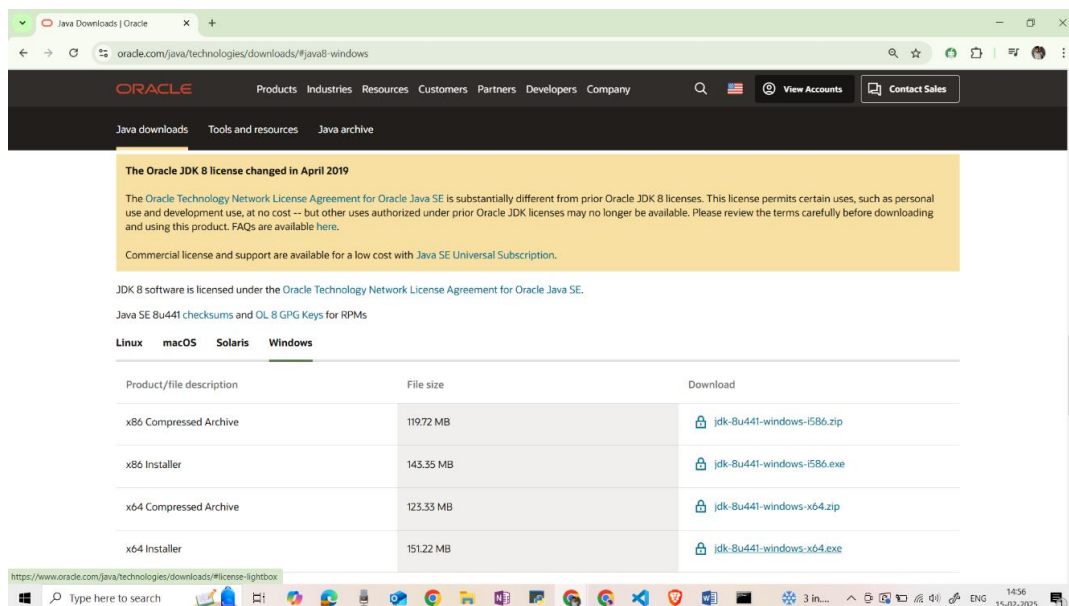


MY LEARNING EXPERIENCE:

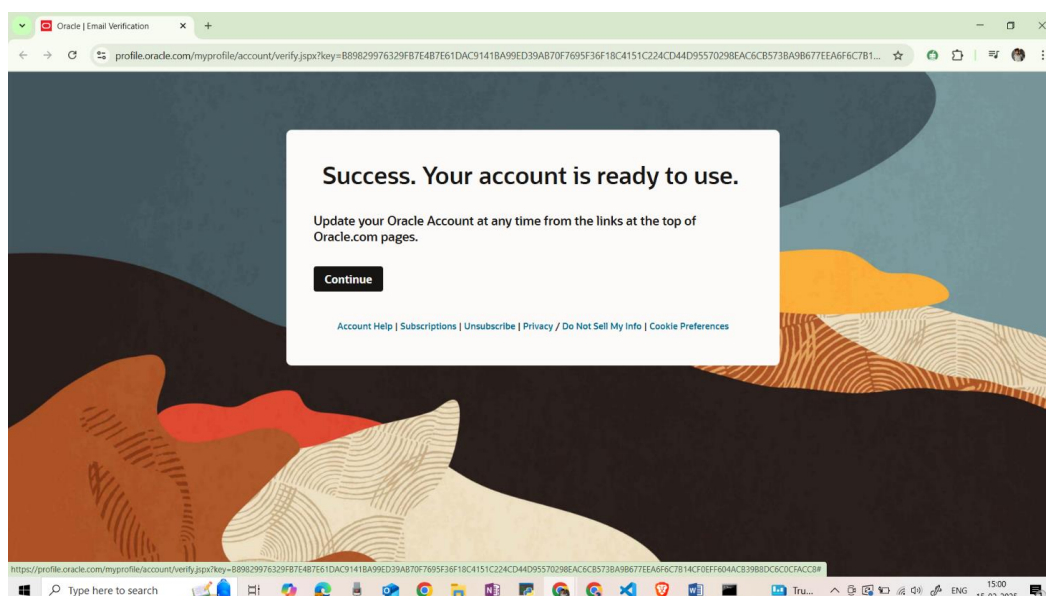
During installation of Hadoop, extraction of "Hadoop.tar.gz " did not work for me. To start, I attempted using WinRAR and 7-Zip, but those did not work for me due to the lack of permission. Rather than trying something else, I went ahead and tried the extraction from PowerShell as administrator, and this helped me unzip the files without any errors. This experience taught me about the strength that the built-in system tools have and made me realize the benefits of utilizing the tools available. Also, when I saw long datanode and namenode codes running in cmd, it felt so satisfying. Felt like I have upgraded my system knowledge.

HADOOP INSTALLATION ON WINDOWS 10

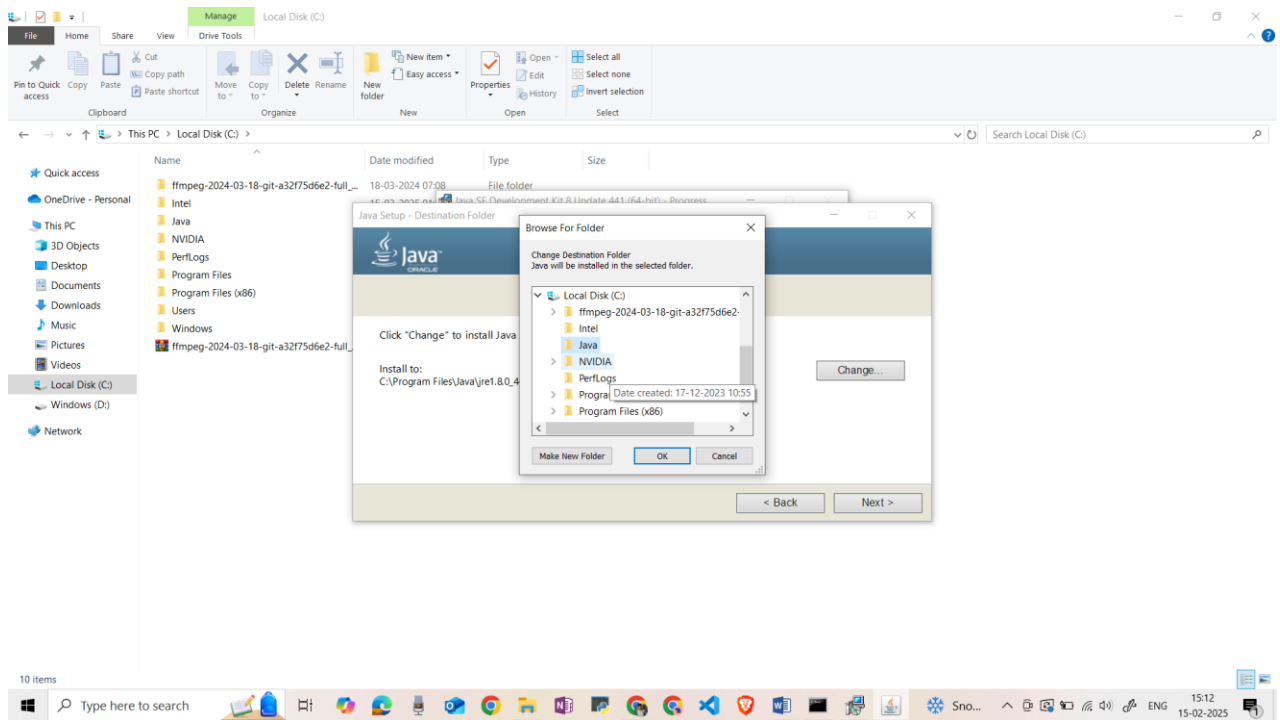
Download JAVA (java version "1.8.0_441")



