

# Cara Install Hadoop

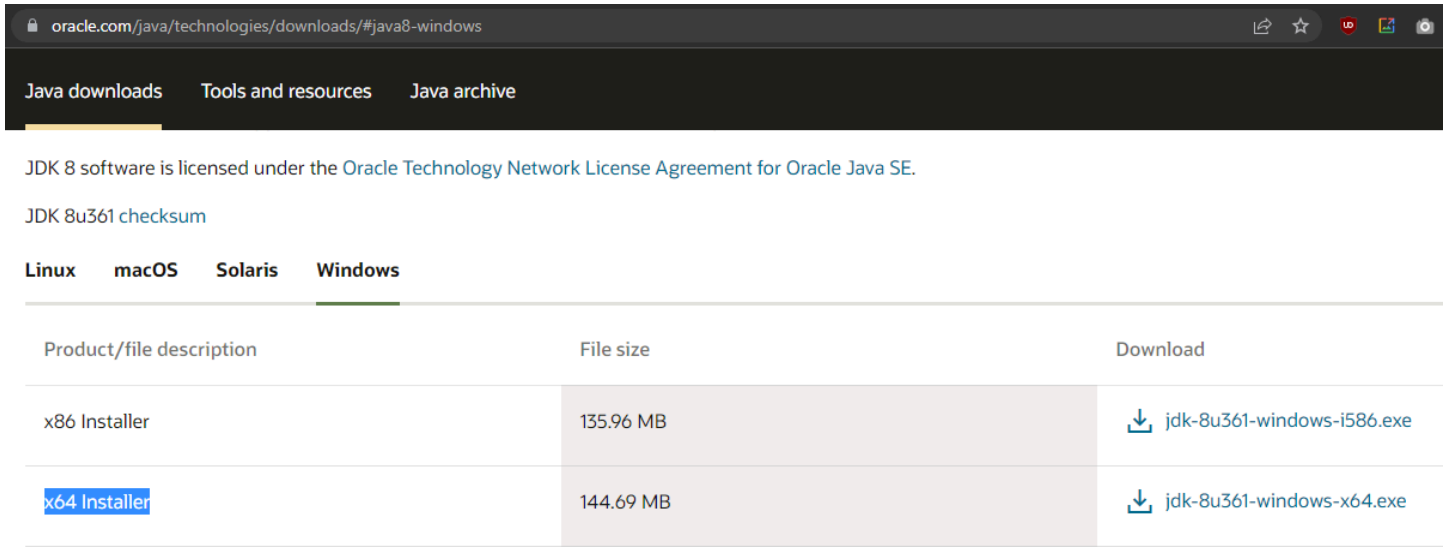
Muhammad Farhan Ananda Mirzah (2011522022), Big Data B

**Referensi yang digunakan** (sudah di tes bisa akses localhost:9870)

<https://www.youtube.com/watch?v=-JIC3h6KexQ&list=PPSV>

## A. Install Java Versi 8

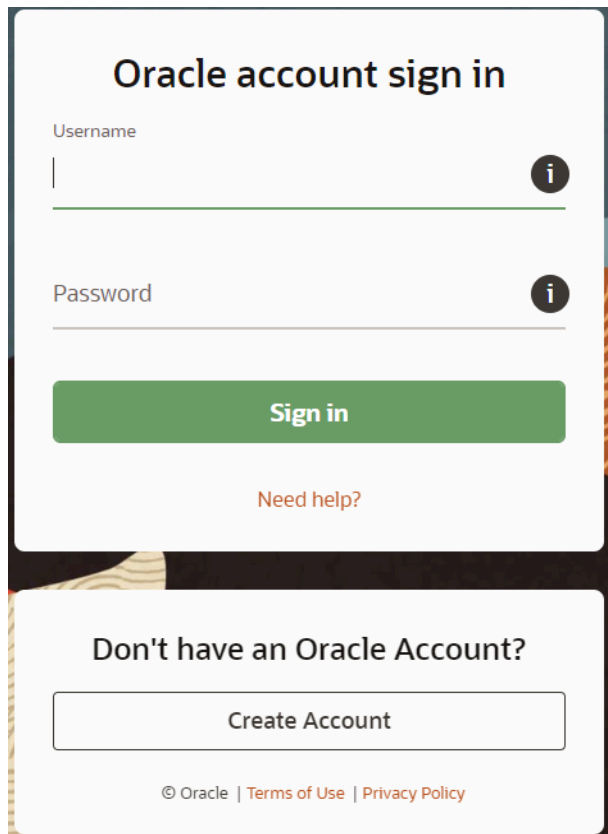
1. Download Java Development Kit 8 (JDK 8) windows x64 dari website berikut (<https://www.oracle.com/java/technologies/downloads/#java8-windows>)



The screenshot shows the Oracle Java Downloads page for JDK 8 on Windows. The page has a dark header with navigation links: "Java downloads", "Tools and resources", and "Java archive". Below the header, there is a section for "JDK 8u361 checksum" with tabs for "Linux", "macOS", "Solaris", and "Windows". The "Windows" tab is selected, showing a table of download links.

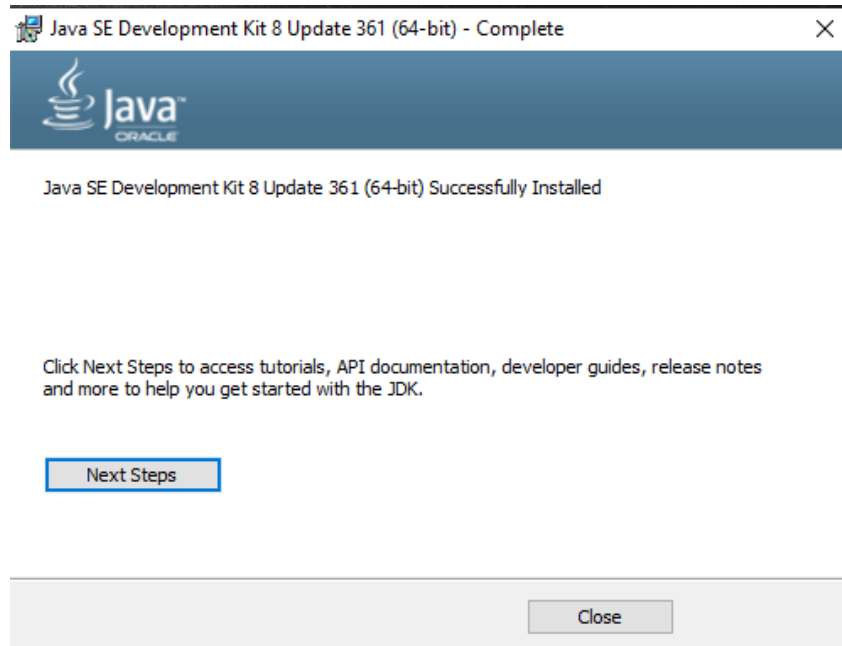
Product/file description	File size	Download
x86 Installer	135.96 MB	<a href="#">jdk-8u361-windows-i586.exe</a>
<b>x64 Installer</b>	144.69 MB	<a href="#">jdk-8u361-windows-x64.exe</a>

