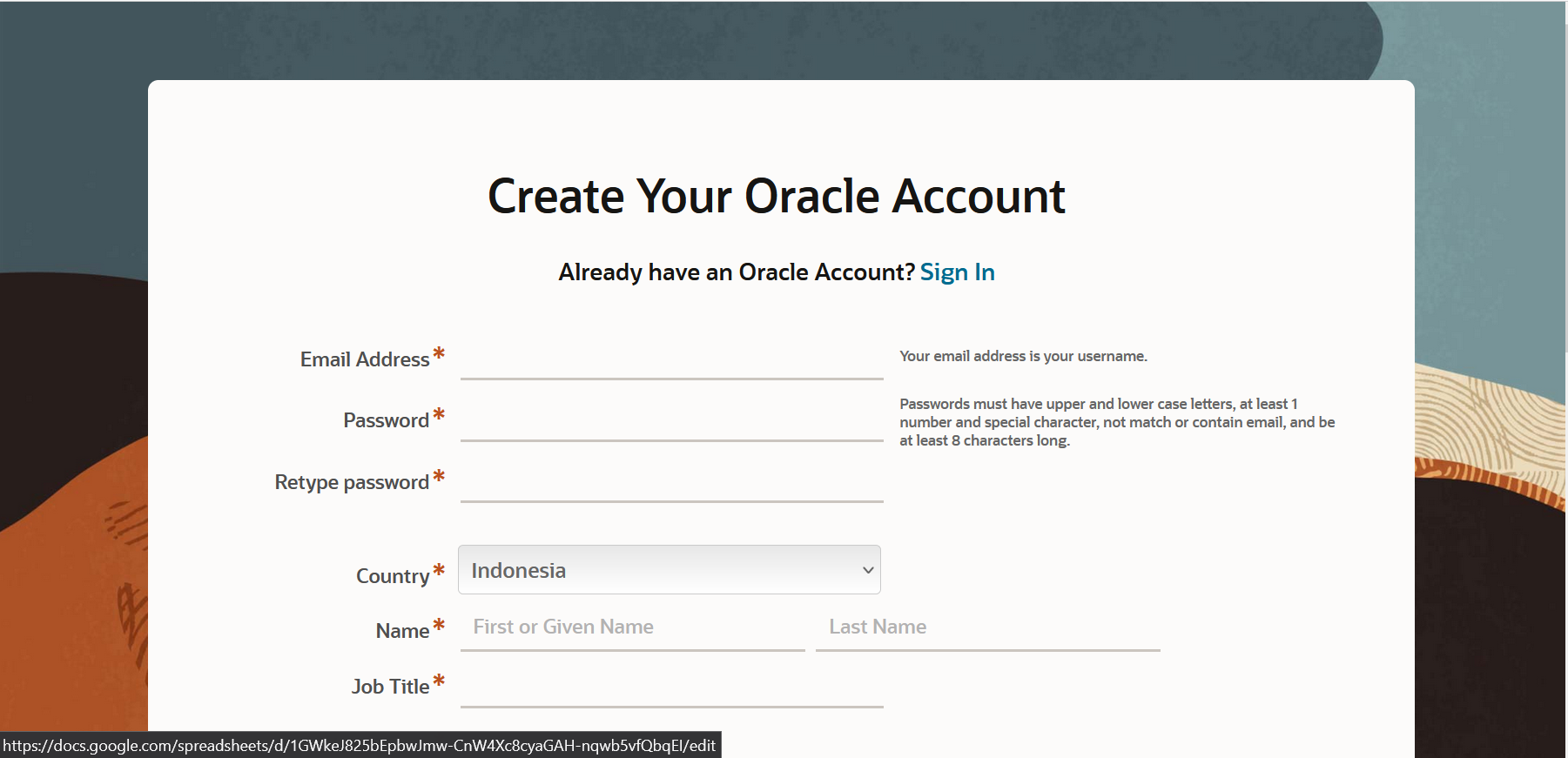
Pada modul praktikum ini, kita akan mempelajari bagaimana cara menginstalasi peralatan untuk big data dimulai dengan hadoop. Singkatnya hadoop merupakan alat untuk memproses data berukuran besar atau big data menjadi beberapa bagian atau disebut cluster untuk referensinya kalian bisa baca di sini.

Sebelum memulai praktikum perlu diperhatikan bahwa modul ini akan menggunakan sistem operasi berbasis Linux khususnya **Ubuntu 20.04 LTS (Menggunakan Virtual Box)** agar dapat stabil menggunakan tools big data nantinya. Apabila menggunakan sistem operasi lain diperbolehkan dan dapat disesuaikan dengan modul ini karena kemungkinan pengaturannya atau menjalankan toolsnya akan berbeda.

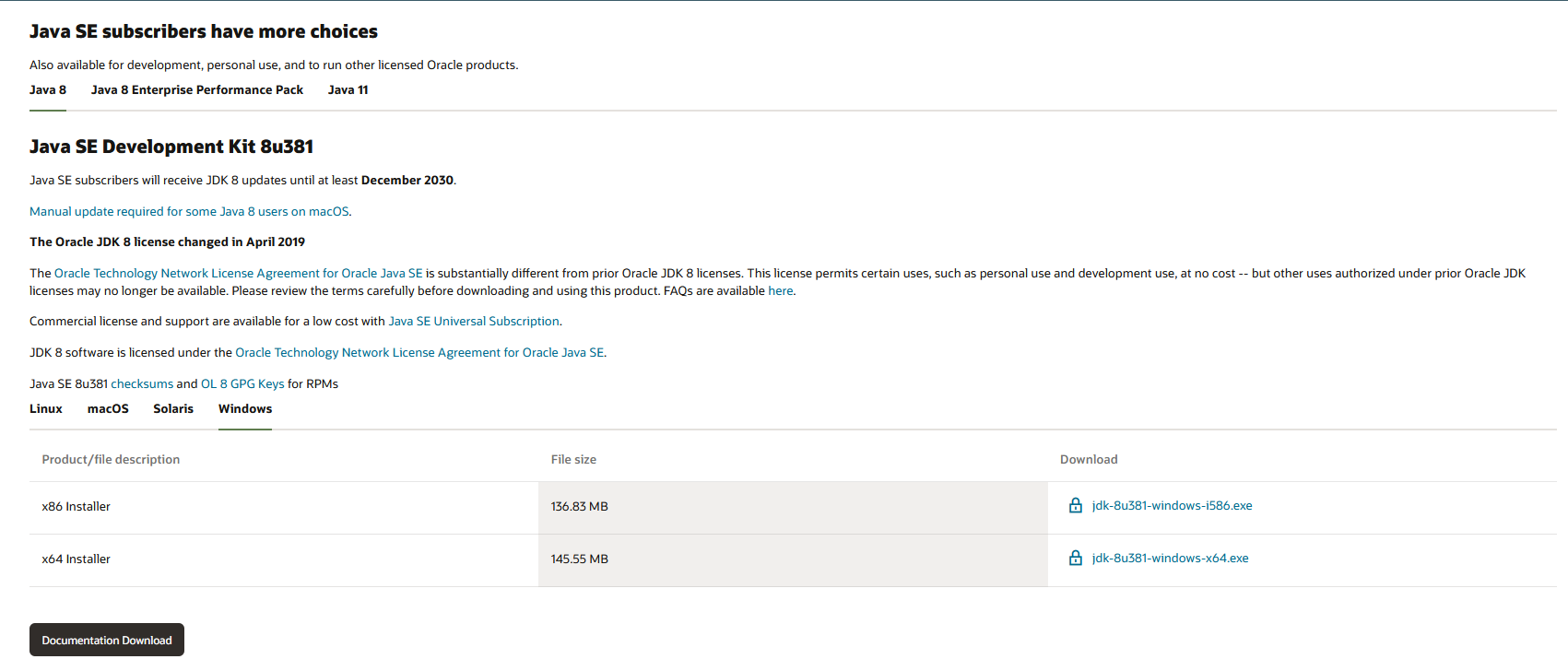
Module kali ini kita akan coba menggunakan system operasi windows untuk instalasi Apache Hadoop 3.2, Apache Hadoop 3.2 hanya di support oleh Java 8 tapi untuk Apache Hadoop 3.3 di support oleh Java 8 dan Java 11 (hanya untuk runtime) compile masih menggunakan java 8

