID 0xf893183c42061da8:
 Processing first storage report for DS-9e4b62e0-b263-4cd5-8797-77ee4ceab5df from datanode DatanodeRegistration(127.0.0.
1:9866, datanodeUuid=ba9470d9-bc45-4728-b7a5-d14f4e820978, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv
=-57;cid=CID-c2a860ab-9738-4e77-a9dd-9db7b88f503f;nsid=1264309249;c=1709476691062)
2024-03-03 21:40:37,292 INFO BlockStateChange: BLOCK* processReport 0x8be113c163ab0c8b with lease ID 0xf893183c42061da8:
 from storage DS-9e4b62e0-b263-4cd5-8797-77ee4ceab5df node DatanodeRegistration(127.0.0.1:9866, datanodeUuid=ba9470d9-bc
45-4728-b7a5-d14f4e820978, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-c2a860ab-9738-4e77-
a9dd-9db7b88f503f;nsid=1264309249;c=1709476691062), blocks: 0, hasStaleStorage: false, processing time: 3 msecs, invalid
atedBlocks: 0
```

- ○ datanode screen


Recycle Bin
Discord
21127456
DBeaver
Docker Desktop
Administrator: Command Prompt
Apache Hadoop Distribution - hadoop namenode
Apache Hadoop Distribution - hadoop datanode

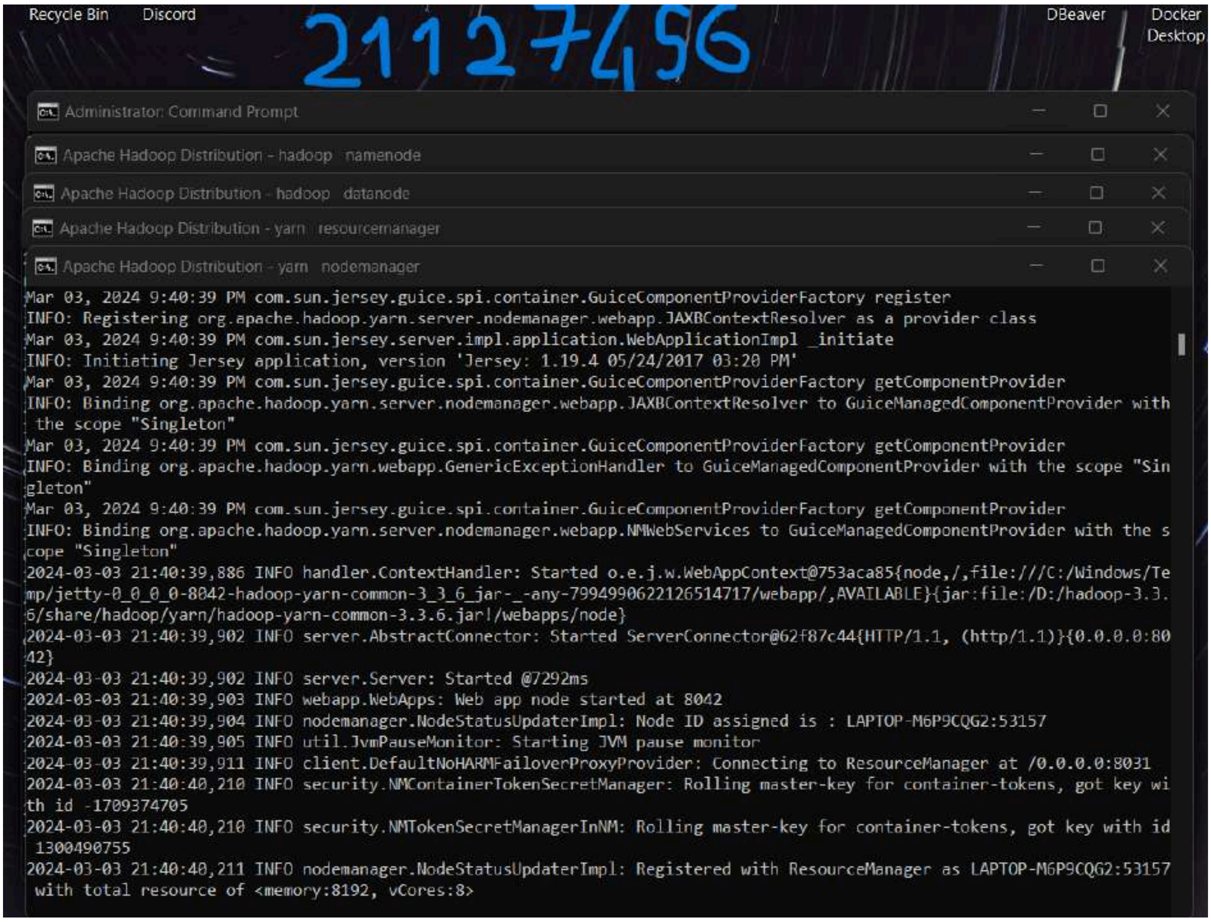
```
68.1.155-1709476691062: 4ms
2024-03-03 21:40:36,978 INFO checker.ThrottledAsyncChecker: Scheduling a check for D:\hadoop-3.3.6\data\datanode
2024-03-03 21:40:36,987 INFO checker.DatasetVolumeChecker: Scheduled health check for volume D:\hadoop-3.3.6\data\datano
de
2024-03-03 21:40:36,990 INFO datanode.VolumeScanner: Now scanning bpid BP-898571220-192.168.1.155-1709476691062 on volum
e D:\hadoop-3.3.6\data\datanode
2024-03-03 21:40:36,992 INFO datanode.VolumeScanner: VolumeScanner(D:\hadoop-3.3.6\data\datanode, DS-9e4b62e0-b263-4cd5-
8797-77ee4ceab5df): finished scanning block pool BP-898571220-192.168.1.155-1709476691062
2024-03-03 21:40:36,995 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to valu
e above 1000 ms/sec. Assuming default value of -1
2024-03-03 21:40:36,995 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 7625106ms
with interval of 21600000ms and throttle limit of -1ms/s
2024-03-03 21:40:37,004 INFO datanode.DataNode: Block pool BP-898571220-192.168.1.155-1709476691062 (Datanode Uuid ba947
0d9-bc45-4728-b7a5-d14f4e820978) service to localhost/127.0.0.1:9000 beginning handshake with NN
2024-03-03 21:40:37,009 INFO datanode.VolumeScanner: VolumeScanner(D:\hadoop-3.3.6\data\datanode, DS-9e4b62e0-b263-4cd5-
8797-77ee4ceab5df): no suitable block pools found to scan.  Waiting 1814399981 ms.
2024-03-03 21:40:37,103 INFO datanode.DataNode: Block pool BP-898571220-192.168.1.155-1709476691062 (Datanode Uuid ba947
0d9-bc45-4728-b7a5-d14f4e820978) service to localhost/127.0.0.1:9000 successfully registered with NN
2024-03-03 21:40:37,105 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 2160
0000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartBeatInterval=3000
2024-03-03 21:40:37,106 INFO datanode.DataNode: Starting IBR Task Handler.
2024-03-03 21:40:37,248 INFO datanode.DataNode: After receiving heartbeat response, updating state of namenode localhost
:9000 to active
2024-03-03 21:40:37,326 INFO datanode.DataNode: Successfully sent block report 0x8be113c163ab0c8b with lease ID 0xf89318
3c42061da8 to namenode: localhost/127.0.0.1:9000,  containing 1 storage report(s), of which we sent 1. The reports had 0
 total blocks and used 1 RPC(s). This took 8 msecs to generate and 68 msecs for RPC and NN processing. Got back one comm
and: FinalizeCommand/5.
2024-03-03 21:40:37,327 INFO datanode.DataNode: Got finalize command for block pool BP-898571220-192.168.1.155-170947669
1062
```

- resource manager screen


Recycle Bin
Discord
21127456
DBeaver
Docker Desktop
Administrator: Command Prompt
Apache Hadoop Distribution - hadoop namenode
Apache Hadoop Distribution - hadoop datanode
Apache Hadoop Distribution - yarn resourcemanager

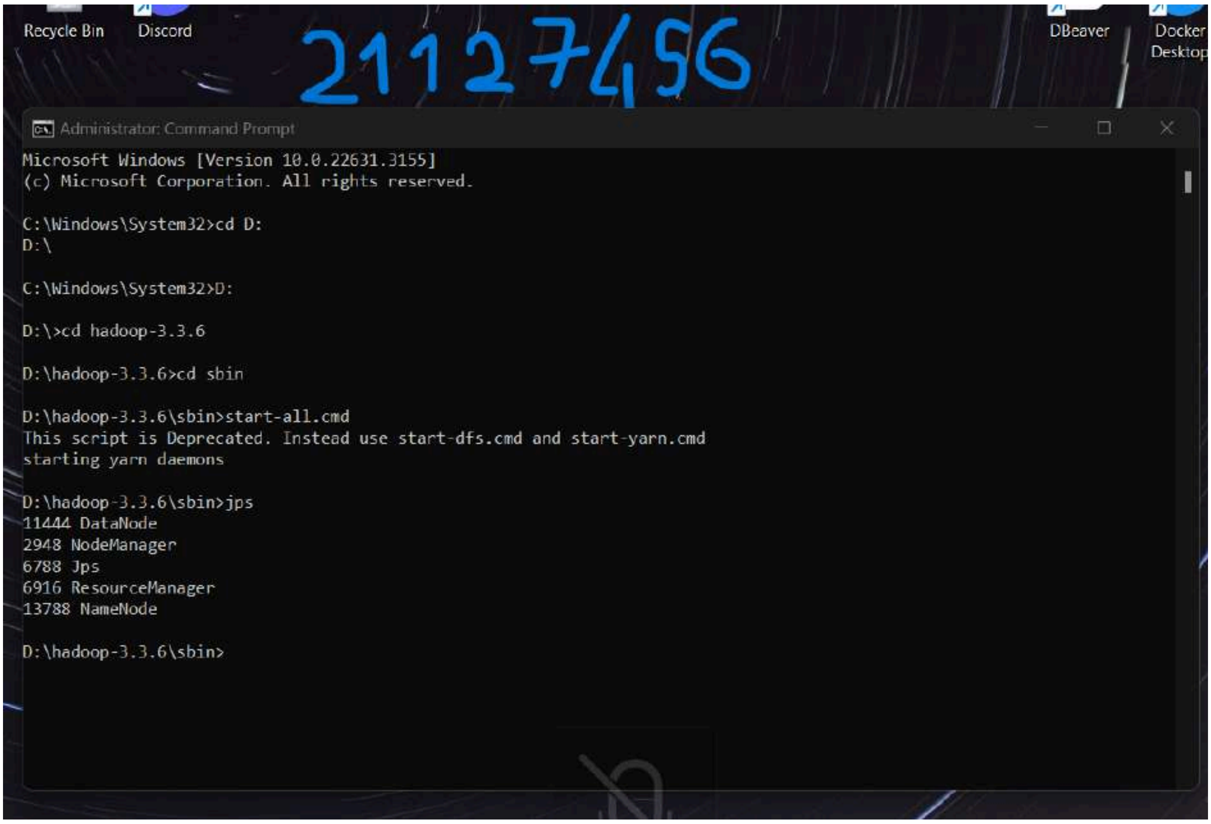
```
2024-03-03 21:40:38,781 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTracke
rPB to the server
2024-03-03 21:40:38,781 INFO ipc.Server: IPC Server listener on 8031: starting
2024-03-03 21:40:38,783 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:40:38,787 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-03-03 21:40:38,800 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-03-03 21:40:38,805 INFO ipc.Server: Listener at 0.0.0.0:8030
2024-03-03 21:40:38,807 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-03-03 21:40:38,814 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProt
ocolPB to the server
2024-03-03 21:40:38,814 INFO ipc.Server: IPC Server listener on 8030: starting
2024-03-03 21:40:38,814 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:40:38,922 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-03-03 21:40:38,923 INFO ipc.Server: Listener at 0.0.0.0:8032
2024-03-03 21:40:38,925 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-03-03 21:40:38,930 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProt
ocolPB to the server
2024-03-03 21:40:38,931 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:40:38,932 INFO ipc.Server: IPC Server listener on 8032: starting
2024-03-03 21:40:39,313 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-33a70ae1-2d12-4e58-b442-d45c4f57cb0a
2024-03-03 21:40:39,364 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2024-03-03 21:40:39,364 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-03-03 21:40:40,192 INFO resourcemanager.ResourceTrackerService: NodeManager from node LAPTOP-M6P9CQG2(cmPort: 53157
 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId LAPTOP-M6P9CQG2:53157
2024-03-03 21:40:40,193 INFO rmnode.RMNodeImpl: LAPTOP-M6P9CQG2:53157 Node Transitioned from NEW to RUNNING
2024-03-03 21:40:40,216 INFO capacity.CapacityScheduler: Added node LAPTOP-M6P9CQG2:53157 clusterResource: <memory:8192,
 vCores:8>
```

- ○ node manager screen


Recycle Bin
Discord
21127456
DBeaver
Docker Desktop
Administrator: Command Prompt
Apache Hadoop Distribution - hadoop namenode
Apache Hadoop Distribution - hadoop datanode
Apache Hadoop Distribution - yarn resourcemanager
Apache Hadoop Distribution - yarn nodemanager

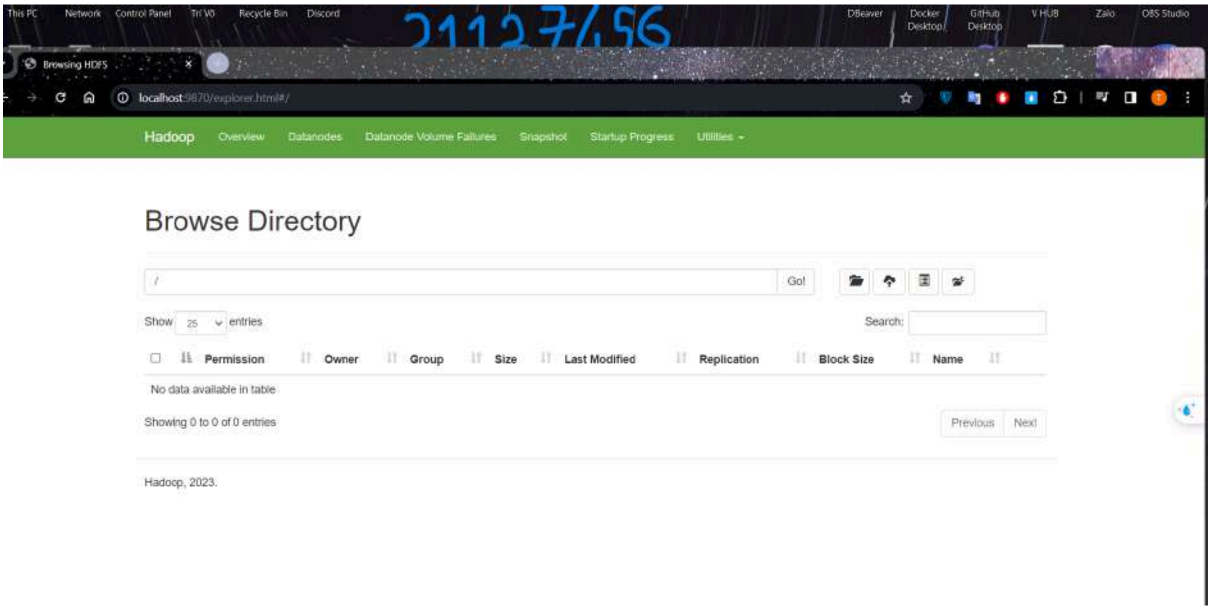
```
Mar 03, 2024 9:40:39 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Mar 03, 2024 9:40:39 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19.4 05/24/2017 03:20 PM'
Mar 03, 2024 9:40:39 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with
 the scope "Singleton"
Mar 03, 2024 9:40:39 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Sin
gleton"
Mar 03, 2024 9:40:39 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the s
cope "Singleton"
2024-03-03 21:40:39,886 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@753aca85{node,/,file:///C:/Windows/Te
mp/jetty-0_0_0_0-8042-hadoop-yarn-common-3_3_6_jar-_-any-7994990622126514717/webapp/,AVAILABLE}{jar:file:/D:/hadoop-3.3.
6/share/hadoop/yarn/hadoop-yarn-common-3.3.6.jar!/webapps/node}
2024-03-03 21:40:39,902 INFO server.AbstractConnector: Started ServerConnector@62f87c44{HTTP/1.1, (http/1.1)}{0.0.0.0:80
42}
2024-03-03 21:40:39,902 INFO server.Server: Started @7292ms
2024-03-03 21:40:39,903 INFO webapp.WebApps: Web app node started at 8042
2024-03-03 21:40:39,904 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : LAPTOP-M6P9CQG2:53157
2024-03-03 21:40:39,905 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-03-03 21:40:39,911 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2024-03-03 21:40:40,210 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key wi
th id -1709374705
2024-03-03 21:40:40,210 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id
 1300490755
2024-03-03 21:40:40,211 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as LAPTOP-M6P9CQG2:53157
 with total resource of <memory:8192, vCores:8>
```

- ○ jps screen



```
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd D:
D:\

C:\Windows\System32>D:

D:\>cd hadoop-3.3.6

D:\hadoop-3.3.6>cd sbin

D:\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop-3.3.6\sbin>jps
11444 DataNode
2948 NodeManager
6788 Jps
6916 ResourceManager
13788 NameNode

D:\hadoop-3.3.6\sbin>
```

- On localhost:9870 of Hadoop File System



Browse Directory

/

Show 25 entries Search:

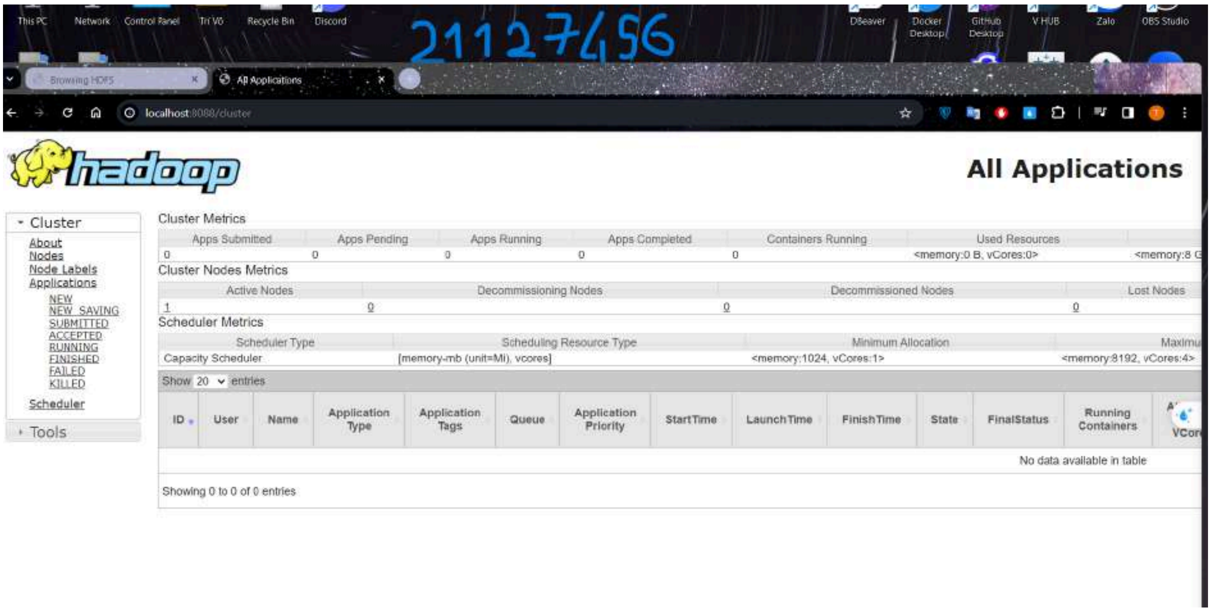
□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
No data available in table								

Showing 0 to 0 of 0 entries

Previous Next

Hadoop, 2023.

- On localhost:8088 is cluster app of hadoop



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:8 G

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

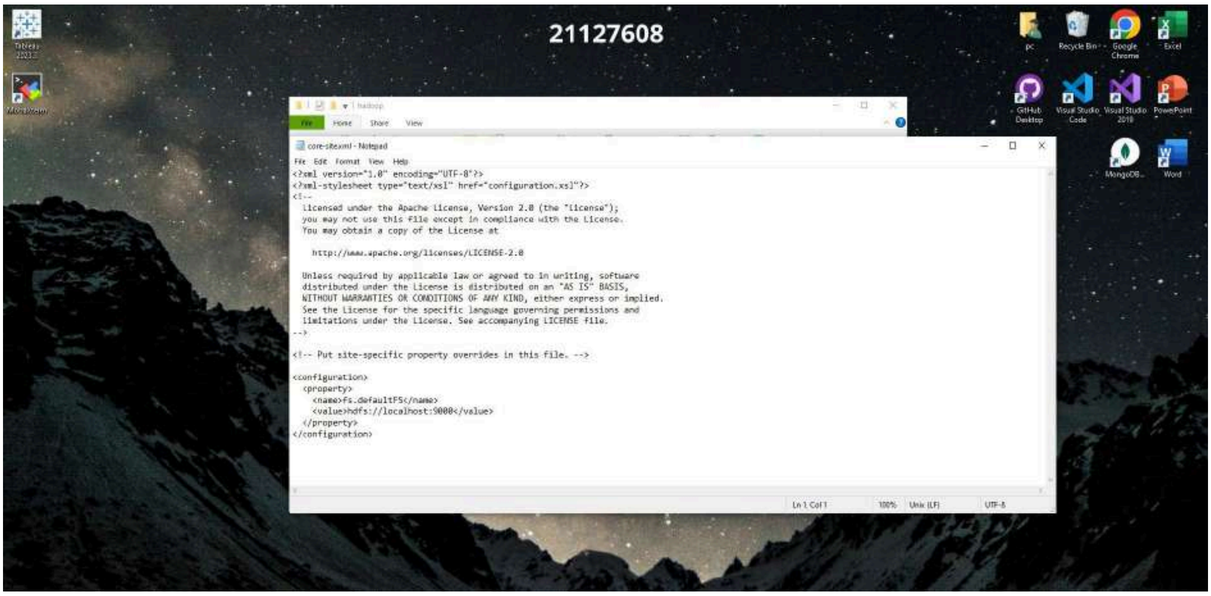
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximu
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	VCor
No data available in table													

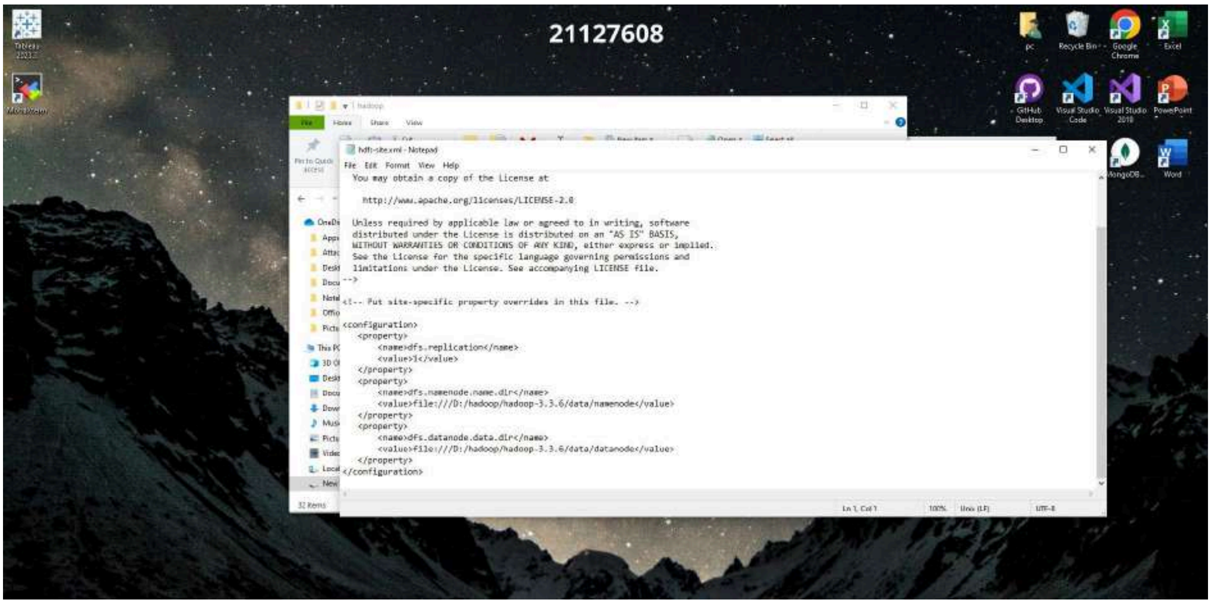
Showing 0 to 0 of 0 entries

- 21127608
  - Set up core-site.xml



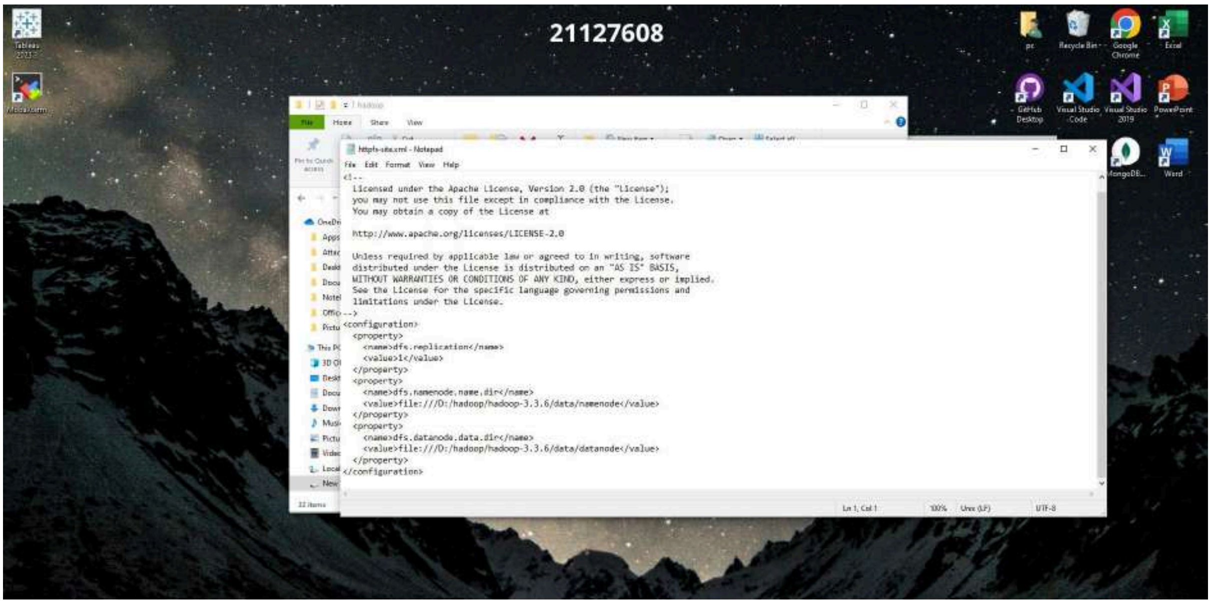
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

- ○ Set up hdfs-site.xml



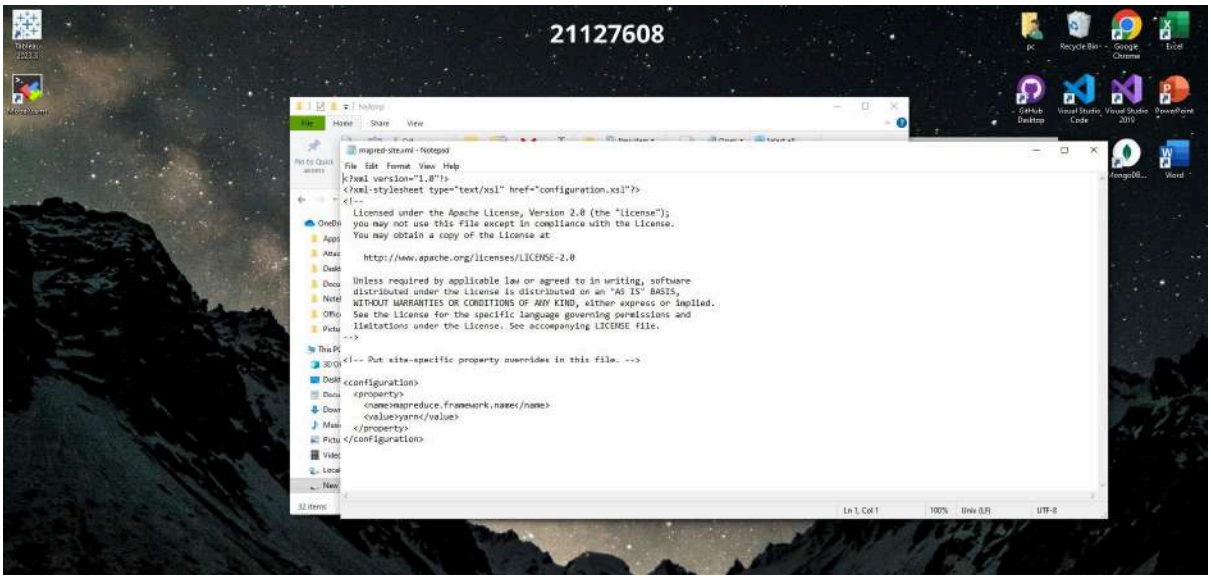
```
  You may obtain a copy of the License at