2. Login dengan akun Oracle. Jika belum ada akun, lakukan registrasi

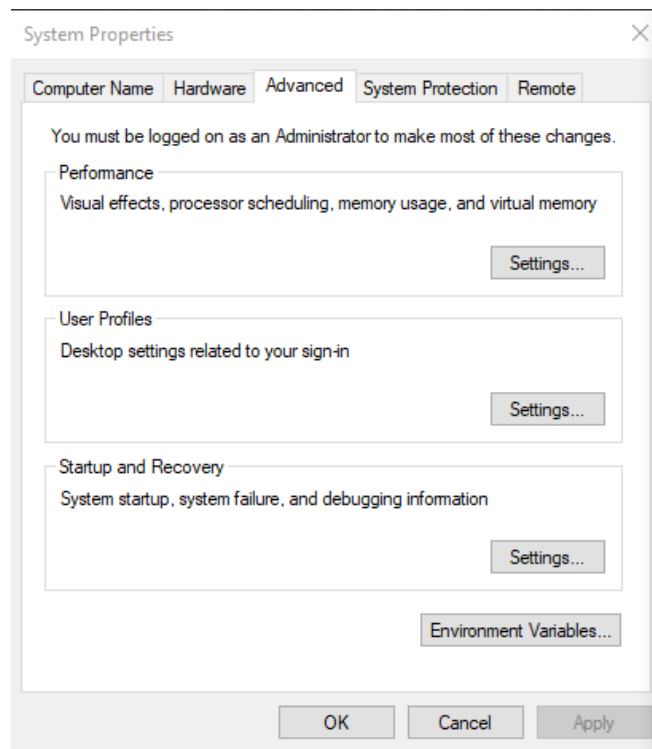


The screenshot shows the Oracle account sign in form. It has a title "Oracle account sign in" and two input fields: "Username" and "Password". Each field has an information icon (i) to its right. Below the input fields is a green "Sign in" button. Below the button is a link "Need help?". At the bottom of the form is a section titled "Don't have an Oracle Account?" with a "Create Account" button. At the very bottom, there is a footer with "© Oracle | Terms of Use | Privacy Policy".

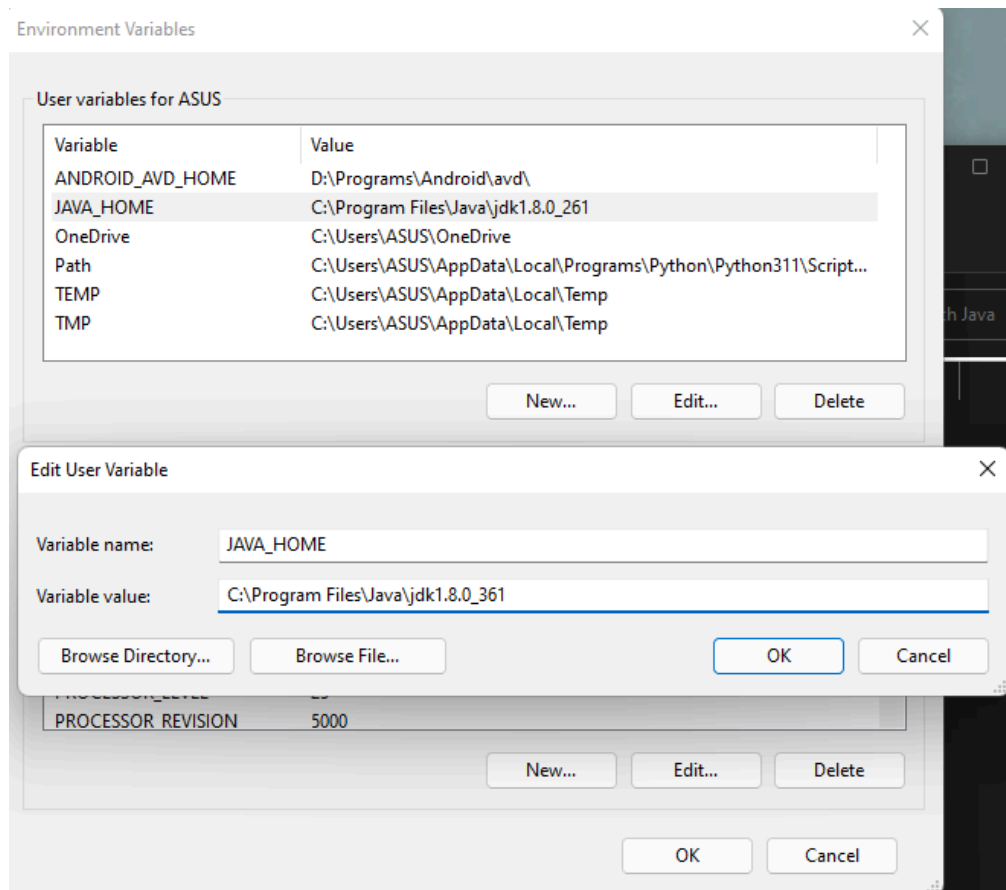
### 3. Install JDK 8 menggunakan file .exe yang di download



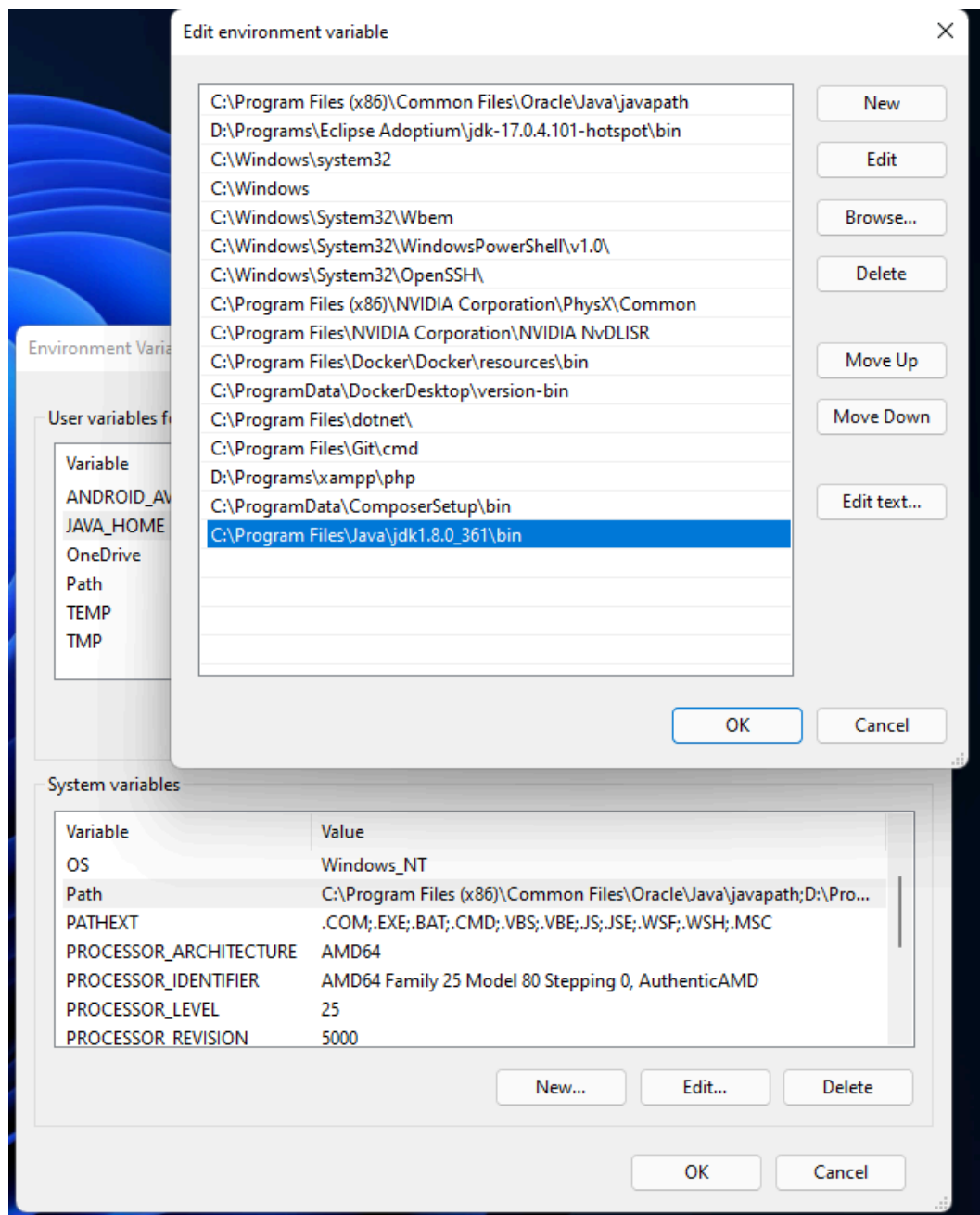
### 4. Lakukan Settings Environment Variables Java. Pertama, buka (Control Panel → System and Security → System → Advanced System Settings). Kemudian akan muncul dialog box System Properties, lalu klik Environment Variables.