Langsung saja kita coba bagaimana cara menginstalasi hadoop di **Windows 10**, berikut langkah-langkahnya :

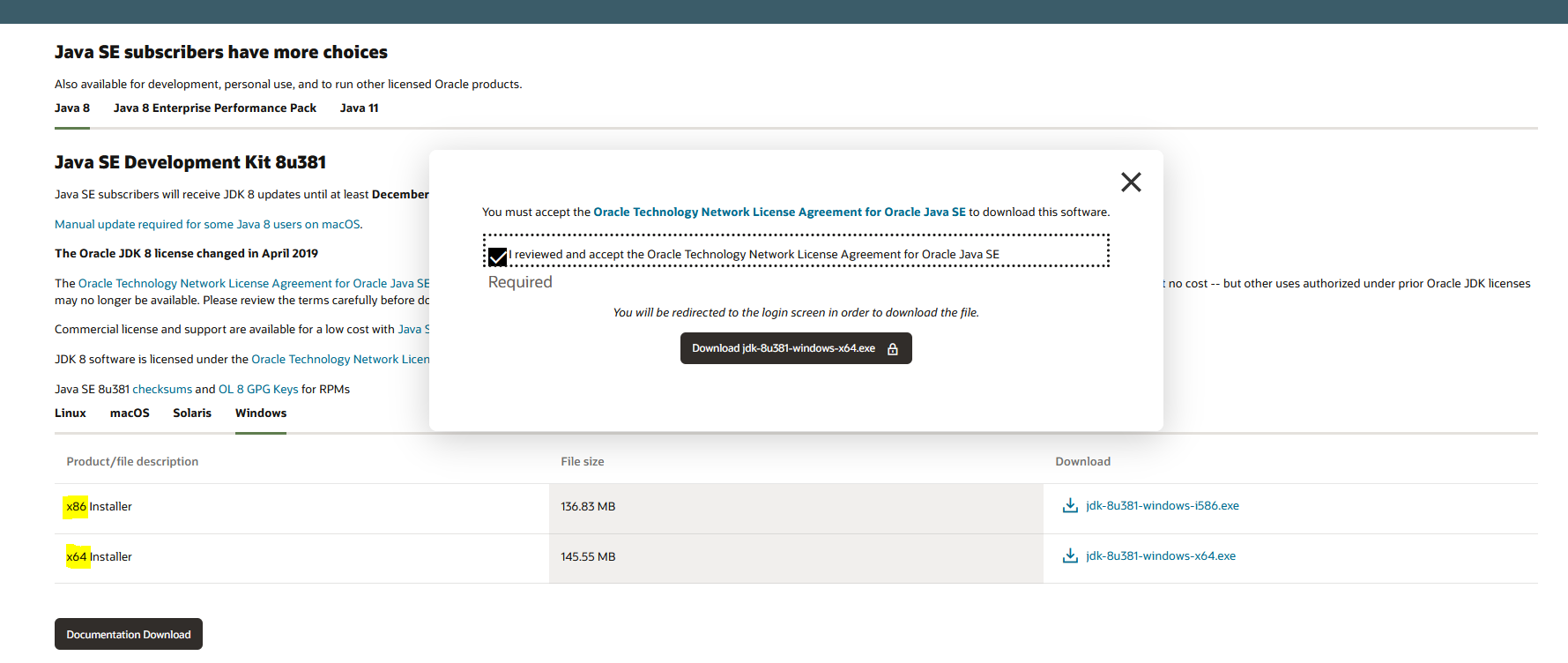
* 1. **Instalasi dan Konfigurasi Java**
     1. Pertama kita harus membuat akun oracle terlebih dahulu di link berikut <https://profile.oracle.com/myprofile/account/create-account.jspx>



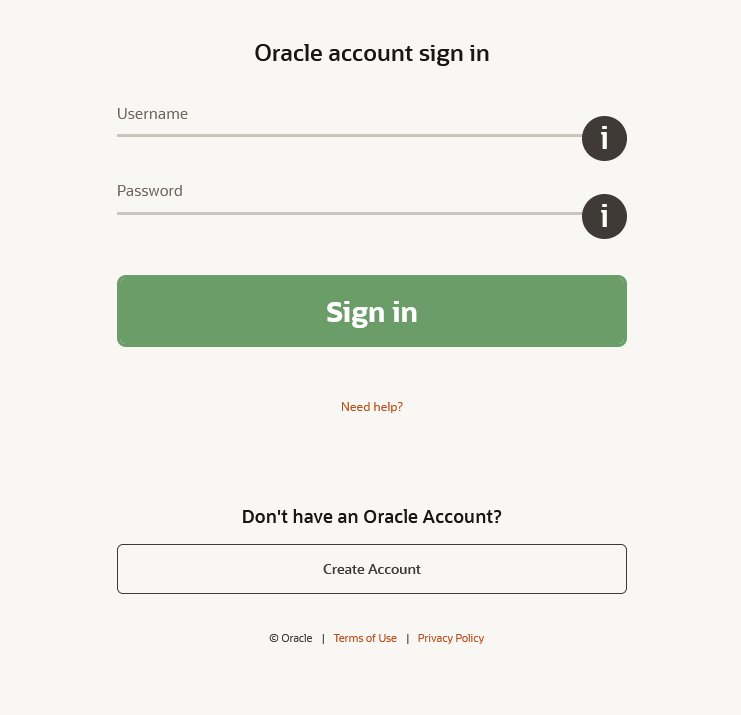
* + 1. download .



