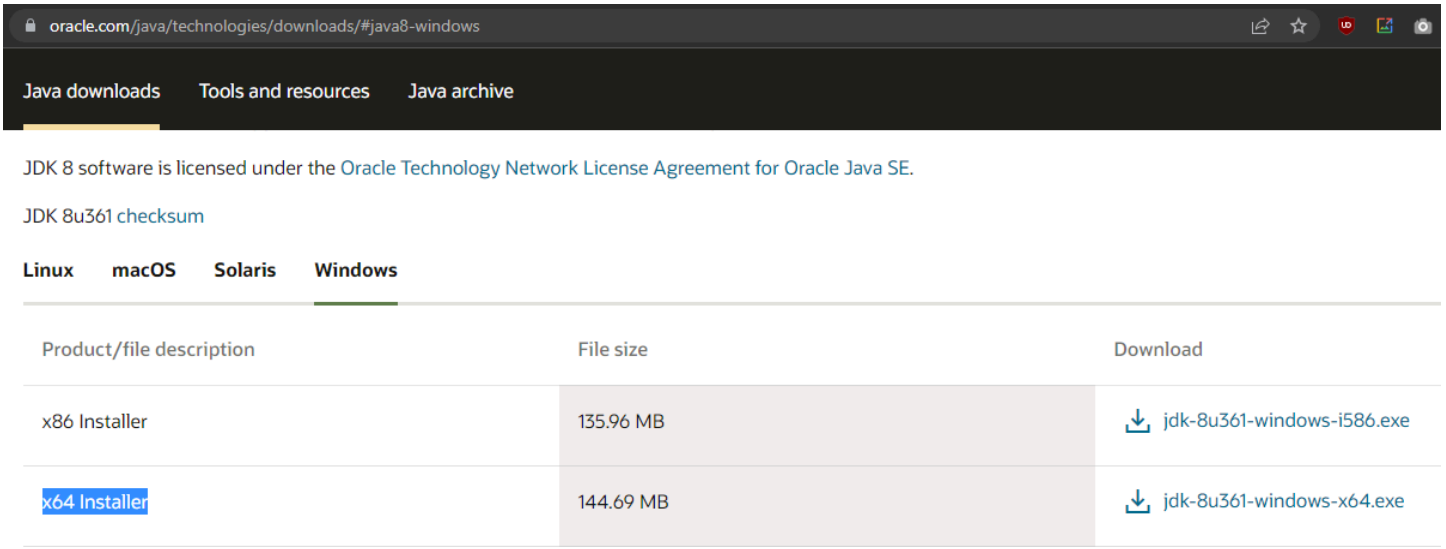
Cara Install Hadoop

Muhammad Farhan Ananda Mirzah (2011522022), Big Data B

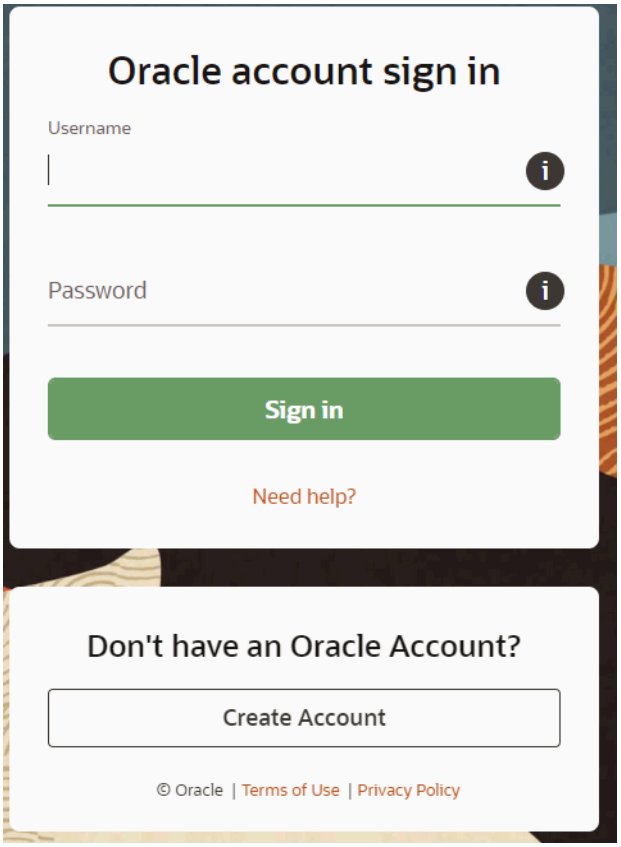
Referensi yang digunakan (sudah di tes bisa akses localhost:9870)
https://www.youtube.com/watch?v=-JlC3h6KexQ&list=PPSV

A. Install Java Versi 8

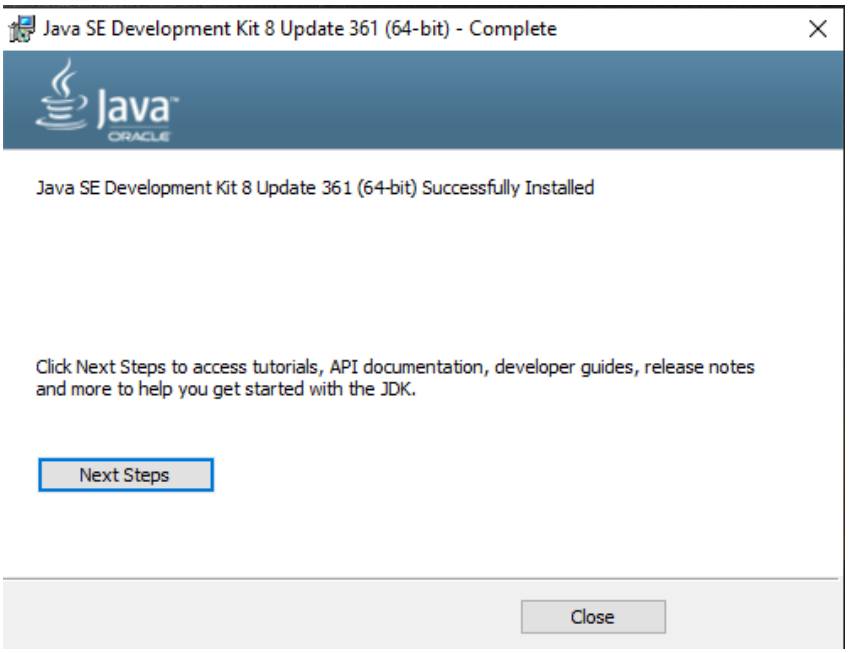
1. Download Java Development Kit 8 (JDK 8) windows x64 dari website berikut (https://www.oracle.com/java/technologies/downloads/#java8-windows)



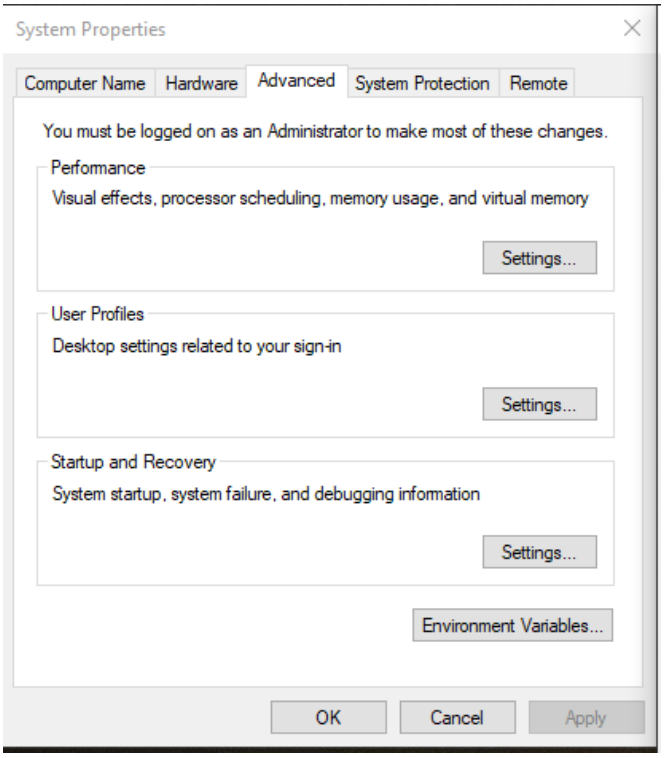
2. Login dengan akun Oracle. Jika belum ada akun, lakukan registrasi