### 5. Atur Home Java. Pada User variables, klik New kemudian isi Variable name dengan **JAVA\_HOME** dan Variable value dengan **direktori install jdk**. Kemudian **klik OK**.



6. Atur path Java. Pada System variables, klik **Path** dua kali. Kemudian klik tombol New yang di atas kanan, dan isi value dengan **direktori jdk\bin**. Lalu **klik OK** pada menu Edit environment variable, Environment Variables dan System Properties.



7. Buka cmd, cek versi java dengan perintah **java -version**. Apabila muncul versi java yang diinstall, maka proses instalasi Java berhasil.

```

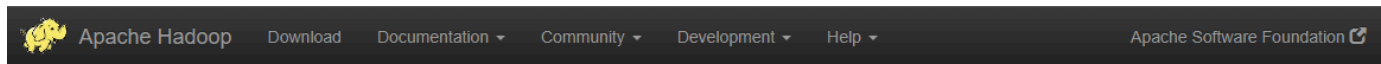
C:\Users\ASUS>java -version
java version "1.8.0_361"
Java(TM) SE Runtime Environment (build 1.8.0_361-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.361-b09, mixed mode)

C:\Users\ASUS>

```

## B. Install Hadoop

1. Download Hadoop dari link berikut (<https://hadoop.apache.org/releases.html>).  
Pilih versi **3.2.4 Binary download**.



### Download

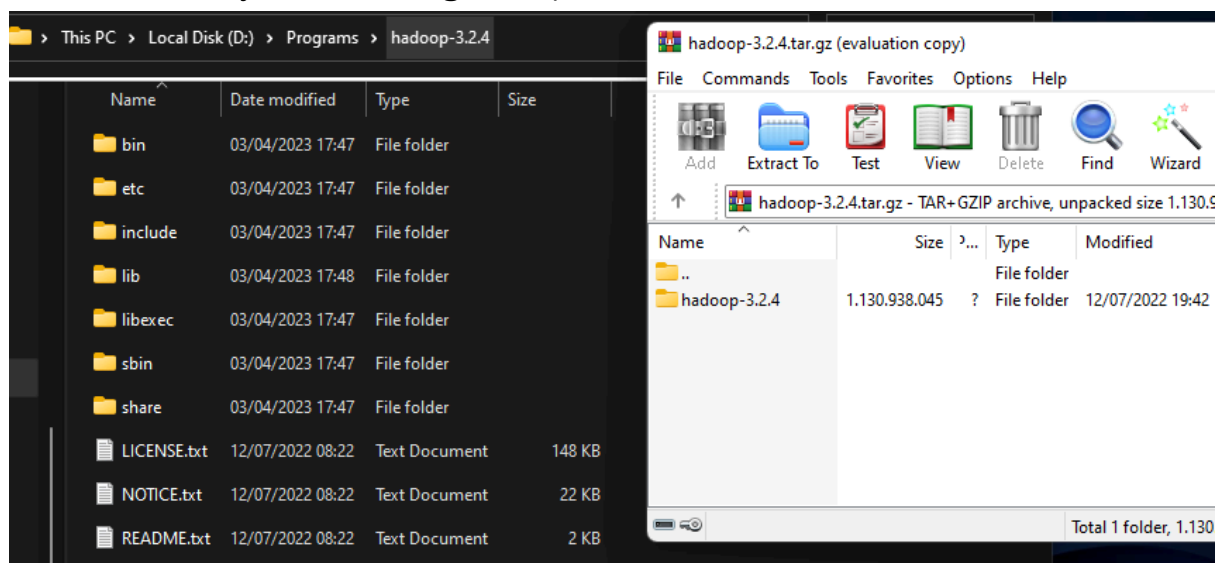
Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
3.3.5	2023 Mar 22	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a> <a href="#">binary-aarch64 (checksum signature)</a>	<a href="#">Announcement</a>
3.2.4	2022 Jul 22	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>
2.10.2	2022 May 31	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>