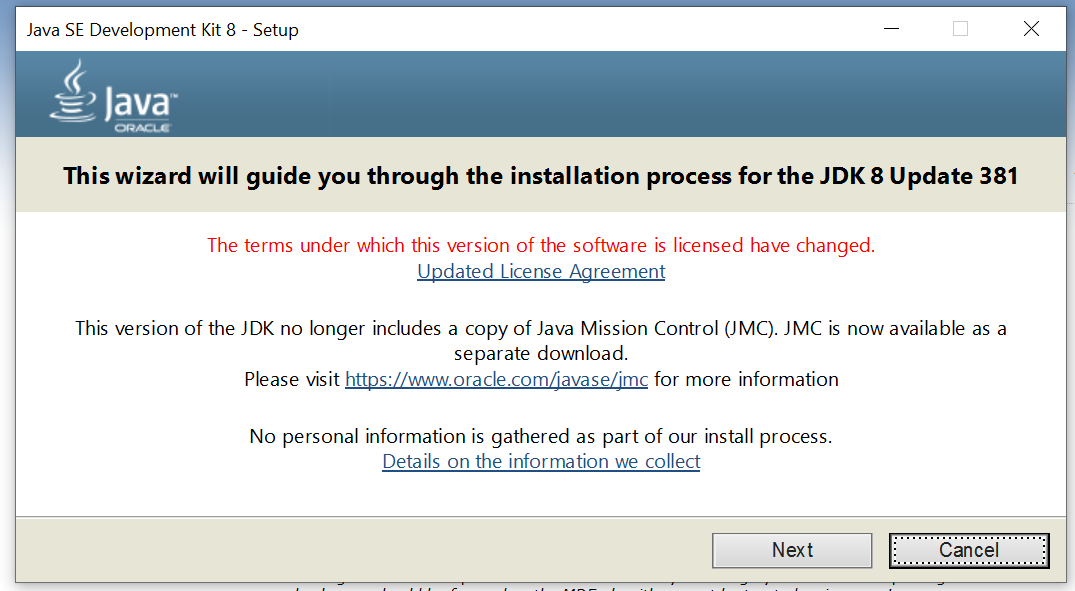
* + 1. Pemilihan jenis unduhan berdasarkan bit windows.



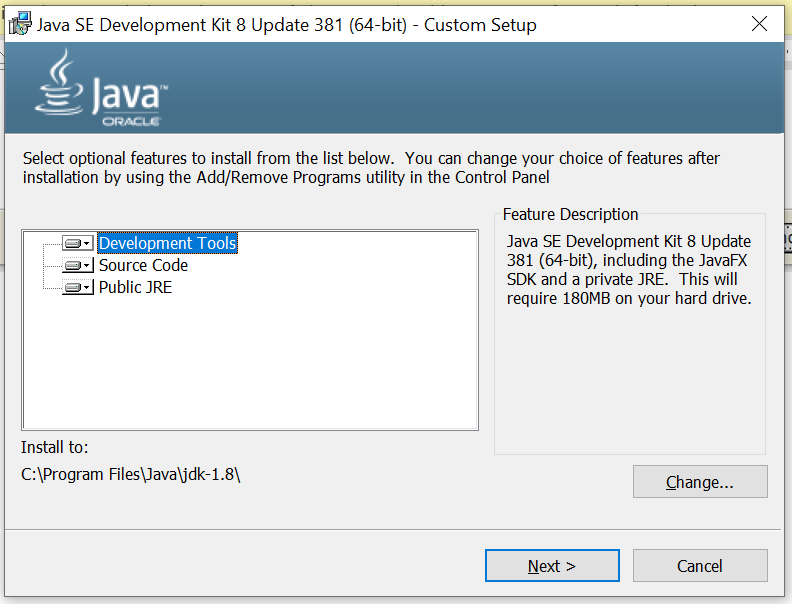
4. lalu login menggunakan akun yang sudah di buat



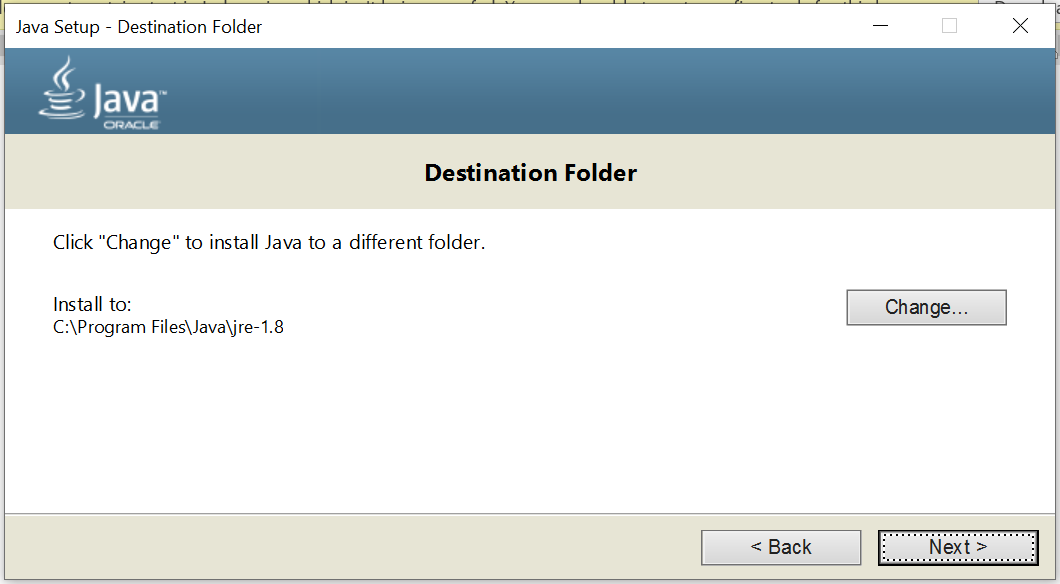
* + 1. Membua Instalasi Java – klick next



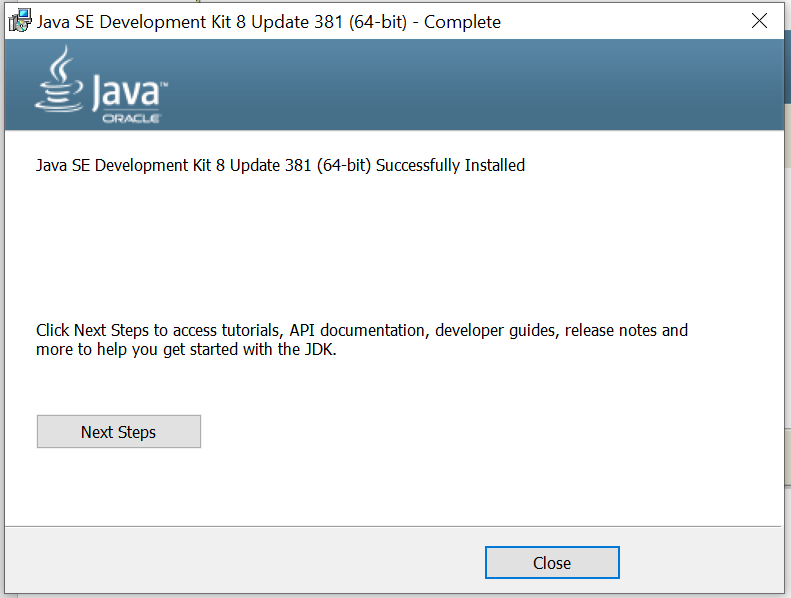
* + 1. Kita akan menginstal Java SE Development Kit 8 kita juga dapat merubah lokasi penginstallan.