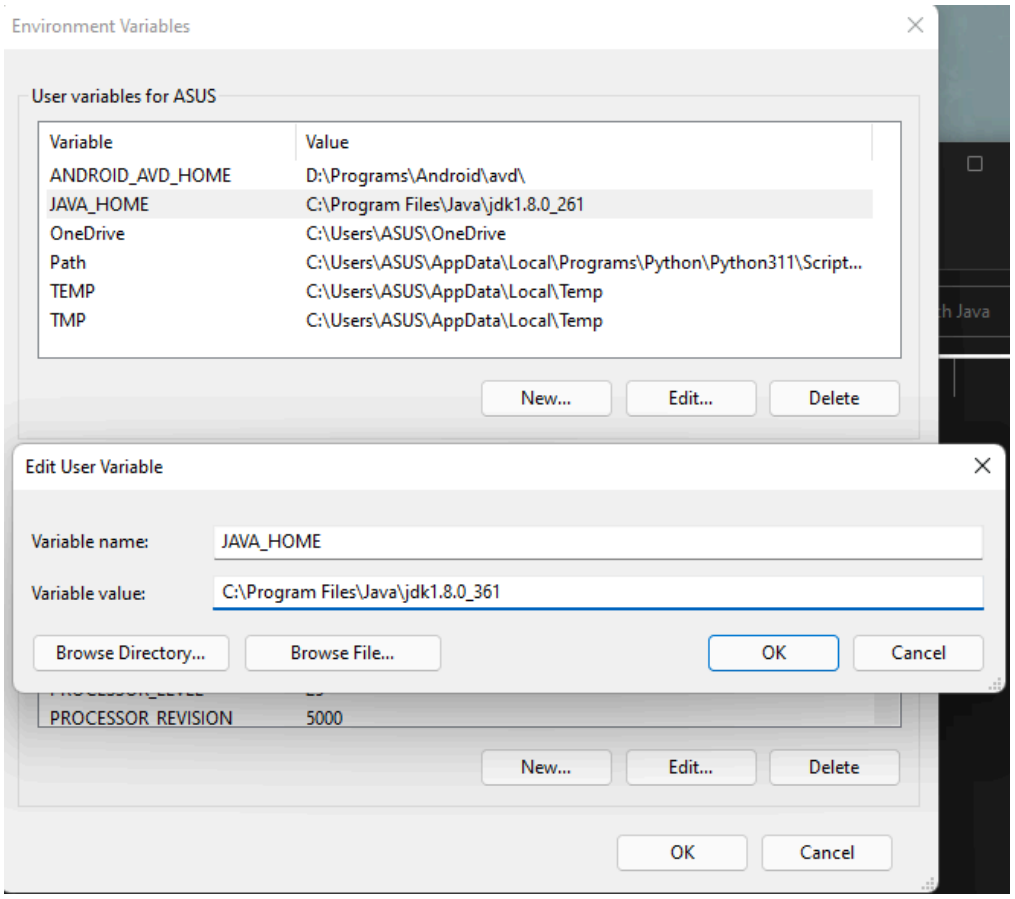
3. Install JDK 8 menggunakan file .exe yang di download



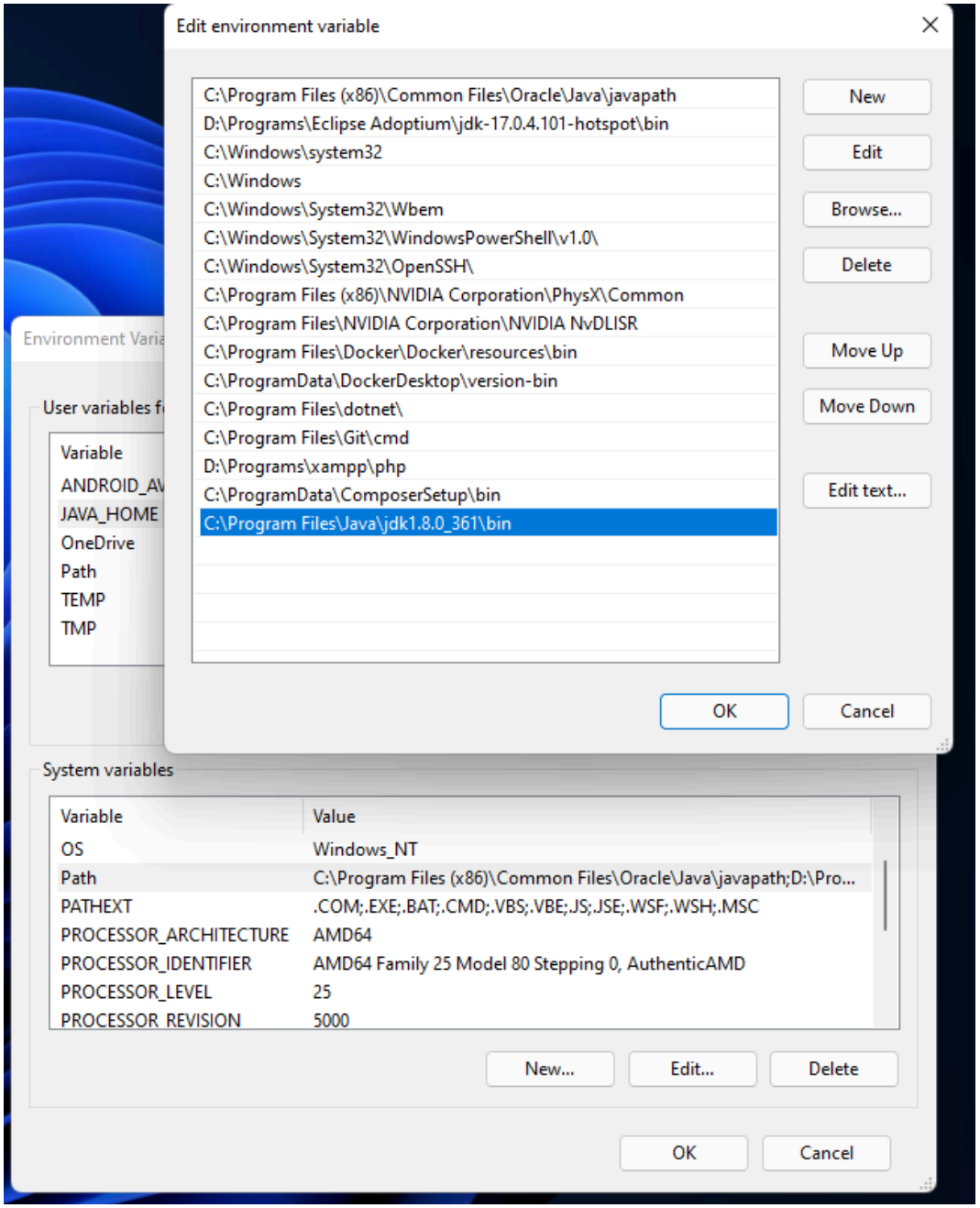
4. Lakukan Settings Environment Variables Java. Pertama, buka (Control Panel → System and Security → System → Advanced System Settings). Kemudian akan muncul dialog box System Properties, lalu klik Environment Variables.



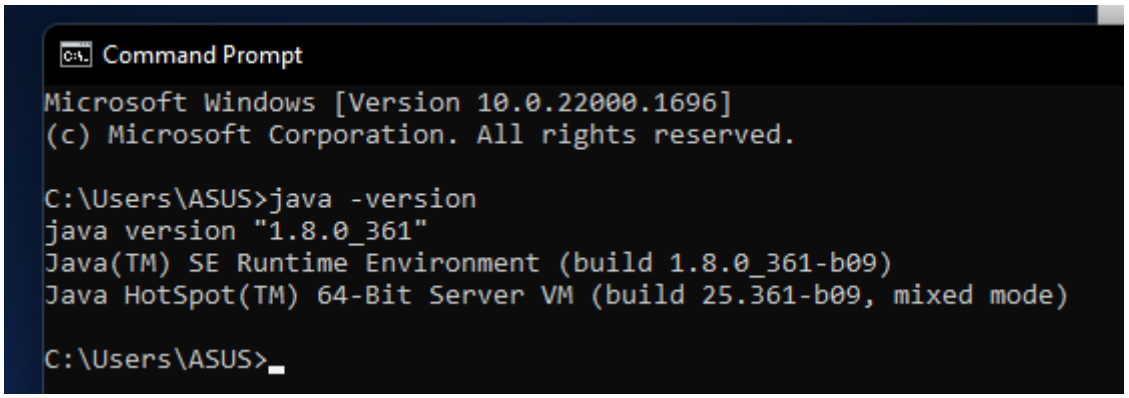
5. Atur Home Java. Pada User variables, klik New kemudian isi Variable name dengan **JAVA_HOME** dan Variable value dengan **direktori install jdk**. Kemudian **klik OK**.



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.