2. Download file **tar.gz** nya



3. Extract file hadoop ke direktori yang anda inginkan (dalam tutorial ini, saya meletakkannya di **D:\Programs**).

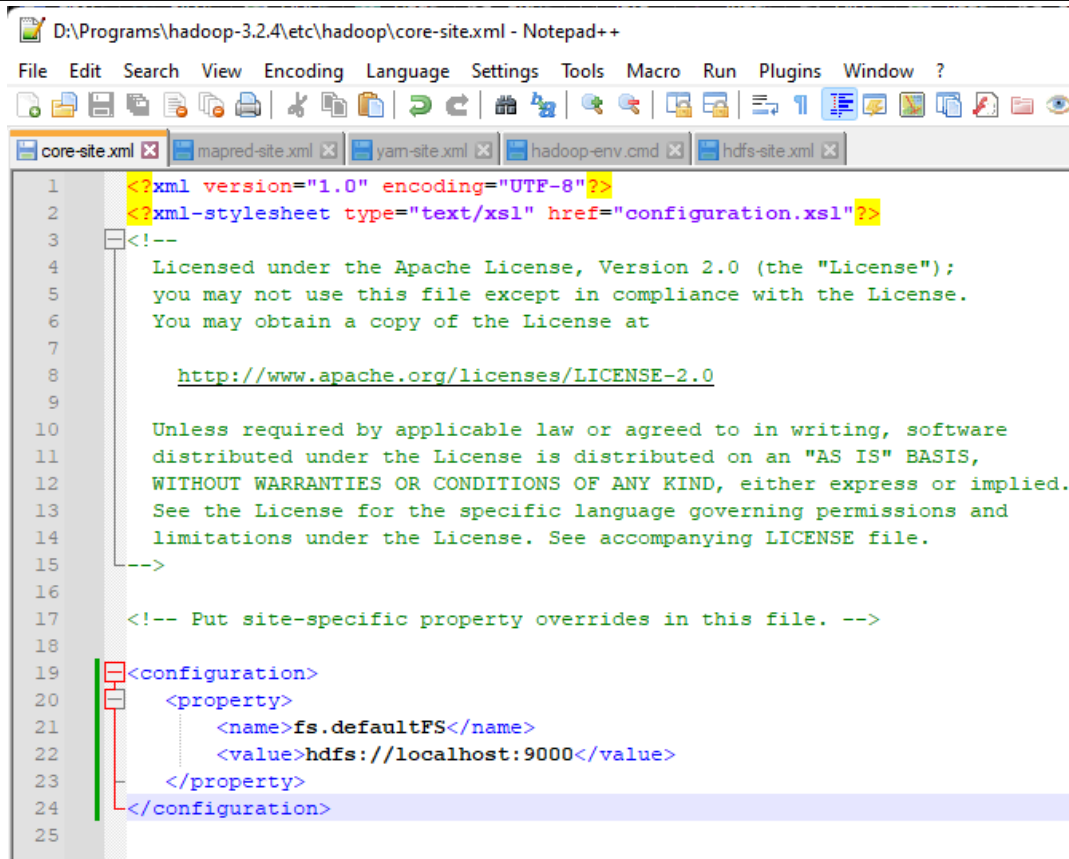


4. Buka file directory `hadoop-3.2.4\etc\hadoop`, lalu buka file **core-site.xml**, **hadoop-env.cmd**, **hdfs-site.xml**, **mapred-site.xml**, dan **yarn-site.xml** menggunakan notepad atau notepad++.



5. Tambahkan code berikut pada file **core-site.xml**

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



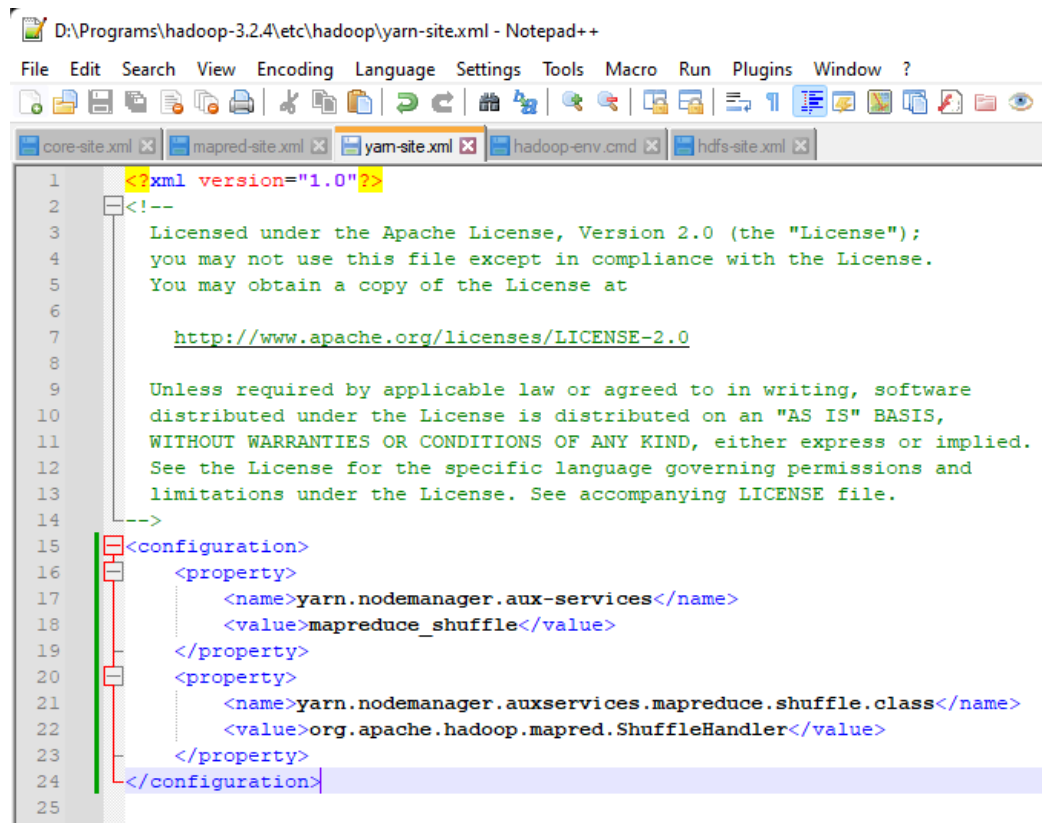
6. Tambahkan code berikut pada file **mapred-site.xml**

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

```
D:\Programs\hadoop-3.2.4\etc\hadoop\mapred-site.xml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
core-site.xml x mapred-site.xml x yarn-site.xml x hadoop-env.cmd x hdfs-site.xml x
1 <?xml version="1.0"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>mapreduce.framework.name</name>
22     <value>yarn</value>
23   </property>
24 </configuration>
25
```

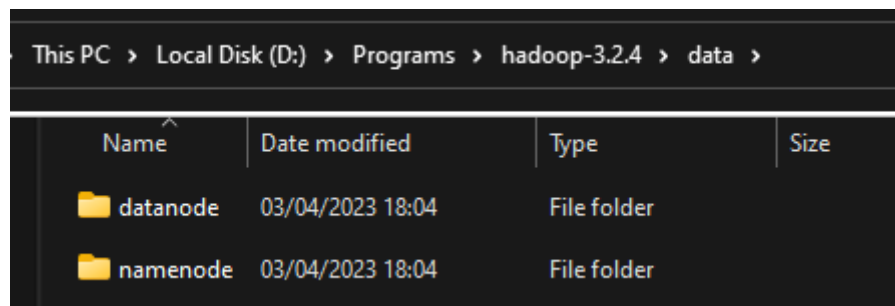
7. Tambahkan code berikut pada file **yarn-site.xml**

```
<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>
```



```
1 <?xml version="1.0"?>
2 <!--
3 Licensed under the Apache License, Version 2.0 (the "License");
4 you may not use this file except in compliance with the License.
5 You may obtain a copy of the License at
6
7 http://www.apache.org/licenses/LICENSE-2.0
8
9 Unless required by applicable law or agreed to in writing, software
10 distributed under the License is distributed on an "AS IS" BASIS,
11 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
12 See the License for the specific language governing permissions and
13 limitations under the License. See accompanying LICENSE file.
14 -->
15 <configuration>
16   <property>
17     <name>yarn.nodemanager.aux-services</name>
18     <value>mapreduce_shuffle</value>
19   </property>
20   <property>
21     <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
22     <value>org.apache.hadoop.mapred.ShuffleHandler</value>
23   </property>
24 </configuration>
```

8. Pada direktori hadoop buatlah folder baru dengan nama data, didalam folder tersebut buat 2 folder baru dengan nama datanode dan namenode.