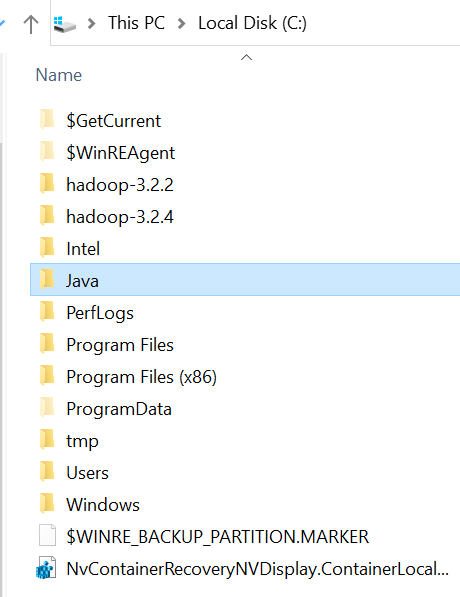
* + 1. Selanjutnya melakukan install java jre



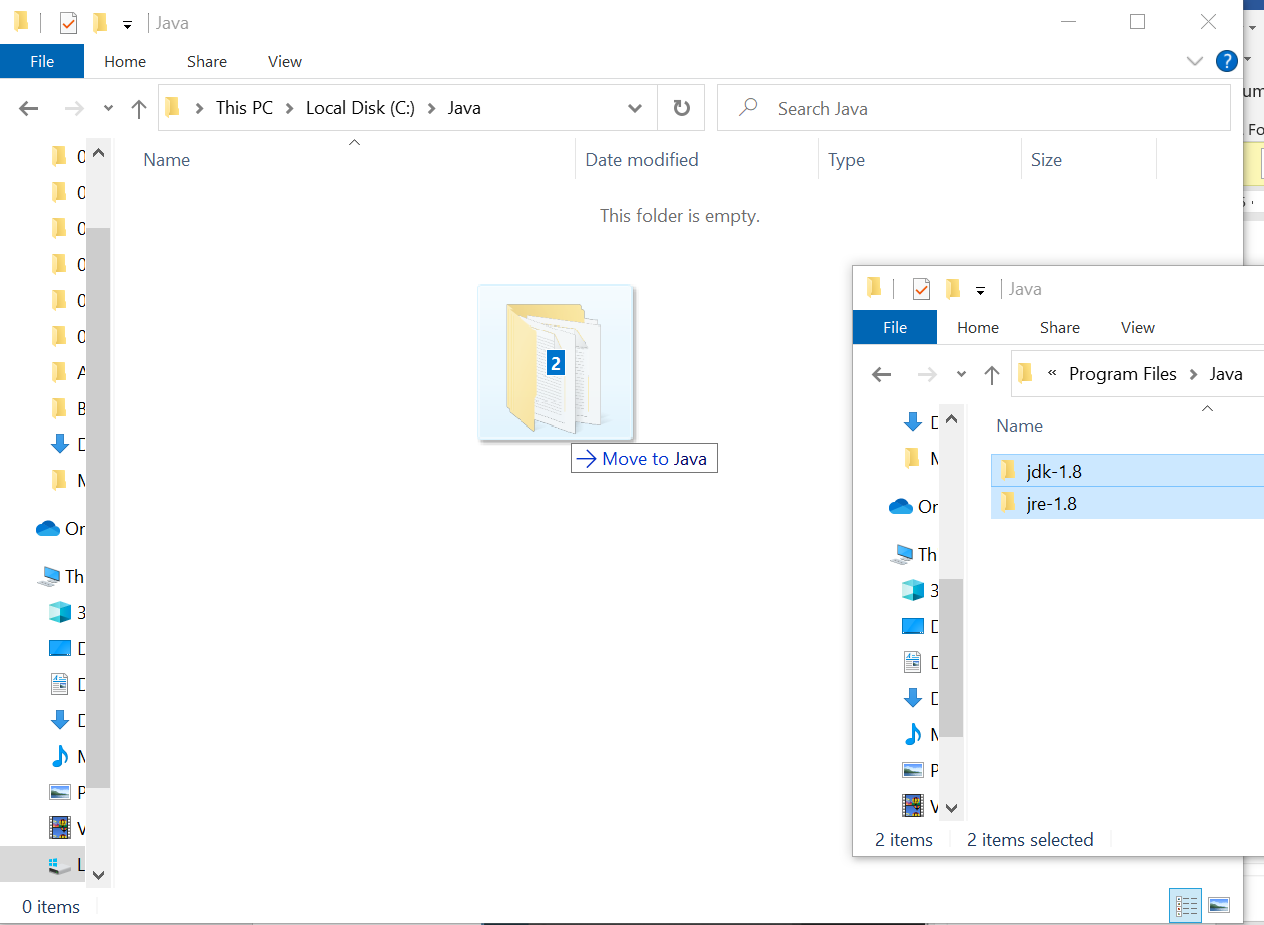
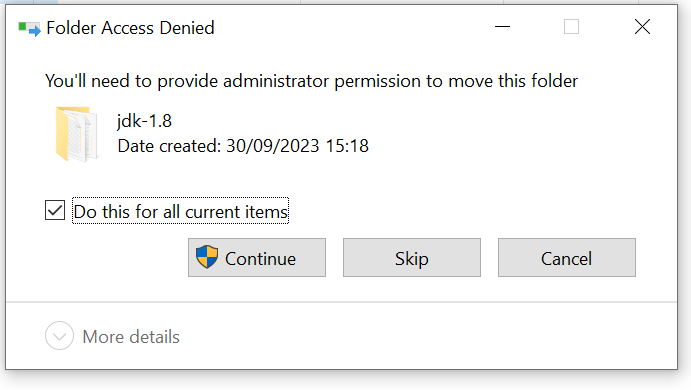
* + 1. Selanjutnya close untuk menutup instalasi java