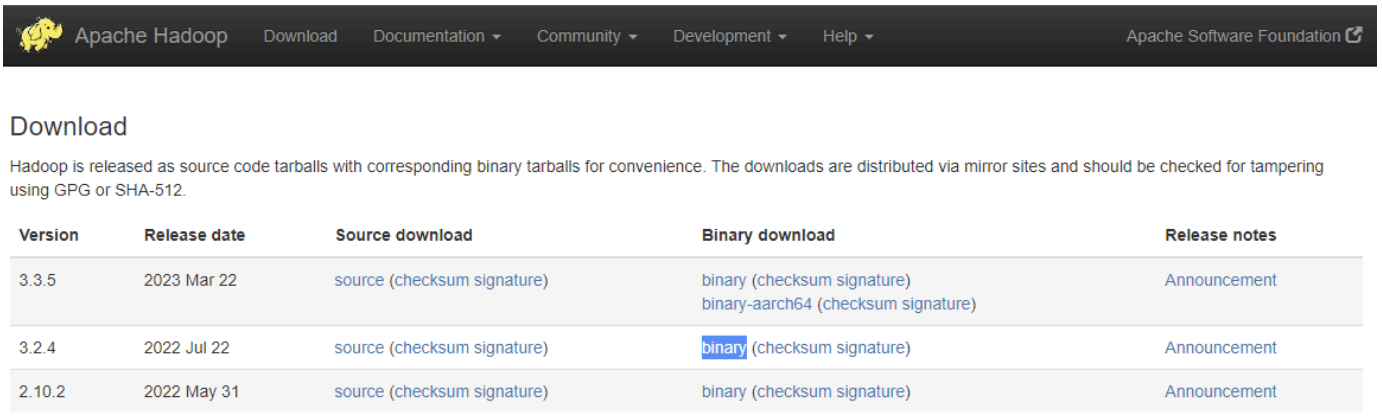


7. Buka cmd, cek versi java dengan perintah **java –version**. Apabila muncul versi java yang diinstall, maka proses instalasi Java berhasil.



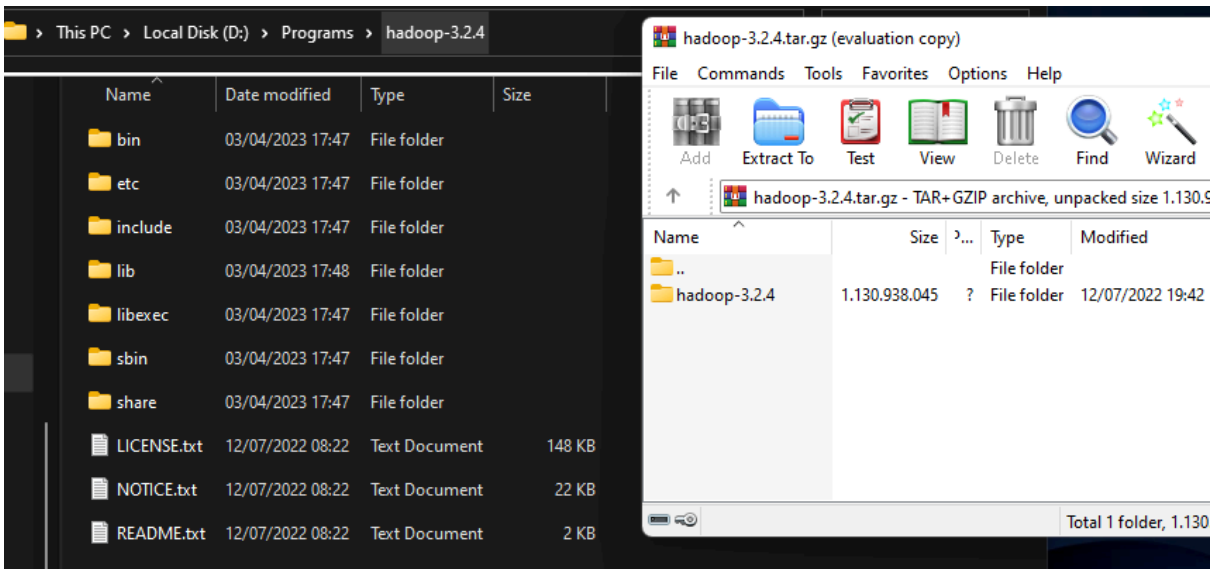
B. Install Hadoop

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

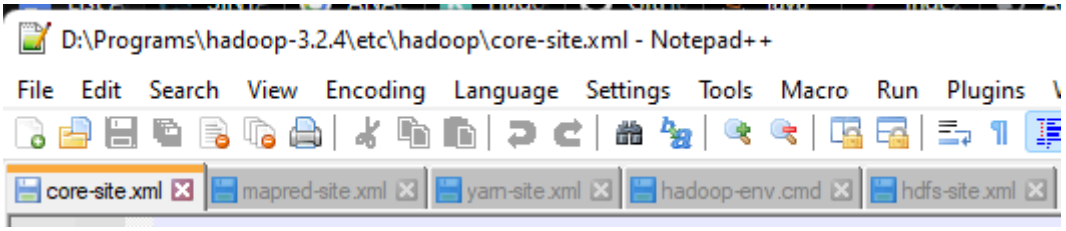


2. Download file **tar.gz** nya

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

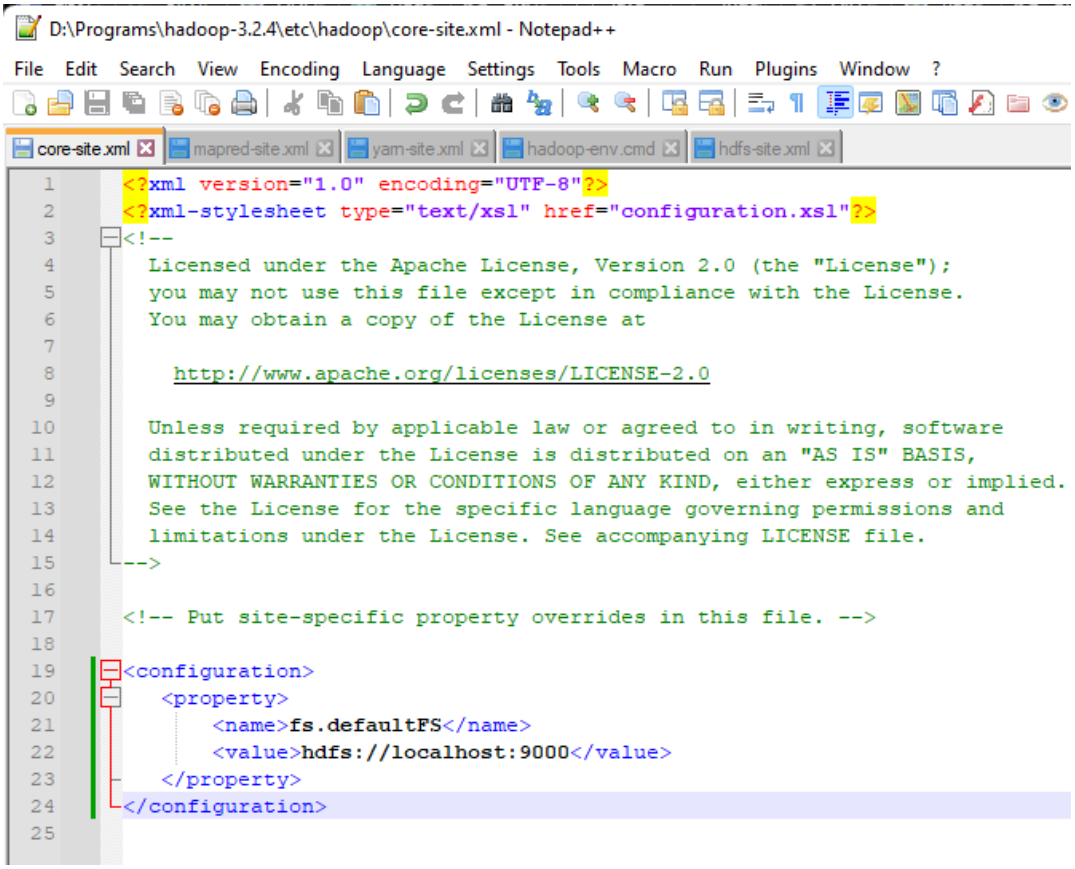


4. Buka file directory hadoop-3.2.4\etc\hadoop, lalu buka file **core-site.xml, hadoop-env.cmd, hdfs-site.xml, mapred-site.xml,** dan **yarn-site.xml** menggunaan notepad atau notepad++.