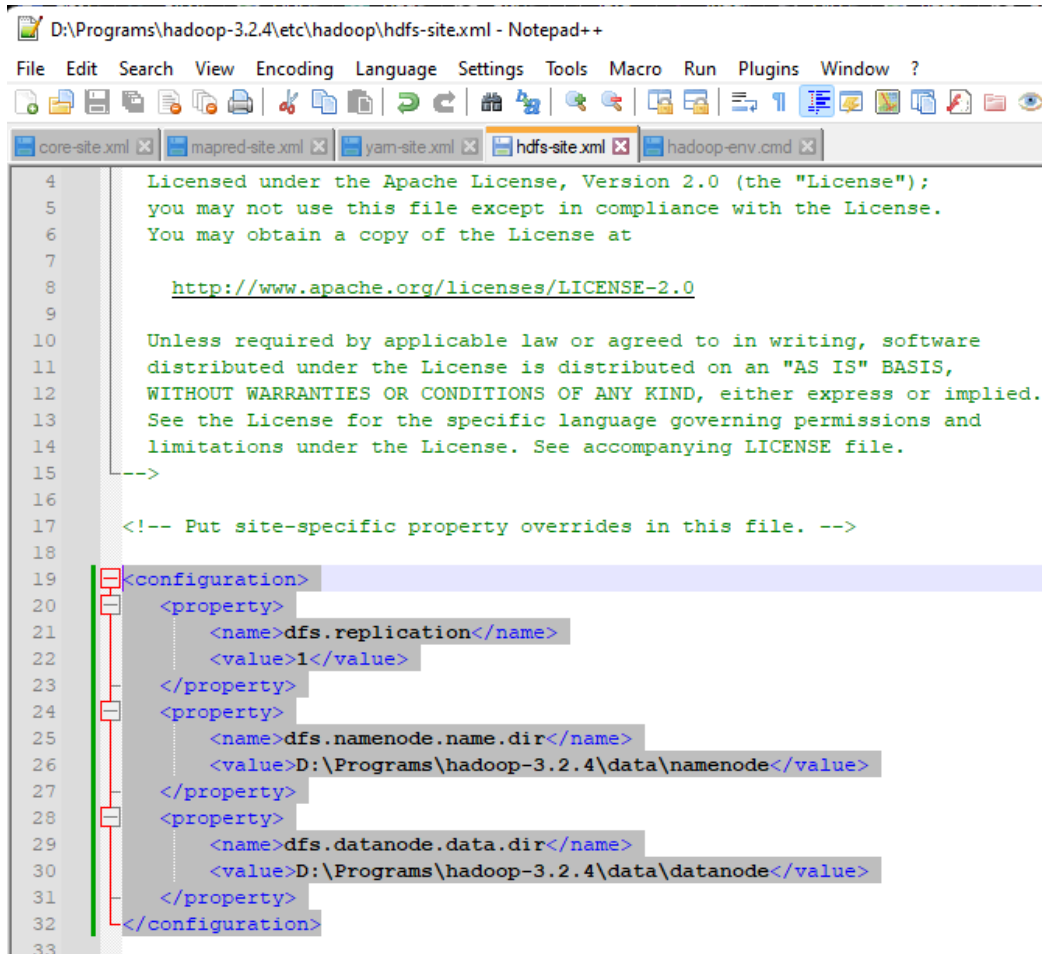


9. Tambahkan code berikut pada file **hdfs-site.xml**. Untuk **value** disesuaikan dengan **direktori** dimana folder **namenode** dan **datanode** dibuat. Untuk instalasi di tutorial ini, value nya seperti ini:

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>D:\Programs\hadoop-3.2.4\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
```

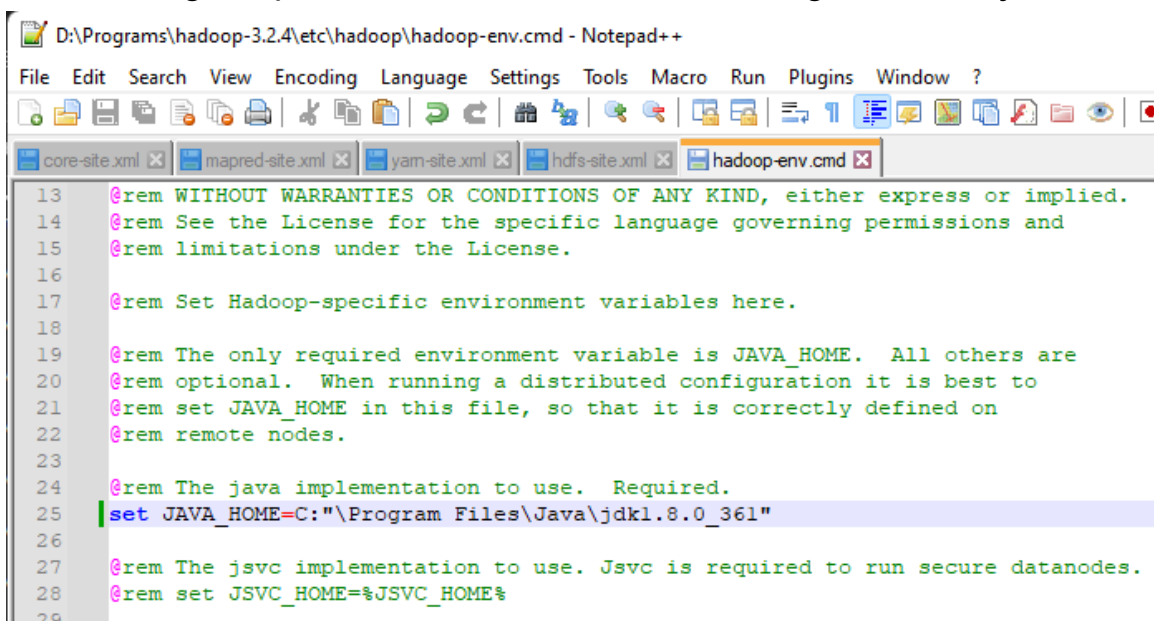