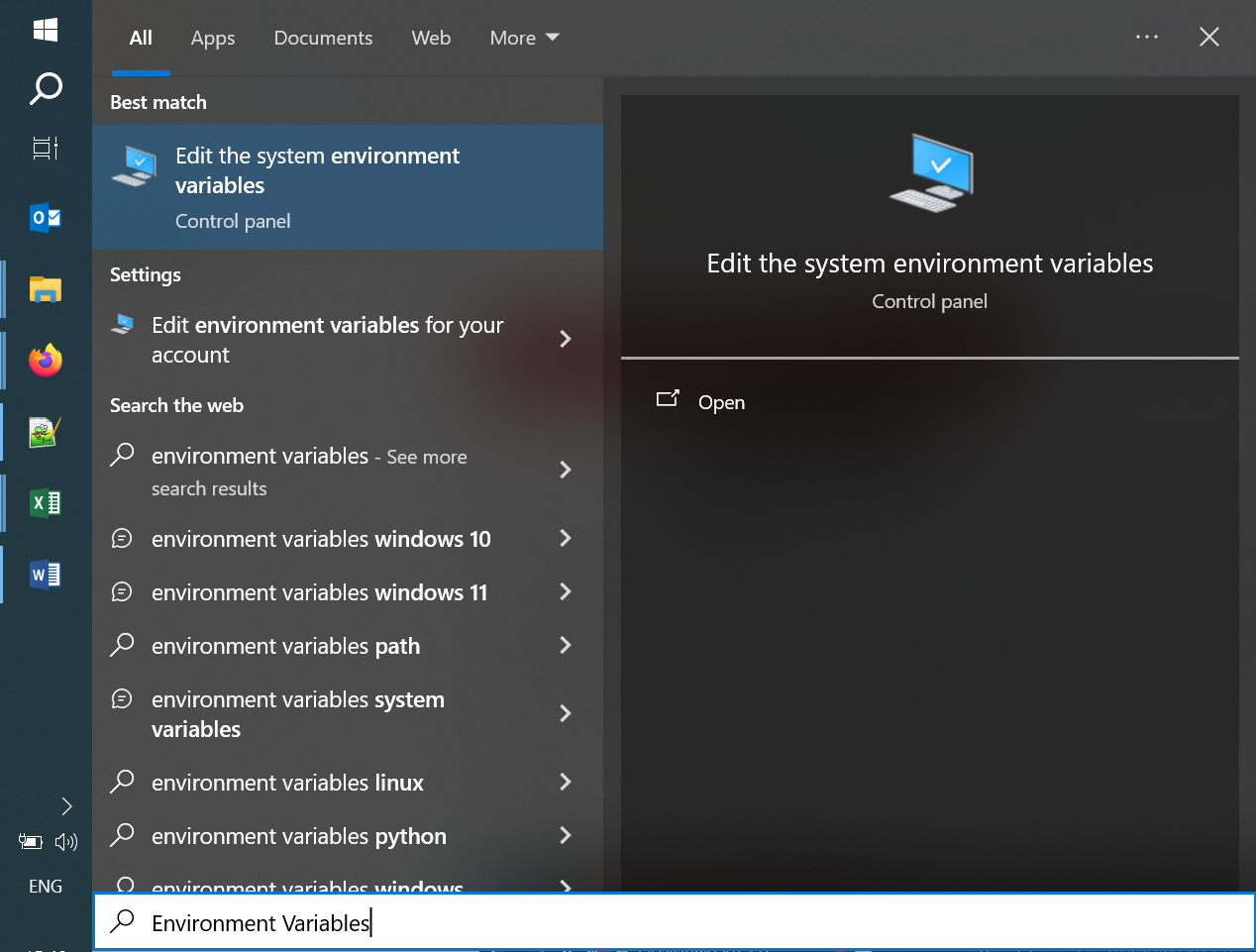
* + 1. Selanjutnya kita akan melakukan pemindahan folder Java karena nantinya akan menyebabkan issue environment karena permission. Pertama buat folder dengan nama “Java” terlebih dahulu di Drive C.



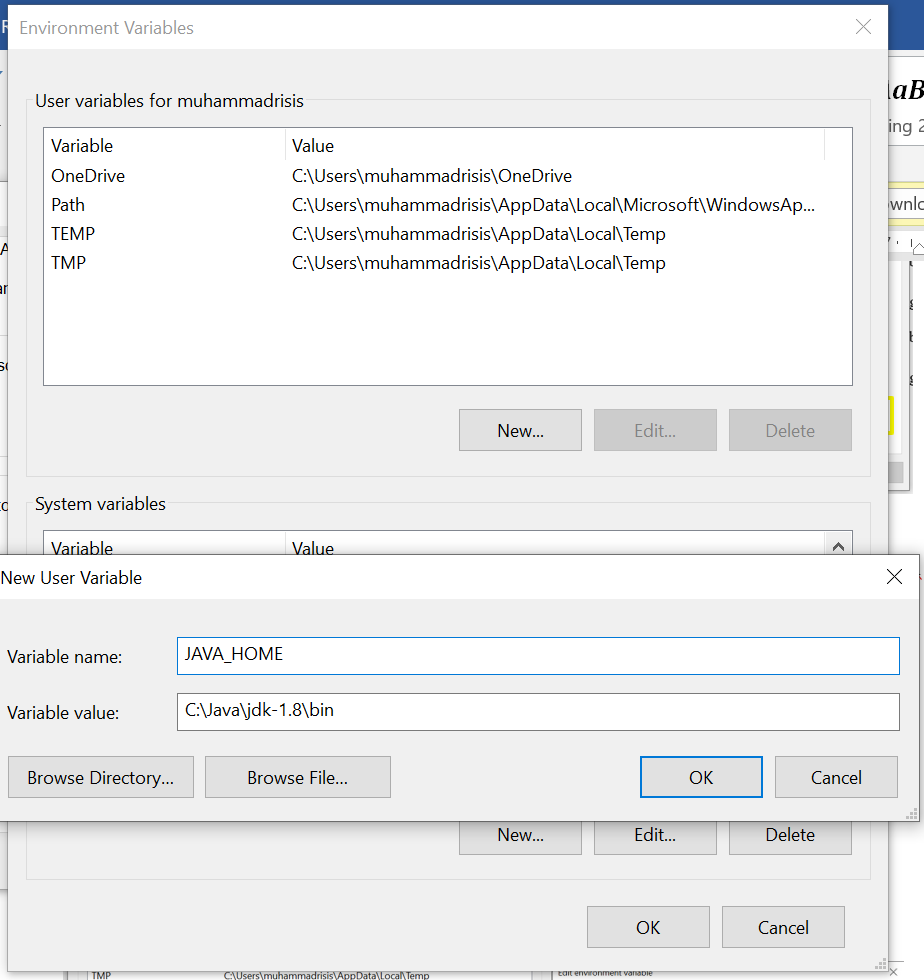
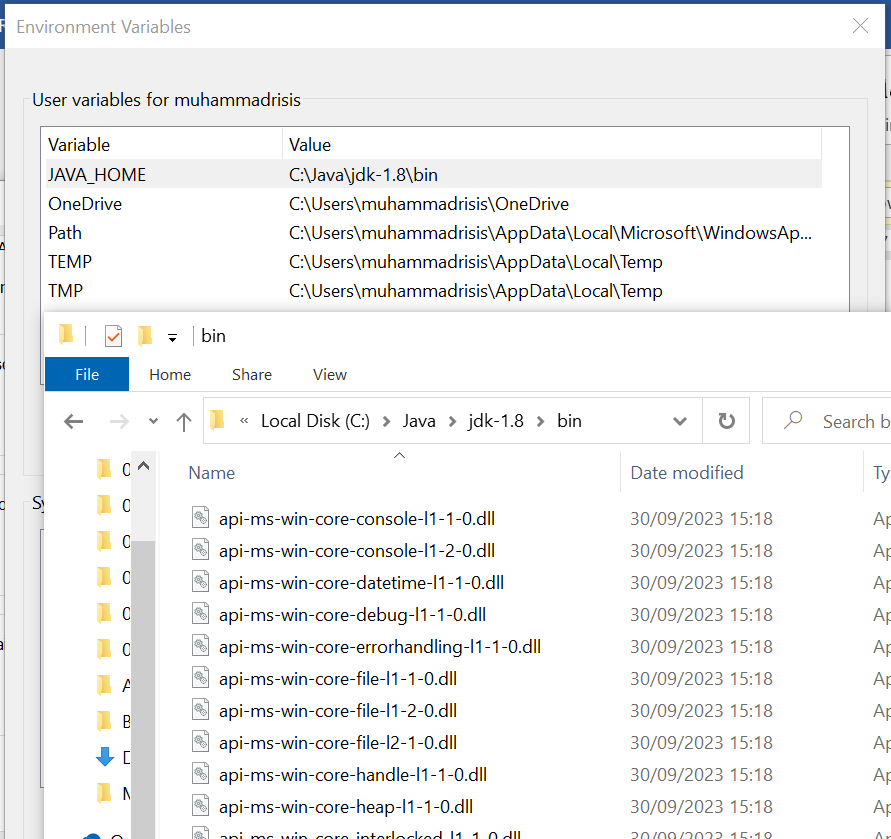
* + 1. Lalu kita pindahkan file jdk dan jre yang berada di “C:\Program Files\Java” ke “C \Java”. WAJIB untuk mencentang (*Do this for all current items*) supaya file java dapat berpindah seluruhnya.