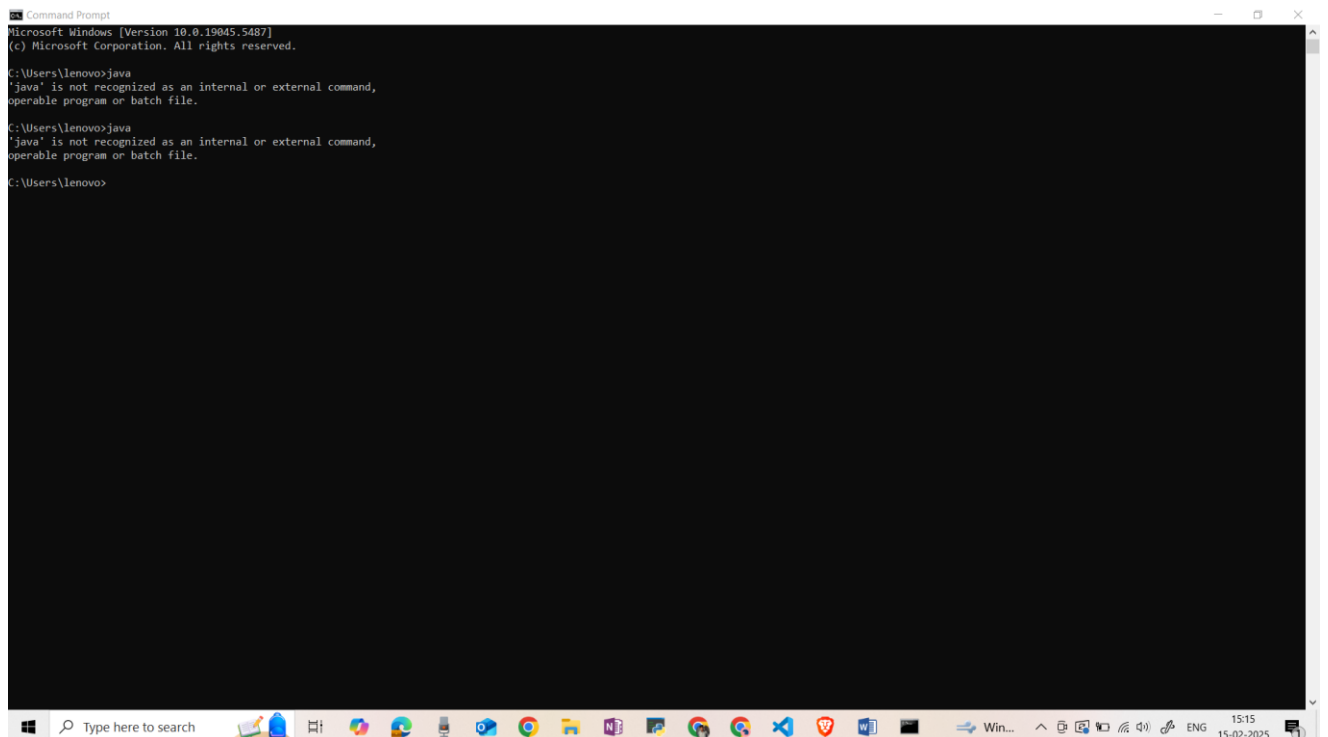
Create an oracle account:



Create a new “Java” folder in local disk c:

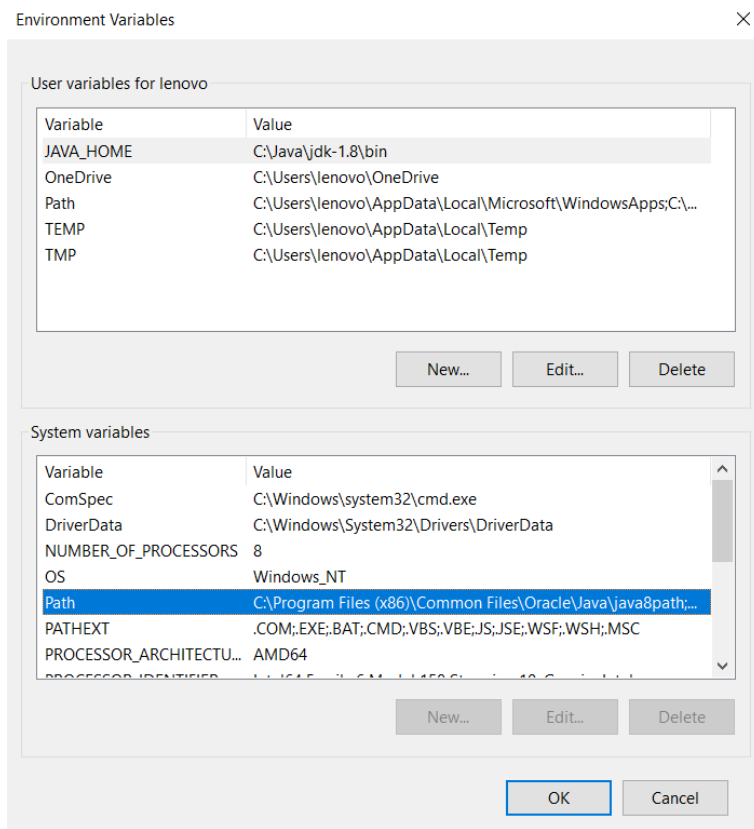


Check for java in cmd:

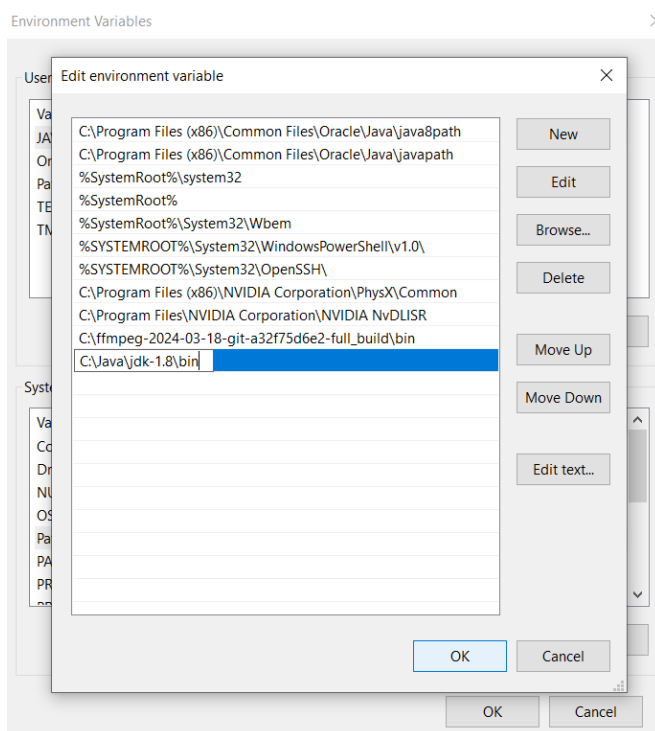


Here, java is not recognized.

Therefore, create JAVA_HOME in environment variables.



Click on path and edit it, copy the java jdk bin path and paste it here



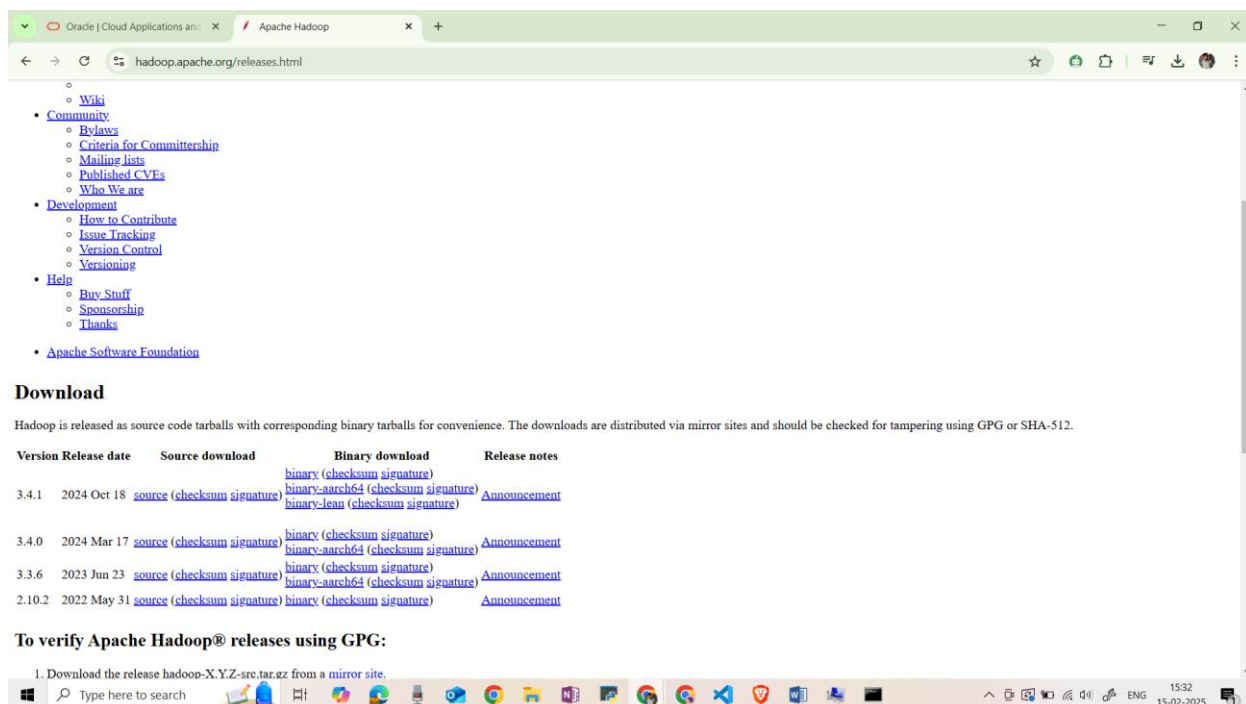
Now, the error has been resolved. Lets check java in cmd.

```
Command Prompt
Microsoft Windows [Version 10.0.19045.5487]
(c) Microsoft Corporation. All rights reserved.

C:\Users\lenovo>java
Usage: java [-options] class [args...]
           (to execute a class)
 or java [-options] -jar jarfile [args...]
           (to execute a jar file)
where options include:
  -d32          use a 32-bit data model if available
  -d64          use a 64-bit data model if available
  -server       to select the "server" VM
                 The default VM is server.

  -cp <class search path of directories and zip/jar files>
  -classpath <class search path of directories and zip/jar files>
                A ; separated list of directories, JAR archives,
                and ZIP archives to search for class files.
  -D<name>=<value>
                set a system property
  -verbose:[class|gc|jni]
                enable verbose output
  -version      print product version and exit
  -version:<value>
                Warning: this feature is deprecated and will be removed
                in a future release.
                require the specified version to run
  -showversion  print product version and continue
  -jre-restrict-search | -no-jre-restrict-search
                Warning: this feature is deprecated and will be removed
```