    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>dfs.replication</name>
       <value>1</value>
   </property>
   <property>
       <name>dfs.namenode.name.dir</name>
       <value>file:///D:/hadoop/hadoop-3.3.6/data/namenode</value>
   </property>
   <property>
       <name>dfs.datanode.data.dir</name>
       <value>file:///D:/hadoop/hadoop-3.3.6/data/datanode</value>
   </property>
</configuration>
```

- ○ Set up httpfs-site.xml



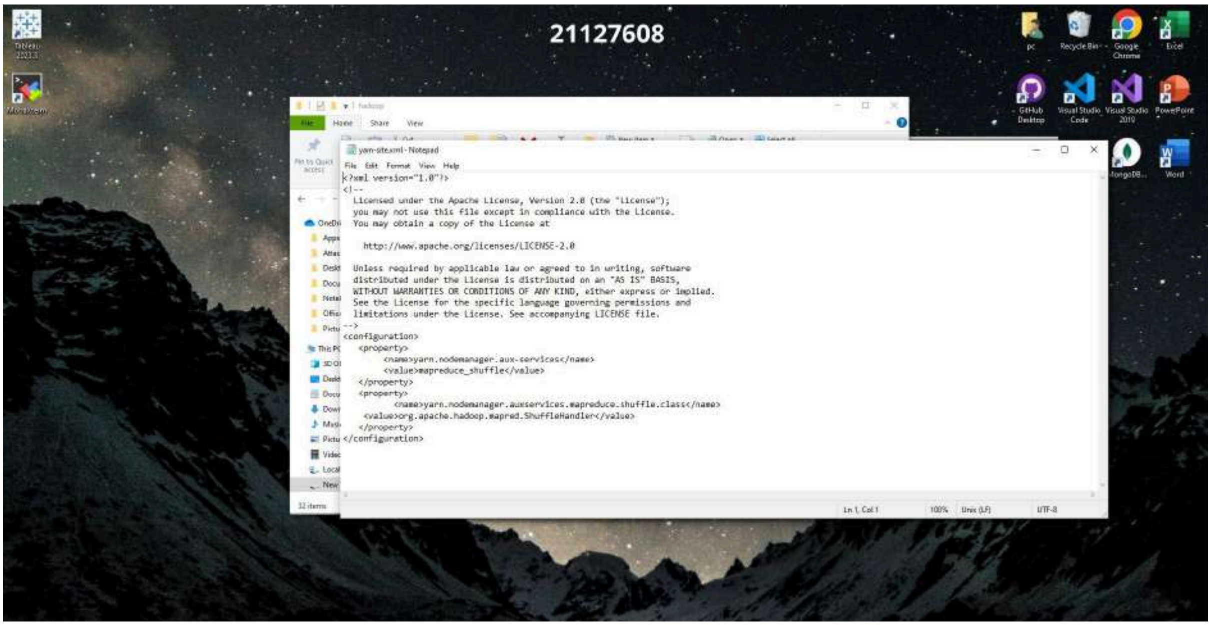
```
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

  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.
-->
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///D:/hadoop/hadoop-3.3.6/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///D:/hadoop/hadoop-3.3.6/data/datanode</value>
  </property>
</configuration>
```

- Set up mapred-site.xml



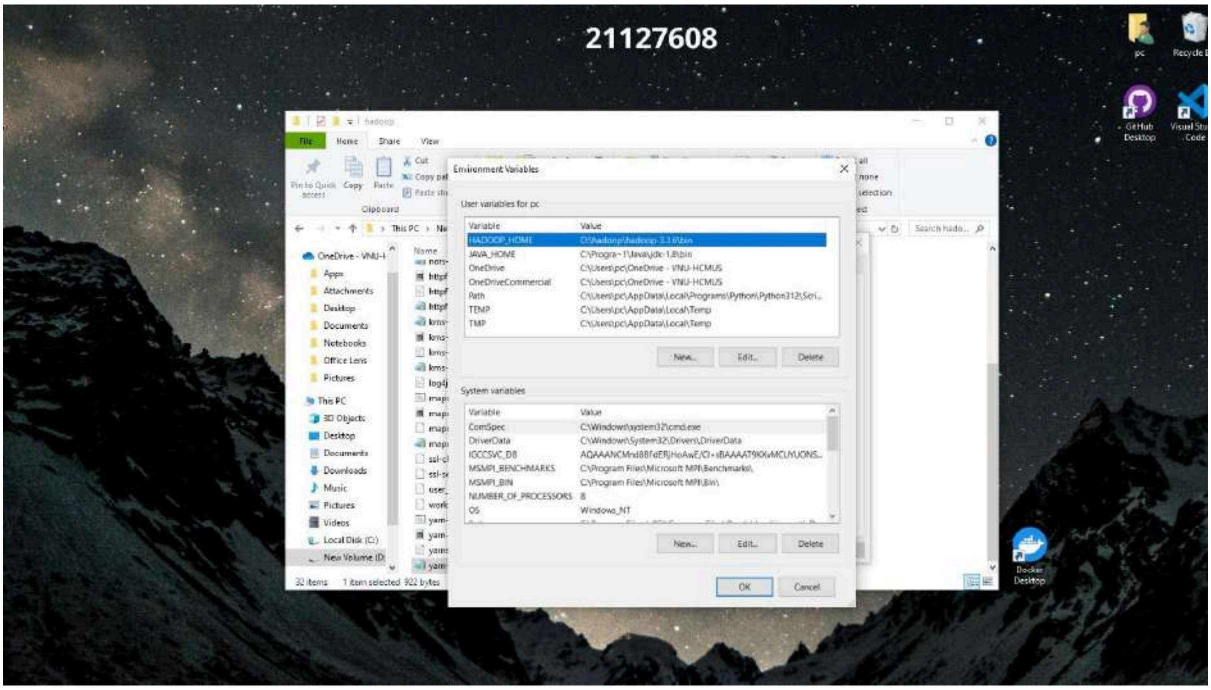
```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

- Set up yarn-site.xml



```
<?xml version="1.0"?>
<!--
  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.
-->
<configuration>
   <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
   </property>
   <property>
          <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
   </property>
</configuration>
```

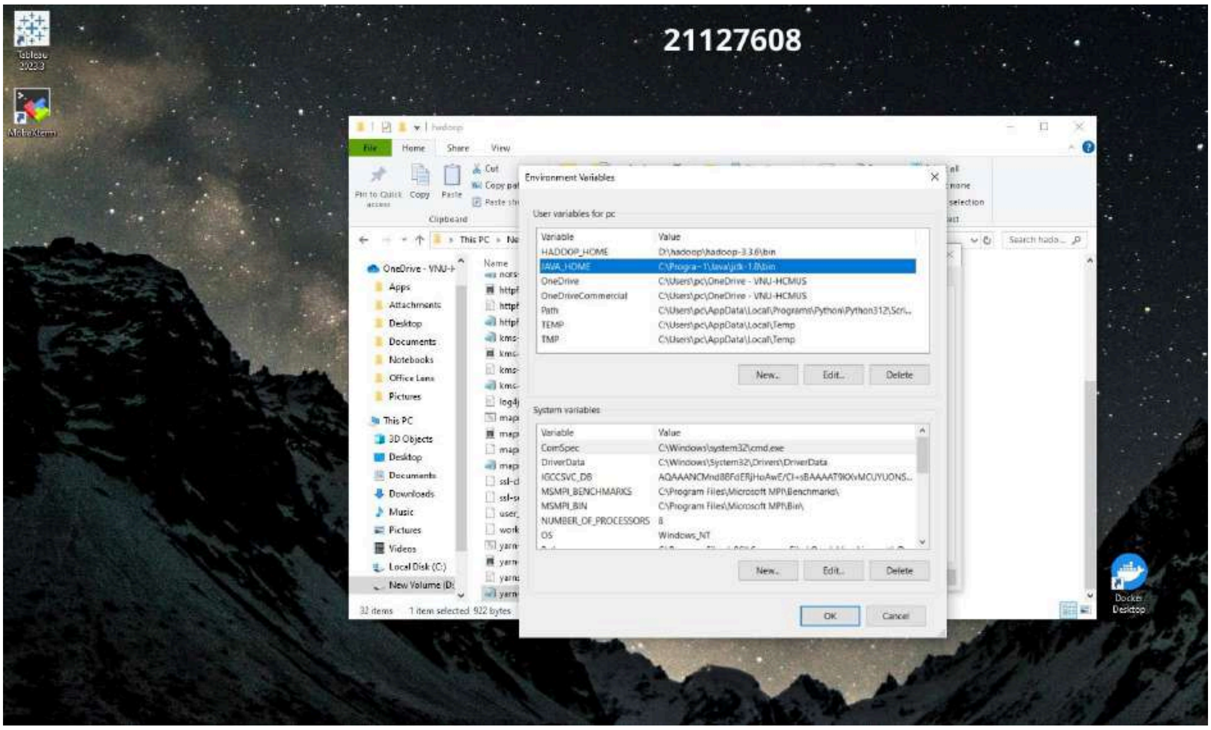
- ○ Set up HADOOP_HOME in Environment Variable



Variable	Value
HADOOP_HOME	C:\hadoop\hadoop-3.3.6\bin
JAVA_HOME	C:\Progra~1\Java\jdk-1.8\bin
OneDrive	C:\Users\pc\OneDrive - VNU-HCMUS
OneDriveCommercial	C:\Users\pc\OneDrive - VNU-HCMUS
Path	C:\Users\pc\AppData\Local\Programs\Python\Python312\Scri...
TEMP	C:\Users\pc\AppData\Local\Temp
TMP	C:\Users\pc\AppData\Local\Temp

Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
IGCCSVC_DB	AQAAANCMnd8BFdERjHoAwE/Cl+sBAAAAT9KKvMCUYUONS...
MSMPI_BENCHMARKS	C:\Program Files\Microsoft MPI\Benchmarks\
MSMPI_BIN	C:\Program Files\Microsoft MPI\Bin\
NUMBER_OF_PROCESSORS	8
OS	Windows_NT

- ○ Set up JAVA_HOME in Environment Variable



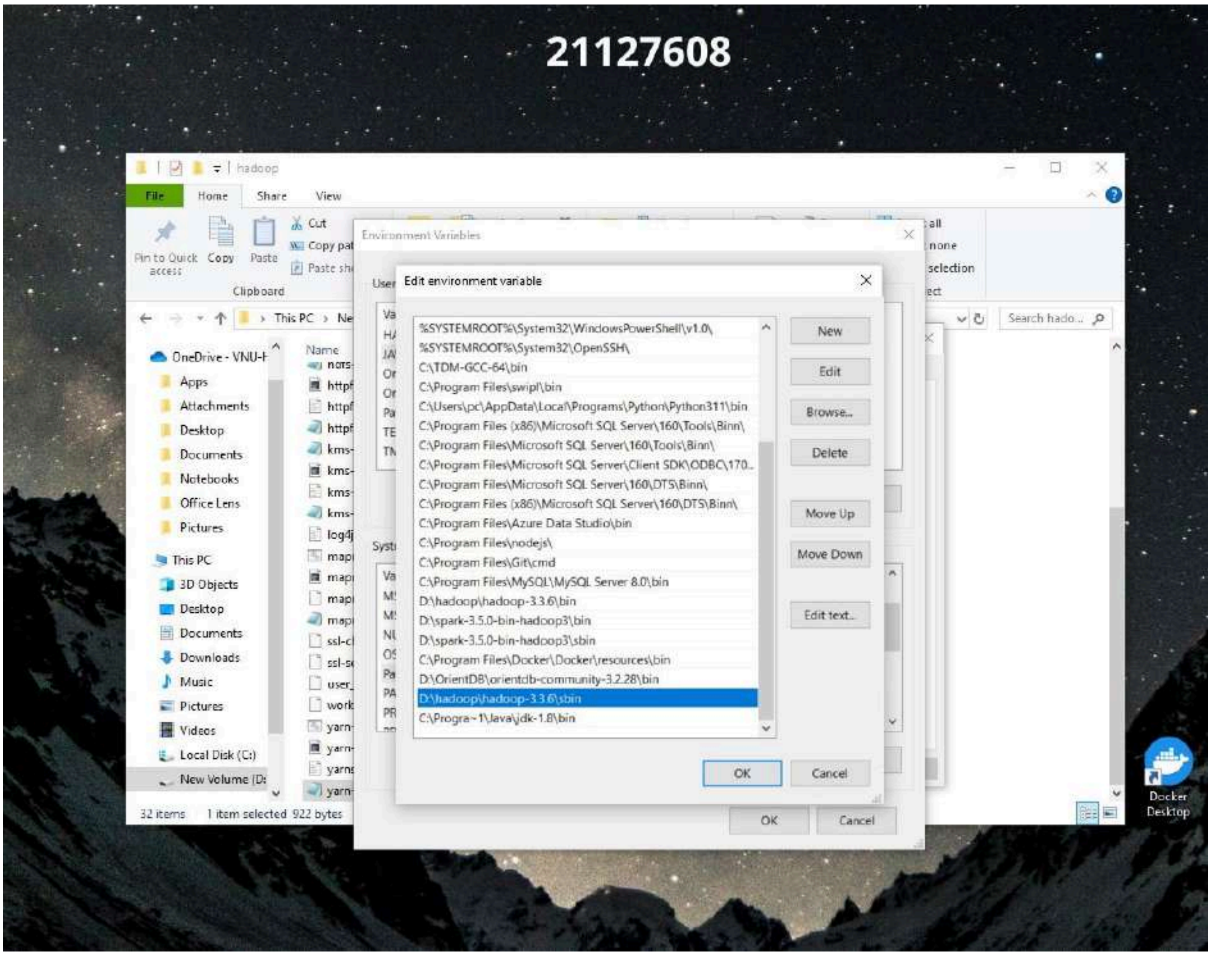
User variables for pc

Variable	Value
HADOOP_HOME	D:\hadoop\hadoop-3.3.6\bin
JAVA_HOME	C:\Progra~1\Java\jdk-1.8\bin
OneDrive	C:\Users\pc\OneDrive - VNU-HCMUS
OneDriveCommercial	C:\Users\pc\OneDrive - VNU-HCMUS
Path	C:\Users\pc\AppData\Local\Programs\Python\Python312\Scri...
TEMP	C:\Users\pc\AppData\Local\Temp
TMP	C:\Users\pc\AppData\Local\Temp

System variables

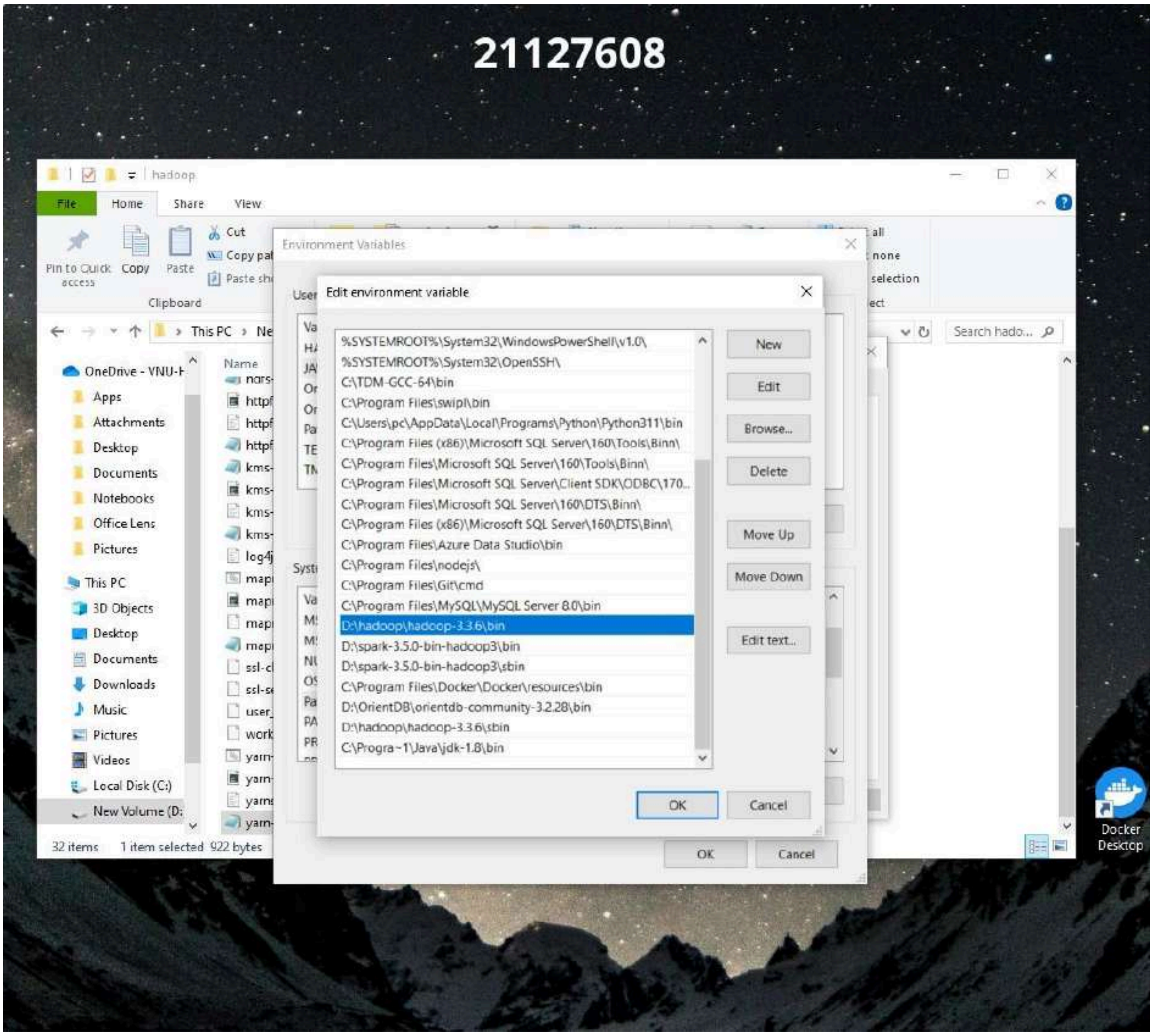
Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
IGCCSVC_DB	AQAAANCMnd8BFdERjHoAwE/Cl+sBAAAAT9KXvMCUYUONS...
MSMPI_BENCHMARKS	C:\Program Files\Microsoft MPI\Benchmarks\
MSMPI_BIN	C:\Program Files\Microsoft MPI\Bin\
NUMBER_OF_PROCESSORS	8
OS	Windows_NT

- Set up path of sbin



```
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\
%SYSTEMROOT%\System32\OpenSSH\
C:\TDM-GCC-64\bin
C:\Program Files\swipl\bin
C:\Users\pc\AppData\Local\Programs\Python\Python311\bin
C:\Program Files (x86)\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\170...
C:\Program Files\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files\Azure Data Studio\bin
C:\Program Files\nodejs\
C:\Program Files\Git\cmd
C:\Program Files\MySQL\MySQL Server 8.0\bin
D:\hadoop\hadoop-3.3.6\bin
D:\spark-3.5.0-bin-hadoop3\bin
D:\spark-3.5.0-bin-hadoop3\sbin
C:\Program Files\Docker\Docker\resources\bin
D:\OrientDB\orientdb-community-3.2.28\bin
D:\hadoop\hadoop-3.3.6\sbin
C:\Progra~1\Java\jdk-1.8\bin
```

- Set up path of bin

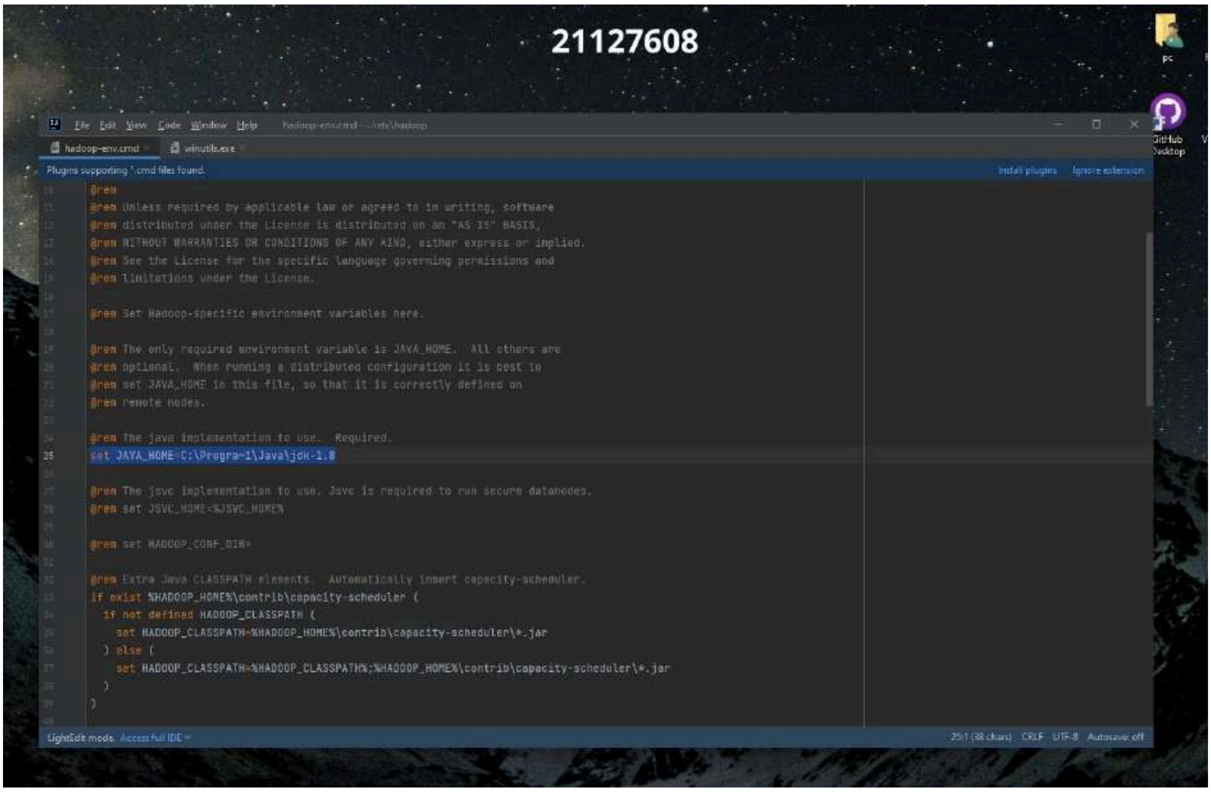


21127608

Edit environment variable

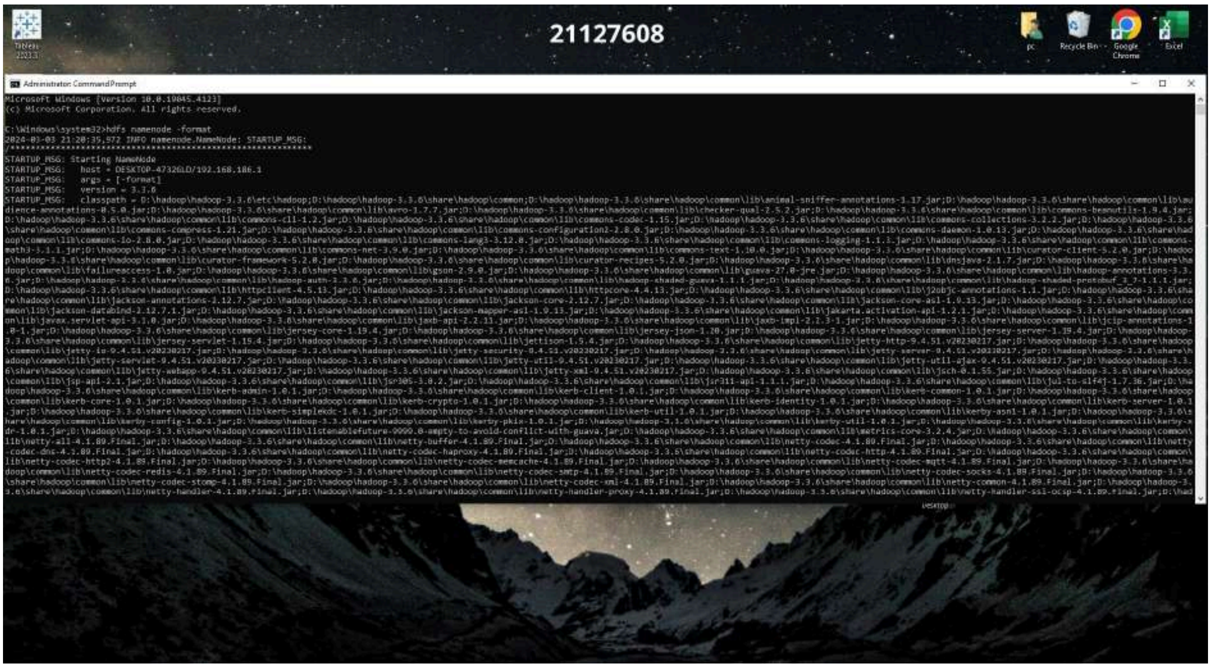
```
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\
%SYSTEMROOT%\System32\OpenSSH\
C:\TDM-GCC-64\bin
C:\Program Files\swipl\bin
C:\Users\pc\AppData\Local\Programs\Python\Python311\bin
C:\Program Files (x86)\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\170...
C:\Program Files\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files\Azure Data Studio\bin
C:\Program Files\nodejs\
C:\Program Files\Git\cmd
C:\Program Files\MySQL\MySQL Server 8.0\bin
D:\hadoop\hadoop-3.3.6\bin
D:\spark-3.5.0-bin-hadoop3\bin
D:\spark-3.5.0-bin-hadoop3\sbin
C:\Program Files\Docker\Docker\resources\bin
D:\OrientDB\orientdb-community-3.2.28\bin
D:\hadoop\hadoop-3.3.6\sbin
C:\Progra~1\Java\jdk-1.8\bin
```