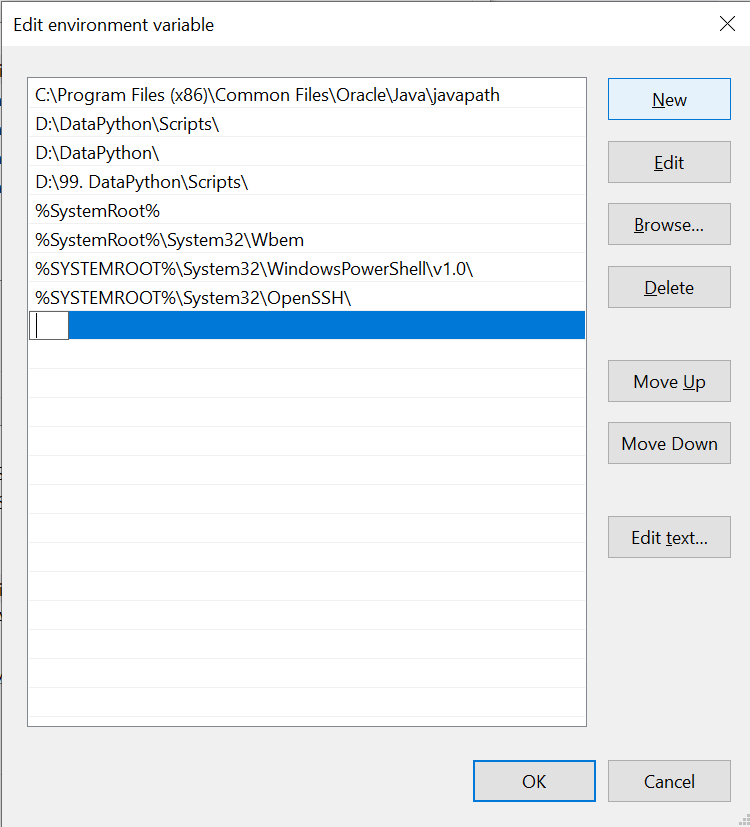
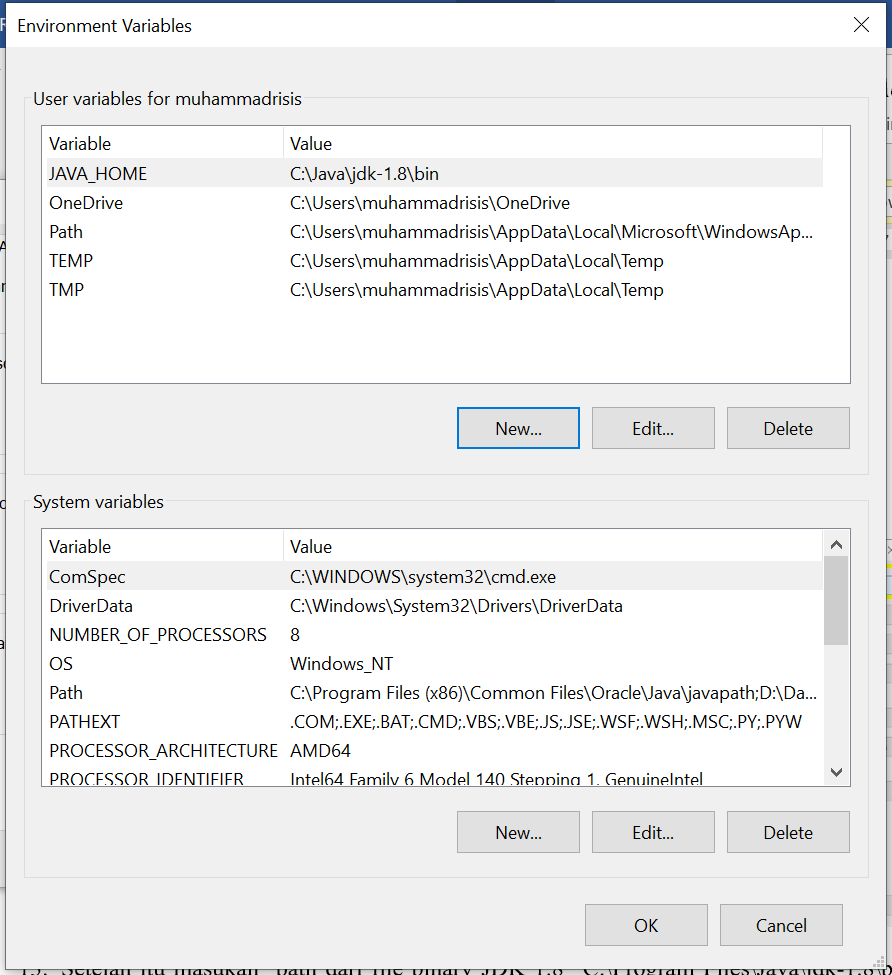
* + 1. Setelah itu kita akan membuat Environment Variables untuk java dengan cara klik icon *search* -> ketik “Environment Variables“ -> pada system properties klik Environment Variables.