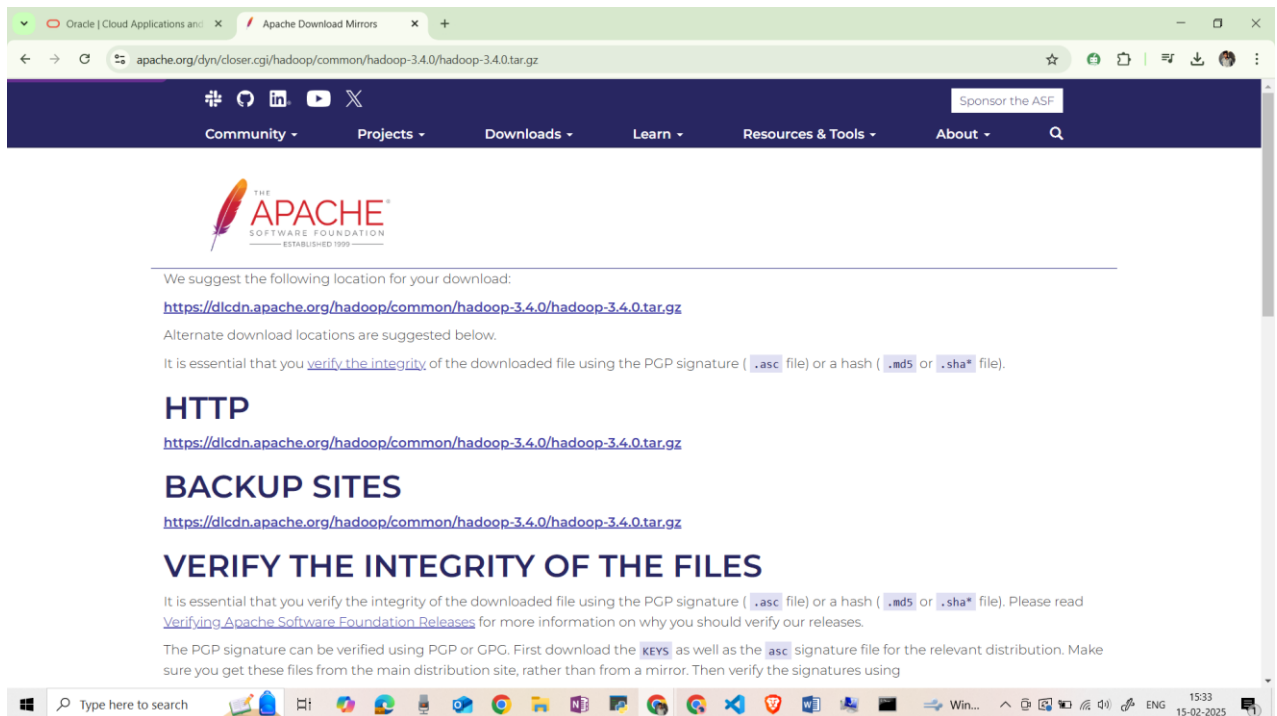
Now, cmd is able to see Java. Lets download Hadoop



The screenshot shows a web browser window with the Apache Hadoop releases page. The address bar shows the URL `hadoop.apache.org/releases.html`. The page has a navigation menu on the left with links to Wiki, Community, Development, Help, and Apache Software Foundation. The main content area is titled "Download" and contains a paragraph about downloading Hadoop as source code tarballs or binary tarballs. Below this is a table with columns for Version, Release date, Source download, Binary download, and Release notes. The table lists four releases: 3.4.1 (2024 Oct 18), 3.4.0 (2024 Mar 17), 3.3.6 (2023 Jun 23), and 2.10.2 (2022 May 31). Each release has links for source, binary, and release notes. The page also includes a section titled "To verify Apache Hadoop® releases using GPG:" with a list of instructions.

Version	Release date	Source download	Binary download	Release notes
3.4.1	2024 Oct 18	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature) binary-lean (checksum signature)	Announcement
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

Download 3.4.0 (stable version)



The screenshot shows the Apache Hadoop 3.4.0 download page. The browser address bar shows the URL: <https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>. The page features the Apache Software Foundation logo and navigation links: Community, Projects, Downloads, Learn, Resources & Tools, and About. The main content area provides instructions for downloading the file, including a direct link to the download page and alternate download locations. It also includes a section for verifying the integrity of the files using PGP signatures or hashes.

We suggest the following location for your download:
<https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>
Alternate download locations are suggested below.
It is essential that you [verify the integrity](#) of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file).

HTTP

<https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>

BACKUP SITES

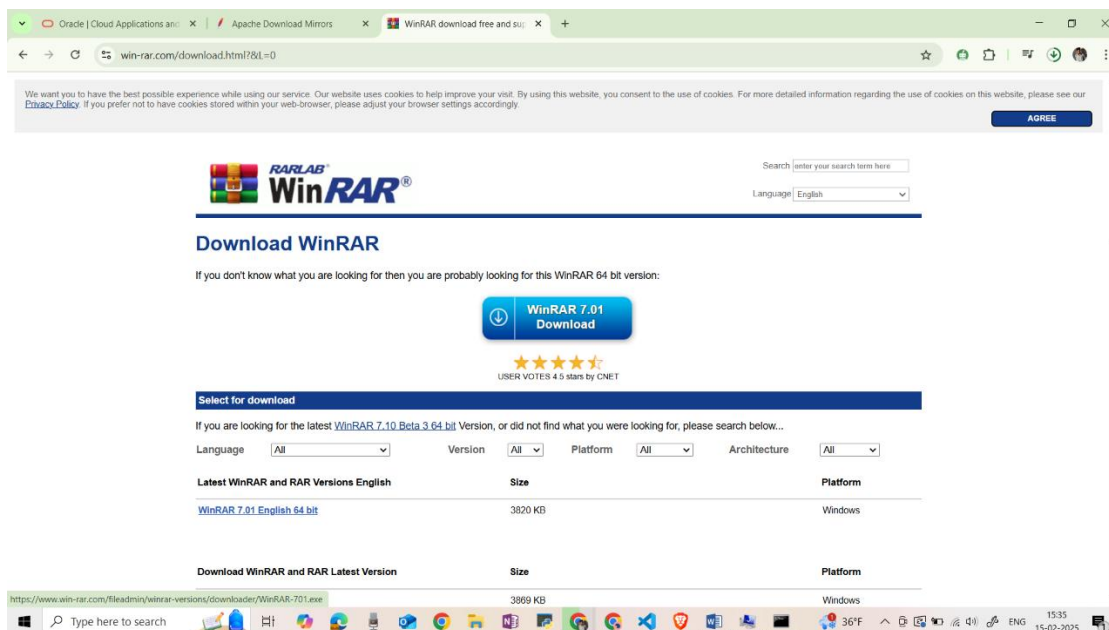
<https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>

VERIFY THE INTEGRITY OF THE FILES

It is essential that you verify the integrity of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file). Please read [Verifying Apache Software Foundation Releases](#) for more information on why you should verify our releases.
The PGP signature can be verified using PGP or GPG. First download the `KEYS` as well as the `.asc` signature file for the relevant distribution. Make sure you get these files from the main distribution site, rather than from a mirror. Then verify the signatures using

Click on the first link and download Hadoop.

Next, download Winrar:



The screenshot shows the WinRAR download page. The browser address bar shows the URL: <http://win-rar.com/download.html?&L=0>. The page features the WinRAR logo and a search bar. The main content area provides instructions for downloading the file, including a direct link to the download page and alternate download locations. It also includes a section for verifying the integrity of the files using PGP signatures or hashes.

We want you to have the best possible experience while using our service. Our website uses cookies to help improve your visit. By using this website, you consent to the use of cookies. For more detailed information regarding the use of cookies on this website, please see our [Privacy Policy](#). If you prefer not to have cookies stored within your web-browser, please adjust your browser settings accordingly.