- Set up JAVA_HOME of hadoop-env.cmd


21127608
File Edit View Code Window Help
hadoop-env.cmd - ...\etc\hadoop
hadoop-env.cmd
winutils.exe
Plugins supporting *.cmd files found.
Install plugins
Ignore extension
LightEdit mode. Access full IDE
25:1 (38 chars) CRLF UTF-8 Autosave: off
pc
GitHub Desktop

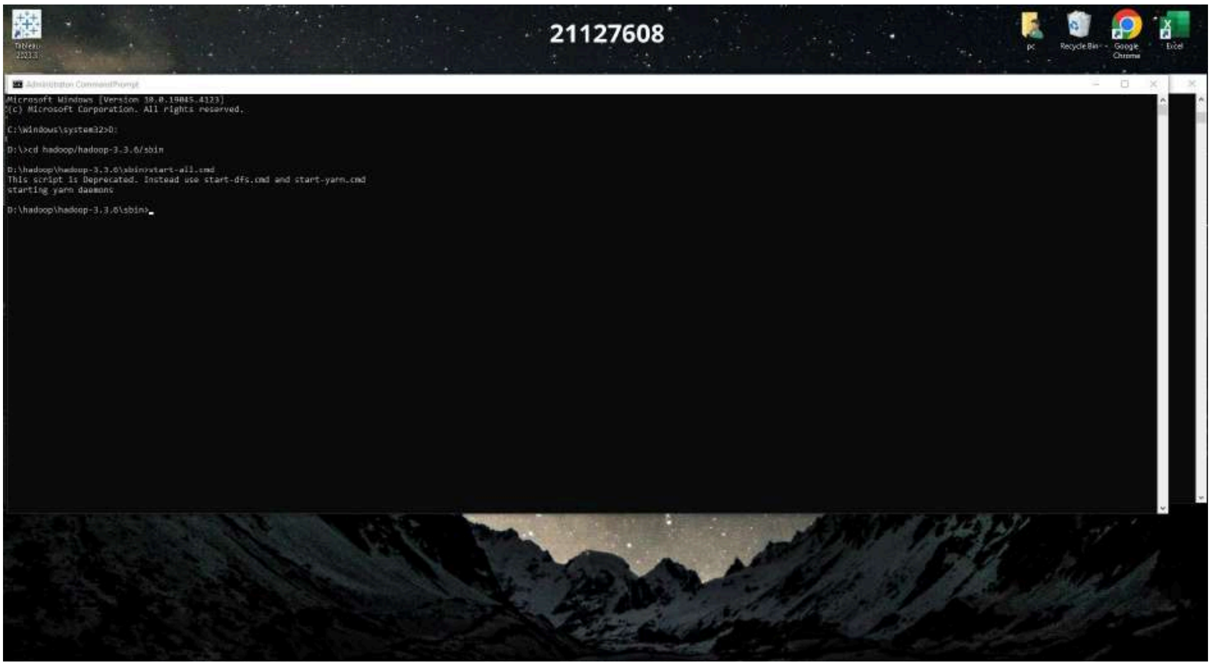
```
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
@rem limitations under the License.

@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA_HOME.  All others are
@rem optional.  When running a distributed configuration it is best to
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.

@rem The java implementation to use.  Required.
set JAVA_HOME=C:\Progra~1\Java\jdk-1.8

@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%

@rem set HADOOP_CONF_DIR=

@rem Extra Java CLASSPATH elements.  Automatically insert capacity-scheduler.
if exist %HADOOP_HOME%\contrib\capacity-scheduler (
  if not defined HADOOP_CLASSPATH (
    set HADOOP_CLASSPATH=%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
  ) else (
    set HADOOP_CLASSPATH=%HADOOP_CLASSPATH%;%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
  )
)
```

- ○ Format namenode of hadoop



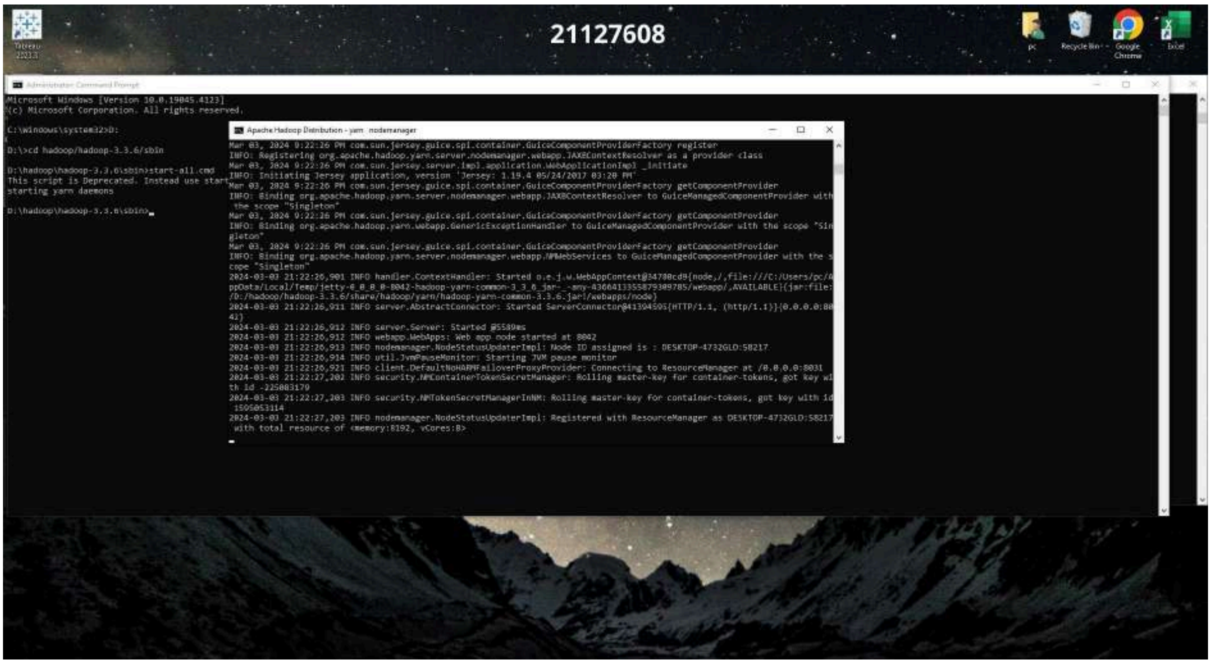
```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>hdfs namenode -format
2024-03-03 21:20:35,972 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = DESKTOP-4732GLD/192.168.186.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = D:\hadoop\hadoop-3.3.6\etc\hadoop;D:\hadoop\hadoop-3.3.6\share\hadoop\common;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\au
dience-annotations-0.5.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\avro-1.7.7.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\checker-qual-2.5.2.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-beanutils-1.9.4.jar;
D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-cli-1.2.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-codec-1.15.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-collections-3.2.2.jar;D:\hadoop\hadoop-3.3.6
\share\hadoop\common\lib\commons-compress-1.21.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-configuration2-2.8.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-daemon-1.0.13.jar;D:\hadoop\hadoop-3.3.6\share\had
oop\common\lib\commons-io-2.8.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-lang3-3.12.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-logging-1.1.3.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-
math3-3.1.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-net-3.9.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\commons-text-1.10.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\curator-client-5.2.0.jar;D:\hadoo
p\hadoop-3.3.6\share\hadoop\common\lib\curator-framework-5.2.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\curator-recipes-5.2.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\dnsjava-2.1.7.jar;D:\hadoop\hadoop-3.3.6\share\ha
doop\common\lib\failureaccess-1.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\gson-2.9.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\guava-27.0-jre.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\hadoop-annotations-3.3.
6.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\hadoop-auth-3.3.6.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\hadoop-shaded-guava-1.1.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\hadoop-shaded-protobuf_3_7-1.1.1.jar;
D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\httpclient-4.5.13.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\httpcore-4.4.13.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\j2objc-annotations-1.1.jar;D:\hadoop\hadoop-3.3.6\sha
re\hadoop\common\lib\jackson-annotations-2.12.7.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jackson-core-2.12.7.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jackson-core-asl-1.9.13.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\co
mmon\lib\jackson-databind-2.12.7.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jackson-mapper-asl-1.9.13.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jakarta.activation-api-1.2.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\comm
on\lib\javax.servlet-api-3.1.0.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jaxb-api-2.2.11.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jaxb-impl-2.2.3-1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jcip-annotations-1
.0-1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jersey-core-1.19.4.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jersey-json-1.20.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jersey-server-1.19.4.jar;D:\hadoop\hadoop-
3.3.6\share\hadoop\common\lib\jersey-servlet-1.19.4.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jettison-1.5.4.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jetty-http-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop
\common\lib\jetty-io-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jetty-security-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jetty-server-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\h
adoop\common\lib\jetty-servlet-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jetty-util-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jetty-util-ajax-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.
6\share\hadoop\common\lib\jetty-webapp-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jetty-xml-9.4.51.v20230217.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jsch-0.1.55.jar;D:\hadoop\hadoop-3.3.6\share\hadoop
\common\lib\jsp-api-2.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jsr305-3.0.2.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jsr311-api-1.1.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\jul-to-slf4j-1.7.36.jar;D:\ha
doop\hadoop-3.3.6\share\hadoop\common\lib\kerb-admin-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-client-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-common-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop
\common\lib\kerb-core-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-crypto-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-identity-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-server-1.0.1
.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-simplekdc-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerb-util-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerby-asn1-1.0.1.jar;D:\hadoop\hadoop-3.3.6\s
hare\hadoop\common\lib\kerby-config-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerby-pkix-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerby-util-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\kerby-x
dr-1.0.1.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\listenablefuture-9999.0-empty-to-avoid-conflict-with-guava.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\metrics-core-3.2.4.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\
lib\netty-all-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-buffer-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty
-codec-dns-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-haproxy-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-http-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\
lib\netty-codec-http2-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-memcache-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-mqtt-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\ha
doop\common\lib\netty-codec-redis-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-smtp-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-socks-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6
\share\hadoop\common\lib\netty-codec-stomp-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-xml-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-common-4.1.89.Final.jar;D:\hadoop\hadoop-3.
3.6\share\hadoop\common\lib\netty-handler-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-handler-proxy-4.1.89.Final.jar;D:\hadoop\hadoop-3.3.6\share\hadoop\common\lib\netty-handler-ssl-ocsp-4.1.89.Final.jar;D:\had
```

- ○ Run all file .cmd in sbin folder



```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>D:

D:\>cd hadoop/hadoop-3.3.6/sbin

D:\hadoop\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop\hadoop-3.3.6\sbin>
```

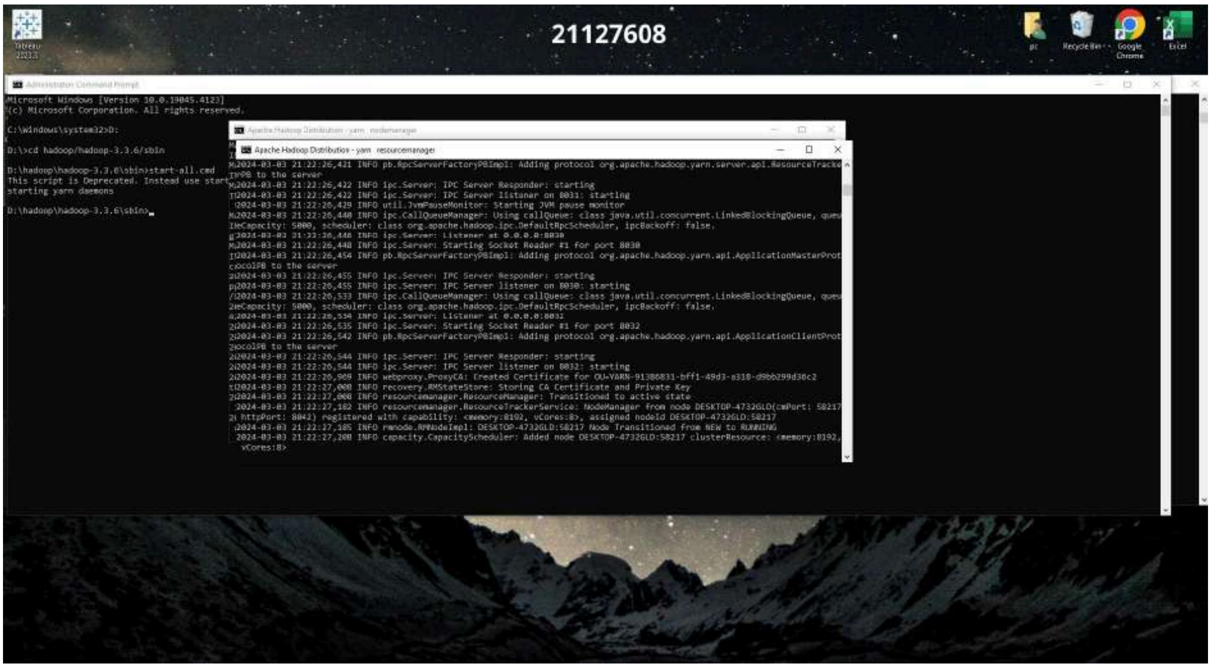
- ○ Yarn nodemanager of sbin


21127608
Tableau 2023.3
pc
Recycle Bin
Google Chrome
Excel
Administrator: Command Prompt
Apache Hadoop Distribution - yarn nodemanager

```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>D:

D:\>cd hadoop/hadoop-3.3.6/sbin

D:\hadoop\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start
starting yarn daemons

D:\hadoop\hadoop-3.3.6\sbin>
```

```
Mar 03, 2024 9:22:26 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Mar 03, 2024 9:22:26 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19.4 05/24/2017 03:20 PM'
Mar 03, 2024 9:22:26 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with
 the scope "Singleton"
Mar 03, 2024 9:22:26 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Sin
gleton"
Mar 03, 2024 9:22:26 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the s
cope "Singleton"
2024-03-03 21:22:26,901 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@34780cd9{node,/,file:///C:/Users/pc/A
ppData/Local/Temp/jetty-0_0_0_0-8042-hadoop-yarn-common-3_3_6_jar-_-any-4366413355879389785/webapp/,AVAILABLE}{jar:file:
/D:/hadoop/hadoop-3.3.6/share/hadoop/yarn/hadoop-yarn-common-3.3.6.jar!/webapps/node}
2024-03-03 21:22:26,911 INFO server.AbstractConnector: Started ServerConnector@41394595{HTTP/1.1, (http/1.1)}{0.0.0.0:80
42}
2024-03-03 21:22:26,912 INFO server.Server: Started @5589ms
2024-03-03 21:22:26,912 INFO webapp.WebApps: Web app node started at 8042
2024-03-03 21:22:26,913 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : DESKTOP-4732GLD:58217
2024-03-03 21:22:26,914 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-03-03 21:22:26,921 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2024-03-03 21:22:27,202 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key wi
th id -225083179
2024-03-03 21:22:27,203 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id
 1595053114
2024-03-03 21:22:27,203 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as DESKTOP-4732GLD:58217
 with total resource of <memory:8192, vCores:8>
```

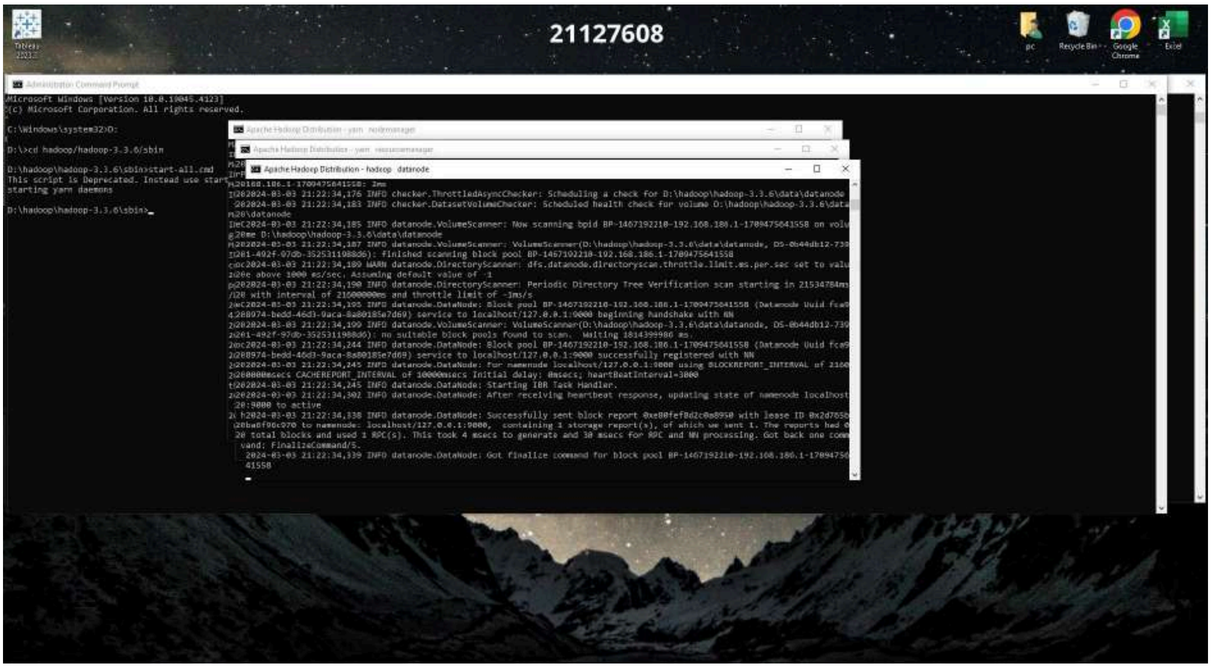
- ○ Yarn resourcemanager of sbin


21127608
Tableau 2023.3
pc
Recycle Bin
Google Chrome
Excel
Administrator: Command Prompt
Apache Hadoop Distribution - yarn nodemanager
Apache Hadoop Distribution - yarn resourcemanager

```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>D:

D:\>cd hadoop/hadoop-3.3.6/sbin

D:\hadoop\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start
starting yarn daemons

D:\hadoop\hadoop-3.3.6\sbin>_
```

```
2024-03-03 21:22:26,421 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTracke
rPB to the server
2024-03-03 21:22:26,422 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:22:26,422 INFO ipc.Server: IPC Server listener on 8031: starting
2024-03-03 21:22:26,429 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-03-03 21:22:26,440 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-03-03 21:22:26,446 INFO ipc.Server: Listener at 0.0.0.0:8030
2024-03-03 21:22:26,448 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-03-03 21:22:26,454 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProt
ocolPB to the server
2024-03-03 21:22:26,455 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:22:26,455 INFO ipc.Server: IPC Server listener on 8030: starting
2024-03-03 21:22:26,533 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-03-03 21:22:26,534 INFO ipc.Server: Listener at 0.0.0.0:8032
2024-03-03 21:22:26,535 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-03-03 21:22:26,542 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProt
ocolPB to the server
2024-03-03 21:22:26,544 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:22:26,544 INFO ipc.Server: IPC Server listener on 8032: starting
2024-03-03 21:22:26,909 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-91386831-bff1-49d3-a318-d9bb299d36c2
2024-03-03 21:22:27,008 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2024-03-03 21:22:27,008 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-03-03 21:22:27,182 INFO resourcemanager.ResourceTrackerService: NodeManager from node DESKTOP-4732GLD(cmPort: 58217
 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId DESKTOP-4732GLD:58217
2024-03-03 21:22:27,185 INFO rmnode.RMNodeImpl: DESKTOP-4732GLD:58217 Node Transitioned from NEW to RUNNING
2024-03-03 21:22:27,208 INFO capacity.CapacityScheduler: Added node DESKTOP-4732GLD:58217 clusterResource: <memory:8192,
 vCores:8>
```

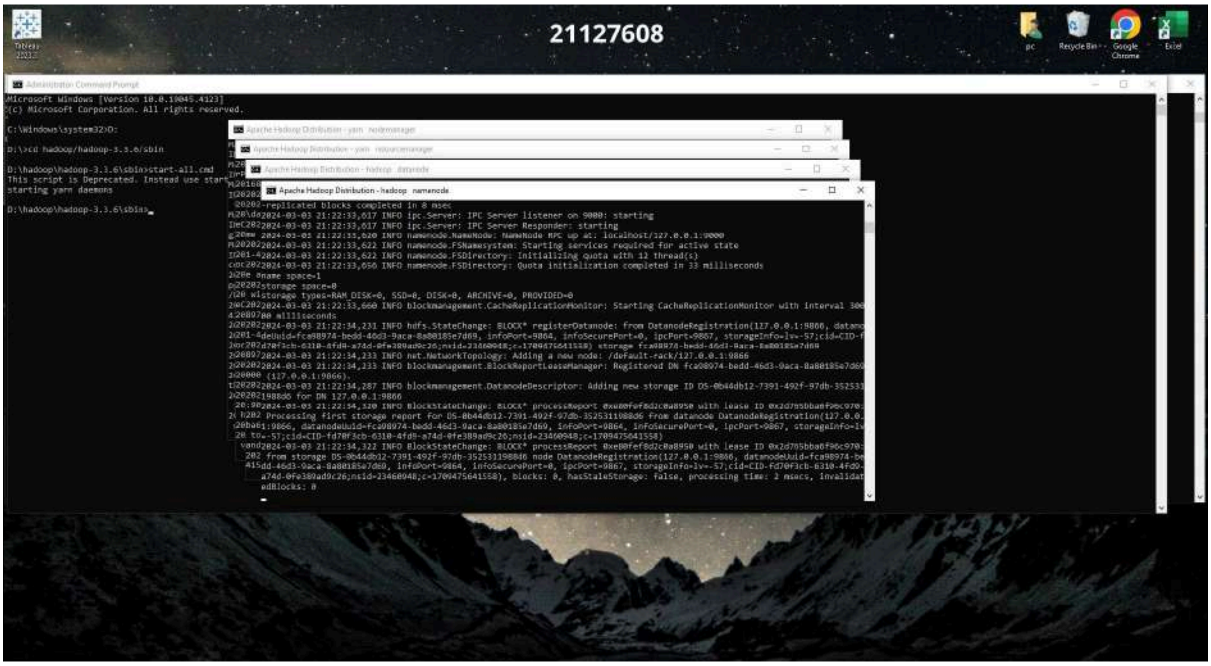
○ Hadoop datanode of sbin


21127608
Tableau 2023.3
pc
Recycle Bin
Google Chrome
Excel
Administrator: Command Prompt
Apache Hadoop Distribution - yarn nodemanager
Apache Hadoop Distribution - yarn resourcemanager
Apache Hadoop Distribution - hadoop datanode

```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>D:

D:\>cd hadoop/hadoop-3.3.6/sbin

D:\hadoop\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start
starting yarn daemons

D:\hadoop\hadoop-3.3.6\sbin>_
```

```
168.186.1-1709475641558: 2ms
2024-03-03 21:22:34,176 INFO checker.ThrottledAsyncChecker: Scheduling a check for D:\hadoop\hadoop-3.3.6\data\datanode
2024-03-03 21:22:34,183 INFO checker.DatasetVolumeChecker: Scheduled health check for volume D:\hadoop\hadoop-3.3.6\data
\datanode
2024-03-03 21:22:34,185 INFO datanode.VolumeScanner: Now scanning bpid BP-1467192210-192.168.186.1-1709475641558 on volu
me D:\hadoop\hadoop-3.3.6\data\datanode
2024-03-03 21:22:34,187 INFO datanode.VolumeScanner: VolumeScanner(D:\hadoop\hadoop-3.3.6\data\datanode, DS-0b44db12-739
1-492f-97db-352531198806): finished scanning block pool BP-1467192210-192.168.186.1-1709475641558
2024-03-03 21:22:34,189 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to valu
e above 1000 ms/sec. Assuming default value of -1
2024-03-03 21:22:34,190 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 21534784ms
 with interval of 21600000ms and throttle limit of -1ms/s
2024-03-03 21:22:34,195 INFO datanode.DataNode: Block pool BP-1467192210-192.168.186.1-1709475641558 (Datanode Uuid fca9
88974-bedd-46d3-9aca-8a80185e7d69) service to localhost/127.0.0.1:9000 beginning handshake with NN
2024-03-03 21:22:34,199 INFO datanode.VolumeScanner: VolumeScanner(D:\hadoop\hadoop-3.3.6\data\datanode, DS-0b44db12-739
1-492f-97db-352531198806): no suitable block pools found to scan.  Waiting 1814399986 ms.
2024-03-03 21:22:34,244 INFO datanode.DataNode: Block pool BP-1467192210-192.168.186.1-1709475641558 (Datanode Uuid fca9
88974-bedd-46d3-9aca-8a80185e7d69) service to localhost/127.0.0.1:9000 successfully registered with NN
2024-03-03 21:22:34,245 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 2160
0000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartBeatInterval=3000
2024-03-03 21:22:34,245 INFO datanode.DataNode: Starting IBR Task Handler.
2024-03-03 21:22:34,302 INFO datanode.DataNode: After receiving heartbeat response, updating state of namenode localhost
:9000 to active
2024-03-03 21:22:34,338 INFO datanode.DataNode: Successfully sent block report 0xe80fef8d2c0a8950 with lease ID 0x2d765b
0ba8f96c970 to namenode: localhost/127.0.0.1:9000,  containing 1 storage report(s), of which we sent 1. The reports had 0
 total blocks and used 1 RPC(s). This took 4 msecs to generate and 30 msecs for RPC and NN processing. Got back one comm
and: FinalizeCommand/5.
2024-03-03 21:22:34,339 INFO datanode.DataNode: Got finalize command for block pool BP-1467192210-192.168.186.1-17094756
41558
_
```