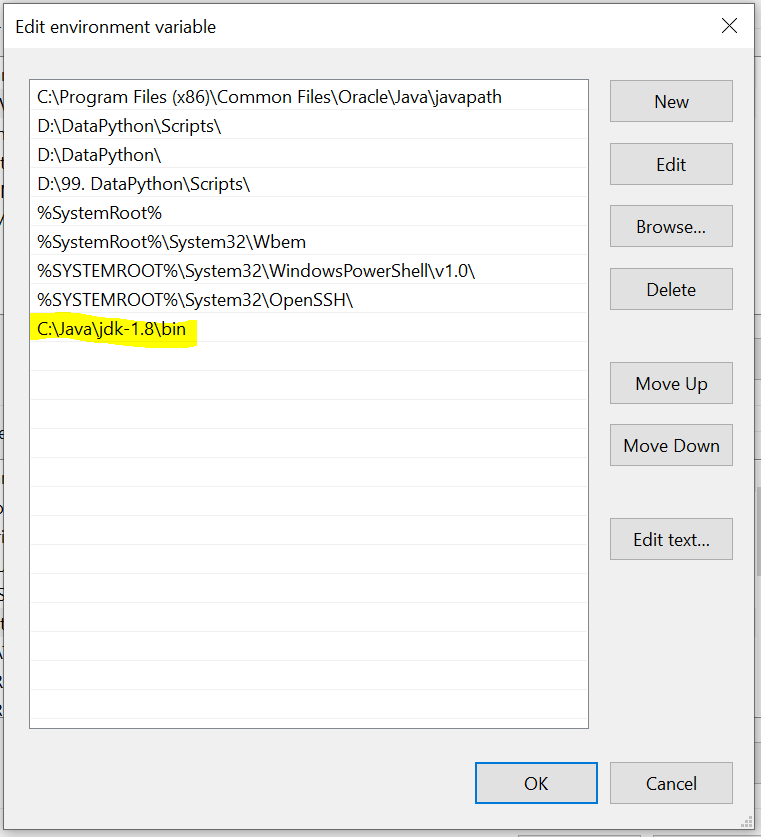
* + 1. Klik new lalu buat Variable baru JAVA\_HOME, dan pada variable value kita masukan *path* folder bin JDK-1.8

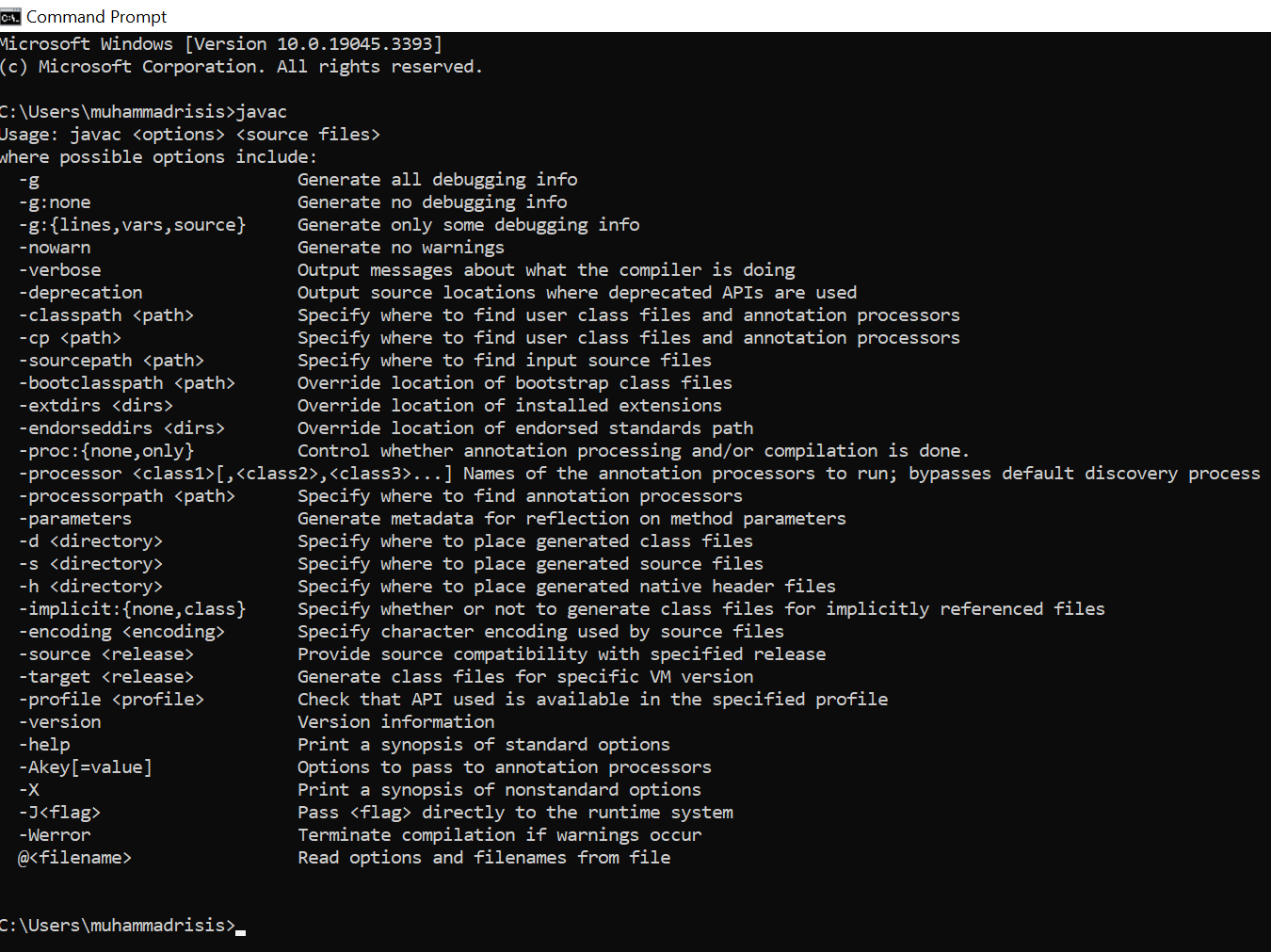
* + 1. Setelah itu mengupdate path system variables