WinRAR

Search
Language

Download WinRAR

If you don't know what you are looking for then you are probably looking for this WinRAR 64 bit version:

WinRAR 7.01 Download

★★★★★
USER VOTES 4.5 stars by CNET

Select for download

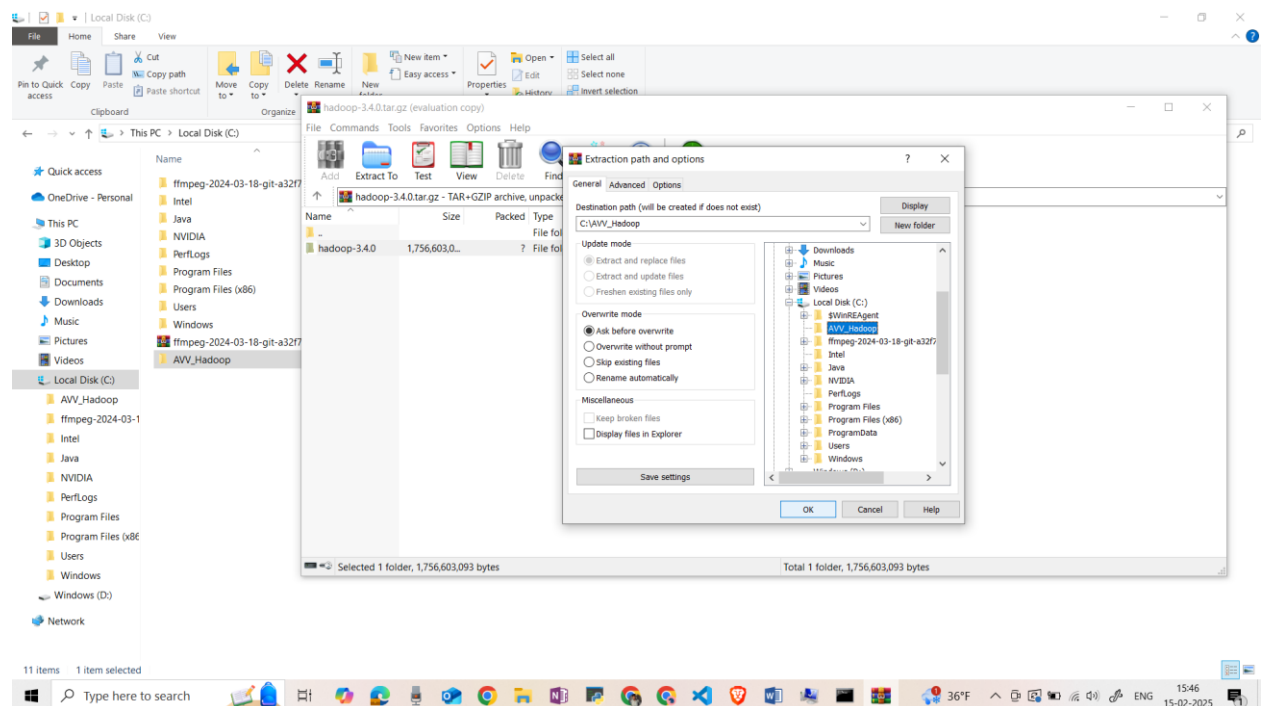
If you are looking for the latest [WinRAR 7.10 Beta 3 64 bit](#) Version, or did not find what you were looking for, please search below...

Language Version Platform Architecture

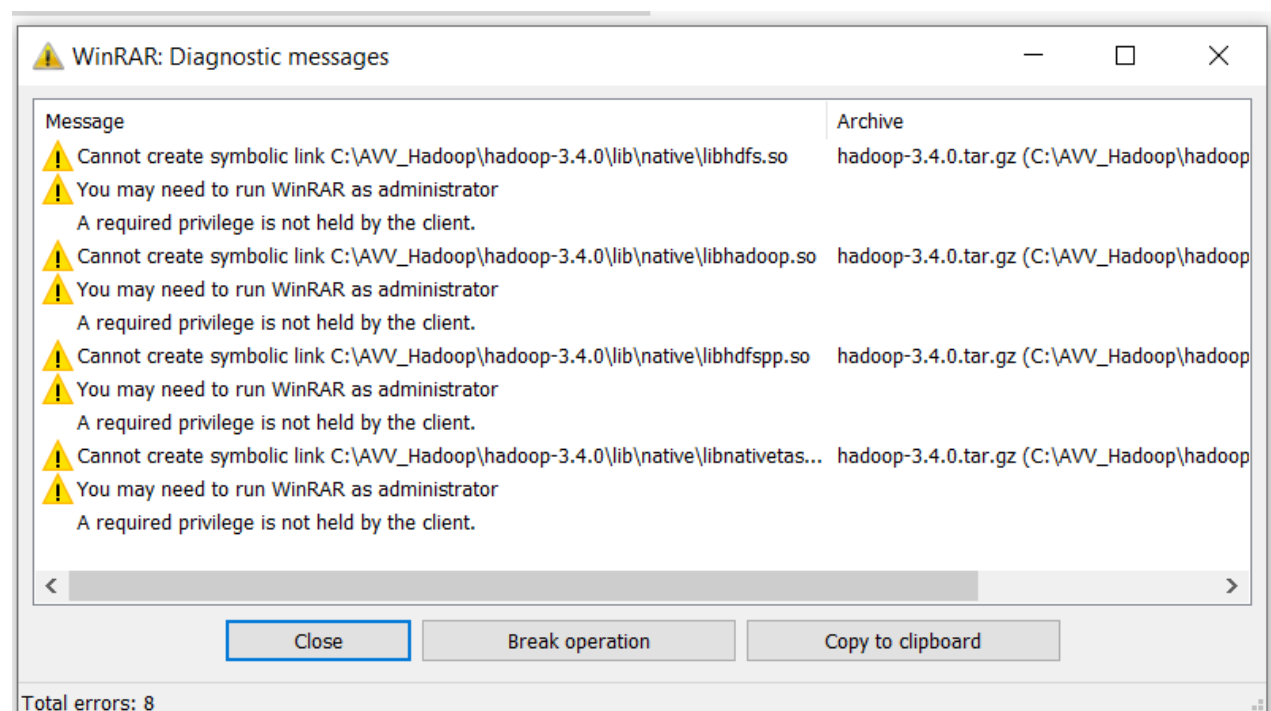
Latest WinRAR and RAR Versions English	Size	Platform
WinRAR 7.01 English 64 bit	3820 KB	Windows

Download WinRAR and RAR Latest Version	Size	Platform
http://www.win-rar.com/fileadmin/winrar-versions/downloader/WinRAR-701.exe	3869 KB	Windows

After downloading Winrar, extract Hadoop in to our newly created folder "AVV_Hadoop" in local disk C with Winrar:



Getting an error when I am trying to extract :



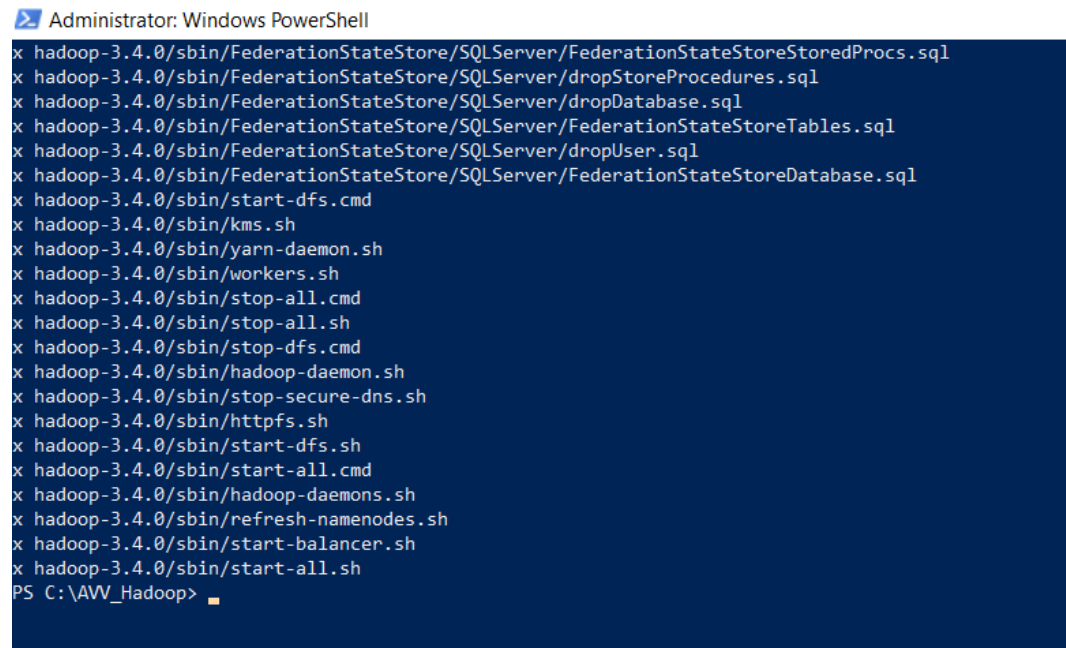
Solution:

I ran these lines in powershell as administrator

1.cd C:\AVV_Hadoop

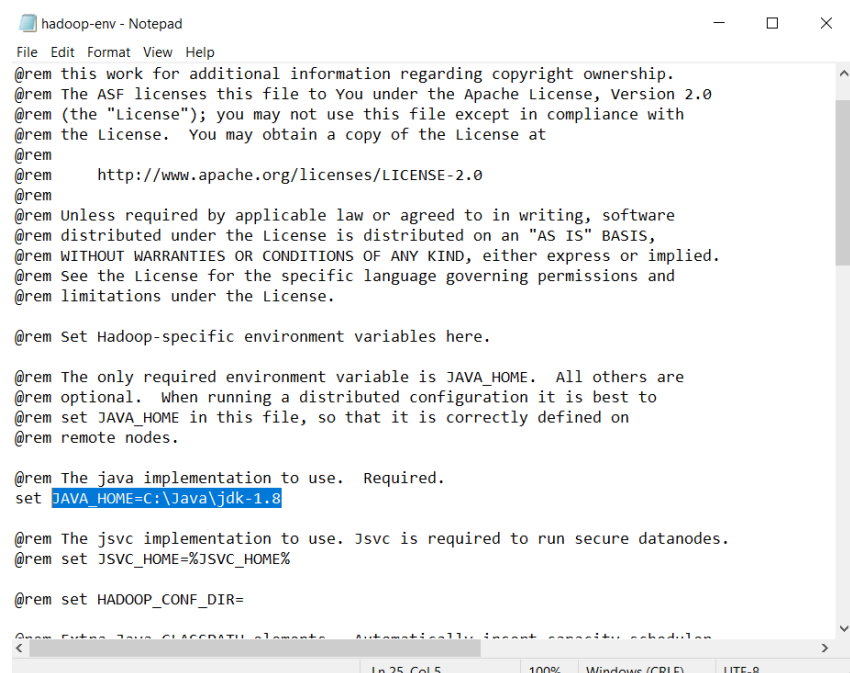
2. tar -xvzf hadoop-3.4.0.tar.gz (for extracting the Hadoop .tar.gz)

Error got rectified.



```
Administrator: Windows PowerShell
x hadoop-3.4.0/sbin/FederationStateStore/SQLServer/FederationStateStoreStoredProcs.sql
x hadoop-3.4.0/sbin/FederationStateStore/SQLServer/dropStoreProcedures.sql
x hadoop-3.4.0/sbin/FederationStateStore/SQLServer/dropDatabase.sql
x hadoop-3.4.0/sbin/FederationStateStore/SQLServer/FederationStateStoreTables.sql
x hadoop-3.4.0/sbin/FederationStateStore/SQLServer/dropUser.sql
x hadoop-3.4.0/sbin/FederationStateStore/SQLServer/FederationStateStoreDatabase.sql
x hadoop-3.4.0/sbin/start-dfs.cmd
x hadoop-3.4.0/sbin/kms.sh
x hadoop-3.4.0/sbin/yarn-daemon.sh
x hadoop-3.4.0/sbin/workers.sh
x hadoop-3.4.0/sbin/stop-all.cmd
x hadoop-3.4.0/sbin/stop-all.sh
x hadoop-3.4.0/sbin/stop-dfs.cmd
x hadoop-3.4.0/sbin/hadoop-daemon.sh
x hadoop-3.4.0/sbin/stop-secure-dns.sh
x hadoop-3.4.0/sbin/httpfs.sh
x hadoop-3.4.0/sbin/start-dfs.sh
x hadoop-3.4.0/sbin/start-all.cmd
x hadoop-3.4.0/sbin/hadoop-daemons.sh
x hadoop-3.4.0/sbin/refresh-namenodes.sh
x hadoop-3.4.0/sbin/start-balancer.sh
x hadoop-3.4.0/sbin/start-all.sh
PS C:\AVV_Hadoop>
```

Open hadoop-env and make JAVA_HOME path as this (jdk path from Java folder, that we have created)



```
hadoop-env - Notepad
File Edit Format View Help
@rem this work for additional information regarding copyright ownership.
@rem The ASF licenses this file to You under the Apache License, Version 2.0
@rem (the "License"); you may not use this file except in compliance with
@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.

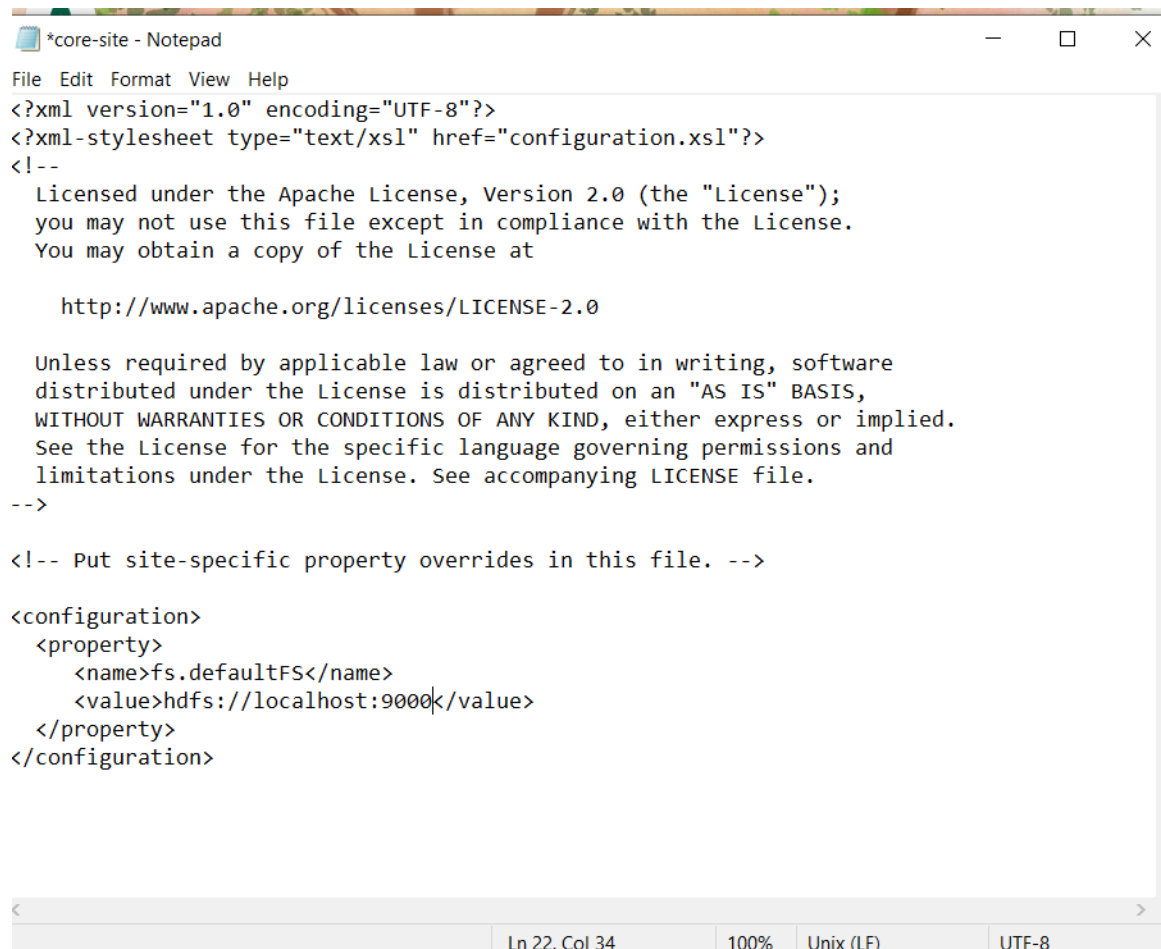
@rem The java implementation to use. Required.
set JAVA_HOME=C:\Java\jdk-1.8

@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
```