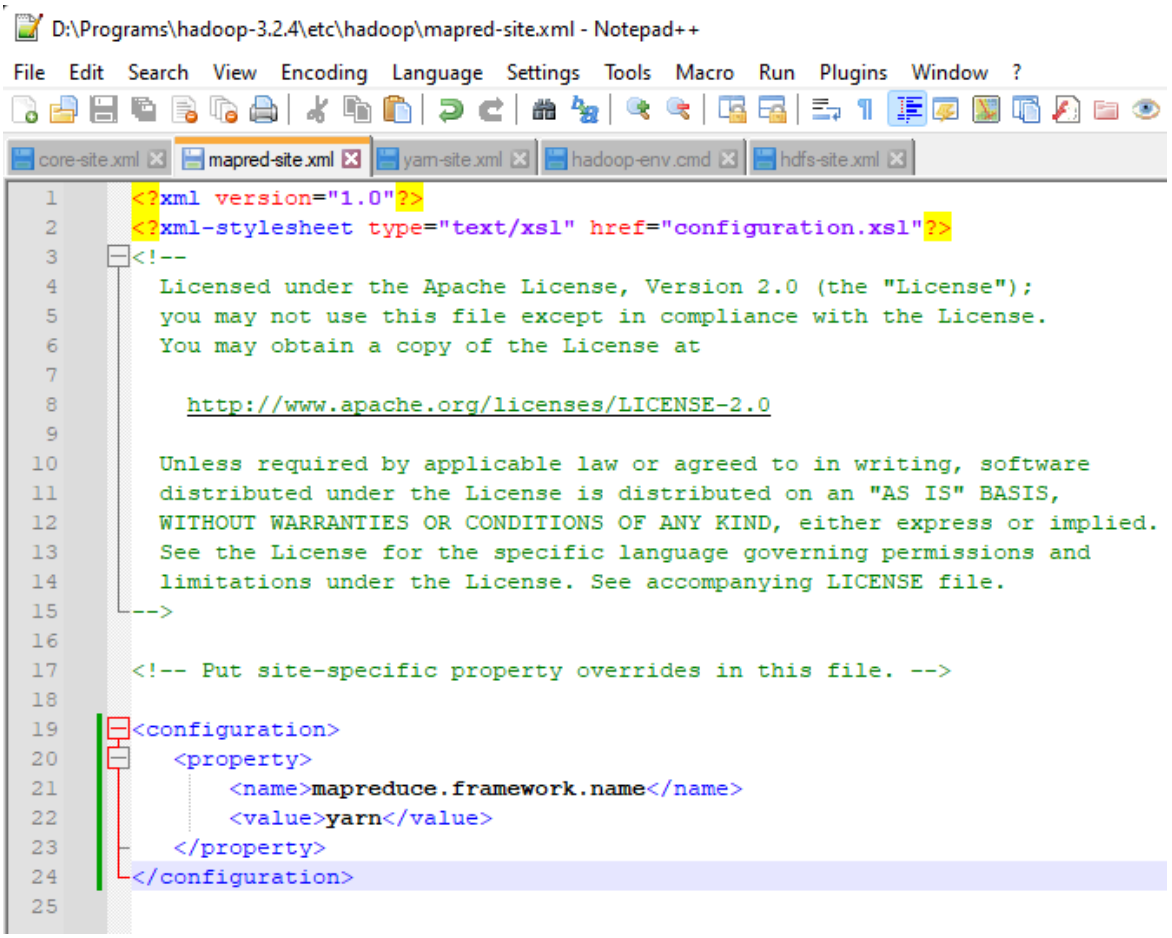
5. Tambahkan code berikut pada file **core-site.xml**

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



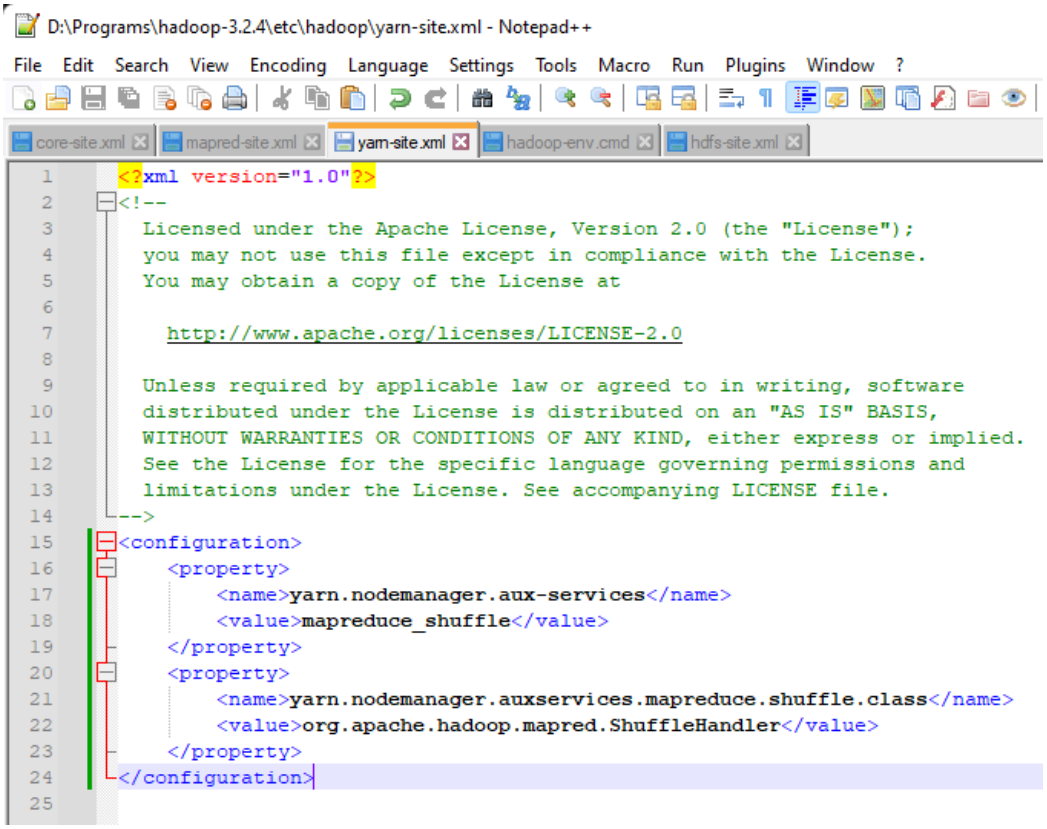
6. Tambahkan code berikut pada file **mapred-site.xml**

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

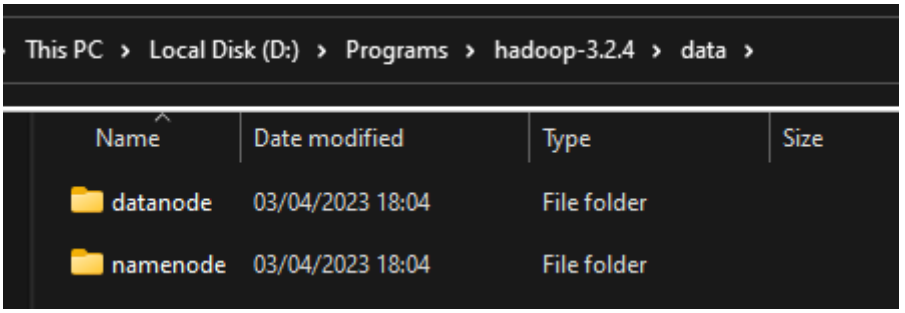


7. Tambahkan code berikut pada file **yarn-site.xml**

```
<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>
```



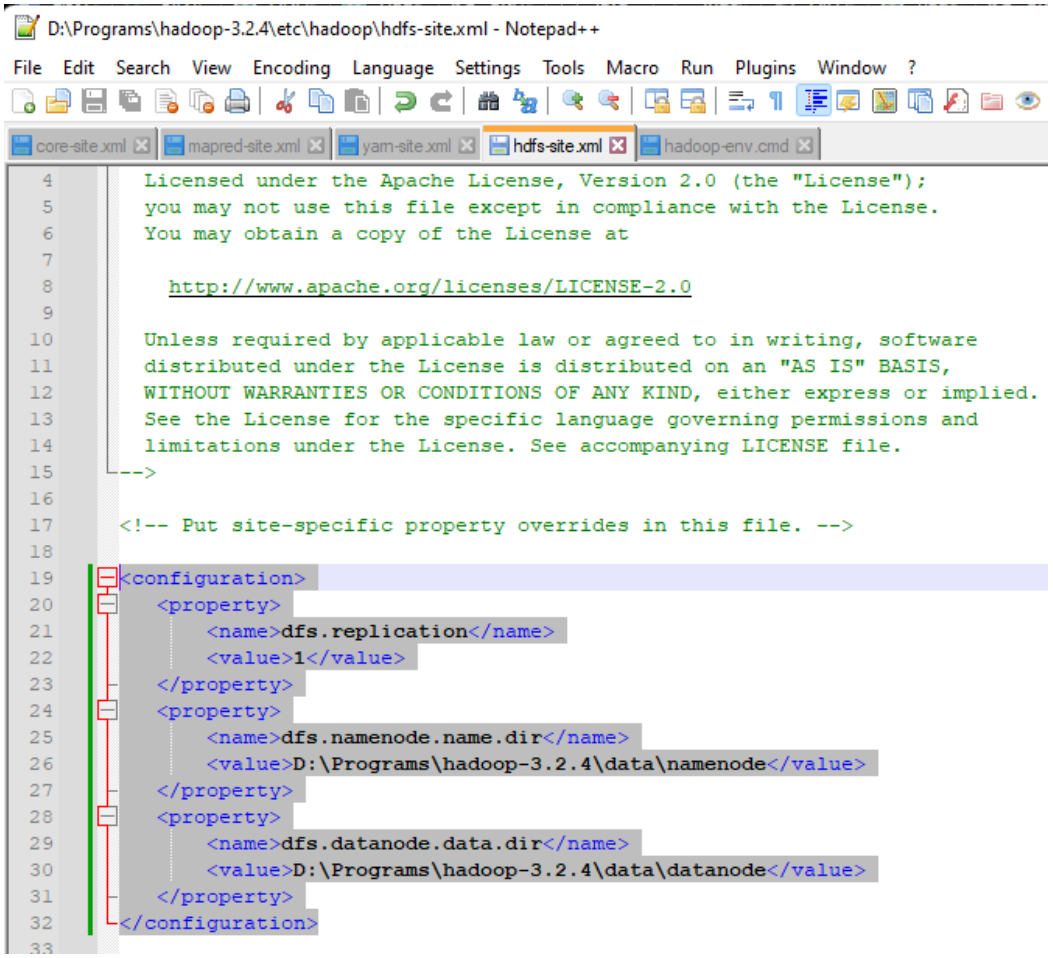
8. Pada direktori hadoop buatlah folder baru dengan nama data, didalam folder tersebut buat 2 folder baru dengan nama datanode dan namenode.