```
<value>D:\Programs\hadoop-3.2.4\data\datanode</value>
</property>
</configuration>
```



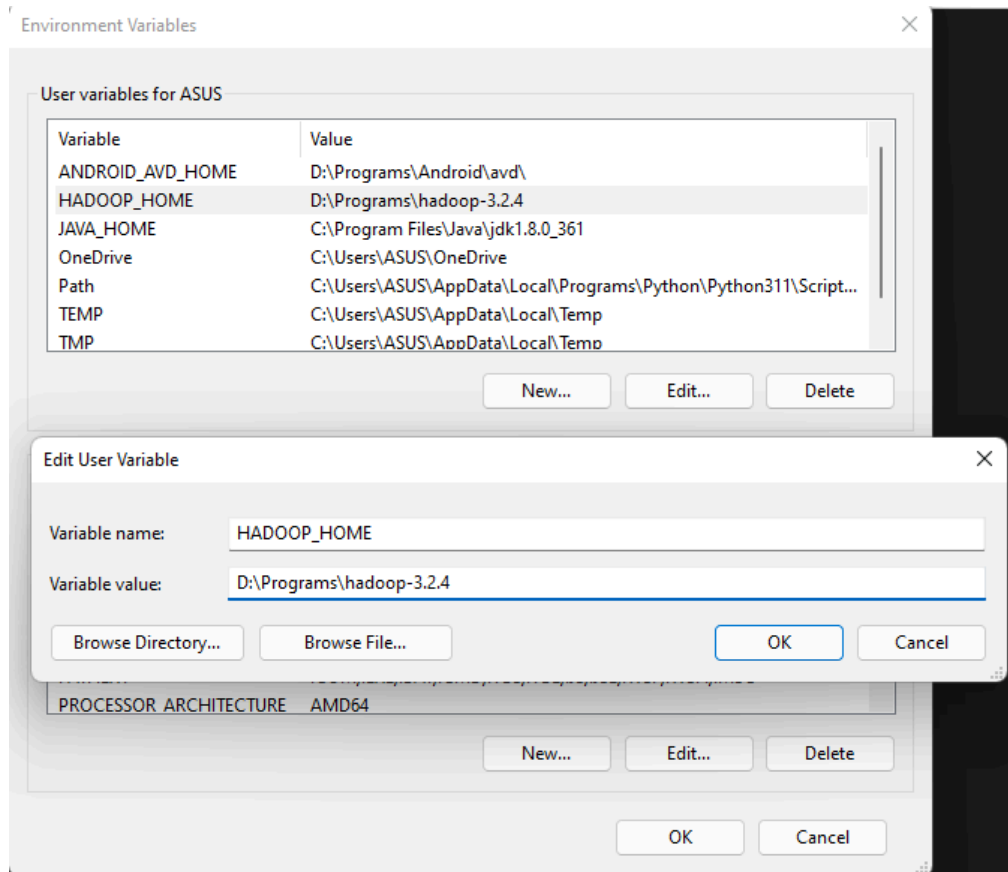
```
D:\Programs\hadoop-3.2.4\etc\hadoop\hdfs-site.xml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
core-site.xml x mapred-site.xml x yam-site.xml x hdfs-site.xml x hadoop-env.cmd x
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 ---
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>dfs.replication</name>
22     <value>1</value>
23   </property>
24   <property>
25     <name>dfs.namenode.name.dir</name>
26     <value>D:\Programs\hadoop-3.2.4\data\namenode</value>
27   </property>
28   <property>
29     <name>dfs.datanode.data.dir</name>
30     <value>D:\Programs\hadoop-3.2.4\data\datanode</value>
31   </property>
32 </configuration>
33
```

10. Pada file **hadoop-env.cmd**, sesuaikan direktori **JAVA\_HOME** dengan direktori **java jdk**. Penulisan direktori JAVA\_HOME sebenarnya tidak boleh terdapat folder yang mengandung spasi, namun apabila ingin tetap terdapat folder dengan spasi, maka ditambahkan “....” agar tidak terjadi error.

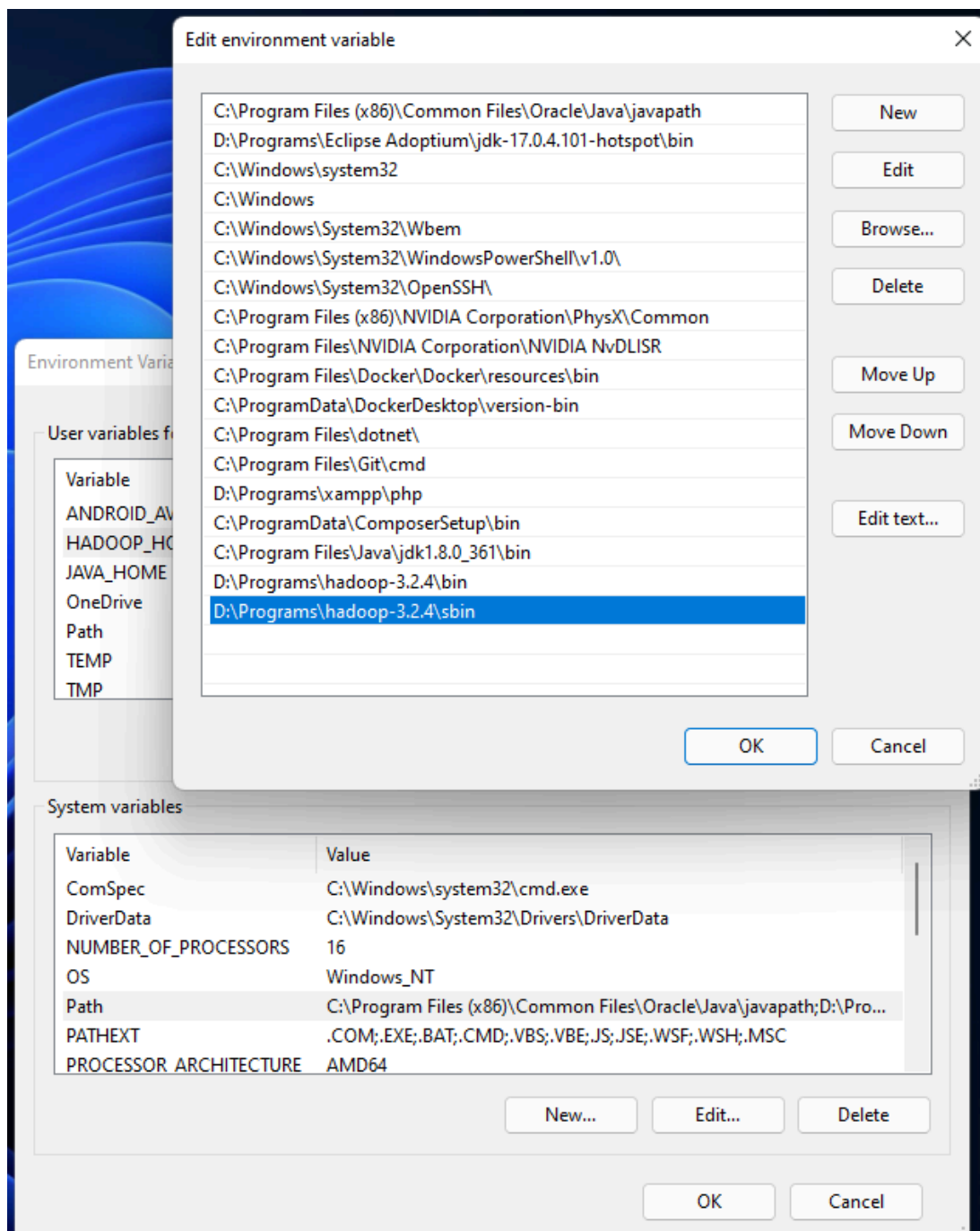