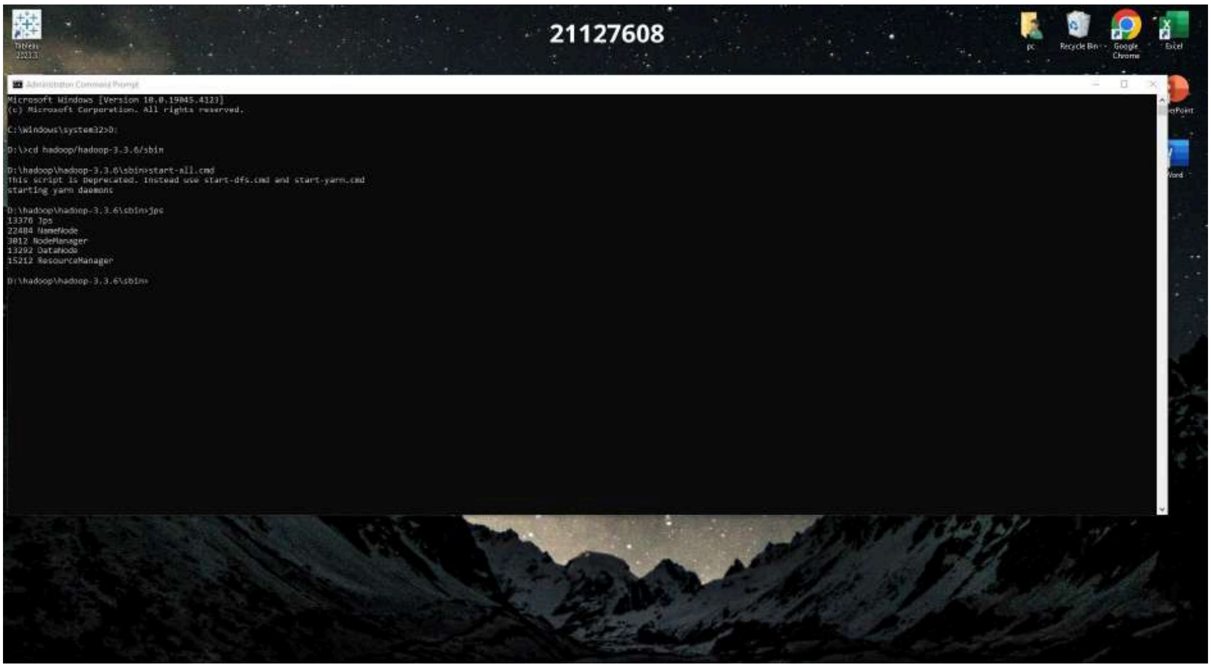
○ Hadoop namenode of sbin


21127608
Administrator: Command Prompt
Apache Hadoop Distribution - yarn nodemanager
Apache Hadoop Distribution - yarn resourcemanager
Apache Hadoop Distribution - hadoop datanode
Apache Hadoop Distribution - hadoop namenode

```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>D:

D:\>cd hadoop/hadoop-3.3.6/sbin

D:\hadoop\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start
starting yarn daemons

D:\hadoop\hadoop-3.3.6\sbin>
```

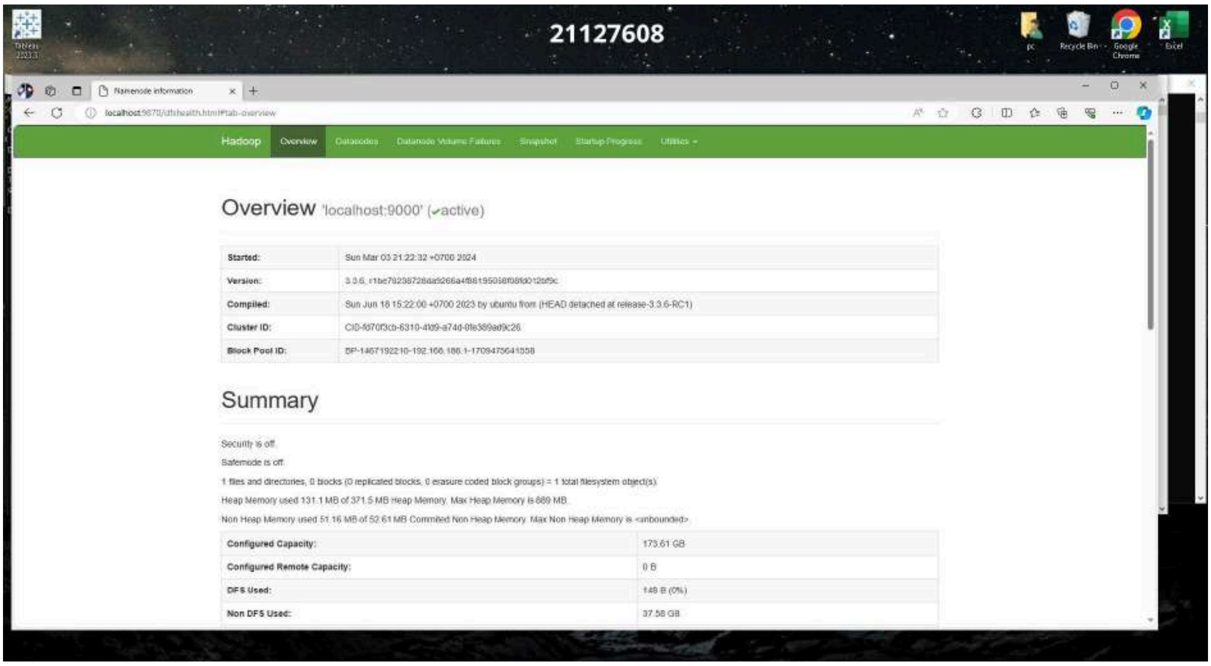
```
-replicated blocks completed in 8 msec
2024-03-03 21:22:33,617 INFO ipc.Server: IPC Server listener on 9000: starting
2024-03-03 21:22:33,617 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 21:22:33,620 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000
2024-03-03 21:22:33,622 INFO namenode.FSNamesystem: Starting services required for active state
2024-03-03 21:22:33,622 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)
2024-03-03 21:22:33,656 INFO namenode.FSDirectory: Quota initialization completed in 33 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2024-03-03 21:22:33,660 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 300
00 milliseconds
2024-03-03 21:22:34,231 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datano
deUuid=fca98974-bedd-46d3-9aca-8a80185e7d69, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-f
d70f3cb-6310-4fd9-a74d-0fe389ad9c26;nsid=23460948;c=1709475641558) storage fca98974-bedd-46d3-9aca-8a80185e7d69
2024-03-03 21:22:34,233 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2024-03-03 21:22:34,233 INFO blockmanagement.BlockReportLeaseManager: Registered DN fca98974-bedd-46d3-9aca-8a80185e7d69
 (127.0.0.1:9866).
2024-03-03 21:22:34,287 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-0b44db12-7391-492f-97db-352531
1988d6 for DN 127.0.0.1:9866
2024-03-03 21:22:34,320 INFO BlockStateChange: BLOCK* processReport 0xe80fef8d2c0a8950 with lease ID 0x2d705bba6f96c970:
 Processing first storage report for DS-0b44db12-7391-492f-97db-3525311988d6 from datanode DatanodeRegistration(127.0.0.
1:9866, datanodeUuid=fca98974-bedd-46d3-9aca-8a80185e7d69, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv
=-57;cid=CID-fd70f3cb-6310-4fd9-a74d-0fe389ad9c26;nsid=23460948;c=1709475641558)
2024-03-03 21:22:34,322 INFO BlockStateChange: BLOCK* processReport 0xe80fef8d2c0a8950 with lease ID 0x2d705bba6f96c970:
 from storage DS-0b44db12-7391-492f-97db-3525311988d6 node DatanodeRegistration(127.0.0.1:9866, datanodeUuid=fca98974-be
dd-46d3-9aca-8a80185e7d69, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-fd70f3cb-6310-4fd9-
a74d-0fe389ad9c26;nsid=23460948;c=1709475641558), blocks: 0, hasStaleStorage: false, processing time: 2 msecs, invalidat
edBlocks: 0
```

- Configuration of hadoop system



```
Microsoft Windows [Version 10.0.19045.4123]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>D:

D:\>cd hadoop/hadoop-3.3.6/sbin

D:\hadoop\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop\hadoop-3.3.6\sbin>jps
13376 Jps
22484 NameNode
3012 NodeManager
13292 DataNode
15212 ResourceManager

D:\hadoop\hadoop-3.3.6\sbin>
```

- On localhost:9870 of Hadoop File System



Overview 'localhost:9000' (✓active)

Started:	Sun Mar 03 21:22:32 +0700 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 15:22:00 +0700 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-fd70f3cb-6310-4fd9-a74d-0fe389ad9c26
Block Pool ID:	BP-1467192210-192.168.186.1-1709475641558

Summary

Security is off.

Safemode is off.

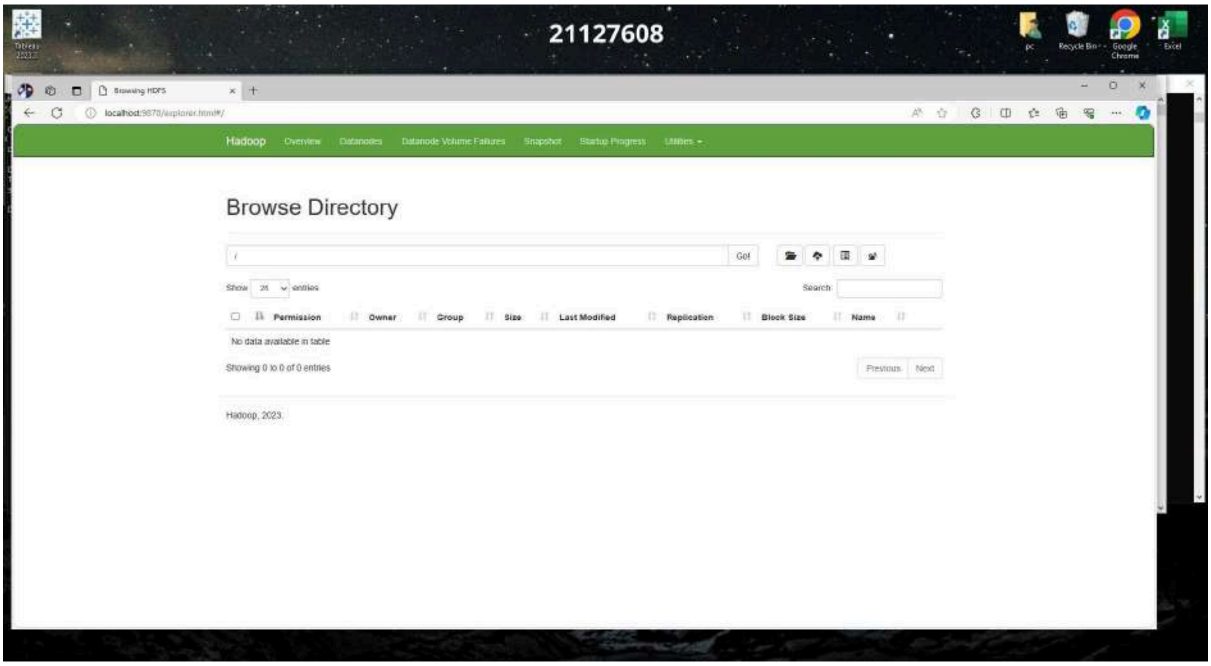
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 131.1 MB of 371.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 51.16 MB of 52.61 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	173.61 GB
Configured Remote Capacity:	0 B
DFS Used:	148 B (0%)
Non DFS Used:	37.58 GB

- ○ On localhost:9870 of Hadoop File System



Browse Directory

/

Show 25 entries

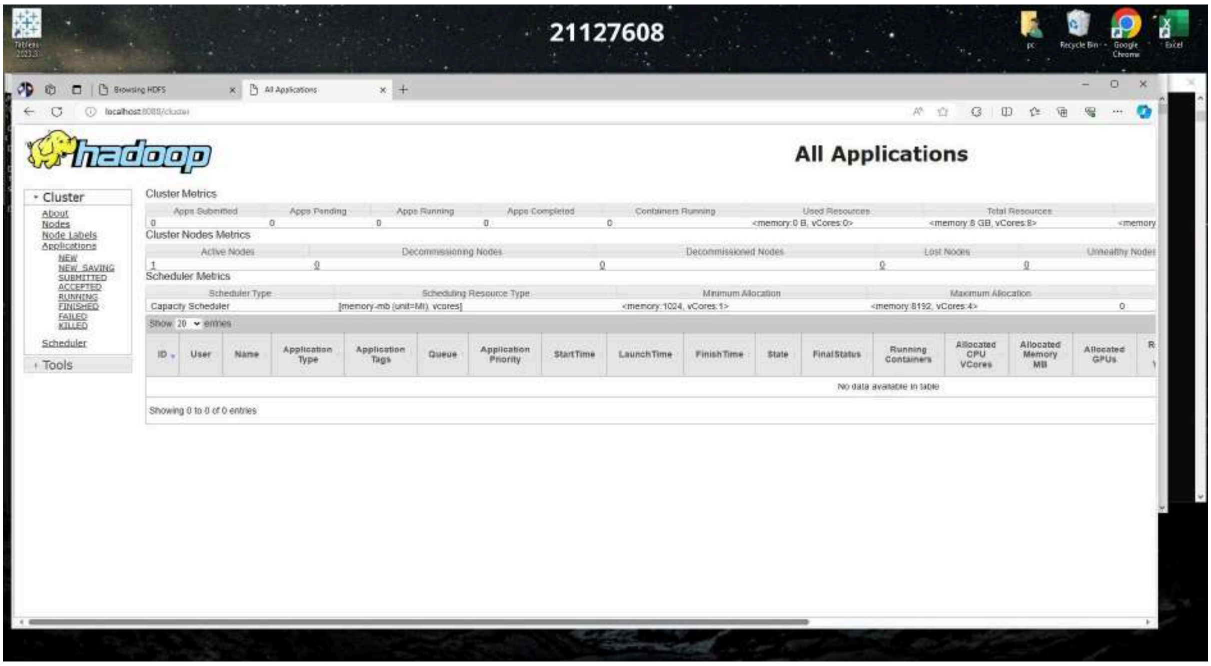
Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
No data available in table								

Showing 0 to 0 of 0 entries

Hadoop, 2023.

- ○ On localhost:8088 is cluster app of hadoop



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>	<memory

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
1	0	0	0	0

Scheduler Metrics

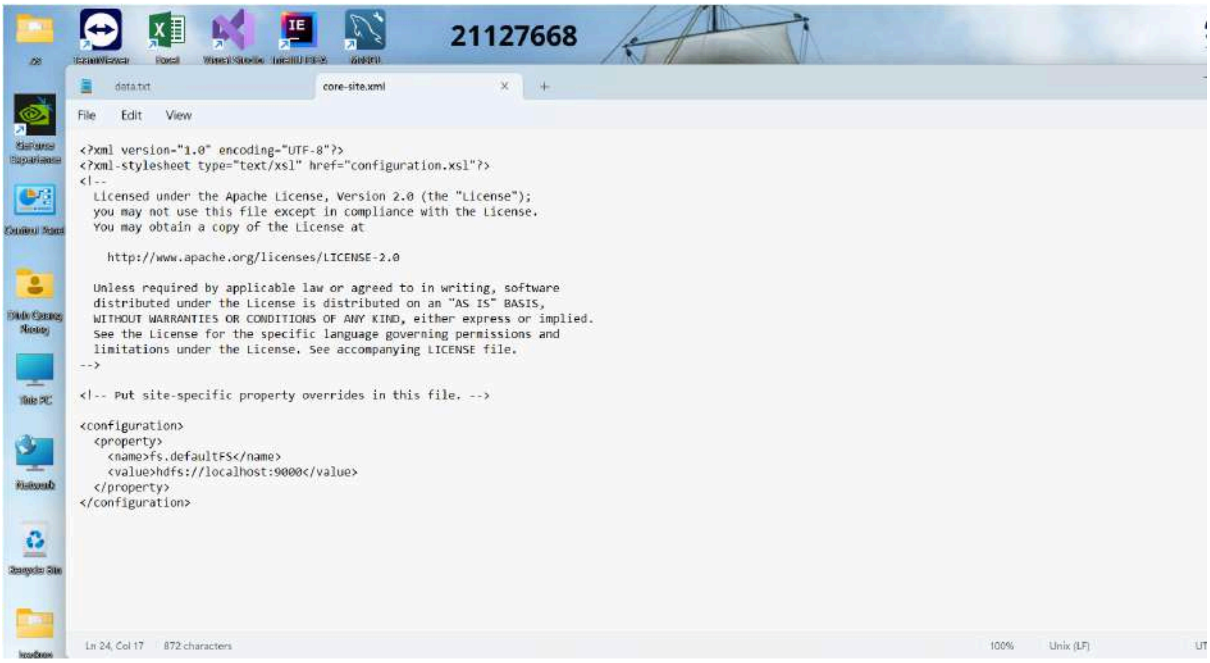
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Allocated GPUs	R
No data available in table																

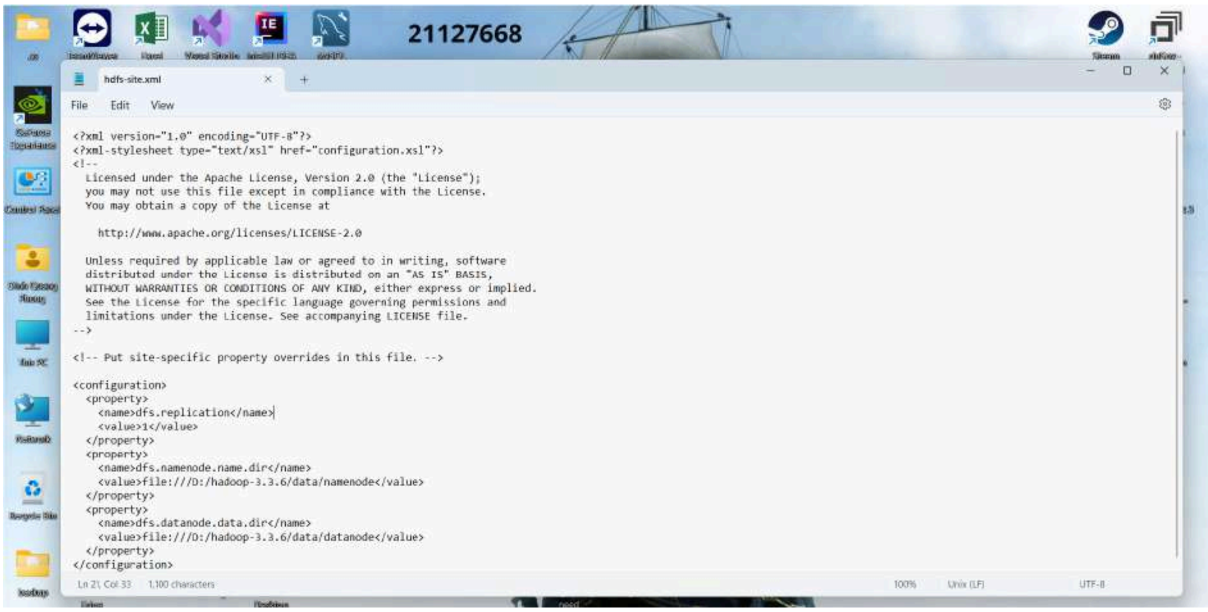
Showing 0 to 0 of 0 entries

- 21127668
    - Set up core-site.xml



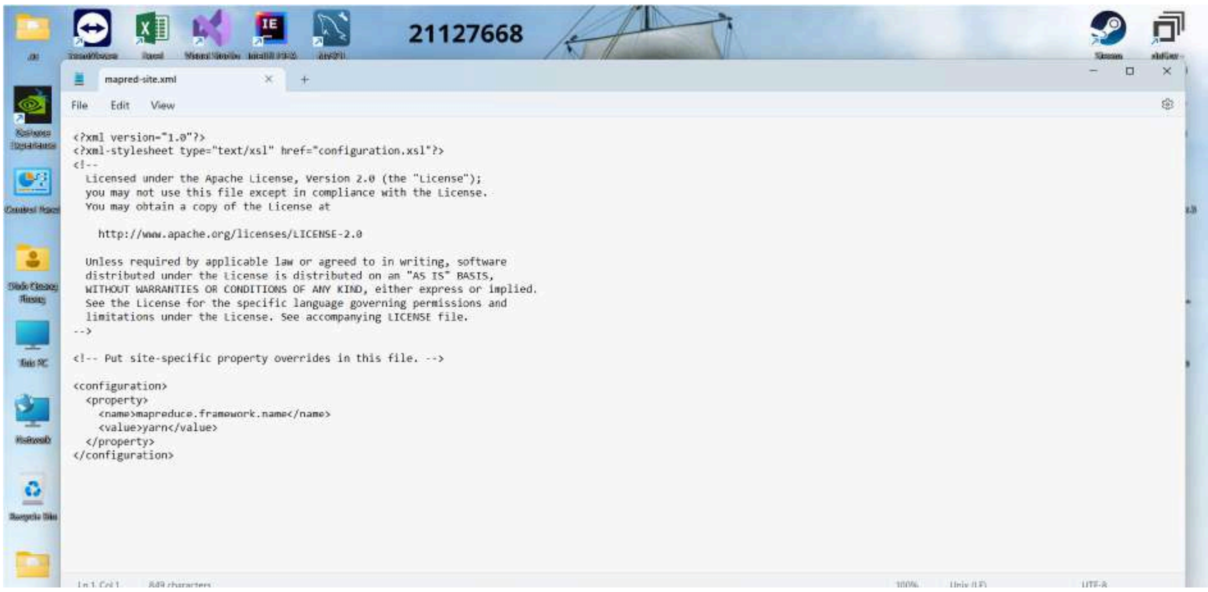
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

- Set up hdfs-site.xml



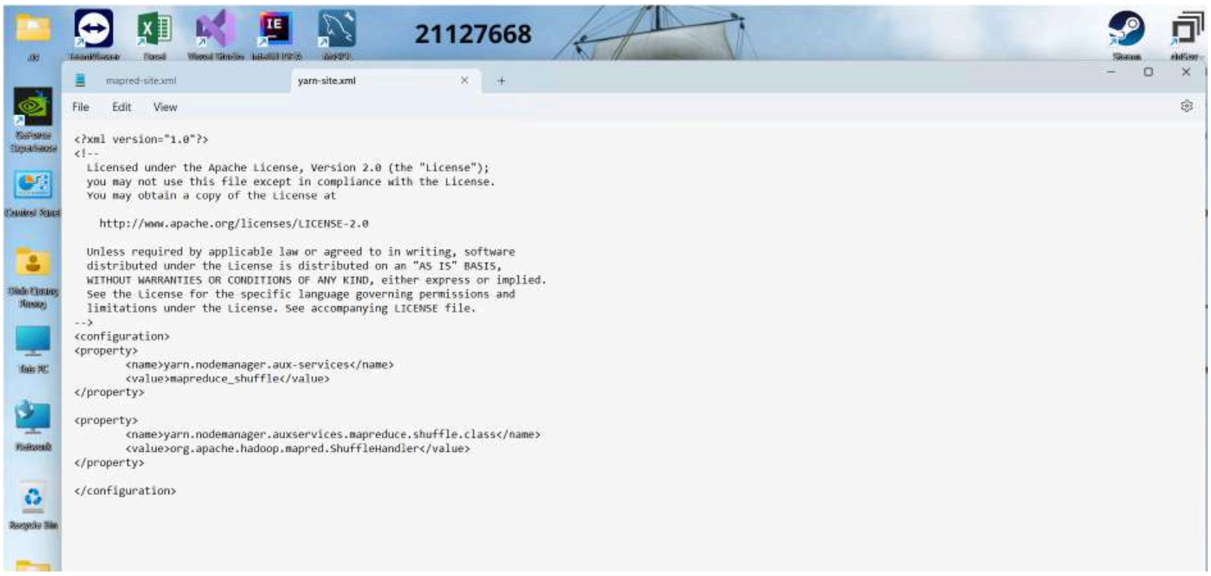
```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///D:/hadoop-3.3.6/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///D:/hadoop-3.3.6/data/datanode</value>
  </property>
</configuration>
```

- Set up mapred-site.xml



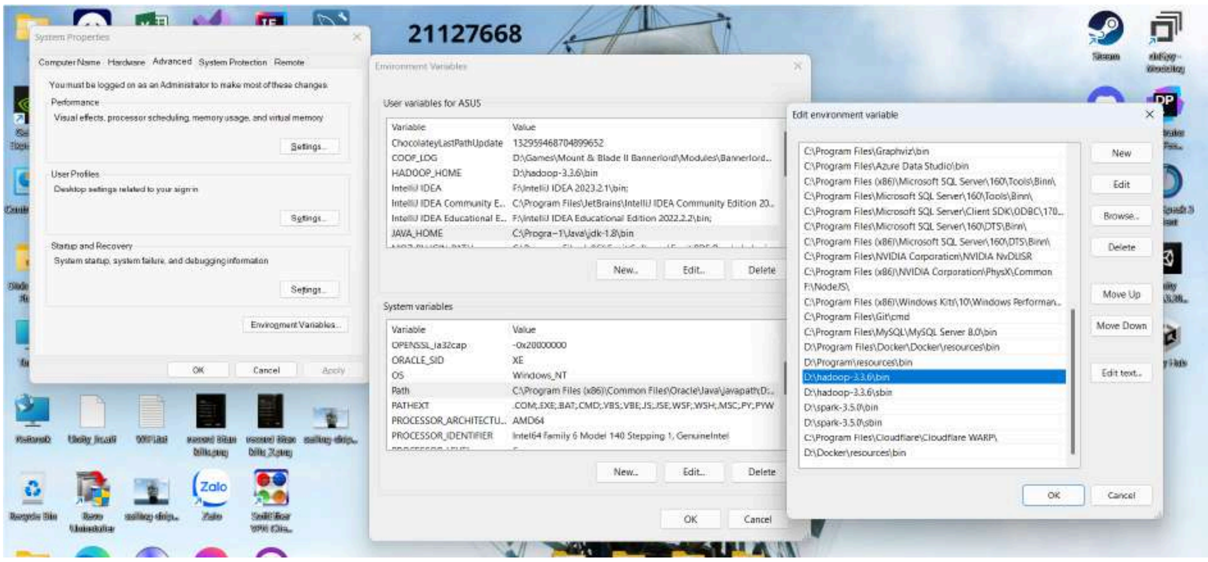
```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

- Set up yarn-site.xml