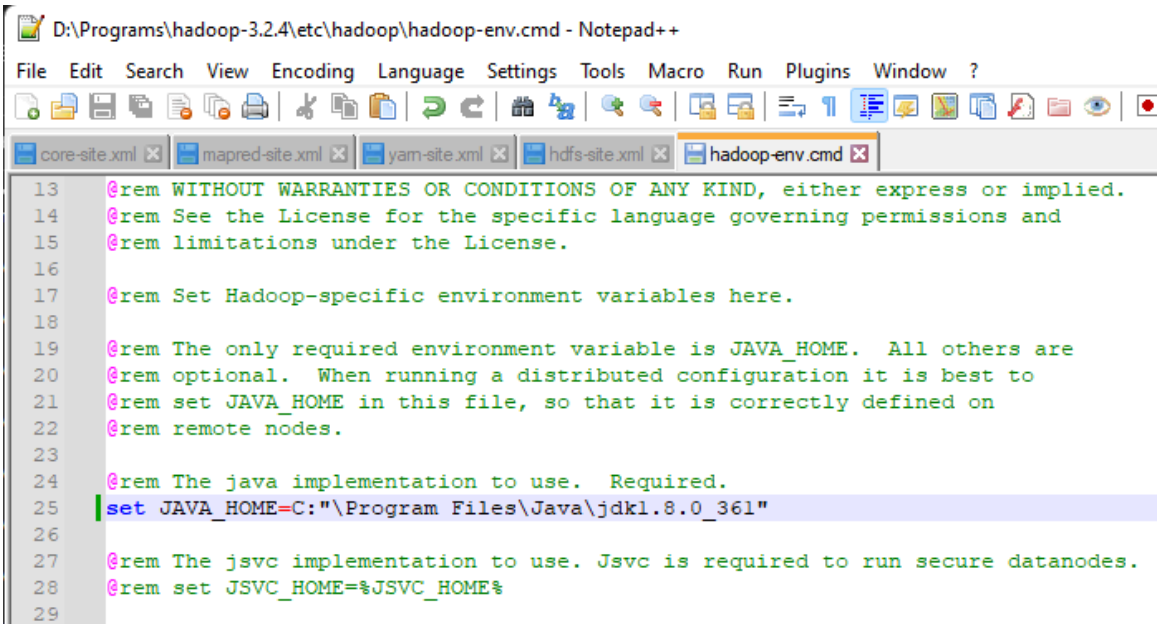
9. Tambahkan code berikut pada file **hdfs-site.xml**. Untuk **value** disesuaikan dengan **direktori** dimana folder **namenode** dan **datanode** dibuat. Untuk instalasi di tutorial ini, value nya seperti ini:

```
<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>
```

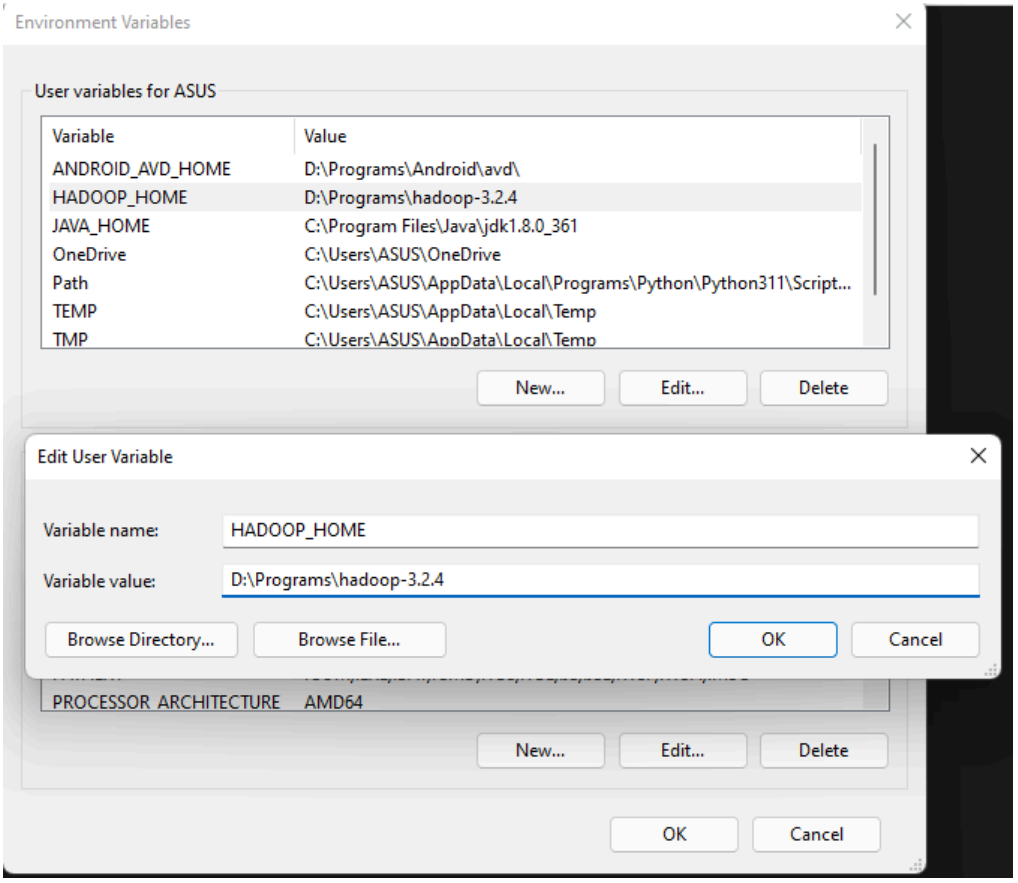


10. Pada file **hadoop-env.cmd**, sesuaikan direktori **JAVA_HOME** dengan **direktori java jdk**. Penulisan direktori JAVA_HOME sebenarnya tidak boleh terdapat folder yang mengandung spasi, namun apabila ingin tetap terdapat folder dengan spasi, maka ditambahkan **"...."** agar tidak terjadi error.