Open core-site and edit it, do this configuration: (This ensures Hadoop interacts with **HDFS** instead of the local filesystem.)



```
*core-site - Notepad
File Edit Format View Help
<?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. -->

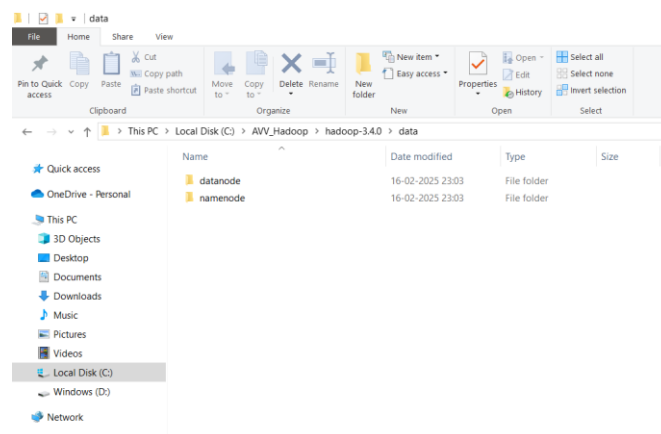
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

Create data folder inside Hadoop folder.

Create two folders inside data folder:

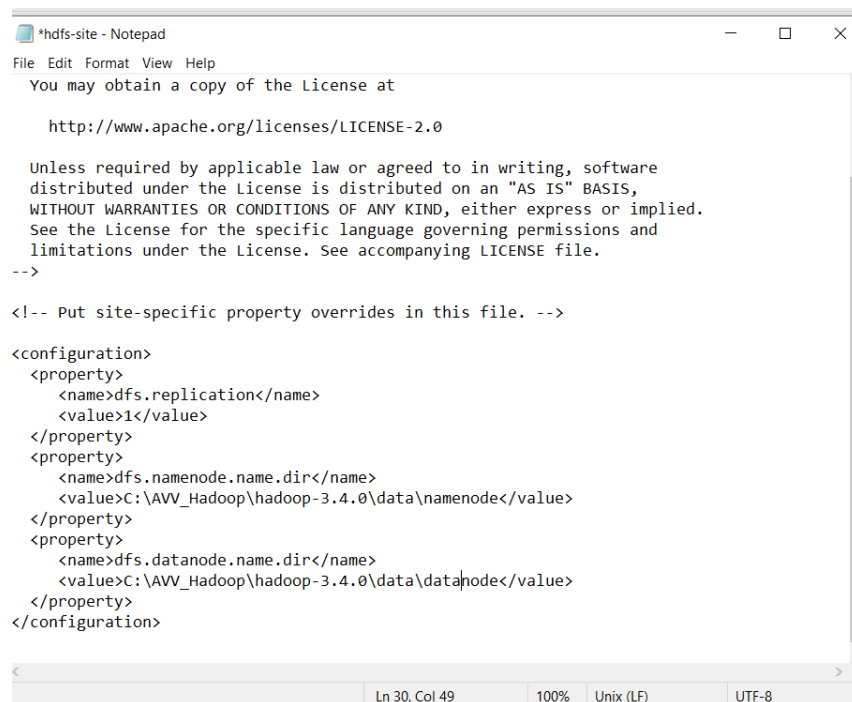
1.datanode

2.namenode



Open hdfs-site.xml and edit it:

This step ensures that data is stored across multiple datanodes. Ensuring fault tolerances by preventing loosing of data.



```
*hdfs-site - Notepad
File Edit Format View Help
You may obtain a copy of the License at

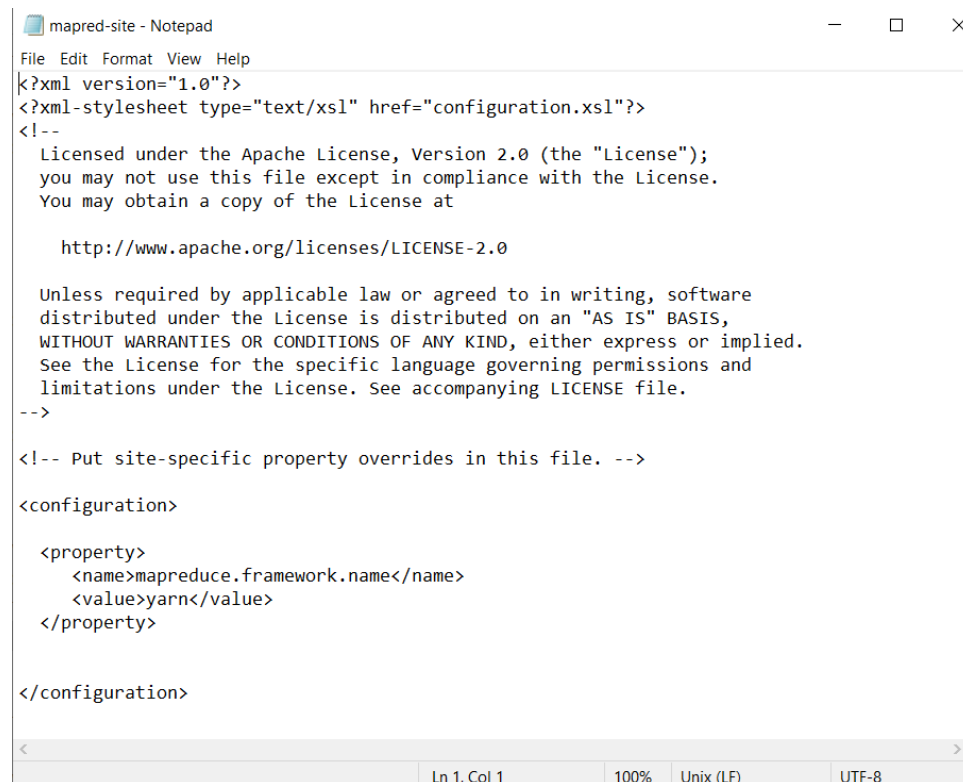
    http://www.apache.org/licenses/LICENSE-2.0

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>C:\AVV_Hadoop\hadoop-3.4.0\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.name.dir</name>
    <value>C:\AVV_Hadoop\hadoop-3.4.0\data\datanode</value>
  </property>
</configuration>
```

Edit mapred-site.xml file : If this change is not performed, Hadoop gets default to single node local execution, this prevents distributed processing.



```
mapred-site - Notepad
File Edit Format View Help
<?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.
-->

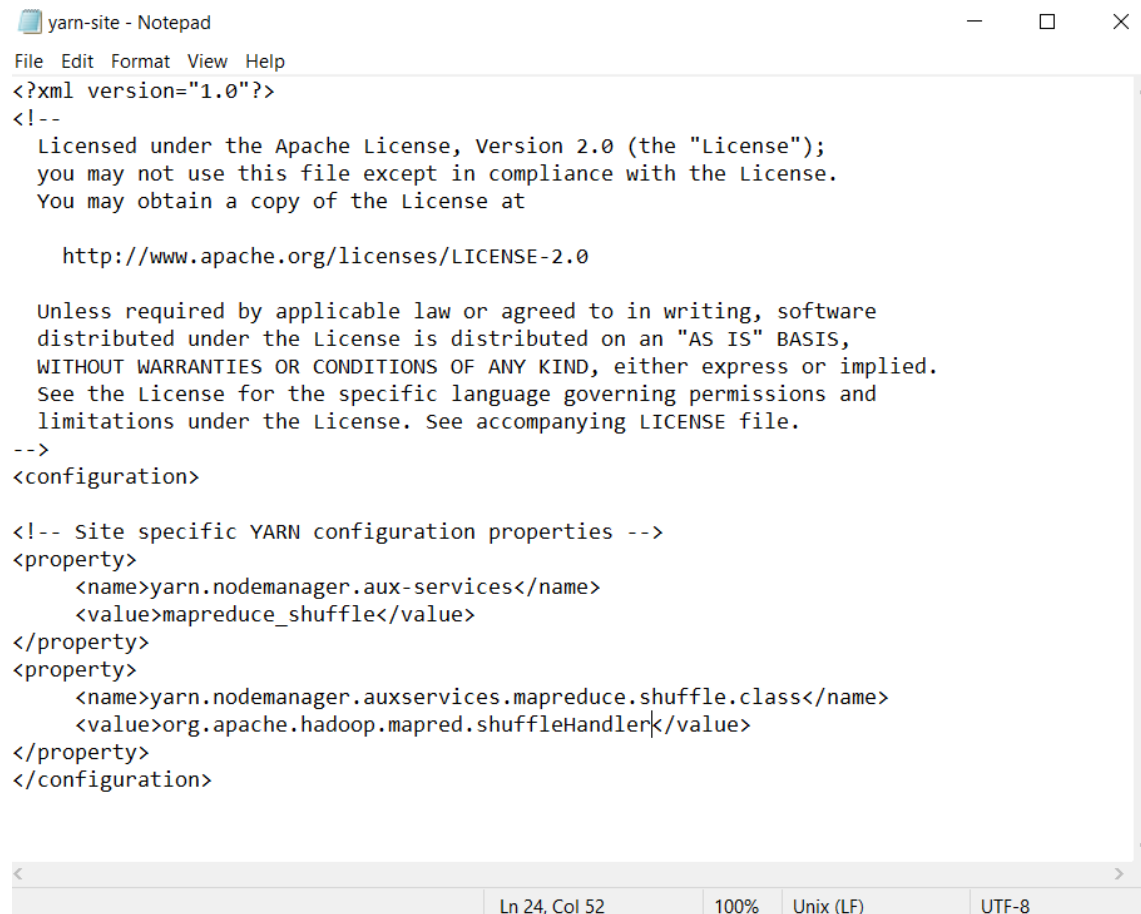
<!-- Put site-specific property overrides in this file. -->

<configuration>

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

</configuration>
```

Edit yarn-site.xml file: This step helps in resource management and job scheduling in Hadoop cluster



```
File Edit Format View Help
<?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>