```
<?xml version="1.0"?>
<!--
  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.
-->
<configuration>
<property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
</property>

<property>
        <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
        <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>

</configuration>
```

- Set up in Environment Variable



You must be logged on as an Administrator to make most of these changes.

Performance
Visual effects, processor scheduling, memory usage, and virtual memory

User Profiles
Desktop settings related to your sign-in

Startup and Recovery
System startup, system failure, and debugging information

User variables for ASUS

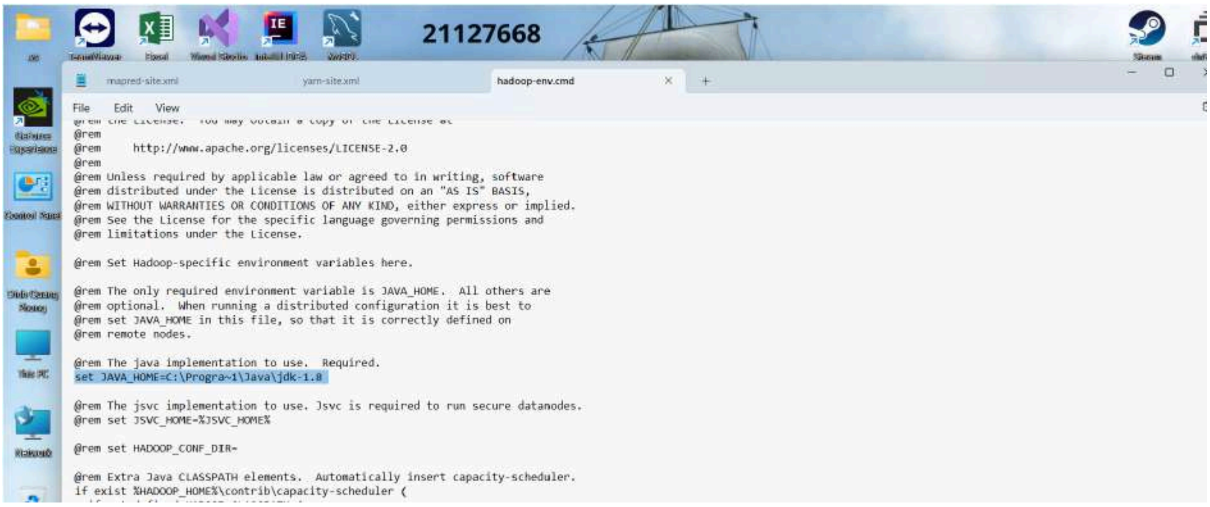
```
Variable                      Value
ChocolateyLastPathUpdate      132959468704899652
COOP_LOG                      D:\Games\Mount & Blade II Bannerlord\Modules\Bannerlord...
HADOOP_HOME                   D:\hadoop-3.3.6\bin
IntelliJ IDEA                 F:\IntelliJ IDEA 2023.2.1\bin;
IntelliJ IDEA Community E...  C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 20...
IntelliJ IDEA Educational E...F:\IntelliJ IDEA Educational Edition 2022.2.2\bin;
JAVA_HOME                     C:\Progra~1\Java\jdk-1.8\bin
```

System variables

```
Variable                      Value
OPENSSL_ia32cap               -0x20000000
ORACLE_SID                    XE
OS                            Windows_NT
Path                          C:\Program Files (x86)\Common Files\Oracle\Java\javapath;D:...
PATHEXT                       .COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.PY;.PYW
PROCESSOR_ARCHITECTU...       AMD64
PROCESSOR_IDENTIFIER          Intel64 Family 6 Model 140 Stepping 1, GenuineIntel
```

```
C:\Program Files\Graphviz\bin
C:\Program Files\Azure Data Studio\bin
C:\Program Files (x86)\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\160\Tools\Binn\
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\170...
C:\Program Files\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\160\DTS\Binn\
C:\Program Files\NVIDIA Corporation\NVIDIA NvDLISR
C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common
F:\NodeJS\
C:\Program Files (x86)\Windows Kits\10\Windows Performan...
C:\Program Files\Git\cmd
C:\Program Files\MySQL\MySQL Server 8.0\bin
D:\Program Files\Docker\Docker\resources\bin
D:\Program\resources\bin
D:\hadoop-3.3.6\bin
D:\hadoop-3.3.6\sbin
D:\spark-3.5.0\bin
D:\spark-3.5.0\sbin
C:\Program Files\Cloudflare\Cloudflare WARP\
D:\Docker\resources\bin
```

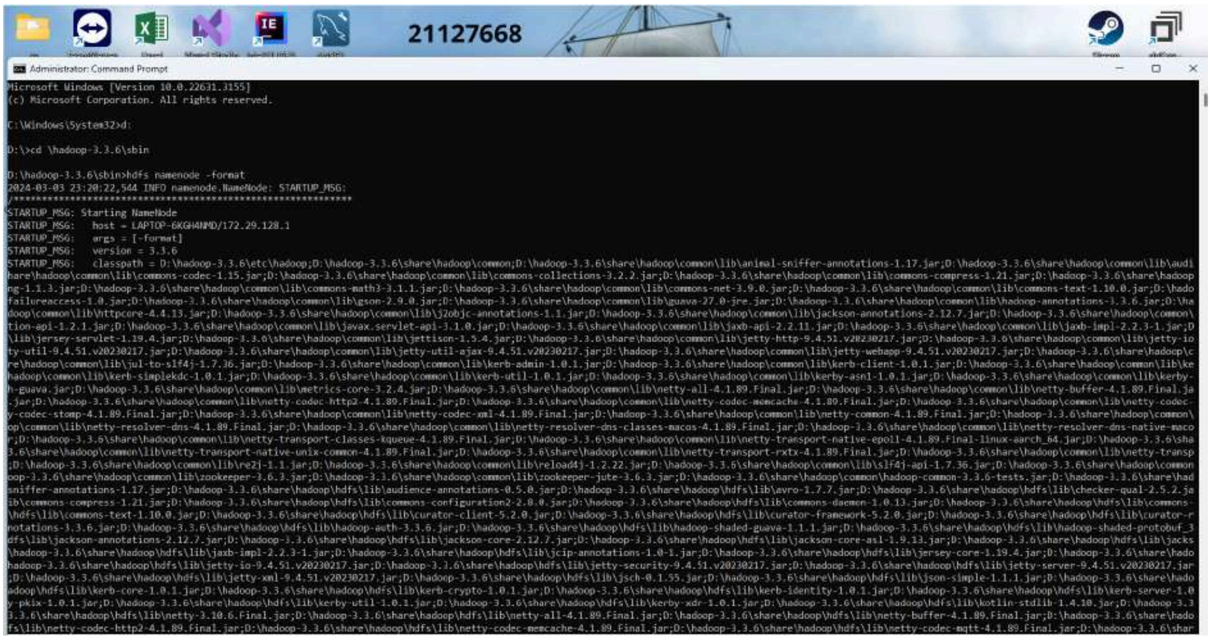
- Set up JAVA_HOME of hadoop-env.cmd



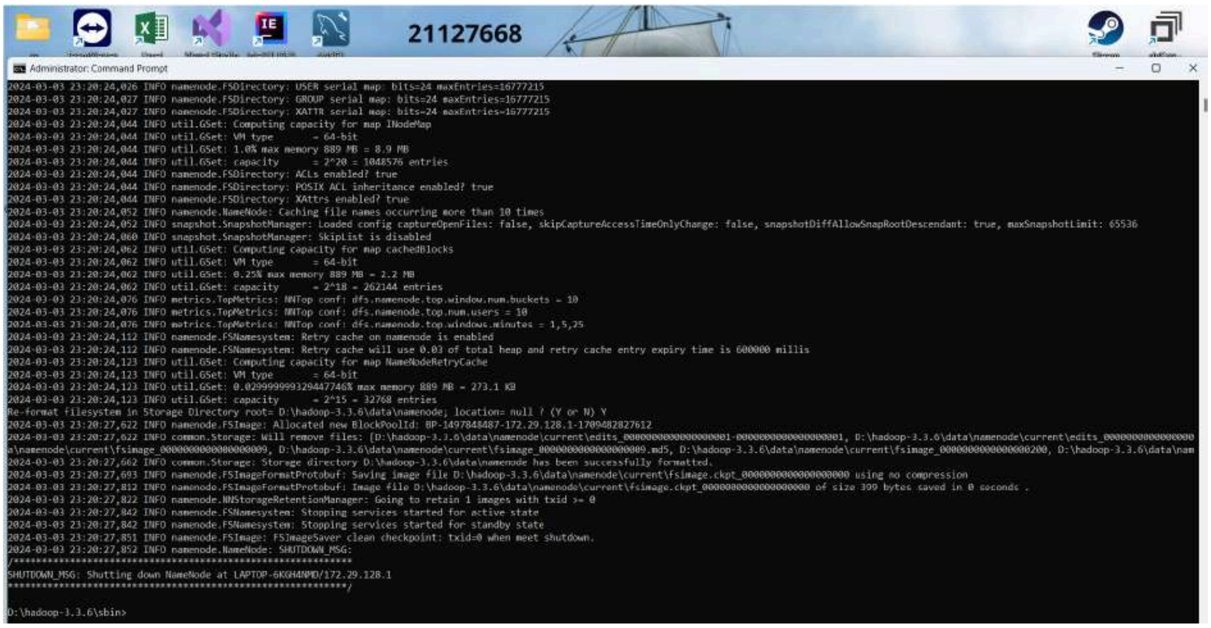
```
@rem the License. You may obtain a copy of the License at
@rem
@rem     http://www.apache.org/licenses/LICENSE-2.0
@rem
@rem Unless required by applicable law or agreed to in writing, software
@rem distributed under the License is distributed on an "AS IS" BASIS,
@rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
@rem See the License for the specific language governing permissions and
@rem limitations under the License.

@rem Set Hadoop-specific environment variables here.

@rem The only required environment variable is JAVA_HOME.  All others are
@rem optional.  When running a distributed configuration it is best to
@rem set JAVA_HOME in this file, so that it is correctly defined on
@rem remote nodes.

@rem The java implementation to use.  Required.
set JAVA_HOME=C:\Progra~1\Java\jdk-1.8

@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%

@rem set HADOOP_CONF_DIR=

@rem Extra Java CLASSPATH elements.  Automatically insert capacity-scheduler.
if exist %HADOOP_HOME%\contrib\capacity-scheduler (
```

Result:

- HDFS namenode format screen

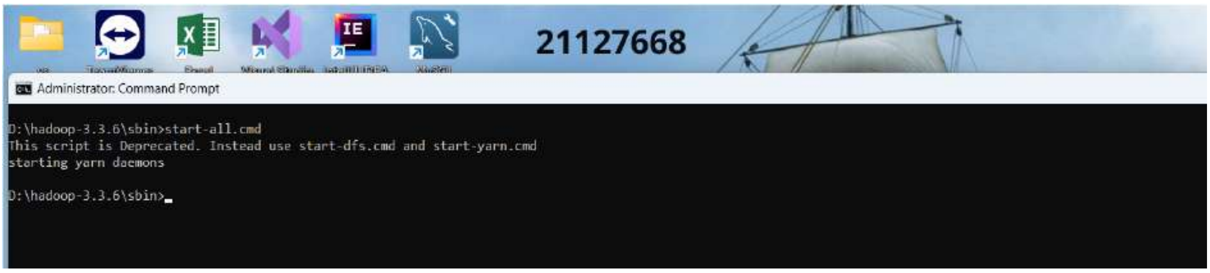


```
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>d:

D:\>cd \hadoop-3.3.6\sbin

D:\hadoop-3.3.6\sbin>hdfs namenode -format
2024-03-03 23:20:22,544 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = LAPTOP-6KGH4NMD/172.29.128.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = D:\hadoop-3.3.6\etc\hadoop;D:\hadoop-3.3.6\share\hadoop\common;D:\hadoop-3.3.6\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\audi
hare\hadoop\common\lib\commons-codec-1.15.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-collections-3.2.2.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-compress-1.21.jar;D:\hadoop-3.3.6\share\hadoop
ng-1.1.3.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-math3-3.1.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-net-3.9.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\commons-text-1.10.0.jar;D:\hado
failureaccess-1.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\gson-2.9.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\guava-27.0-jre.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\hadoop-annotations-3.3.6.jar;D:\ha
doop\common\lib\httpcore-4.4.13.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\j2objc-annotations-1.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jackson-annotations-2.12.7.jar;D:\hadoop-3.3.6\share\hadoop\common\
tion-api-1.2.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\javax.servlet-api-3.1.0.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jaxb-api-2.2.11.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jaxb-impl-2.2.3-1.jar;D
\lib\jersey-servlet-1.19.4.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jettison-1.5.4.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jetty-http-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jetty-io
ty-util-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jetty-util-ajax-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\jetty-webapp-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\c
re\hadoop\common\lib\jul-to-slf4j-1.7.36.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\kerb-admin-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\kerb-client-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\ke
hadoop\common\lib\kerb-simplekdc-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\kerb-util-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\kerby-asn1-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\kerby-
h-guava.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\metrics-core-3.2.4.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-all-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-buffer-4.1.89.Final.ja
.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-http2-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-memcache-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-
y-codec-stomp-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-codec-xml-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-common-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\
op\common\lib\netty-resolver-dns-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-resolver-dns-classes-macos-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-resolver-dns-native-maco
r;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-transport-classes-kqueue-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-transport-native-epoll-4.1.89.Final-linux-aarch_64.jar;D:\hadoop-3.3.6\sha
3.6\share\hadoop\common\lib\netty-transport-native-unix-common-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-transport-rxtx-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\netty-transp
;D:\hadoop-3.3.6\share\hadoop\common\lib\re2j-1.1.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\reload4j-1.2.22.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\slf4j-api-1.7.36.jar;D:\hadoop-3.3.6\share\hadoop\common
oop-3.3.6\share\hadoop\common\lib\zookeeper-3.6.3.jar;D:\hadoop-3.3.6\share\hadoop\common\lib\zookeeper-jute-3.6.3.jar;D:\hadoop-3.3.6\share\hadoop\common\hadoop-common-3.3.6-tests.jar;D:\hadoop-3.3.6\share\had
sniffer-annotations-1.17.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\audience-annotations-0.5.0.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\avro-1.7.7.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\checker-qual-2.5.2.ja
ib\commons-compress-1.21.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\commons-configuration2-2.8.0.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\commons-daemon-1.0.13.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\commons-
\hdfs\lib\commons-text-1.10.0.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\curator-client-5.2.0.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\curator-framework-5.2.0.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\curator-r
notations-3.3.6.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\hadoop-auth-3.3.6.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\hadoop-shaded-guava-1.1.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\hadoop-shaded-protobuf_3
dfs\lib\jackson-annotations-2.12.7.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jackson-core-2.12.7.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jackson-core-asl-1.9.13.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jacks
\hadoop-3.3.6\share\hadoop\hdfs\lib\jaxb-impl-2.2.3-1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jcip-annotations-1.0-1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jersey-core-1.19.4.jar;D:\hadoop-3.3.6\share\hado
hadoop-3.3.6\share\hadoop\hdfs\lib\jetty-io-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jetty-security-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jetty-server-9.4.51.v20230217.jar
;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jetty-xml-9.4.51.v20230217.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\jsch-0.1.55.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\json-simple-1.1.1.jar;D:\hadoop-3.3.6\share\hado
adoop\hdfs\lib\kerb-core-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\kerb-crypto-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\kerb-identity-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\kerb-server-1.0
y-pkix-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\kerby-util-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\kerby-xdr-1.0.1.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\kotlin-stdlib-1.4.10.jar;D:\hadoop-3.3
3.3.6\share\hadoop\hdfs\lib\netty-3.10.6.Final.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\netty-all-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\netty-buffer-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hado
fs\lib\netty-codec-http2-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\netty-codec-memcache-4.1.89.Final.jar;D:\hadoop-3.3.6\share\hadoop\hdfs\lib\netty-codec-mqtt-4.1.89.Final.jar;D:\hadoop-3.3.6\shar
```



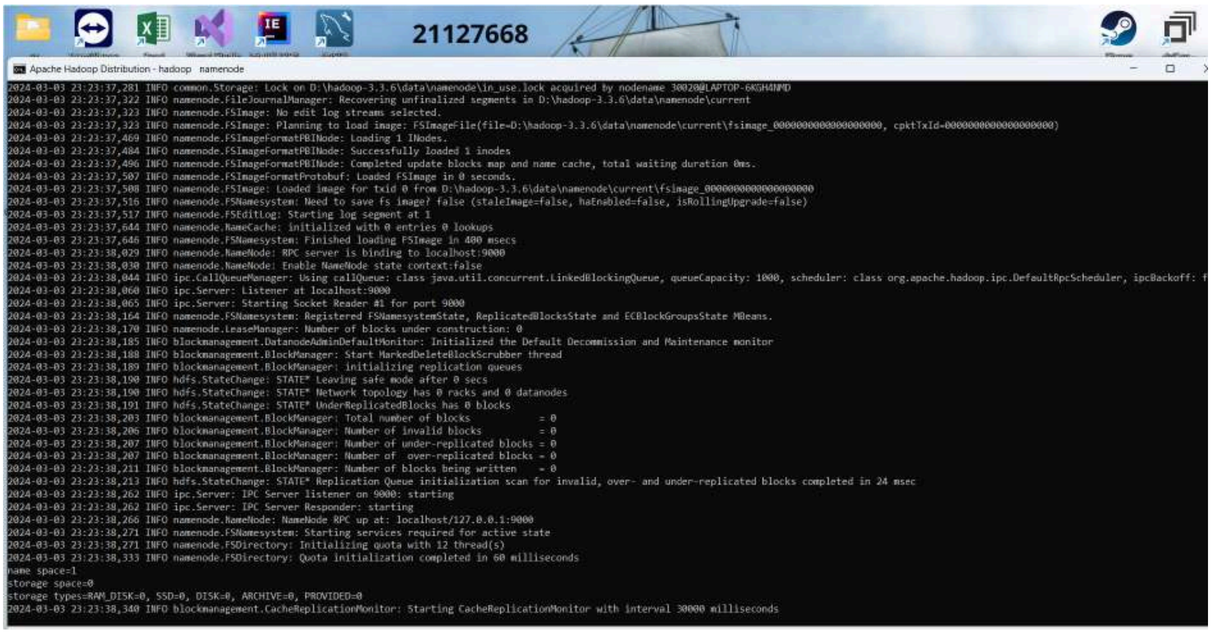
```
2024-03-03 23:20:24,026 INFO namenode.FSDirectory: USER serial map: bits=24 maxEntries=16777215
2024-03-03 23:20:24,027 INFO namenode.FSDirectory: GROUP serial map: bits=24 maxEntries=16777215
2024-03-03 23:20:24,027 INFO namenode.FSDirectory: XATTR serial map: bits=24 maxEntries=16777215
2024-03-03 23:20:24,044 INFO util.GSet: Computing capacity for map INodeMap
2024-03-03 23:20:24,044 INFO util.GSet: VM type       = 64-bit
2024-03-03 23:20:24,044 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
2024-03-03 23:20:24,044 INFO util.GSet: capacity      = 2^20 = 1048576 entries
2024-03-03 23:20:24,044 INFO namenode.FSDirectory: ACLs enabled? true
2024-03-03 23:20:24,044 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2024-03-03 23:20:24,044 INFO namenode.FSDirectory: XAttrs enabled? true
2024-03-03 23:20:24,052 INFO namenode.NameNode: Caching file names occurring more than 10 times
2024-03-03 23:20:24,052 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAccessTimeOnlyChange: false, snapshotDiffAllowSnapRootDescendant: true, maxSnapshotLimit: 65536
2024-03-03 23:20:24,060 INFO snapshot.SnapshotManager: SkipList is disabled
2024-03-03 23:20:24,062 INFO util.GSet: Computing capacity for map cachedBlocks
2024-03-03 23:20:24,062 INFO util.GSet: VM type       = 64-bit
2024-03-03 23:20:24,062 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
2024-03-03 23:20:24,062 INFO util.GSet: capacity      = 2^18 = 262144 entries
2024-03-03 23:20:24,076 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2024-03-03 23:20:24,076 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-03-03 23:20:24,076 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-03-03 23:20:24,112 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-03-03 23:20:24,112 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2024-03-03 23:20:24,123 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-03-03 23:20:24,123 INFO util.GSet: VM type       = 64-bit
2024-03-03 23:20:24,123 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
2024-03-03 23:20:24,123 INFO util.GSet: capacity      = 2^15 = 32768 entries
Re-format filesystem in Storage Directory root= D:\hadoop-3.3.6\data\namenode; location= null ? (Y or N) Y
2024-03-03 23:20:27,622 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1497848487-172.29.128.1-1709482827612
2024-03-03 23:20:27,622 INFO common.Storage: Will remove files: [D:\hadoop-3.3.6\data\namenode\current\edits_0000000000000000001-0000000000000000001, D:\hadoop-3.3.6\data\namenode\current\edits_0000000000000000
a\namenode\current\fsimage_0000000000000000009, D:\hadoop-3.3.6\data\namenode\current\fsimage_0000000000000000009.md5, D:\hadoop-3.3.6\data\namenode\current\fsimage_0000000000000000200, D:\hadoop-3.3.6\data\nam
2024-03-03 23:20:27,662 INFO common.Storage: Storage directory D:\hadoop-3.3.6\data\namenode has been successfully formatted.
2024-03-03 23:20:27,693 INFO namenode.FSImageFormatProtobuf: Saving image file D:\hadoop-3.3.6\data\namenode\current\fsimage.ckpt_0000000000000000000 using no compression
2024-03-03 23:20:27,812 INFO namenode.FSImageFormatProtobuf: Image file D:\hadoop-3.3.6\data\namenode\current\fsimage.ckpt_0000000000000000000 of size 399 bytes saved in 0 seconds .
2024-03-03 23:20:27,822 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-03-03 23:20:27,842 INFO namenode.FSNamesystem: Stopping services started for active state
2024-03-03 23:20:27,842 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-03-03 23:20:27,851 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-03-03 23:20:27,852 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at LAPTOP-6KGH4NMD/172.29.128.1
************************************************************/