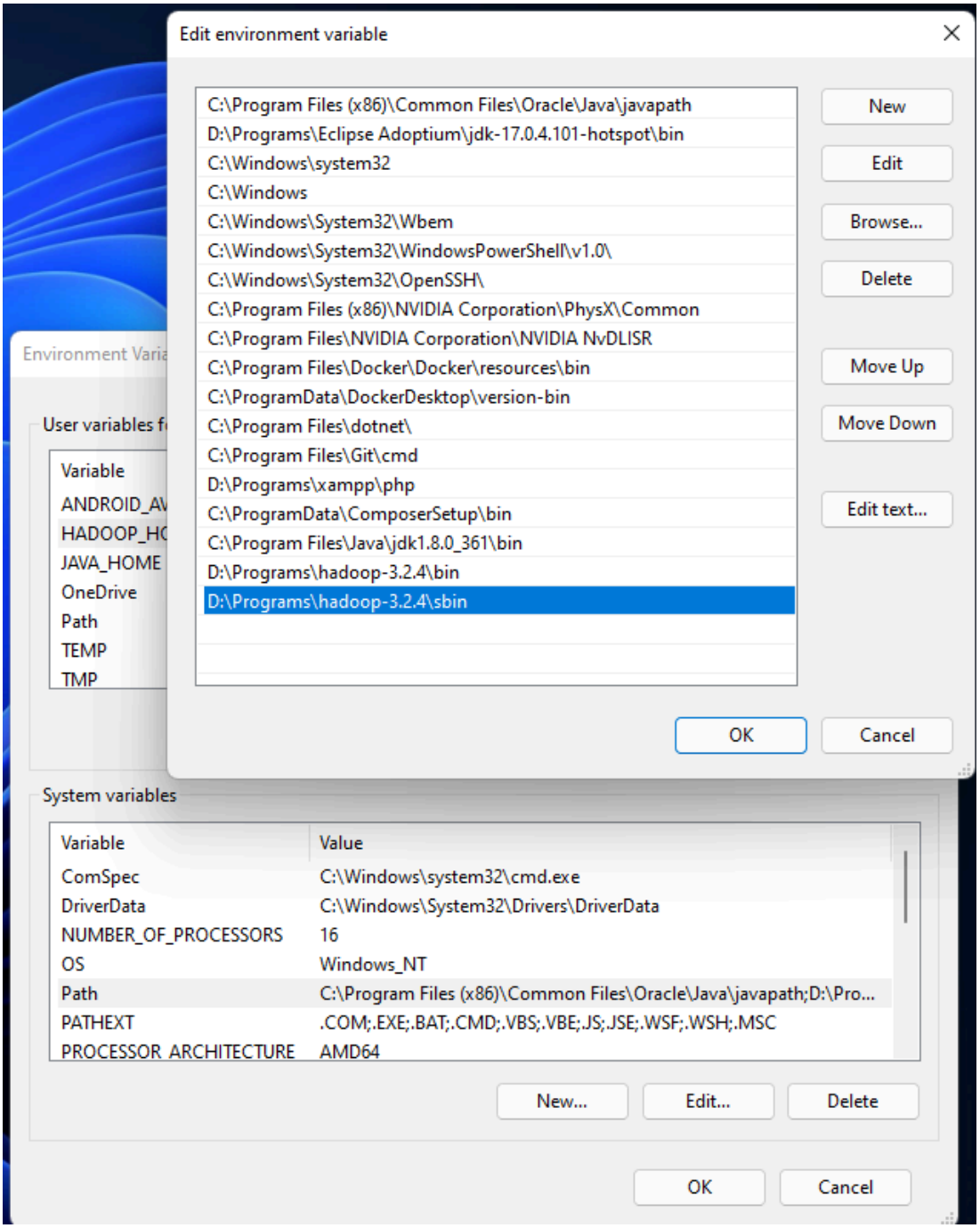


11. Setelah melakukan konfigurasi, sekarang Kita akan mengatur Environment Variables pada Hadoop. Buka (Control Panel → System and Security → System Advanced System Settings). Kemudian akan muncul dialog box System Properties, lalu klik Environment Variables.

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**.

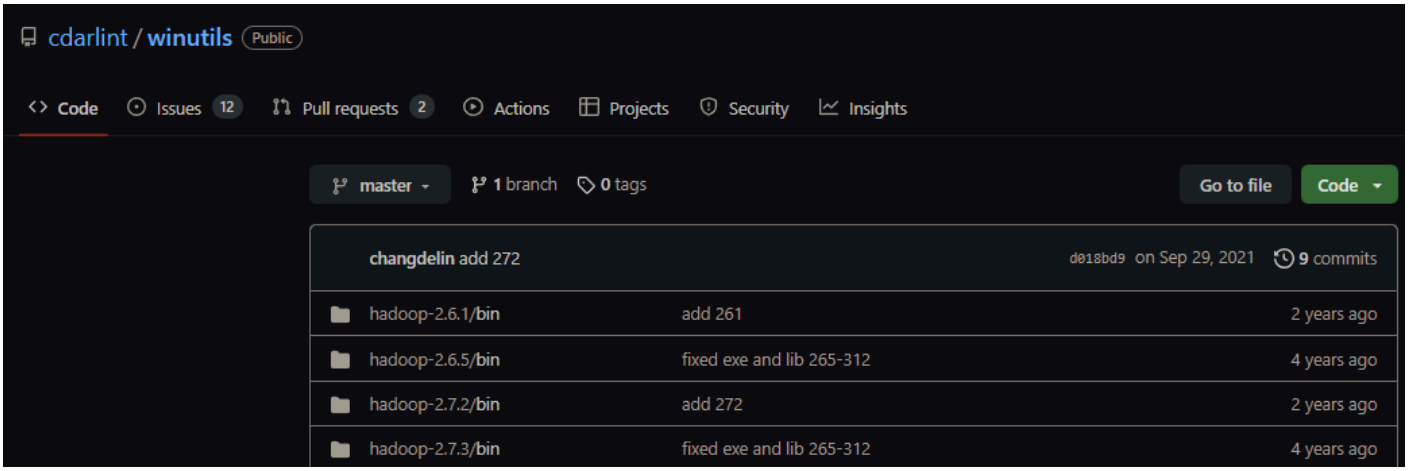


13. Atur Path Hadoop. Pada System variables, klik **Path**. Kemudian klik New dan isi dengan direktori **hadoop\bin** dan **hadoop\sbin**. Lalu **klik OK** pada Edit environment variable, Environment Variables dan System Properties.

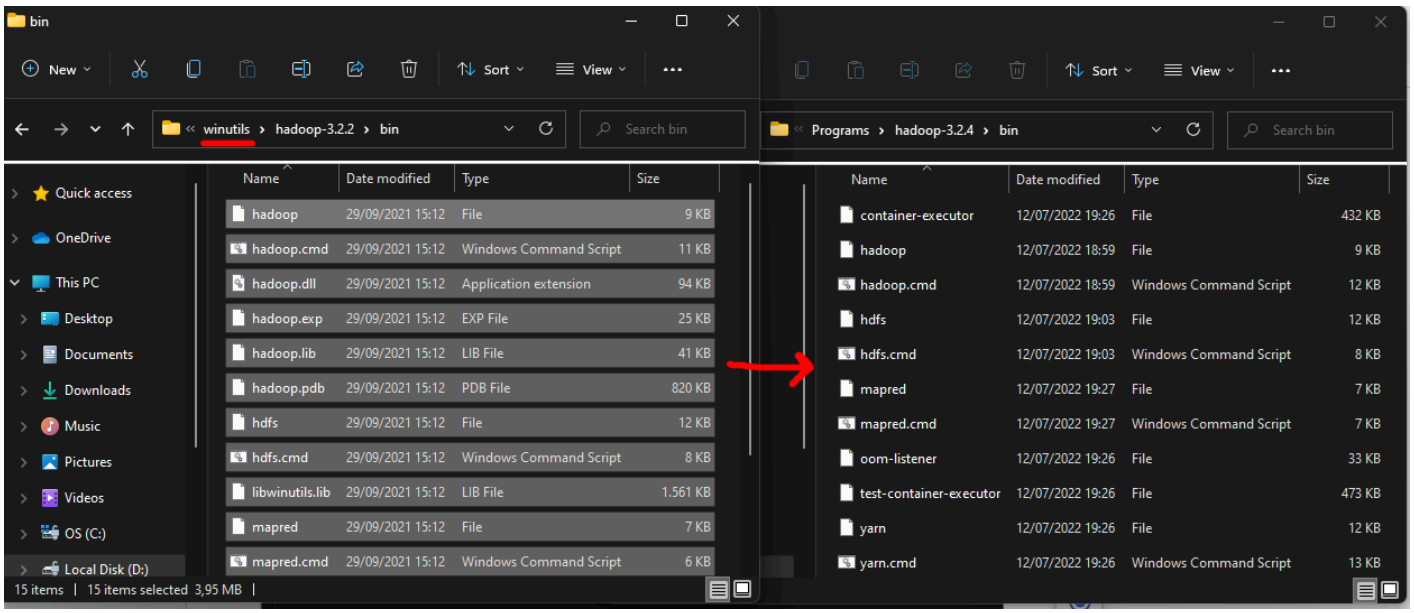


14. Agar Hadoop dapat dioperasikan pada OS Windows, diperlukan patch file **winutils**. File tersebut dapat di download melalui link berikut:
    - https://github.com/cdarlint/winutils

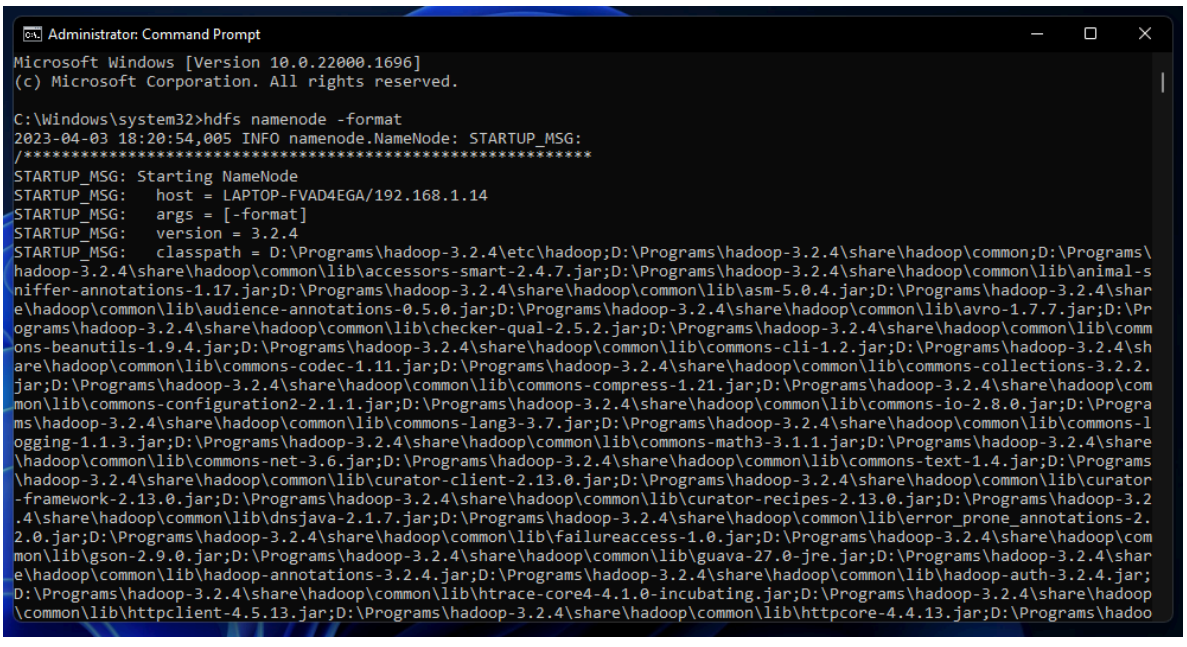
    Untuk sekarang, kita akan memilih folder yang versi paling dekat dengan hadoop-3.2.4, yaitu **3.2.2**