<!-- Site specific YARN configuration properties -->
<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>
```

Ln 24, Col 52 100% Unix (LF) UTF-8

Download the fixed bin folder.

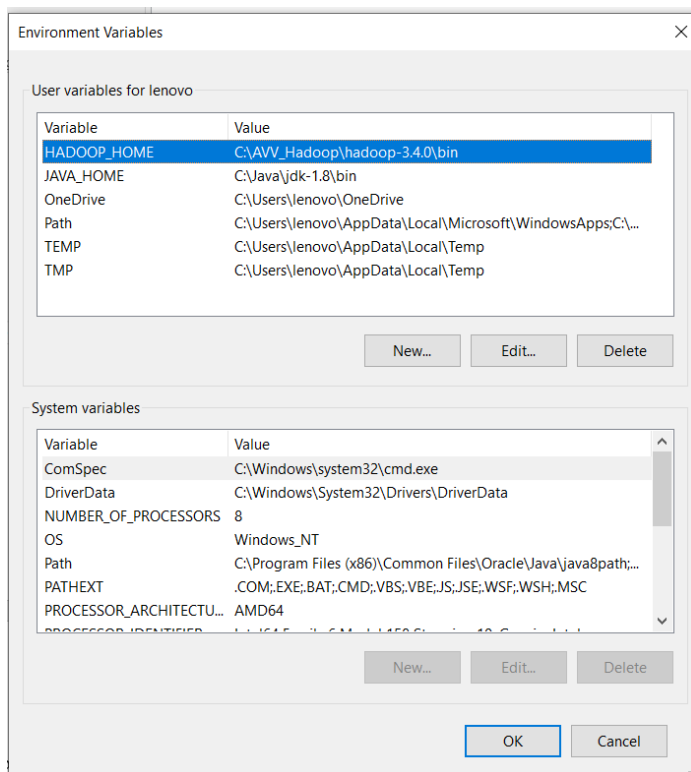
Delete the existing bin folder and replace it with the fixed one.

Go to the fixed bin and double click on winutils. No error.

Now, open cmd and type “hdfs namenode -format”

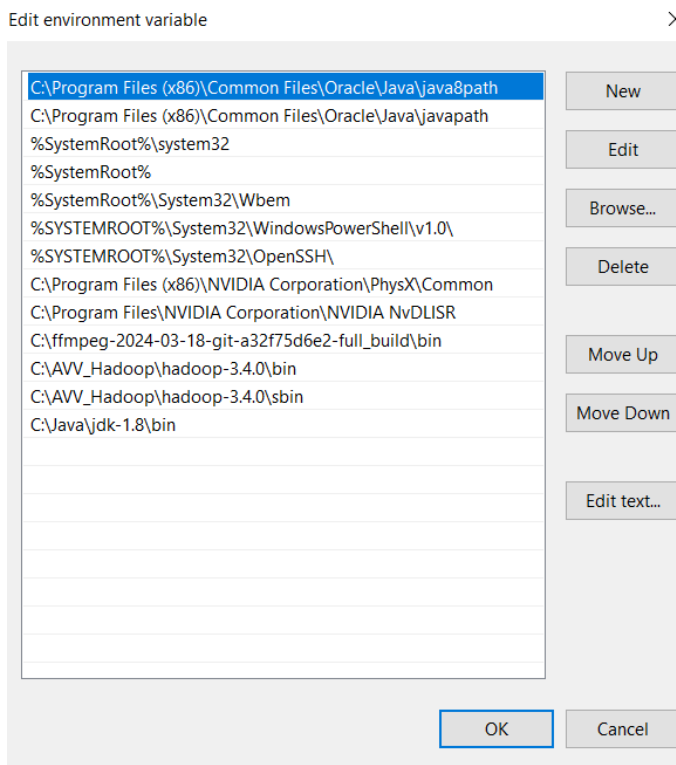
```
'hdfs' is not recognized as an internal or external command,
operable program or batch file.
```

To fix this, lets open environmental variables and make HADOOP_HOME:



Select path and edit :

Paste the paths of “bin” and “sbin” from Hadoop:



Open cmd as administrator and again type “hdfs namenode -format”

```
C:\ Select Administrator: Command Prompt
2025-02-17 00:00:05,706 INFO namenode.FSImage: Allocated new BlockPoolId: BP-731252915-10.0.0.214-1739768405684
2025-02-17 00:00:05,810 INFO common.Storage: Storage directory C:\AVV_Hadoop\hadoop-3.4.0\data\namenode has been success
fully formatted.
2025-02-17 00:00:05,951 INFO namenode.FSImageFormatProtobuf: Saving image file C:\AVV_Hadoop\hadoop-3.4.0\data\namenode\
current\fsimage.ckpt_000000000000000000 using no compression
2025-02-17 00:00:06,396 INFO namenode.FSImageFormatProtobuf: Image file C:\AVV_Hadoop\hadoop-3.4.0\data\namenode\current
\fsimage.ckpt_000000000000000000 of size 401 bytes saved in 0 seconds .
2025-02-17 00:00:06,437 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2025-02-17 00:00:06,458 INFO blockmanagement.DatanodeManager: Slow peers collection thread shutdown
2025-02-17 00:00:06,504 INFO namenode.FSNamesystem: Stopping services started for active state
2025-02-17 00:00:06,505 INFO namenode.FSNamesystem: Stopping services started for standby state
2025-02-17 00:00:06,520 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2025-02-17 00:00:06,522 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-T6GTQM0/10.0.0.214
*****/
C:\Windows\system32>
```

Can clearly see namenode has been successfully formatted. Resolved this error.

Later check jps:

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19045.5487]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>jps
1976 Jps

C:\Windows\system32>
```

Now, open cmd as administrator and run this command :

```
” C:\Windows\system32>cd \
```

```
C:\>cd AVV_Hadoop
```

```
C:\AVV_Hadoop>cd hadoop-3.4.0
```

```
C:\AVV_Hadoop\hadoop-3.4.0>cd sbin
```

```
C:\AVV_Hadoop\hadoop-3.4.0\sbin>start-dfs.cmd
```

```
C:\AVV_Hadoop\hadoop-3.4.0\sbin>jps
```

```
2672 NameNode
```

```
18004 DataNode
```

```
17160 Jps
```

```
C:\AVV_Hadoop\hadoop-3.4.0\sbin>start-yarn.cmd
```

```
starting yarn daemons
```

```
C:\AVV_Hadoop\hadoop-3.4.0\sbin>
```

```
”
```

Then, namenode and datanode will run.

The image shows a Windows desktop environment. At the top, there is a taskbar with several application icons. The main area contains three windows. The leftmost window is a Command Prompt titled 'Administrator: Command Prompt', showing a series of commands and their outputs. The commands include setting the path for Hadoop binaries, running 'hadoop fs -ls /', and starting the Hadoop NameNode and DataNode services. The middle window is titled 'Apache Hadoop Distribution - namenode' and displays the logs for the NameNode startup, showing various configuration parameters and the successful completion of the initialization process. The rightmost window is titled 'Apache Hadoop Distribution - hadoop - datanode' and shows the logs for the DataNode startup, including a deprecation warning and the successful completion of the startup process. The bottom of the screen shows the Windows taskbar with the search bar and several pinned application icons.

Administrator Command Prompt

Microsoft Windows [Version 10.0.19045.5487]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>jps
3132 Jps

C:\Windows\system32>cd \

C:\>cd AWV_Hadoop

C:\AWV_Hadoop>cd sbin
The system cannot find the path specified.

C:\AWV_Hadoop>cd hadoop-3.4.0

C:\AWV_Hadoop\hadoop-3.4.0>sbin
'sbin' is not recognized as an internal or external command,
operable program or batch file.

C:\AWV_Hadoop\hadoop-3.4.0>cd sbin

C:\AWV_Hadoop\hadoop-3.4.0\sbin>start-dfs.cmd

C:\AWV_Hadoop\hadoop-3.4.0\sbin>jps
2672 NameNode
18004 DataNode
17160 Jps

C:\AWV_Hadoop\hadoop-3.4.0\sbin>

namenode

DS-9b36b9c7-c619-4429-8743-a0bae560c025 from datanode DatanodeRegistration(127.0.0.1:874f-b65b-a513bf19966d, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv-c-0756a35b60a6;nsid=1225535821;c=1739768405684)

lateChange: BLOCK* processReport 0xfbf15504b5dfe3795 with Lease ID 0x09b367de4dc4808;8743-a0bae560c025 node DatanodeRegistration(127.0.0.1:8066, datanodeUId-ab943d21-70-9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv-c-57;cid=CID-f882ace4-393e-4669-1739768405684), blocks: 0, hasStaleStorage: false, processing time: 4 msecs, invalid

host/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 2160
; heartBeatInterval=3000
artbeat response, updating state of namenode local host