```
D:\Programs\hadoop-3.2.4\etc\hadoop\hadoop-env.cmd - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
core-site.xml x mapred-site.xml x yam-site.xml x hdfs-site.xml x hadoop-env.cmd x
13 @rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 @rem See the License for the specific language governing permissions and
15 @rem limitations under the License.
16
17 @rem Set Hadoop-specific environment variables here.
18
19 @rem The only required environment variable is JAVA_HOME. All others are
20 @rem optional. When running a distributed configuration it is best to
21 @rem set JAVA_HOME in this file, so that it is correctly defined on
22 @rem remote nodes.
23
24 @rem The java implementation to use. Required.
25 set JAVA_HOME=C:\Program Files\Java\jdk1.8.0_361
26
27 @rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
28 @rem set JSVC_HOME=%JSVC_HOME%
29
```

11. Setelah melakukan konfigurasi, sekarang Kita akan mengatur Environment Variables pada Hadoop. Buka (Control Panel → System and Security → System Advanced System Settings). Kemudian akan muncul dialog box System Properties, lalu klik Environment Variables.
12. Atur Home Hadoop. Pada User variables, klik New kemudian isi Variable name dengan **HADOOP\_HOME** dan variable value dengan **direktori hadoop**. Kemudian **klik OK**.



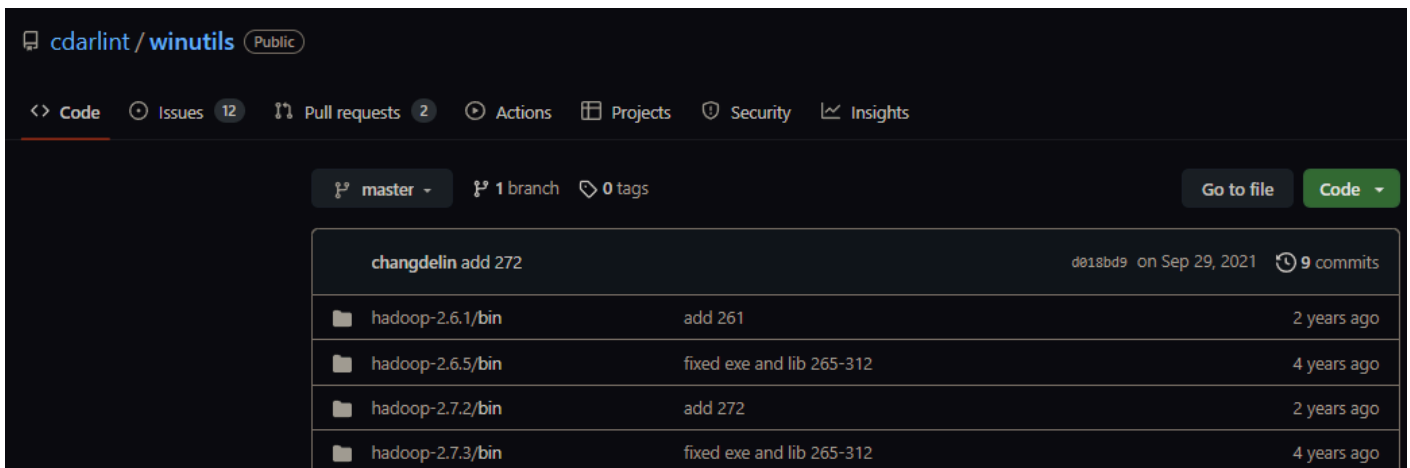
13. Atur Path Hadoop. Pada System variables, klik **Path**. Kemudian klik New dan isi dengan direktori **hadoop\bin** dan **hadoop\sbin**. Lalu **klik OK** pada Edit environment variable, Environment Variables dan System Properties.



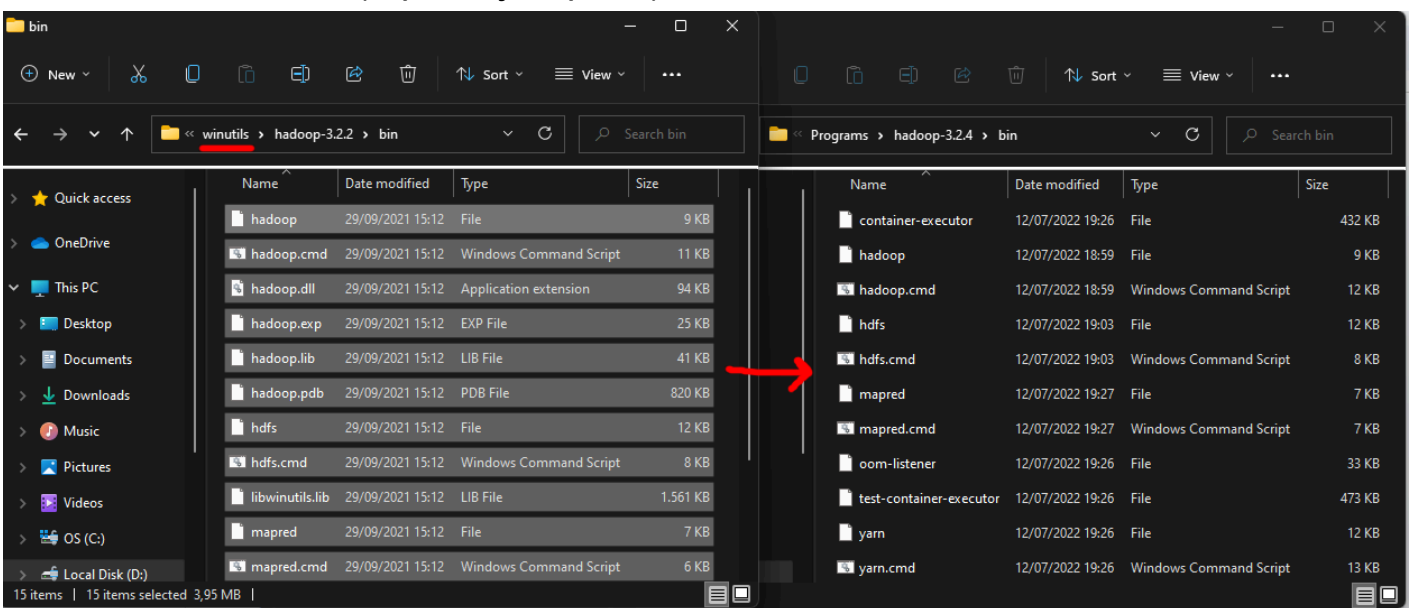
14. Agar Hadoop dapat dioperasikan pada OS Windows, diperlukan patch file **winutils**. File tersebut dapat di download melalui link berikut:

- <https://github.com/cdarlint/winutils>

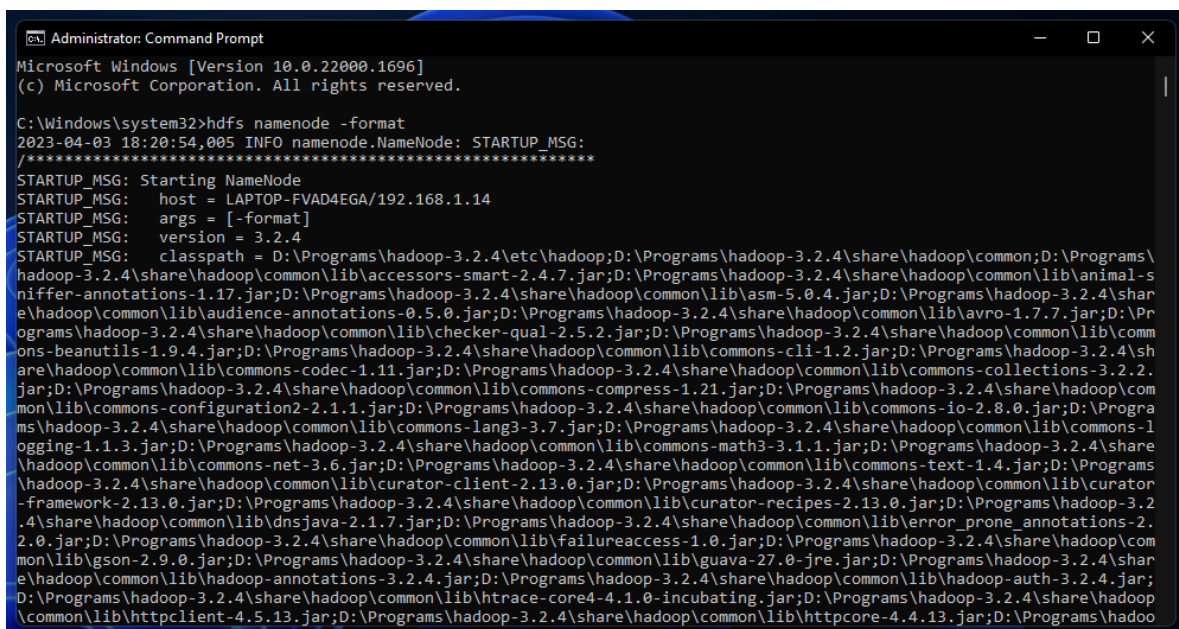
Untuk sekarang, kita akan memilih folder yang versi paling dekat dengan hadoop-3.2.4, yaitu **3.2.2**



15. Ekstrak file tersebut, kemudian copy isi folder bin yang di download ke dalam folder (replace jika perlu)



16. Buka CMD sebagai administrator. Kemudian ketikkan perintah **hdfs namenode -format** (jika ada error "hdfs is not recognized", pastikan semua Environment Variables sudah sesuai dan sudah di klik ok)



Kalau berhasil, akan ada “namenode has been successfully formatted”