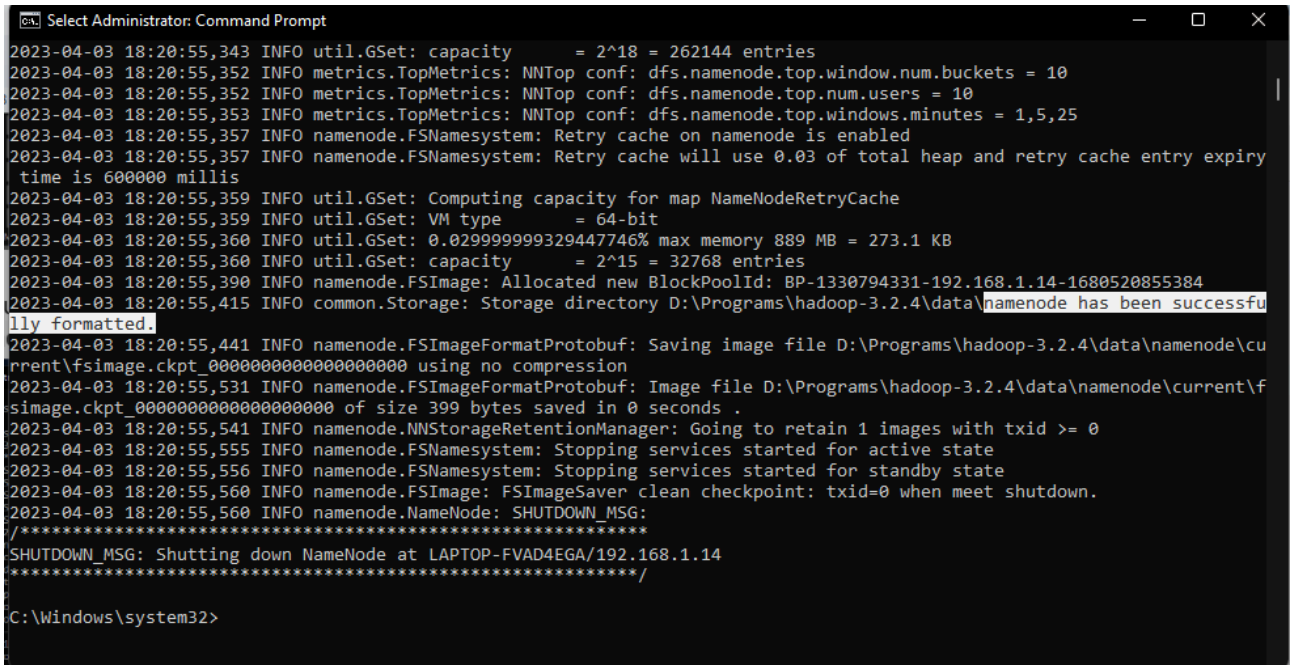
15. Ekstrak file tersebut, kemudian copy isi folder bin yang di download ke dalam folder (replace jika perlu)



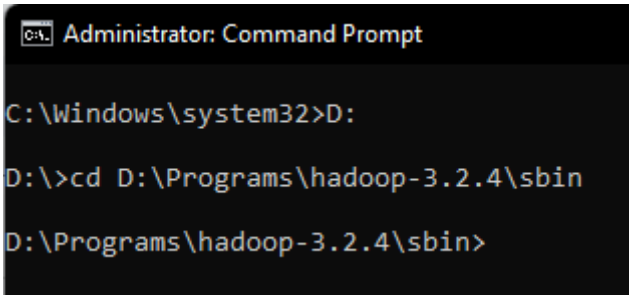
16. Buka CMD sebagai administrator. Kemudian ketikkan perintah **hdfs namenode –format** (jika ada error “hdfs is not recognized”, pastikan semua Environment Variables sudah sesuai dan sudah di **klik ok**)



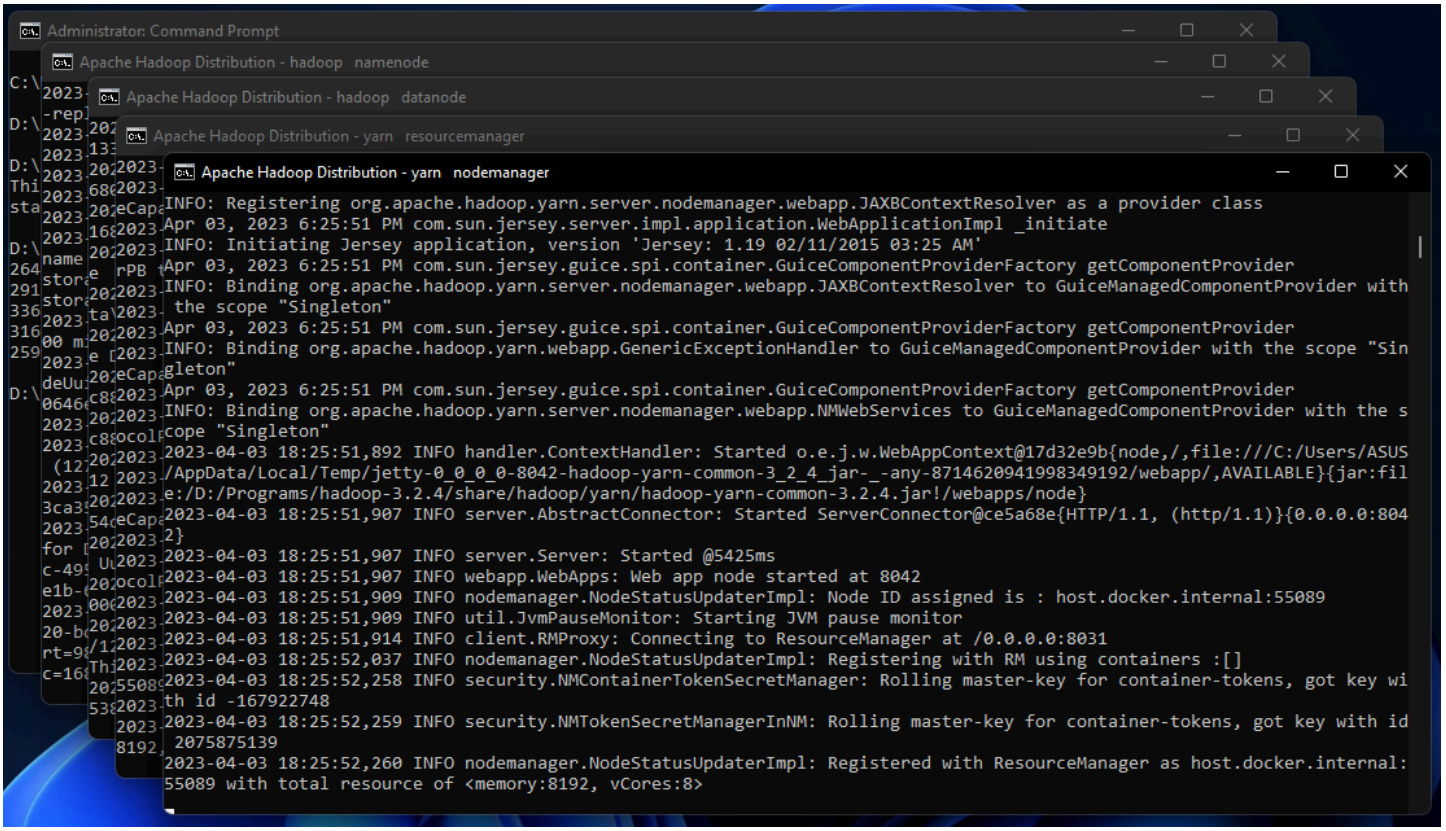
Kalau berhasil, akan ada “**namenode has been successfully formatted**”