Block report 0xfbf15504b5dfe3795 with Lease ID 0x99b367
Storage report(s), of which we sent 1. The reports had 0
d 126 msecs for RPC and NN processing. Got back one co

nd for block pool BP-731252915-10.0.0.214-173976840568

2 items

Type here to search

33°F

00:29
17-02-2025

```
Administrator Command Prompt
Microsoft Windows [Version 10.0.19045.5487]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>jps
3132 jps

C:\Windows\system32>cd \

C:\>cd AW_Hadoop

C:\AW_Hadoop>cd sbin
The system cannot find the path specified.

C:\AW_Hadoop>cd hadoop-3.4.0

C:\AW_Hadoop\hadoop-3.4.0>sbin
'sbin' is not recognized as an internal or external command,
operable program or batch file.

C:\AW_Hadoop\hadoop-3.4.0>cd sbin

C:\AW_Hadoop\hadoop-3.4.0>sbinstart-dfs.cmd

C:\AW_Hadoop\hadoop-3.4.0>sbin>jps
2672 NameNode
18004 DataNode
17160 Jps

C:\AW_Hadoop\hadoop-3.4.0>sbinstart-yarn.cmd
Starting yarn daemons

Apache Hadoop Distribution - yarn nodemanager
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Feb 17, 2025 12:32:54 AM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19.4 05/24/2017 03:20 PM'
Feb 17, 2025 12:32:54 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider
the scope "Singleton"
Feb 17, 2025 12:32:54 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.GenericExceptionHandler to GuiceManagedComponentProvider with the scope
"Singleton"
Feb 17, 2025 12:32:55 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the
scope "Singleton"
2025-02-17 00:32:55,343 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@45bb52ff{/node/,file:///C:/Users/
V/AppData/Local/Temp/jetty-0.0.0-8042-hadoop-yarn-common-3.4.0-jar_-any-449213434918624815/webapp/,AVAILABLE[}
file:///C:/AW_Hadoop\hadoop-3.4.0\share\hadoop-yarn\hadoop-yarn-common-3.4.0-jar\webapps\{node/
2025-02-17 00:32:55,359 INFO server.AbstractConnector: Started ServerConnector@10358c32(HTTP/1.1,(http/1.1)){0.0.0.
42}
2025-02-17 00:32:55,359 INFO server.Server: Started @20627ms
2025-02-17 00:32:55,363 INFO webapp.WebApps: Web app node started at 8042
2025-02-17 00:32:55,364 INFO server.NMStatusUpdaterImpl: Node ID assigned is : 10.0.0.214:56480.
2025-02-17 00:32:55,367 INFO util.NMPauseMonitor: Starting JVM pause monitor
2025-02-17 00:32:55,381 INFO client.DefaultHARMAFailureProxyProvider: Connecting to ResourceManager at /0.0.0.0:8
2025-02-17 00:32:55,466 INFO nodemanager.NMStatusUpdaterImpl: Running Applications Size : 0
2025-02-17 00:32:56,045 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got k
th id 2001969280
2025-02-17 00:32:56,046 INFO security.NMTokenSecretManagerImpl: Rolling master-key for container-tokens, got key wi
810485303
nodemanager.NMStatusUpdaterImpl: Registered with ResourceManager as 10.0.0.214:56480
cores:8>
```

The screenshot shows the Hadoop DFS Health Overview page. The browser address bar indicates the URL is localhost:9870/dfshealth.html#tab-overview. The page has a green header with navigation links: Hadoop, Overview (selected), Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Overview 'localhost:9000' (✓active)'. Below this is a table with the following data:

Started:	Mon Feb 17 00:27:42 -0500 2025
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaec760
Compiled:	Mon Mar 04 01:35:00 -0500 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-f883ace4-393e-4669-906c-0756a35b60a6
Block Pool ID:	BP-731252915-10.0.0.214-1739768405684

Below the table is a 'Summary' section. It states: '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 77.27 MB of 228 MB Heap Memory. Max Heap Memory is 889 MB.', and 'Non Heap Memory used 53.83 MB of 55.88 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.' At the bottom of the summary, there is a table showing capacity information:

Configured Capacity:	232.25 GB
Configured Remote Capacity:	0 B

Now, Go to localhost:8088 (opens the **YARN ResourceManager Web UI**)

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>

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application F
Capacity Scheduler	[memory-mb (unit=M), 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
No data available in table													

Showing 0 to 0 of 0 entries

Uploading files :

I have downloaded 5 files from : <https://www.ncei.noaa.gov/oa/global-historical-climatology-network/index.html#hourly/access/by-station/> into “hadoop_files” folder in my pc.

hadoop_files

File Home Share View

Navigation pane: Preview pane, Details pane

Layout: Extra large icons, Large icons, Medium icons, Small icons, List, Tiles, Content

Current view: Group by, Add columns, Size all columns to fit

Show/hide: Item check boxes, File name extensions, Hidden items, Hide selected items, Options

Name	Date modified	Type	Size
GHCHh_AAI0000TNCa_por.psv	18-02-2025 10:15	Text Document	2,65,538 KB
GHCHh_ACU55-00193_por.psv	18-02-2025 10:13	Text Document	2,980 KB
GHCHh_AEI0000OMAW_por.psv	18-02-2025 10:13	Text Document	5,337 KB
GHCHh_AFA00409120_por.psv	18-02-2025 10:14	Text Document	9 KB
GHCHh_AFA00409731_por.psv	18-02-2025 10:13	Text Document	443 KB

To start Hadoop cluster:

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.19045.5487]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>hadoop version
Hadoop 3.4.0
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled by root on 2024-03-04T06:35Z
Compiled on platform linux-x86_64
Compiled with protoc 3.21.12
From source with checksum f7fe694a3613358b38812ae9c31114e
This command was run using /C:/AVV_Hadoop/hadoop-3.4.0/share/hadoop/common/hadoop-common-3.4.0.jar

C:\Windows\system32>start-dfs.cmd
```

After this, namenode and datanode ran successfully.

```
ber of blocks processed: 0/0  
2025-02-18 11:16:27,465 INFO ipc.Server: IPC Server Responder: starting  
2025-02-18 11:16:27,466 INFO ipc.Server: IPC Server listener on 9000: starting  
2025-02-18 11:16:27,479 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000.  
2025-02-18 11:16:27,481 INFO namenode.FSNamesystem: Starting services required for active state  
2025-02-18 11:16:27,482 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)  
  
name s  
storage DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
2025-02-18 11:16:21,051 INFO datanode.DataNode: STARTUP_MSG:  
*****  
00 mil STARTUP_MSG: Starting DataN  
2025-0 STARTUP_MSG: host = DESK  
deUuiD STARTUP_MSG: args = [] Microsoft Windows [Version 10.0.19045.5487]  
883acc STARTUP_MSG: version = 3.(c) Microsoft Corporation. All rights reserved.  
2025-0 STARTUP_MSG: classpath =  
2025-0 doop\hadoop-3.4.0\share\hc\C:\Windows\system32>hadoop version  
(127 ommon\lib\audience-annotat:Hadoop 3.4.0  
2025-0 p\hadoop-3.4.0\share\hadoopSource code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7791783192ee7a5dfaec760  
60c025 r-qual-2.5.2.jar;C:\AVV_HacCompiled by root on 2024-03-04T06:35Z  
2025-0 _0\share\hadoop\common\lib\Compiled on platform linux-x86_64  
Proce ar;C:\AVV_Hadoop\hadoop-3.4.0\Compiled with protoc 3.21.12  
1:9866 oop\common\lib\commons-compFrom source with checksum f7fe694a3613358b38812ae9c31114e  
=-57;c .0.jar;C:\AVV_Hadoop\hadoopThis command was run using /C:/AVV_Hadoop/hadoop-3.4.0/share/hadoop/common/hadoop-common-3.4.0.jar  
2025-0 oop\common\lib\commons-io-?  
from Hadoop\hadoop-3.4.0\share\HC:\Windows\system32>start-dfs.cmd  
e0-474 commons-math3-3.6.1.jar;C:\A  
906c-0 .0\share\hadoop\common\lib\C:\Windows\system32>  
atedB1 .0.jar;C:\AVV_Hadoop\hadoop  
doop\common\lib\curator-re  
doop\hadoop-3.4.0\share\he  
-2.9.0.jar;C:\AVV_Hadoop\ha  
\common\lib\hadoop-annotati  
_Hadoop\hadoop-3.4.0\share\  
n\lib\hadoop-shaded-protob  
AWV_Hadoop\hadoop-3.4.0\sh  
20bjc-annotations-1.1.jar;C
```

Next steps: (for uploading files)

1.check jps (list all **Java processes** running on a system, primarily in Hadoop environments.)

2.Run start-yarn.cmd

3.Run : hdfs dfs -ls / (used to verify that HDFS is working correctly)

4.Upload the folder consisting of 5 files. Using “-put” command.

“C:\Windows\system32>hdfs dfs -put C:\Users\lenovo\Downloads\hadoop_files /”

```
Administrator: Command Prompt
C:\Windows\system32>start-dfs.cmd
C:\Windows\system32>jps
19056 Jps
20312 DataNode
18668 NameNode

C:\Windows\system32>start-yarn.cmd
starting yarn daemons

C:\Windows\system32>jps
11780 Jps
14196 ResourceManager
16564 NodeManager
20312 DataNode
18668 NameNode

C:\Windows\system32>hdfs dfs -ls /

C:\Windows\system32>hdfs fs -mkdir /avv_hadoop_files
Error: Could not find or load main class fs

C:\Windows\system32>hdfs dfs -mkdir /avv_hadoop_files

C:\Windows\system32>hdfs dfs -put C:\Users\lenovo\Downloads\hadoop_files /

C:\Windows\system32>
```

After running the last command I can now see the files in the localhost:9870 inside the folder “hadoop_files”.