```
Select Administrator: Command Prompt
2023-04-03 18:20:55,343 INFO util.GSet: capacity = 2^18 = 262144 entries
2023-04-03 18:20:55,352 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2023-04-03 18:20:55,352 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2023-04-03 18:20:55,353 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2023-04-03 18:20:55,357 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2023-04-03 18:20:55,357 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry
time is 600000 millis
2023-04-03 18:20:55,359 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2023-04-03 18:20:55,359 INFO util.GSet: VM type = 64-bit
2023-04-03 18:20:55,360 INFO util.GSet: 0.0299999999329447746% max memory 889 MB = 273.1 KB
2023-04-03 18:20:55,360 INFO util.GSet: capacity = 2^15 = 32768 entries
2023-04-03 18:20:55,390 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1330794331-192.168.1.14-1680520855384
2023-04-03 18:20:55,415 INFO common.Storage: Storage directory D:\Programs\hadoop-3.2.4\data\namenode has been successfu
lly formatted.
2023-04-03 18:20:55,441 INFO namenode.FSImageFormatProtobuf: Saving image file D:\Programs\hadoop-3.2.4\data\namenode\cu
rrent\fsimage.ckpt_000000000000000000 using no compression
2023-04-03 18:20:55,531 INFO namenode.FSImageFormatProtobuf: Image file D:\Programs\hadoop-3.2.4\data\namenode\current\fs
image.ckpt_000000000000000000 of size 399 bytes saved in 0 seconds .
2023-04-03 18:20:55,541 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2023-04-03 18:20:55,555 INFO namenode.FSNamesystem: Stopping services started for active state
2023-04-03 18:20:55,556 INFO namenode.FSNamesystem: Stopping services started for standby state
2023-04-03 18:20:55,560 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2023-04-03 18:20:55,560 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at LAPTOP-FVAD4EGA/192.168.1.14
*****/
C:\Windows\system32>
```

17. Gunakan command **cd** untuk masuk ke folder **hadoop\sbin**

```
Administrator: Command Prompt

C:\Windows\system32>D:

D:\>cd D:\Programs\hadoop-3.2.4\sbin

D:\Programs\hadoop-3.2.4\sbin>
```

18. Ketikkan command **start-all.cmd**. Jika berhasil, 4 window baru akan muncul (**namenode**, **datanode**, **resourcemanager**, dan **nodemanager**)

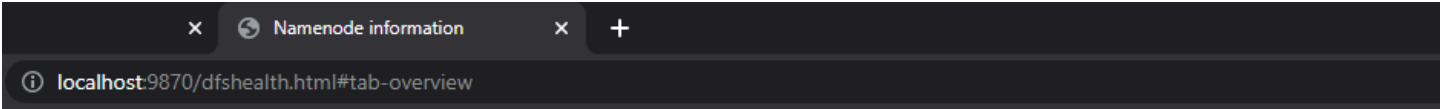


```
Administrator: Command Prompt
C:\> Apache Hadoop Distribution - hadoop namenode
D:\> Apache Hadoop Distribution - hadoop datanode
D:\> Apache Hadoop Distribution - yarn resourcemanager
D:\> Apache Hadoop Distribution - yarn nodemanager
2023-04-03 18:25:51,892 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@17d32e9b{node/,file:///C:/Users/ASUS
/AppData/Local/Temp/jetty-0_0_0_8042-hadoop-yarn-common-3_2_4_jar-_any-8714620941998349192/webapp/,AVAILABLE}{jar:fil
e:/D:/Programs/hadoop-3.2.4/share/hadoop/yarn/hadoop-yarn-common-3.2.4.jar!/webapps/node}
2023-04-03 18:25:51,907 INFO server.AbstractConnector: Started ServerConnector@ce5a68e{HTTP/1.1,(http/1.1)}{0.0.0.0:804
2}
2023-04-03 18:25:51,907 INFO server.Server: Started @5425ms
2023-04-03 18:25:51,907 INFO webapp.WebApps: Web app node started at 8042
2023-04-03 18:25:51,909 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : host.docker.internal:55089
2023-04-03 18:25:51,909 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2023-04-03 18:25:51,914 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8031
2023-04-03 18:25:52,037 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers :[]
2023-04-03 18:25:52,258 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key wi
th id -167922748
2023-04-03 18:25:52,259 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id
2075875139
2023-04-03 18:25:52,260 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as host.docker.internal:
55089 with total resource of <memory:8192, vCores:8>
```

19. Gunakan command **jps** untuk melihat apakah Hadoop berjalan dengan benar (seharusnya 5 ini berjalan)

```
Administrator: Command Prompt
D:\Programs\hadoop-3.2.4\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
D:\Programs\hadoop-3.2.4\sbin>jps
26496 ResourceManager
29120 NameNode
33680 NodeManager
31652 Jps
25992 DataNode
D:\Programs\hadoop-3.2.4\sbin>_
```

20. Coba buka
- a. <http://localhost:9870>



## Overview 'localhost:9000' (active)

Started:	Mon Apr 03 18:25:48 +0700 2023
Version:	3.2.4, r7e5d9983b388e372fe640f21f048f2f2ae6e9eba
Compiled:	Tue Jul 12 18:58:00 +0700 2022 by ubuntu from branch-3.2.4
Cluster ID:	CID-30646e0c-2682-4cd3-9e1b-67063ed09b73
Block Pool ID:	BP-1330794331-192.168.1.14-1680520855384

## 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 91.73 MB of 184.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 55.15 MB of 56.55 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.


Configured Capacity:	341.8 GB
Configured Remote Capacity:	0 B
DFS Used:	321 B (0%)
Non DFS Used:	305.9 GB
DFS Remaining:	35.89 GB (10.5%)
Block Pool Used:	321 B (0%)

b. <http://localhost:8088>

All Applications

New Tab

localhost:8088/cluster



Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

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>

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>

Show 20 entries

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

Showing 0 to 0 of 0 entries

21. Jika dapat dibuka, maka Hadoop sudah berhasil di install.