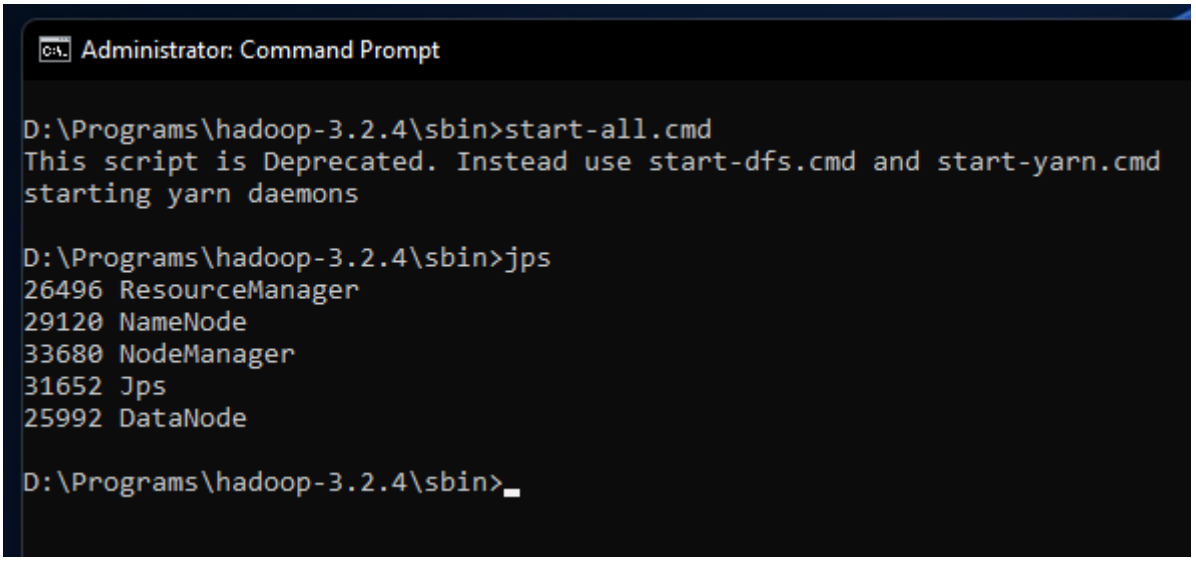
17. Gunakan command **cd** untuk masuk ke folder **hadoop\sbin**



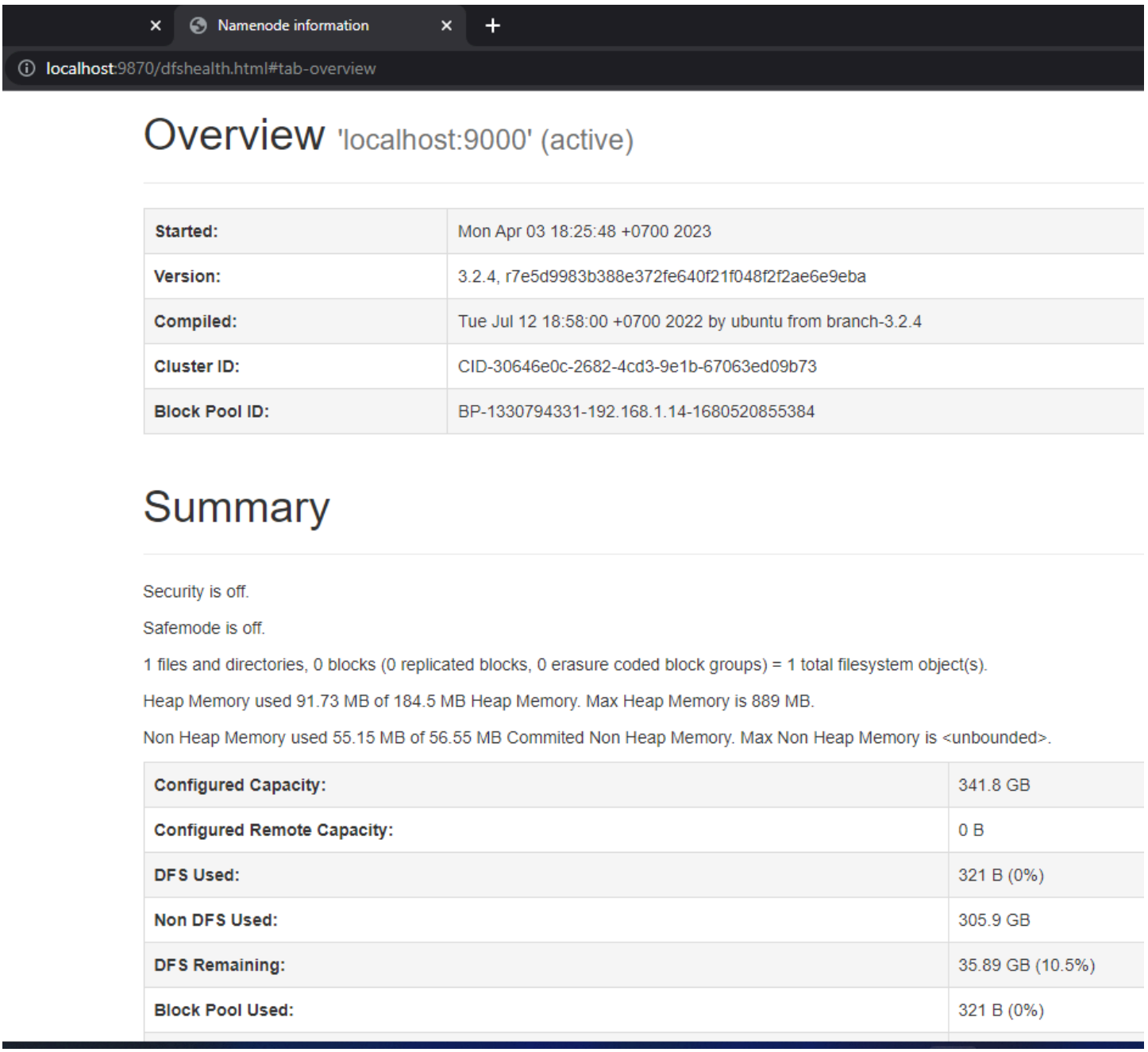
18. Ketikkan command **start-all.cmd**. Jika berhasil, 4 window baru akan muncul (**namenode, datanode, resourcemanager,** dan **nodemanager**)



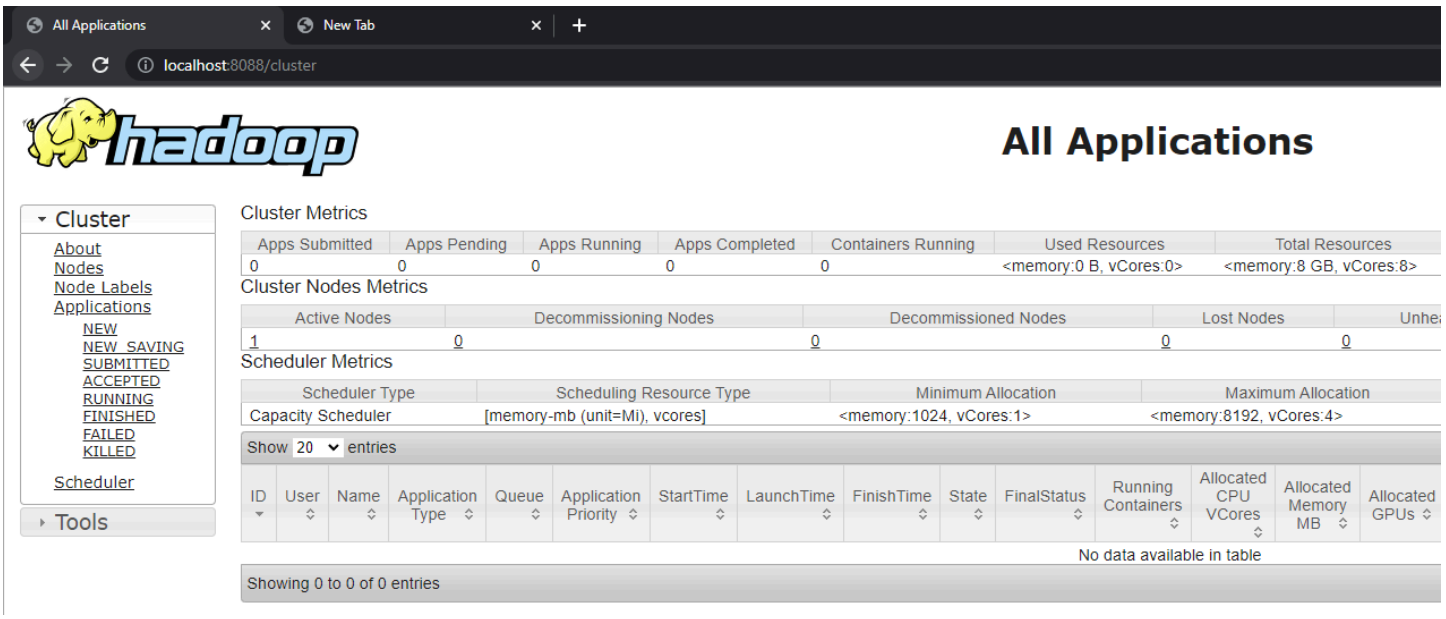
19. Gunakan command **jps** untuk melihat apakah Hadoop berjalan dengan benar (seharusnya 5 ini berjalan)



20. Coba buka
    a. http://localhost:9870



b. http://localhost:8088



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