D:\hadoop-3.3.6\sbin>
```

○ start-all.cmd screen



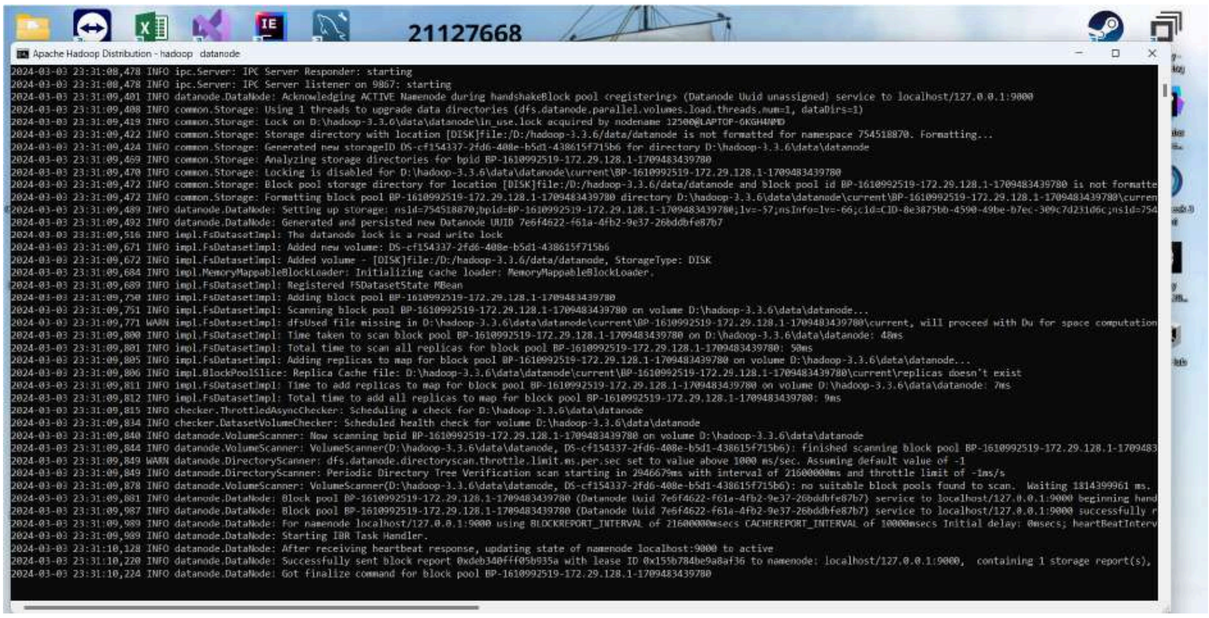
```
D:\hadoop-3.3.6\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop-3.3.6\sbin>
```

○ namenode screen



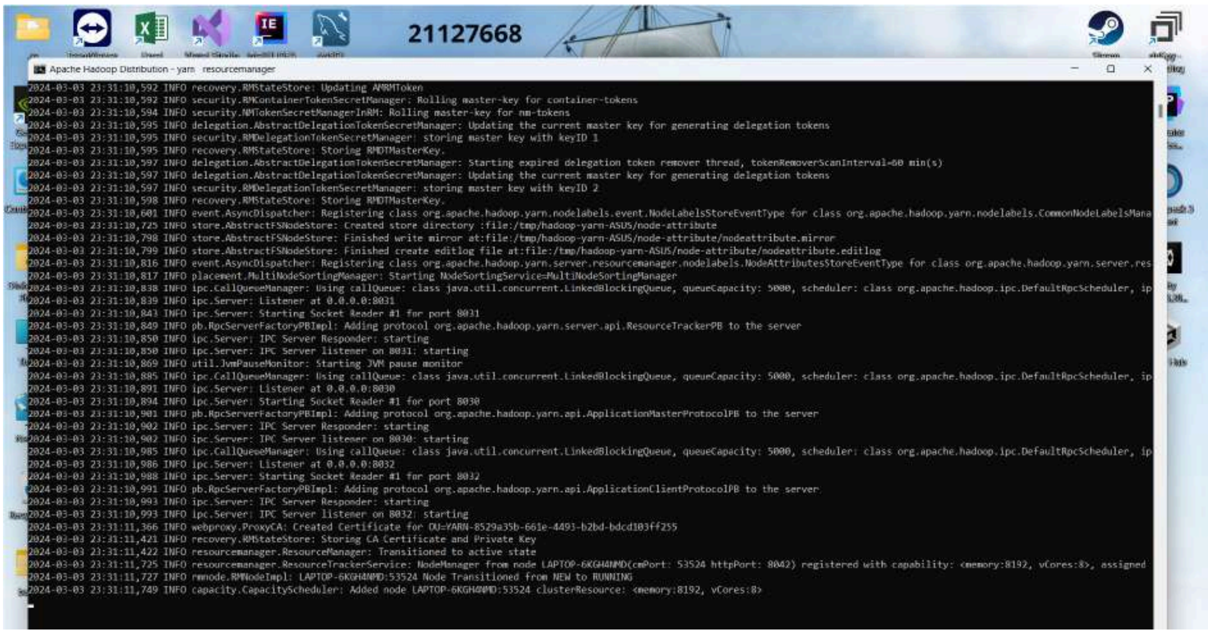
```
2024-03-03 23:23:37,281 INFO common.Storage: Lock on D:\hadoop-3.3.6\data\namenode\in_use.lock acquired by nodename 30020@LAPTOP-6KGH4NMD
2024-03-03 23:23:37,322 INFO namenode.FileJournalManager: Recovering unfinalized segments in D:\hadoop-3.3.6\data\namenode\current
2024-03-03 23:23:37,323 INFO namenode.FSImage: No edit log streams selected.
2024-03-03 23:23:37,323 INFO namenode.FSImage: Planning to load image: FSImageFile(file=D:\hadoop-3.3.6\data\namenode\current\fsimage_0000000000000000000, cpktTxId=0000000000000000000)
2024-03-03 23:23:37,469 INFO namenode.FSImageFormatPBINode: Loading 1 INodes.
2024-03-03 23:23:37,484 INFO namenode.FSImageFormatPBINode: Successfully loaded 1 inodes
2024-03-03 23:23:37,496 INFO namenode.FSImageFormatPBINode: Completed update blocks map and name cache, total waiting duration 0ms.
2024-03-03 23:23:37,507 INFO namenode.FSImageFormatProtobuf: Loaded FSImage in 0 seconds.
2024-03-03 23:23:37,508 INFO namenode.FSImage: Loaded image for txid 0 from D:\hadoop-3.3.6\data\namenode\current\fsimage_0000000000000000000
2024-03-03 23:23:37,516 INFO namenode.FSNamesystem: Need to save fs image? false (staleImage=false, haEnabled=false, isRollingUpgrade=false)
2024-03-03 23:23:37,517 INFO namenode.FSEditLog: Starting log segment at 1
2024-03-03 23:23:37,644 INFO namenode.NameCache: initialized with 0 entries 0 lookups
2024-03-03 23:23:37,646 INFO namenode.FSNamesystem: Finished loading FSImage in 400 msecs
2024-03-03 23:23:38,029 INFO namenode.NameNode: RPC server is binding to localhost:9000
2024-03-03 23:23:38,030 INFO namenode.NameNode: Enable NameNode state context:false
2024-03-03 23:23:38,044 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 1000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: f
2024-03-03 23:23:38,060 INFO ipc.Server: Listener at localhost:9000
2024-03-03 23:23:38,065 INFO ipc.Server: Starting Socket Reader #1 for port 9000
2024-03-03 23:23:38,164 INFO namenode.FSNamesystem: Registered FSNamesystemState, ReplicatedBlocksState and ECBlockGroupsState MBeans.
2024-03-03 23:23:38,170 INFO namenode.LeaseManager: Number of blocks under construction: 0
2024-03-03 23:23:38,185 INFO blockmanagement.DatanodeAdminDefaultMonitor: Initialized the Default Decommission and Maintenance monitor
2024-03-03 23:23:38,188 INFO blockmanagement.BlockManager: Start MarkedDeleteBlockScrubber thread
2024-03-03 23:23:38,189 INFO blockmanagement.BlockManager: initializing replication queues
2024-03-03 23:23:38,190 INFO hdfs.StateChange: STATE* Leaving safe mode after 0 secs
2024-03-03 23:23:38,190 INFO hdfs.StateChange: STATE* Network topology has 0 racks and 0 datanodes
2024-03-03 23:23:38,191 INFO hdfs.StateChange: STATE* UnderReplicatedBlocks has 0 blocks
2024-03-03 23:23:38,203 INFO blockmanagement.BlockManager: Total number of blocks            = 0
2024-03-03 23:23:38,206 INFO blockmanagement.BlockManager: Number of invalid blocks          = 0
2024-03-03 23:23:38,207 INFO blockmanagement.BlockManager: Number of under-replicated blocks = 0
2024-03-03 23:23:38,207 INFO blockmanagement.BlockManager: Number of  over-replicated blocks = 0
2024-03-03 23:23:38,211 INFO blockmanagement.BlockManager: Number of blocks being written    = 0
2024-03-03 23:23:38,213 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over- and under-replicated blocks completed in 24 msec
2024-03-03 23:23:38,262 INFO ipc.Server: IPC Server listener on 9000: starting
2024-03-03 23:23:38,262 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 23:23:38,266 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000
2024-03-03 23:23:38,271 INFO namenode.FSNamesystem: Starting services required for active state
2024-03-03 23:23:38,271 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)
2024-03-03 23:23:38,333 INFO namenode.FSDirectory: Quota initialization completed in 60 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2024-03-03 23:23:38,340 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds
```

- datanode screen



```
2024-03-03 23:31:08,478 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 23:31:08,478 INFO ipc.Server: IPC Server listener on 9867: starting
2024-03-03 23:31:09,401 INFO datanode.DataNode: Acknowledging ACTIVE Namenode during handshakeBlock pool <registering> (Datanode Uuid unassigned) service to localhost/127.0.0.1:9000
2024-03-03 23:31:09,408 INFO common.Storage: Using 1 threads to upgrade data directories (dfs.datanode.parallel.volumes.load.threads.num=1, dataDirs=1)
2024-03-03 23:31:09,419 INFO common.Storage: Lock on D:\hadoop-3.3.6\data\datanode\in_use.lock acquired by nodename 12500@LAPTOP-6KGH4NMD
2024-03-03 23:31:09,422 INFO common.Storage: Storage directory with location [DISK]file:/D:/hadoop-3.3.6/data/datanode is not formatted for namespace 754518870. Formatting...
2024-03-03 23:31:09,424 INFO common.Storage: Generated new storageID DS-cf154337-2fd6-408e-b5d1-438615f715b6 for directory D:\hadoop-3.3.6\data\datanode
2024-03-03 23:31:09,469 INFO common.Storage: Analyzing storage directories for bpid BP-1610992519-172.29.128.1-1709483439780
2024-03-03 23:31:09,470 INFO common.Storage: Locking is disabled for D:\hadoop-3.3.6\data\datanode\current\BP-1610992519-172.29.128.1-1709483439780
2024-03-03 23:31:09,472 INFO common.Storage: Block pool storage directory for location [DISK]file:/D:/hadoop-3.3.6/data/datanode and block pool id BP-1610992519-172.29.128.1-1709483439780 is not formatte
2024-03-03 23:31:09,472 INFO common.Storage: Formatting block pool BP-1610992519-172.29.128.1-1709483439780 directory D:\hadoop-3.3.6\data\datanode\current\BP-1610992519-172.29.128.1-1709483439780\curren
2024-03-03 23:31:09,489 INFO datanode.DataNode: Setting up storage: nsid=754518870;bpid=BP-1610992519-172.29.128.1-1709483439780;lv=-57;nsInfo=lv=-66;cid=CID-8e3875bb-4590-49be-b7ec-309c7d231d6c;nsid=754
2024-03-03 23:31:09,492 INFO datanode.DataNode: Generated and persisted new Datanode UUID 7e6f4622-f61a-4fb2-9e37-26bddbfe87b7
2024-03-03 23:31:09,516 INFO impl.FsDatasetImpl: The datanode lock is a read write lock
2024-03-03 23:31:09,671 INFO impl.FsDatasetImpl: Added new volume: DS-cf154337-2fd6-408e-b5d1-438615f715b6
2024-03-03 23:31:09,672 INFO impl.FsDatasetImpl: Added volume - [DISK]file:/D:/hadoop-3.3.6/data/datanode, StorageType: DISK
2024-03-03 23:31:09,684 INFO impl.MemoryMappableBlockLoader: Initializing cache loader: MemoryMappableBlockLoader.
2024-03-03 23:31:09,689 INFO impl.FsDatasetImpl: Registered FSDatasetState MBean
2024-03-03 23:31:09,750 INFO impl.FsDatasetImpl: Adding block pool BP-1610992519-172.29.128.1-1709483439780
2024-03-03 23:31:09,751 INFO impl.FsDatasetImpl: Scanning block pool BP-1610992519-172.29.128.1-1709483439780 on volume D:\hadoop-3.3.6\data\datanode...
2024-03-03 23:31:09,771 WARN impl.FsDatasetImpl: dfsUsed file missing in D:\hadoop-3.3.6\data\datanode\current\BP-1610992519-172.29.128.1-1709483439780\current, will proceed with Du for space computation
2024-03-03 23:31:09,800 INFO impl.FsDatasetImpl: Time taken to scan block pool BP-1610992519-172.29.128.1-1709483439780 on D:\hadoop-3.3.6\data\datanode: 48ms
2024-03-03 23:31:09,801 INFO impl.FsDatasetImpl: Total time to scan all replicas for block pool BP-1610992519-172.29.128.1-1709483439780: 50ms
2024-03-03 23:31:09,805 INFO impl.FsDatasetImpl: Adding replicas to map for block pool BP-1610992519-172.29.128.1-1709483439780 on volume D:\hadoop-3.3.6\data\datanode...
2024-03-03 23:31:09,806 INFO impl.BlockPoolSlice: Replica Cache file: D:\hadoop-3.3.6\data\datanode\current\BP-1610992519-172.29.128.1-1709483439780\current\replicas doesn't exist
2024-03-03 23:31:09,811 INFO impl.FsDatasetImpl: Time to add replicas to map for block pool BP-1610992519-172.29.128.1-1709483439780 on volume D:\hadoop-3.3.6\data\datanode: 7ms
2024-03-03 23:31:09,812 INFO impl.FsDatasetImpl: Total time to add all replicas to map for block pool BP-1610992519-172.29.128.1-1709483439780: 9ms
2024-03-03 23:31:09,815 INFO checker.ThrottledAsyncChecker: Scheduling a check for D:\hadoop-3.3.6\data\datanode
2024-03-03 23:31:09,834 INFO checker.DatasetVolumeChecker: Scheduled health check for volume D:\hadoop-3.3.6\data\datanode
2024-03-03 23:31:09,840 INFO datanode.VolumeScanner: Now scanning bpid BP-1610992519-172.29.128.1-1709483439780 on volume D:\hadoop-3.3.6\data\datanode
2024-03-03 23:31:09,844 INFO datanode.VolumeScanner: VolumeScanner(D:\hadoop-3.3.6\data\datanode, DS-cf154337-2fd6-408e-b5d1-438615f715b6): finished scanning block pool BP-1610992519-172.29.128.1-1709483
2024-03-03 23:31:09,849 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to value above 1000 ms/sec. Assuming default value of -1
2024-03-03 23:31:09,849 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 2946679ms with interval of 21600000ms and throttle limit of -1ms/s
2024-03-03 23:31:09,878 INFO datanode.VolumeScanner: VolumeScanner(D:\hadoop-3.3.6\data\datanode, DS-cf154337-2fd6-408e-b5d1-438615f715b6): no suitable block pools found to scan.  Waiting 1814399961 ms.
2024-03-03 23:31:09,881 INFO datanode.DataNode: Block pool BP-1610992519-172.29.128.1-1709483439780 (Datanode Uuid 7e6f4622-f61a-4fb2-9e37-26bddbfe87b7) service to localhost/127.0.0.1:9000 beginning hand
2024-03-03 23:31:09,987 INFO datanode.DataNode: Block pool BP-1610992519-172.29.128.1-1709483439780 (Datanode Uuid 7e6f4622-f61a-4fb2-9e37-26bddbfe87b7) service to localhost/127.0.0.1:9000 successfully r
2024-03-03 23:31:09,989 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 21600000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartBeatInterv
2024-03-03 23:31:09,989 INFO datanode.DataNode: Starting IBR Task Handler.
2024-03-03 23:31:10,128 INFO datanode.DataNode: After receiving heartbeat response, updating state of namenode localhost:9000 to active
2024-03-03 23:31:10,220 INFO datanode.DataNode: Successfully sent block report 0xdeb340fff05b935a with lease ID 0x155b784be9a8af36 to namenode: localhost/127.0.0.1:9000,  containing 1 storage report(s),
2024-03-03 23:31:10,224 INFO datanode.DataNode: Got finalize command for block pool BP-1610992519-172.29.128.1-1709483439780
```

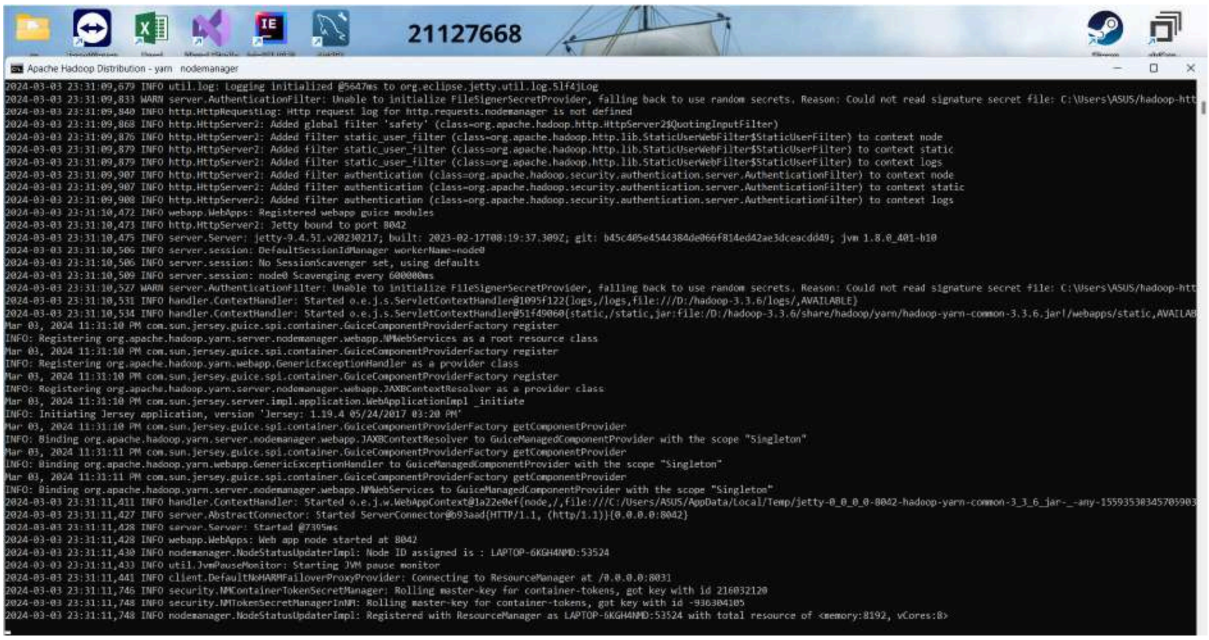
- resource manager screen



Apache Hadoop Distribution - yarn resourcemanager

```
2024-03-03 23:31:10,592 INFO recovery.RMStateStore: Updating AMRMToken
2024-03-03 23:31:10,592 INFO security.RMContainerTokenSecretManager: Rolling master-key for container-tokens
2024-03-03 23:31:10,594 INFO security.NMTokenSecretManagerInRM: Rolling master-key for nm-tokens
2024-03-03 23:31:10,595 INFO delegation.AbstractDelegationTokenSecretManager: Updating the current master key for generating delegation tokens
2024-03-03 23:31:10,595 INFO security.RMDelegationTokenSecretManager: storing master key with keyID 1
2024-03-03 23:31:10,595 INFO recovery.RMStateStore: Storing RMDTMasterKey.
2024-03-03 23:31:10,597 INFO delegation.AbstractDelegationTokenSecretManager: Starting expired delegation token remover thread, tokenRemoverScanInterval=60 min(s)
2024-03-03 23:31:10,597 INFO delegation.AbstractDelegationTokenSecretManager: Updating the current master key for generating delegation tokens
2024-03-03 23:31:10,597 INFO security.RMDelegationTokenSecretManager: storing master key with keyID 2
2024-03-03 23:31:10,598 INFO recovery.RMStateStore: Storing RMDTMasterKey.
2024-03-03 23:31:10,601 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.nodelabels.event.NodeLabelsStoreEventType for class org.apache.hadoop.yarn.nodelabels.CommonNodeLabelsMana
2024-03-03 23:31:10,725 INFO store.AbstractFSNodeStore: Created store directory :file:/tmp/hadoop-yarn-ASUS/node-attribute
2024-03-03 23:31:10,798 INFO store.AbstractFSNodeStore: Finished write mirror at:file:/tmp/hadoop-yarn-ASUS/node-attribute/nodeattribute.mirror
2024-03-03 23:31:10,799 INFO store.AbstractFSNodeStore: Finished create editlog file at:file:/tmp/hadoop-yarn-ASUS/node-attribute/nodeattribute.editlog
2024-03-03 23:31:10,816 INFO event.AsyncDispatcher: Registering class org.apache.hadoop.yarn.server.resourcemanager.nodelabels.NodeAttributesStoreEventType for class org.apache.hadoop.yarn.server.res
2024-03-03 23:31:10,817 INFO placement.MultiNodeSortingManager: Starting NodeSortingService=MultiNodeSortingManager
2024-03-03 23:31:10,838 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ip
2024-03-03 23:31:10,839 INFO ipc.Server: Listener at 0.0.0.0:8031
2024-03-03 23:31:10,843 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2024-03-03 23:31:10,849 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
2024-03-03 23:31:10,850 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 23:31:10,850 INFO ipc.Server: IPC Server listener on 8031: starting
2024-03-03 23:31:10,869 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-03-03 23:31:10,885 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ip
2024-03-03 23:31:10,891 INFO ipc.Server: Listener at 0.0.0.0:8030
2024-03-03 23:31:10,894 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-03-03 23:31:10,901 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2024-03-03 23:31:10,902 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 23:31:10,902 INFO ipc.Server: IPC Server listener on 8030: starting
2024-03-03 23:31:10,985 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ip
2024-03-03 23:31:10,986 INFO ipc.Server: Listener at 0.0.0.0:8032
2024-03-03 23:31:10,988 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-03-03 23:31:10,991 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2024-03-03 23:31:10,993 INFO ipc.Server: IPC Server Responder: starting
2024-03-03 23:31:10,993 INFO ipc.Server: IPC Server listener on 8032: starting
2024-03-03 23:31:11,366 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-8529a35b-661e-4493-b2bd-bdcd103ff255
2024-03-03 23:31:11,421 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2024-03-03 23:31:11,422 INFO resourcemanager.ResourceManager: Transitioned to active state
2024-03-03 23:31:11,725 INFO resourcemanager.ResourceTrackerService: NodeManager from node LAPTOP-6KGH4NMD(cmPort: 53524 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned
2024-03-03 23:31:11,727 INFO rmnode.RMNodeImpl: LAPTOP-6KGH4NMD:53524 Node Transitioned from NEW to RUNNING
2024-03-03 23:31:11,749 INFO capacity.CapacityScheduler: Added node LAPTOP-6KGH4NMD:53524 clusterResource: <memory:8192, vCores:8>
```

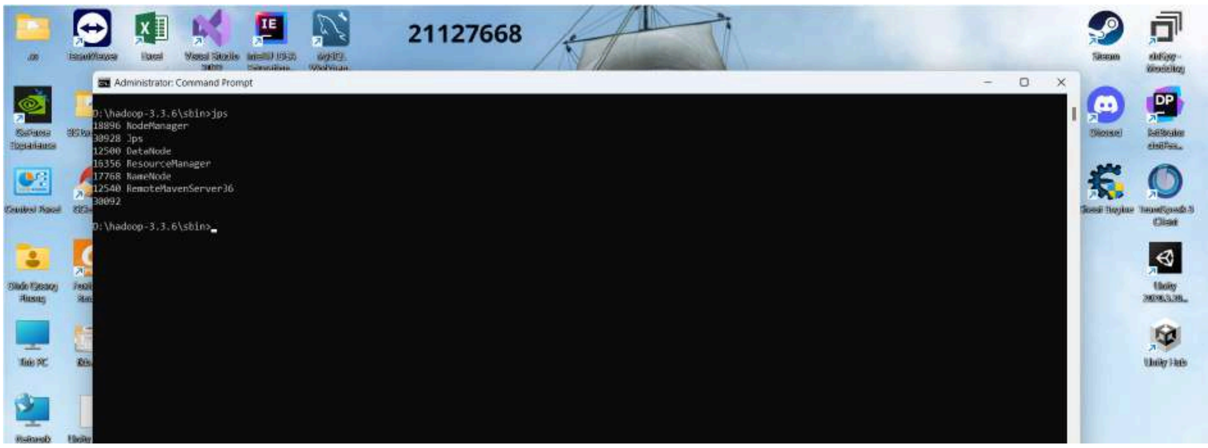
- node manager screen



Apache Hadoop Distribution - yarn nodemanager

```
2024-03-03 23:31:09,679 INFO util.log: Logging initialized @5647ms to org.eclipse.jetty.util.log.Slf4jLog
2024-03-03 23:31:09,833 WARN server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets. Reason: Could not read signature secret file: C:\Users\ASUS/hadoop-htt
2024-03-03 23:31:09,840 INFO http.HttpRequestLog: Http request log for http.requests.nodemanager is not defined
2024-03-03 23:31:09,868 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingInputFilter)
2024-03-03 23:31:09,876 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context node
2024-03-03 23:31:09,879 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context static
2024-03-03 23:31:09,879 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
2024-03-03 23:31:09,907 INFO http.HttpServer2: Added filter authentication (class=org.apache.hadoop.security.authentication.server.AuthenticationFilter) to context node
2024-03-03 23:31:09,907 INFO http.HttpServer2: Added filter authentication (class=org.apache.hadoop.security.authentication.server.AuthenticationFilter) to context static
2024-03-03 23:31:09,908 INFO http.HttpServer2: Added filter authentication (class=org.apache.hadoop.security.authentication.server.AuthenticationFilter) to context logs
2024-03-03 23:31:10,472 INFO webapp.WebApps: Registered webapp guice modules
2024-03-03 23:31:10,473 INFO http.HttpServer2: Jetty bound to port 8042
2024-03-03 23:31:10,475 INFO server.Server: jetty-9.4.51.v20230217; built: 2023-02-17T08:19:37.309Z; git: b45c405e4544384de066f814ed42ae3dceacdd49; jvm 1.8.0_401-b10
2024-03-03 23:31:10,506 INFO server.session: DefaultSessionIdManager workerName=node0
2024-03-03 23:31:10,506 INFO server.session: No SessionScavenger set, using defaults
2024-03-03 23:31:10,509 INFO server.session: node0 Scavenging every 600000ms
2024-03-03 23:31:10,527 WARN server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets. Reason: Could not read signature secret file: C:\Users\ASUS/hadoop-htt
2024-03-03 23:31:10,531 INFO handler.ContextHandler: Started o.e.j.s.ServletContextHandler@1095f122{logs,/logs,file:///D:/hadoop-3.3.6/logs/,AVAILABLE}
2024-03-03 23:31:10,534 INFO handler.ContextHandler: Started o.e.j.s.ServletContextHandler@51f49060{static,/static,jar:file:/D:/hadoop-3.3.6/share/hadoop/yarn/hadoop-yarn-common-3.3.6.jar!/webapps/static,AVAILAB
Mar 03, 2024 11:31:10 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices as a root resource class
Mar 03, 2024 11:31:10 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
Mar 03, 2024 11:31:10 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Mar 03, 2024 11:31:10 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19.4 05/24/2017 03:20 PM'
Mar 03, 2024 11:31:10 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Mar 03, 2024 11:31:11 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Mar 03, 2024 11:31:11 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
2024-03-03 23:31:11,411 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@1a22e0ef{node,/,file:///C:/Users/ASUS/AppData/Local/Temp/jetty-0_0_0_0-8042-hadoop-yarn-common-3_3_6_jar-_-any-15593530345705903
2024-03-03 23:31:11,427 INFO server.AbstractConnector: Started ServerConnector@b93aad{HTTP/1.1, (http/1.1)}{0.0.0.0:8042}
2024-03-03 23:31:11,428 INFO server.Server: Started @7395ms
2024-03-03 23:31:11,428 INFO webapp.WebApps: Web app node started at 8042
2024-03-03 23:31:11,430 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : LAPTOP-6KGH4NMD:53524
2024-03-03 23:31:11,433 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-03-03 23:31:11,441 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2024-03-03 23:31:11,746 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 216032120
2024-03-03 23:31:11,748 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id -936304105
2024-03-03 23:31:11,748 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as LAPTOP-6KGH4NMD:53524 with total resource of <memory:8192, vCores:8>
```

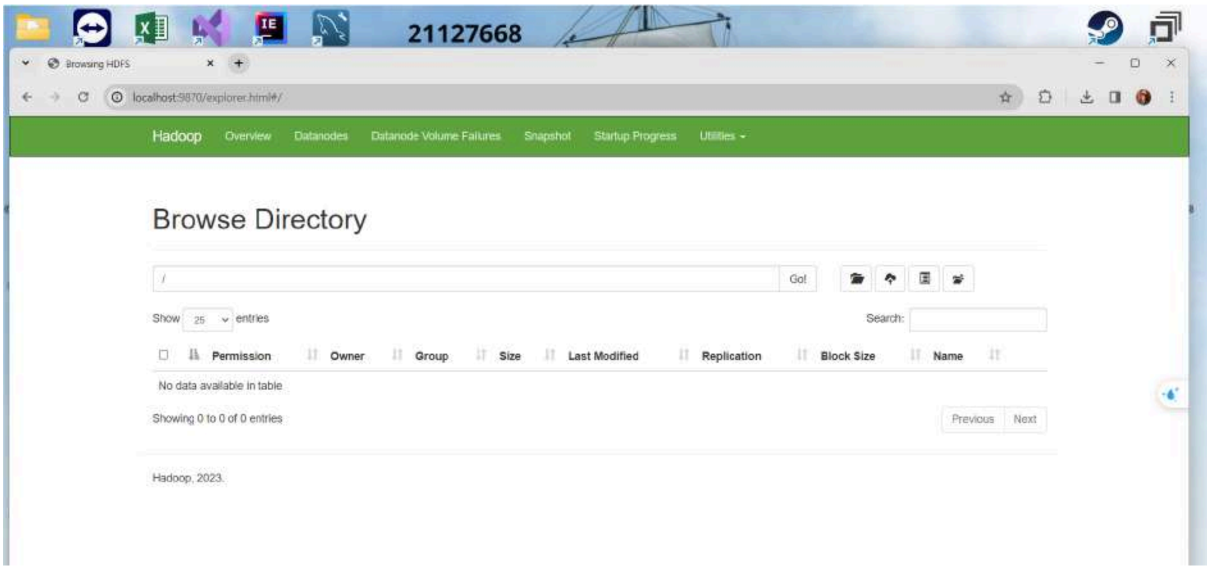
- jps screen



Administrator: Command Prompt

```
D:\hadoop-3.3.6\sbin>jps
18896 NodeManager
30928 Jps
12500 DataNode
16356 ResourceManager
17768 NameNode
12540 RemoteMavenServer36
30092

D:\hadoop-3.3.6\sbin>
```

- On localhost:9870 of Hadoop File System



Browse Directory

/

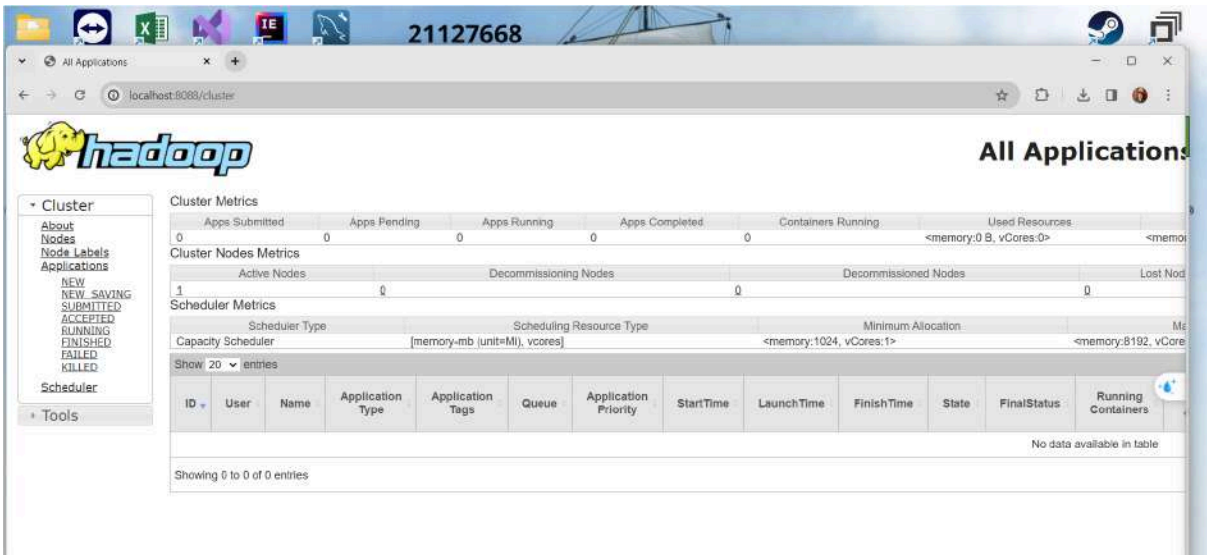
Show 25 entries Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
No data available in table								

Showing 0 to 0 of 0 entries

Hadoop, 2023.

- On localhost:8088 is cluster app of hadoop



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	
0	0	0	0	0	<memory:0 B, vCores:0>	<memor

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nod
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Ma
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCore

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers
No data available in table												

Showing 0 to 0 of 0 entries

b. The problem and solution

NO	Problem	Solution
1	Hadoop's bin folder is incomplete, leading to errors during execution and automatic shutdown of the server	Search on stackoverflow and hadoop related websites to add components in bin folder for completeness
2	The version of Java does not match hadoop leading to	Find the version of Java suitable for Hadoop: Java

	ERROR namenode.NameNode: Failed to start namenode.	1.8.0, The Oracle JDK 8 license changed in April 2019
3	File yarn.cmd of sbin can not run when executing the start-all.cmd command	Execute the command on the terminal with administrator rights

2. Introduction to MapReduce:

a. How do the input keys-values, the intermediate keys-values, and the output keys-values relate?

- The computation takes a set of input key/value pairs, and produces a set of output key/value pairs. The user of the MapReduce library expresses the computation as two functions: Map and Reduce.
  - Map, written by the user, takes an input pair and produces a set of intermediate key/value pairs. The MapReduce library groups together all intermediate values associated with the same intermediate key I and passes them to the Reduce function.
  - The Reduce function, also written by the user, accepts an intermediate key I and a set of values for that key. It merges together these values to form a possibly smaller set of values. Typically just zero or one output value is produced per Reduce invocation. The intermediate values are supplied to the user's reduce function via an iterator. This allows us to handle lists of values that are too large to fit in memory.

➔ In summary, the input keys-values are transformed into intermediate keys-values by the map tasks, and then these intermediate key-value pairs are processed by the reduce tasks to produce the final output keys-values. The relationship between them is characterized by the flow of data through the MapReduce computation pipeline.

b. How does MapReduce deal with node failures?

- The master periodically pings workers to monitor their status. If a worker does not respond for a certain period of time, it is marked as failed by the master.
- Any map tasks that were in progress or completed on the failed worker are reset to their initial idle state and become eligible for rescheduling on other workers. This avoids losing work due to failures.
- Completed reduce tasks do not need to be re-executed since their output is stored in a global file system not local to any worker node. This output is not impacted by worker failures.

- When a map task is re-executed on a new worker after the original worker failed, all reduced tasks are notified so they fetch the updated output rather than potentially outdated data from the failed worker.
- Large scale failures of groups of workers can be tolerated as failed work is simply rescheduled on remaining healthy nodes, allowing jobs to complete even if large portions of the cluster temporarily fail.
- The use of atomic task outputs and failure notifications ensures the overall results meet the expected semantics - they are equivalent to non-faulty sequential execution if tasks are deterministic, or provide reasonable semantics otherwise.

c. What is the meaning and implication of locality? What does it use?

- The workings of the MapReduce framework within a computing environment utilizing the Google File System (GFS), "locality" refers to the principle of executing computational tasks (specifically map tasks in this case) on machines where the required input data is already stored locally, i.e., on the same physical node or within the same network proximity. By leveraging data locality, tasks can read their input directly from local storage rather than requiring data transfer over the network.
- Implication: By processing data locally, MapReduce avoids the need to transfer large datasets across the network, which can be slow and resource-intensive. This leads to:
    - Reduce network traffic: Transferring massive amounts of data across the network can be slow and resource-intensive. Locality minimizes this by processing data on the node where it resides, significantly reducing network bandwidth usage.
    - Improve performance: By avoiding unnecessary data movement, locality leads to faster overall job execution time. This is crucial for large-scale data processing where efficiency is paramount.
    - Increase scalability: By reducing network load, MapReduce can handle larger datasets and workloads more efficiently, contributing to better scalability of the system.
- Usage: The MapReduce scheduler aims to place tasks on nodes with local input replicas. If not possible locally, it tries to schedule tasks on nearby nodes, e.g. same network switch/rack, to improve data locality.

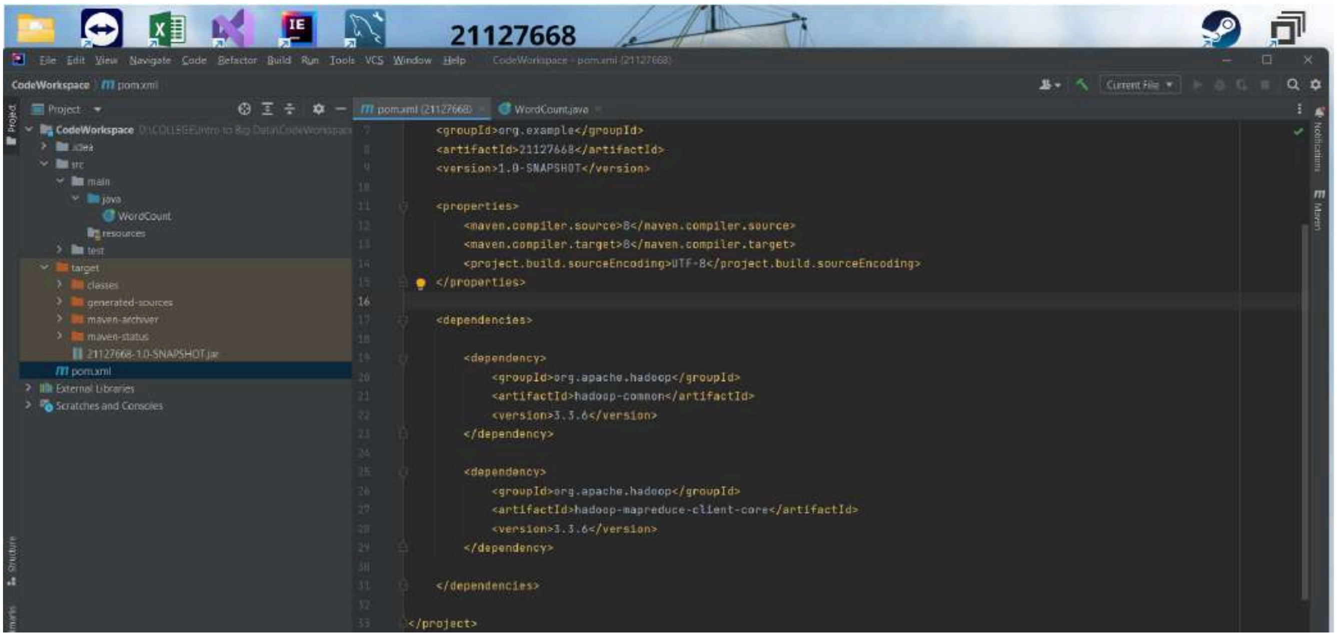
d. Which problem is addressed by introducing a combiner function to the MapReduce model?

- Introducing a combiner function to the MapReduce model addresses the problem of excessive data transfer over the network, especially when there is significant repetition in the intermediate keys produced by each map task.

- In MapReduce, after the map phase, intermediate key-value pairs are shuffled and sorted before being passed to the reduce tasks. However, in scenarios like word counting, where many identical intermediate key-value pairs are generated by map tasks (e.g., <the, 1>), sending all these pairs over the network to a single reduced task can result in substantial network traffic and potentially lead to network congestion and increased processing latency.
- The combiner function acts as a partial aggregation stage before sending data over the network. It reduces the number of key-value pairs with the same key by combining their values (e.g., summing "the" counts from different mappers). This significantly reduces network traffic and improves overall performance.

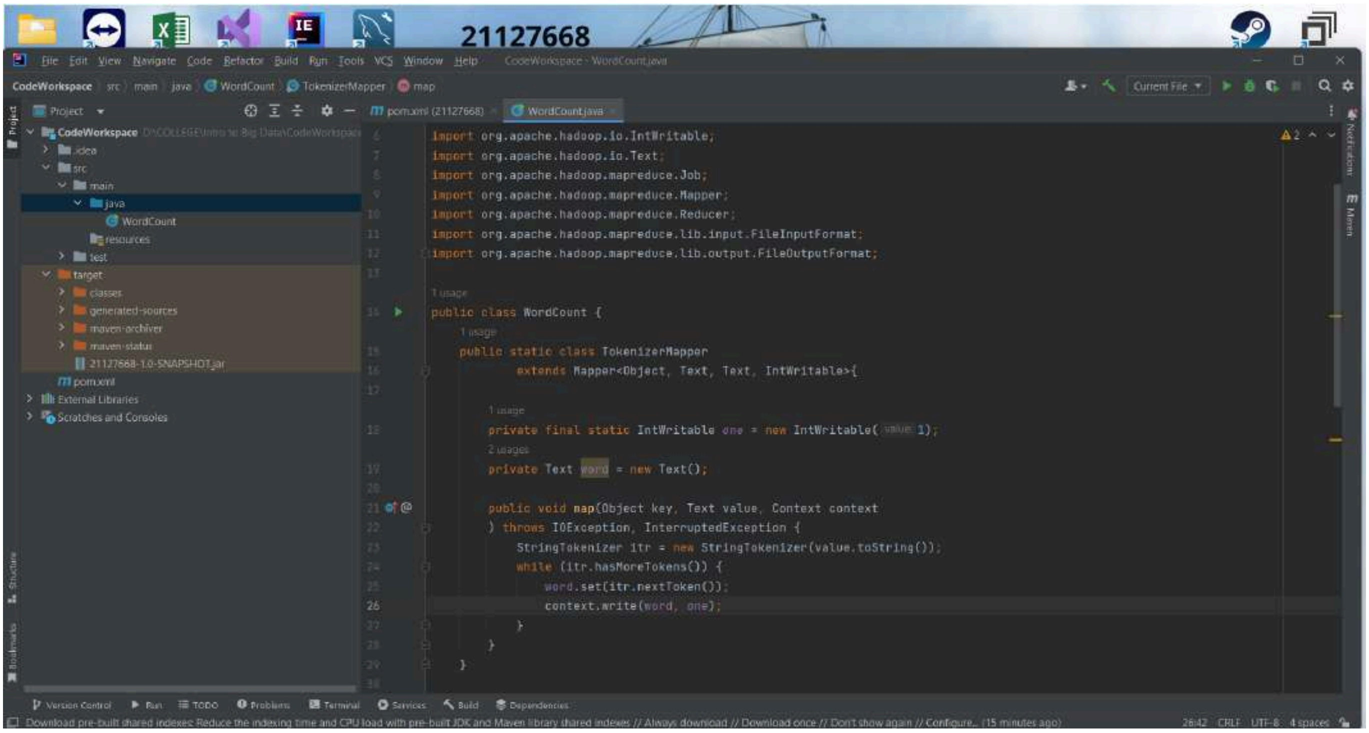
3. Running a warm-up problem: Word Count

- Add dependencies to Maven pom.xml file after creating the Maven project.



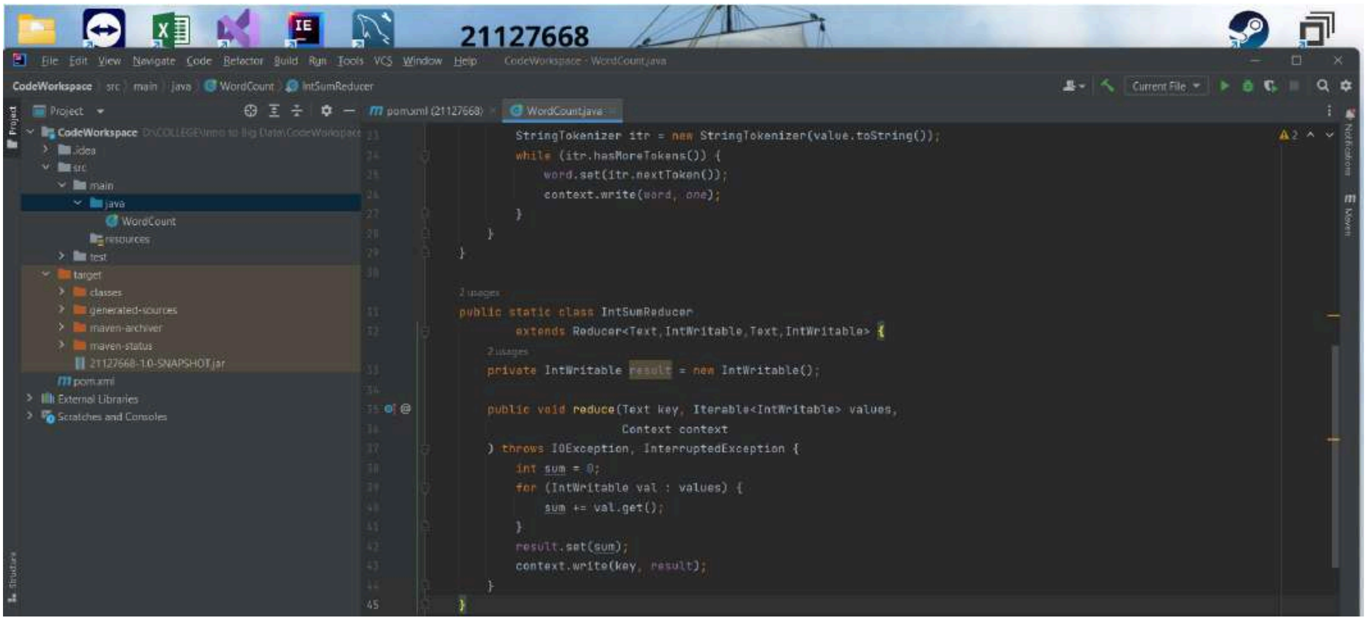
```
    <groupId>org.example</groupId>
    <artifactId>21127668</artifactId>
    <version>1.0-SNAPSHOT</version>

    <properties>
        <maven.compiler.source>8</maven.compiler.source>
        <maven.compiler.target>8</maven.compiler.target>
        <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
    </properties>

    <dependencies>

        <dependency>
            <groupId>org.apache.hadoop</groupId>
            <artifactId>hadoop-common</artifactId>
            <version>3.3.6</version>
        </dependency>

        <dependency>
            <groupId>org.apache.hadoop</groupId>
            <artifactId>hadoop-mapreduce-client-core</artifactId>
            <version>3.3.6</version>
        </dependency>

    </dependencies>

</project>
```

- Write Mapper class. In the context of the WordCount program, the input data is a text file. The Mapper takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Specifically, the Mapper reads a line of text, breaks it into words, and for each word, it emits a key-value pair, where the key is the word and the value is 1.



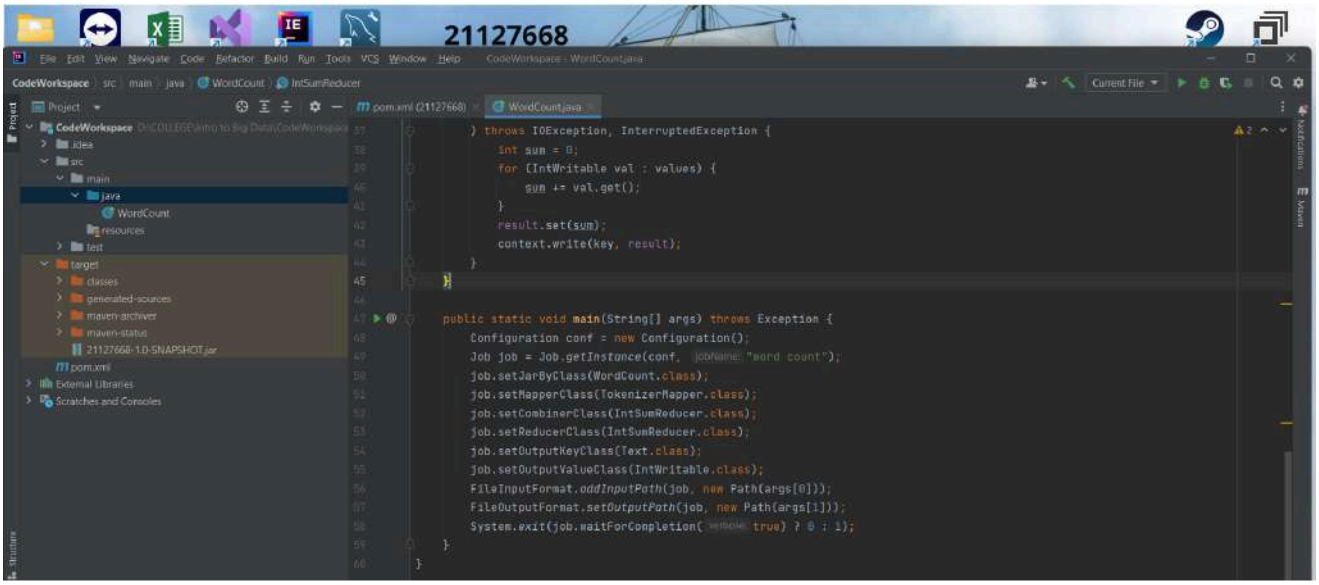
```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {
    public static class TokenizerMapper
            extends Mapper<Object, Text, Text, IntWritable>{

        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Context context
        ) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }
```

- Write Reducer class to take the output from the Mapper as input and combine those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job. In the context of the WordCount program, the Reducer takes the input from the Mapper (the key-value pairs), sums up the values for each unique key (the word), and emits a key-value pair, where the key is the word and the value is the total count



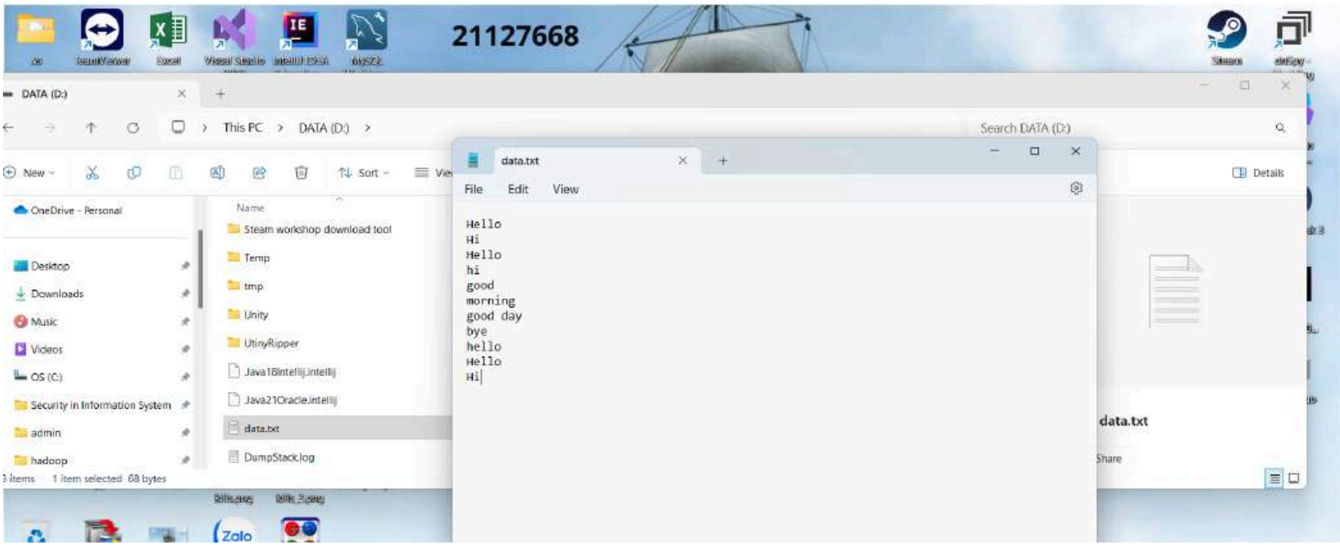
```
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }

    public static class IntSumReducer
            extends Reducer<Text,IntWritable,Text,IntWritable> {
        private IntWritable result = new IntWritable();

        public void reduce(Text key, Iterable<IntWritable> values,
                           Context context
        ) throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }
```

- Write WordCount main function, sets up the configuration for the job, specifies the input and output paths, sets the Mapper and Reducer classes, and submits the job to the Hadoop cluster. Once the job is submitted, Hadoop takes care of distributing the data, scheduling and running the map and reduce tasks, and collecting the results.



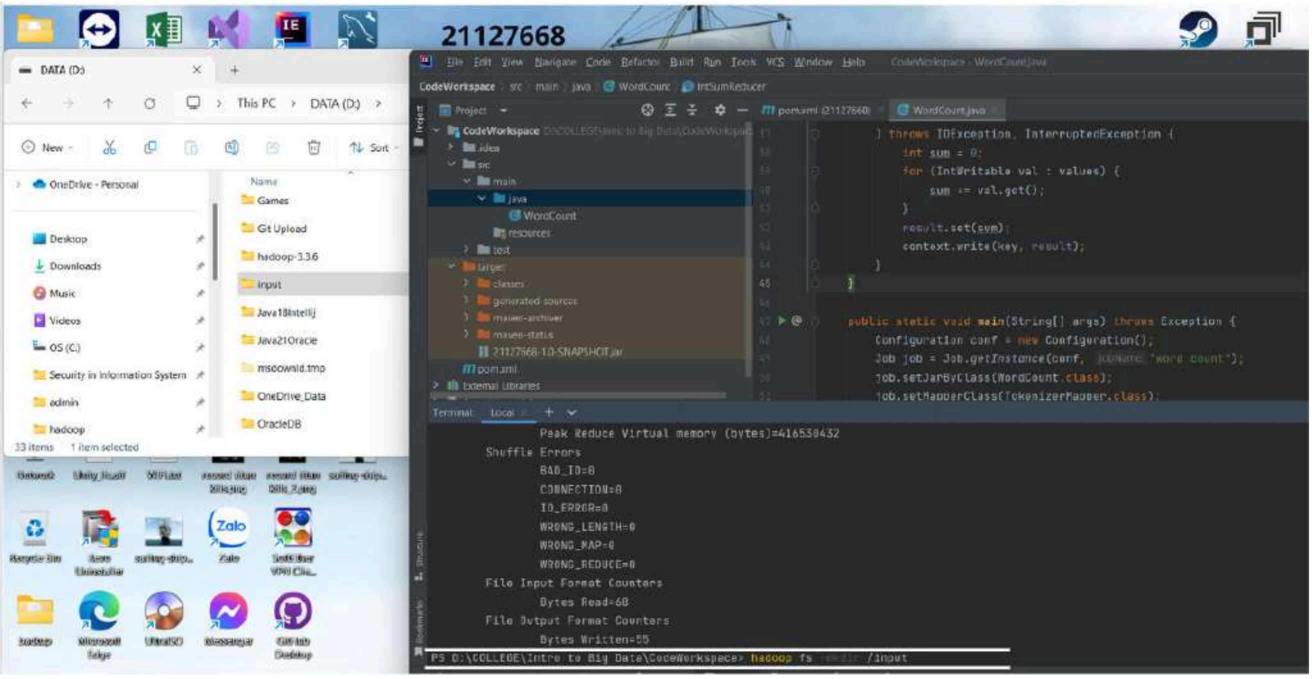
```
        ) throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }

    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "word count");
        job.setJarByClass(WordCount.class);
        job.setMapperClass(TokenizerMapper.class);
        job.setCombinerClass(IntSumReducer.class);
        job.setReducerClass(IntSumReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

- Create **data.txt** file



```
Hello
Hi
Hello
hi
good
morning
good day
bye
hello
Hello
Hi
```

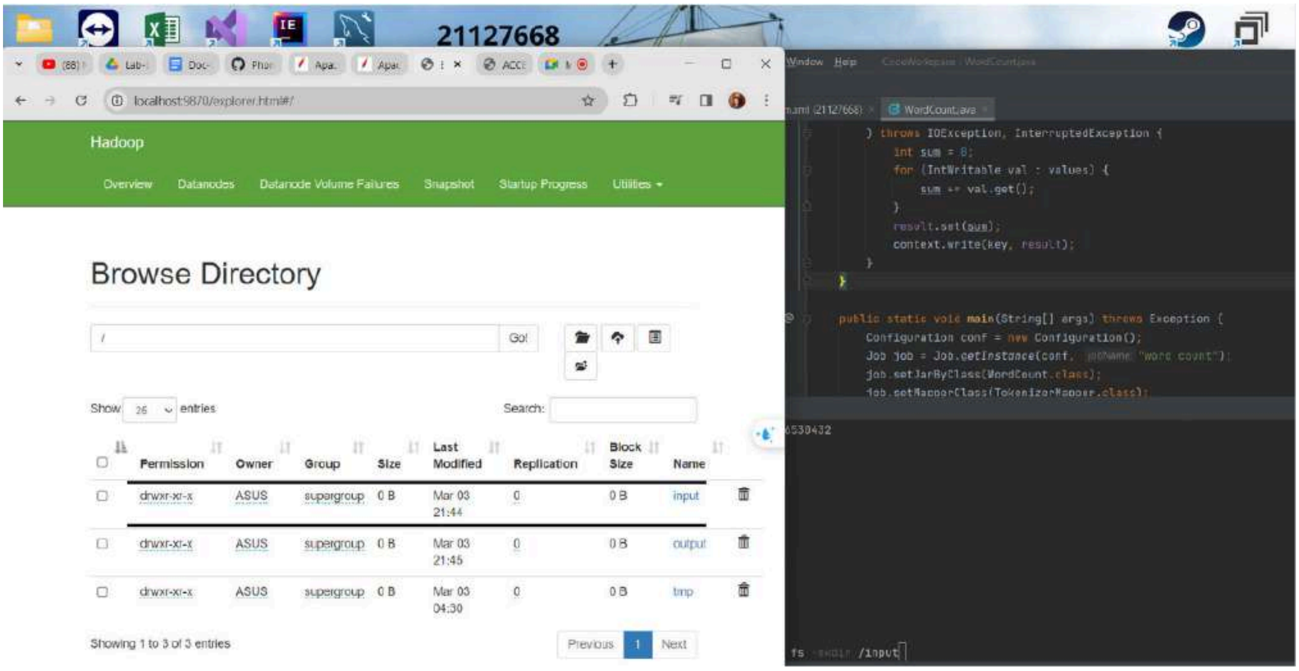
- Create **input** folder in the Hadoop File System



```
    } throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, jobName: "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
```

```
                Peak Reduce Virtual memory (bytes)=416530432
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=68
        File Output Format Counters
                Bytes Written=55
PS D:\COLLEGE\Intro to Big Data\CodeWorkspace> hadoop fs -mkdir /input
```

On **localhost:9870**

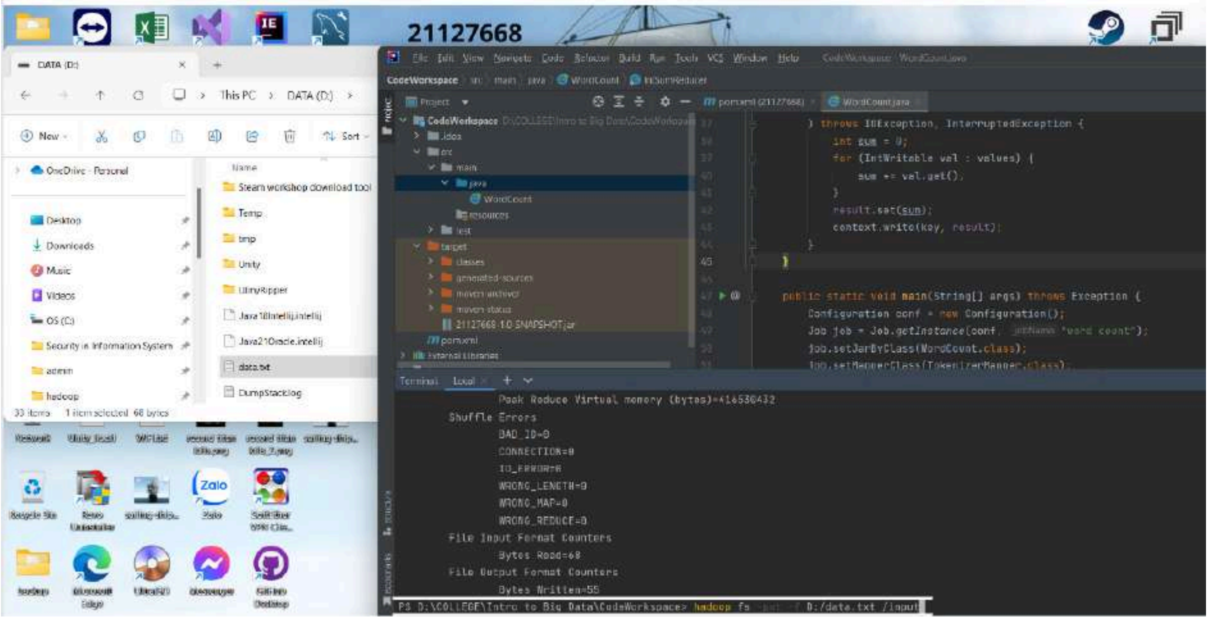


Browse Directory

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	ASUS	supergroup	0 B	Mar 03 21:44	0	0 B	input
□	drwxr-xr-x	ASUS	supergroup	0 B	Mar 03 21:45	0	0 B	output
□	drwxr-xr-x	ASUS	supergroup	0 B	Mar 03 04:30	0	0 B	tmp

Showing 1 to 3 of 3 entries

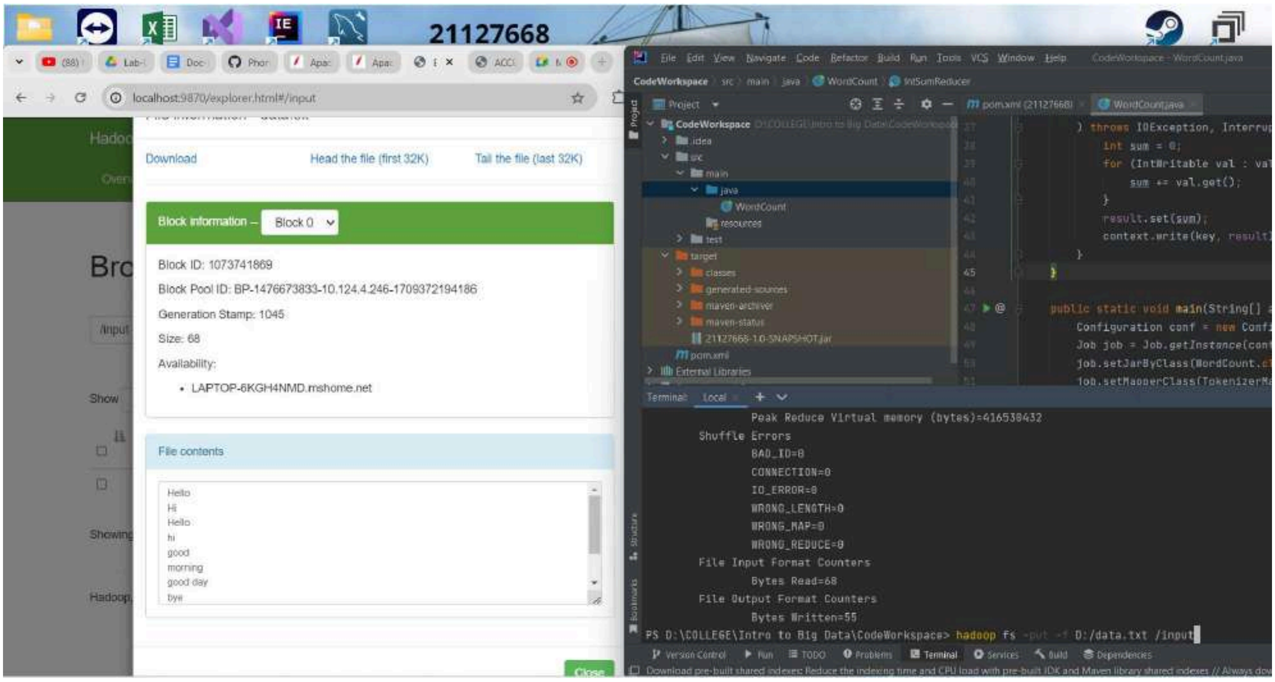
- Put **data.txt** into input folder in the Hadoop File System



```java
    ) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, jobName: "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
```

```
                Peak Reduce Virtual memory (bytes)=416530432
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=68
        File Output Format Counters
                Bytes Written=55
PS D:\COLLEGE\Intro to Big Data\CodeWorkspace> hadoop fs -put -f D:/data.txt /input
```

On **localhost:9870**



localhost:9870/explorer.html#/input

Download   Head the file (first 32K)   Tail the file (last 32K)

Block information – Block 0

Block ID: 1073741869

Block Pool ID: BP-1476673833-10.124.4.246-1709372194186

Generation Stamp: 1045

Size: 68

Availability:

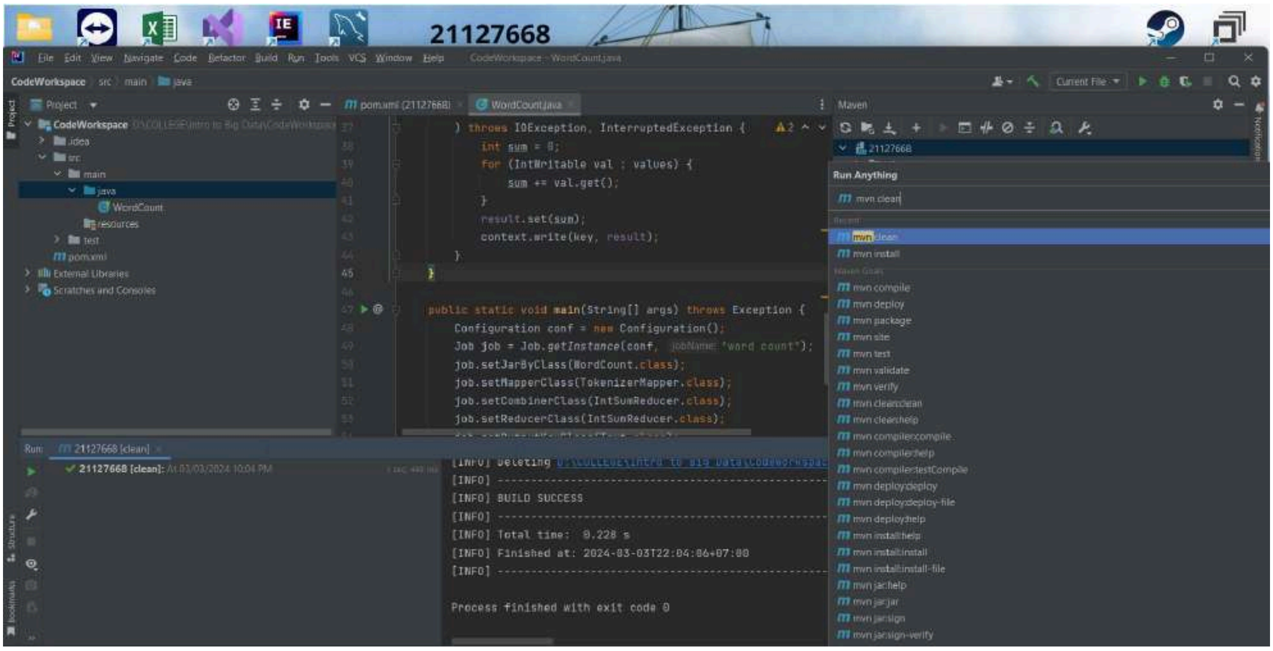
- LAPTOP-6KGH4NMD.mshome.net

File contents

```
Hello
Hi
Hello
hi
good
morning
good day
bye
```

```
                Peak Reduce Virtual memory (bytes)=416530432
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=68
        File Output Format Counters
                Bytes Written=55
PS D:\COLLEGE\Intro to Big Data\CodeWorkspace> hadoop fs -put -f D:/data.txt /input
```

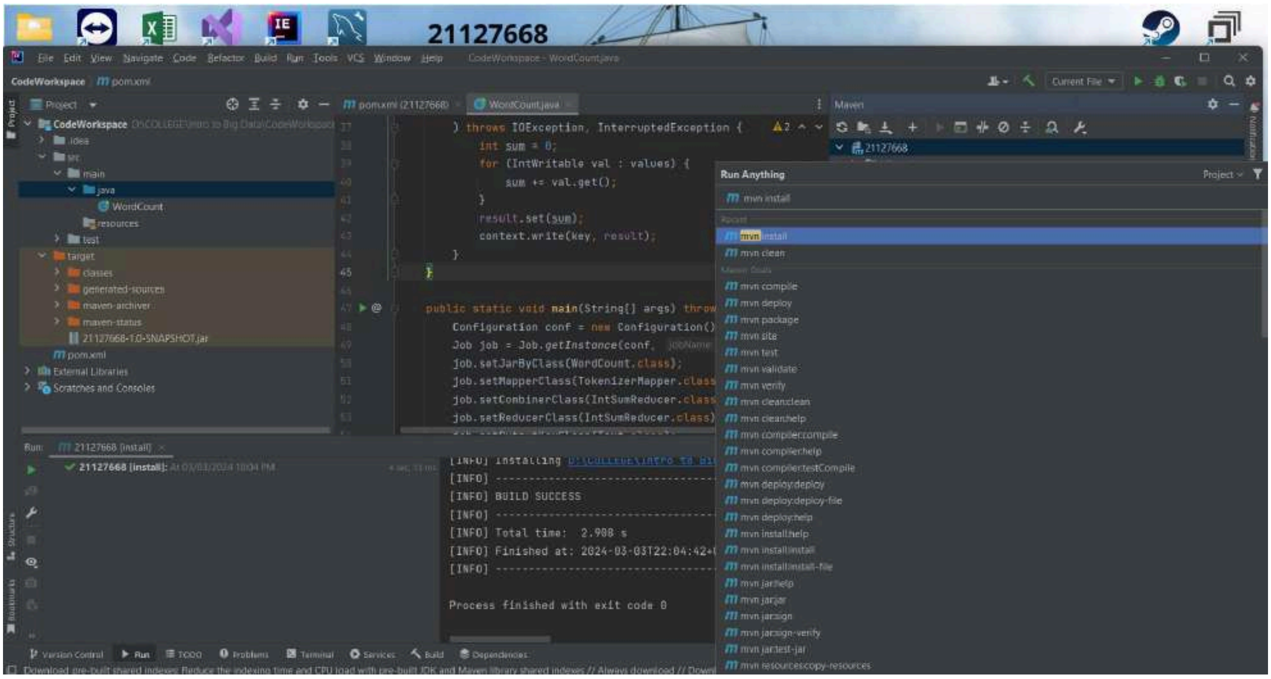
- Create **jar file** of maven



```
    ) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, jobName: "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
```

21127668 [clean]: At 03/03/2024 10:04 PM

```
[INFO] Deleting D:\COLLEGE\Intro to Big Data\CodeWorkspac
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  0.228 s
[INFO] Finished at: 2024-03-03T22:04:06+07:00
[INFO] ------------------------------------------------------------------------

Process finished with exit code 0
```

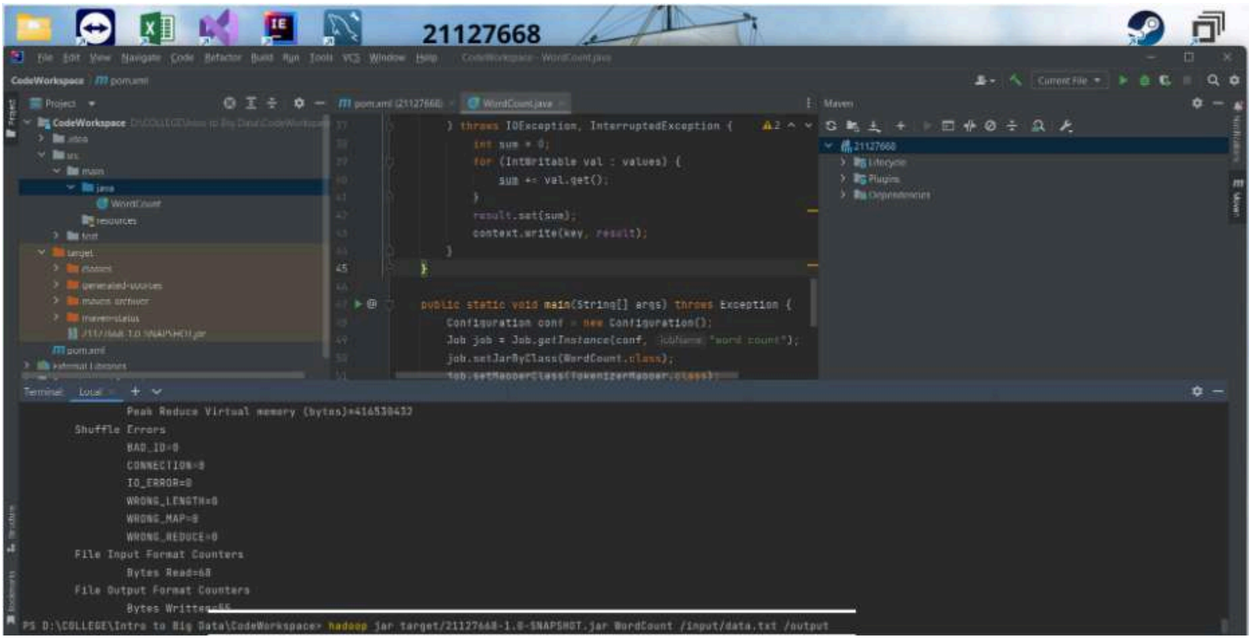


```
    ) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

public static void main(String[] args) throw
    Configuration conf = new Configuration()
    Job job = Job.getInstance(conf, jobName
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class
    job.setCombinerClass(IntSumReducer.class
    job.setReducerClass(IntSumReducer.class)
```

21127668 [install]: At 03/03/2024 10:04 PM

```
[INFO] Installing D:\COLLEGE\Intro to Bi
[INFO] -----------------------------------------------
[INFO] BUILD SUCCESS
[INFO] -----------------------------------------------
[INFO] Total time:  2.988 s
[INFO] Finished at: 2024-03-03T22:04:42+0
[INFO] -----------------------------------------------

Process finished with exit code 0
```

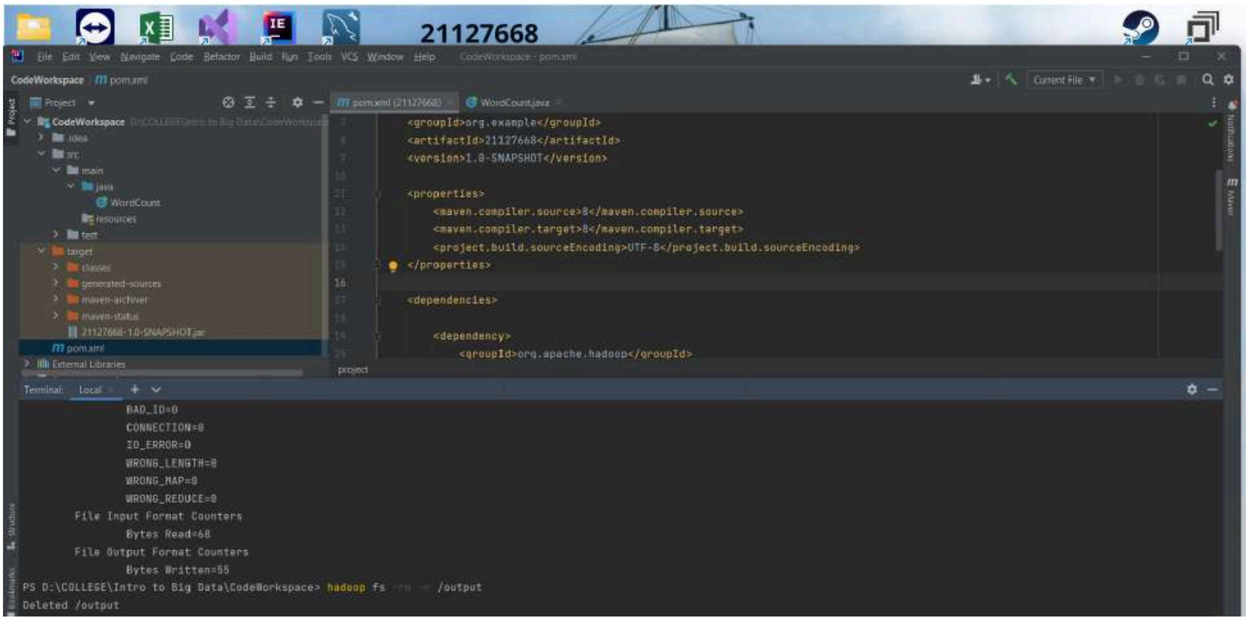
- Run the **jar file** using hadoop jar command



```
    } throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, jobName: "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
```

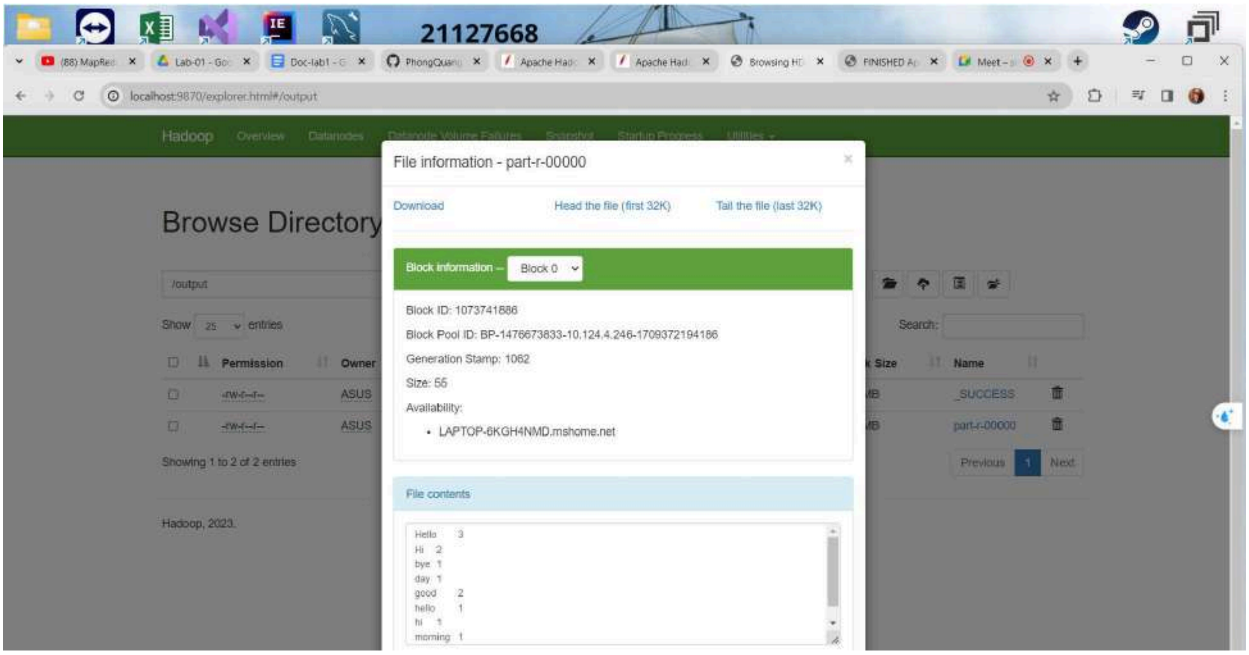
```
                Peak Reduce Virtual memory (bytes)=416530432
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=68
        File Output Format Counters
                Bytes Written=55
PS D:\COLLEGE\Intro to Big Data\CodeWorkspace> hadoop jar target/21127668-1.0-SNAPSHOT.jar WordCount /input/data.txt /output
```

- Eventually an output folder will be created and on **localhost:9870** at **part-r-00000** file which represents the final result
  If the output folder is already exist then we need to delete it



```
<groupId>org.example</groupId>
<artifactId>21127668</artifactId>
<version>1.0-SNAPSHOT</version>

<properties>
    <maven.compiler.source>8</maven.compiler.source>
    <maven.compiler.target>8</maven.compiler.target>
    <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
</properties>

<dependencies>

    <dependency>
        <groupId>org.apache.hadoop</groupId>
```

```
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=68
        File Output Format Counters
                Bytes Written=55
PS D:\COLLEGE\Intro to Big Data\CodeWorkspace> hadoop fs -rm -r /output
Deleted /output
```



Permission	Owner	Name
-rw-r--r--	ASUS	_SUCCESS
-rw-r--r--	ASUS	part-r-00000

Showing 1 to 2 of 2 entries

Hadoop, 2023.

File information - part-r-00000

Download | Head the file (first 32K) | Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741886

Block Pool ID: BP-1476673833-10.124.4.246-1709372194186

Generation Stamp: 1062

Size: 55

Availability:

- LAPTOP-6KGH4NMD.mshome.net

File contents

```
Hello	3
Hi	2
bye	1
day	1
good	2
hello	1
hi	1
morning	1
```

4. Reference

[1]https://www.youtube.com/watch?v=knAS0w-jiUk

[2]https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/ClusterSetup.html

[3]https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/SecureMode.html

[4]https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html

[5]https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html#Example%3A_WordCount_v1.0

[6]https://research.google/pubs/mapreduce-simplified-data-processing-on-large-clusters/

[7]https://storage.googleapis.com/gweb-research2023-media/pubtools/pdf/16cb30b4b92fd4989b8619a61752a2387c6dd474.pdf

[8]https://www.slideshare.net/oom65/ hadoop-security-architecture.

[9]Word Count

[10]Install Hadoop