* + 1. Setelah itu masukan path dari file binary JDK 1.8 “C:\Program Files\Java\jdk-1.8\bin” lalu klik oke

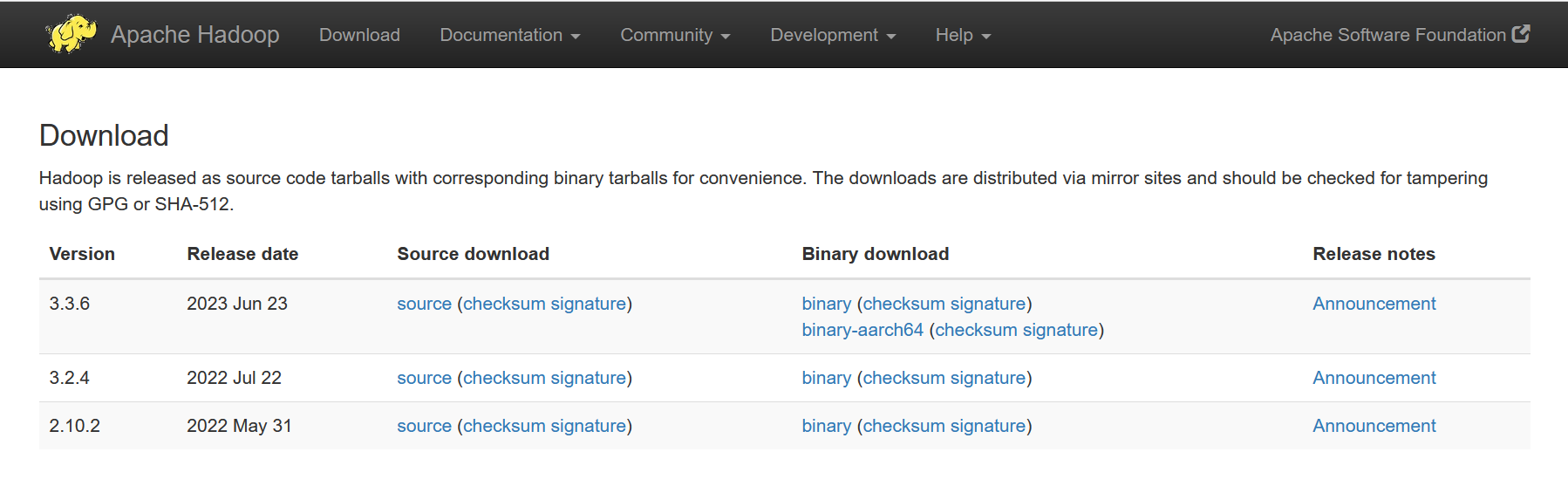


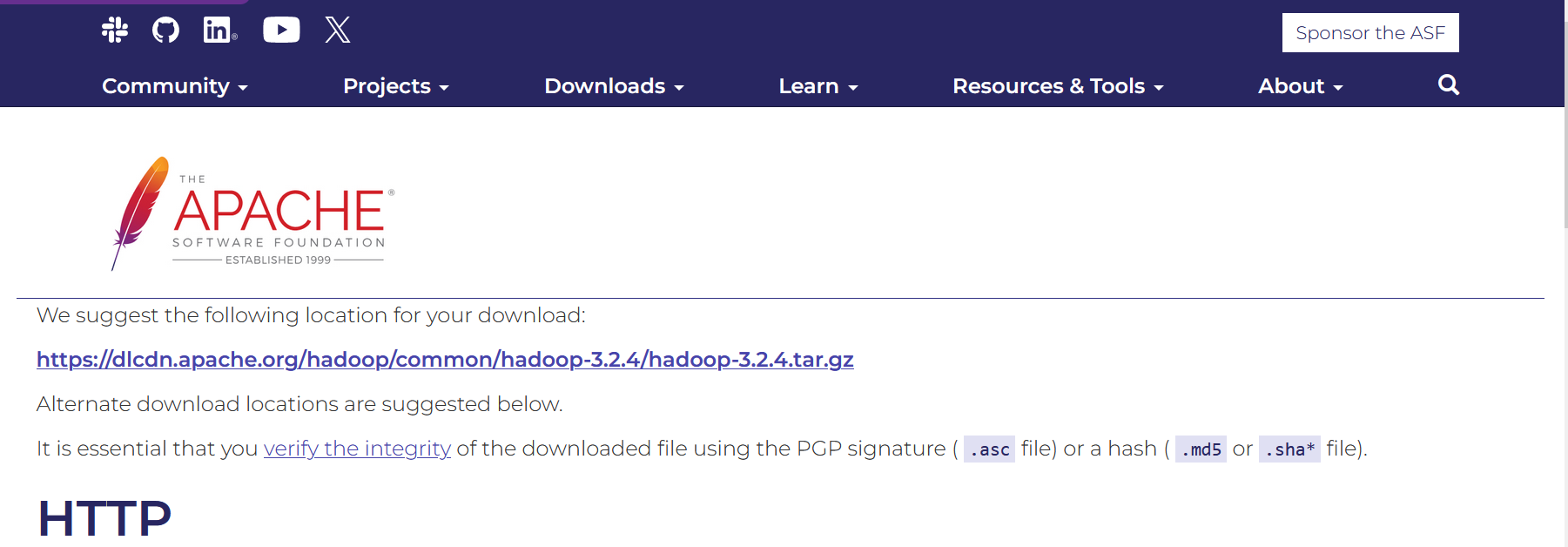
* + 1. Setelah itu validasi instalasi java kita dengan cara membuka CMD dan ketikan “javac” jika sudah benar maka akan tampil seperti di bawah



**B. Instalasi Hadoop**

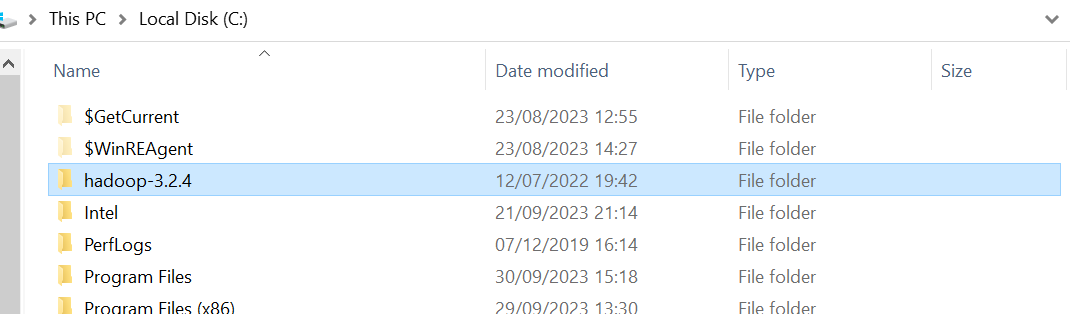
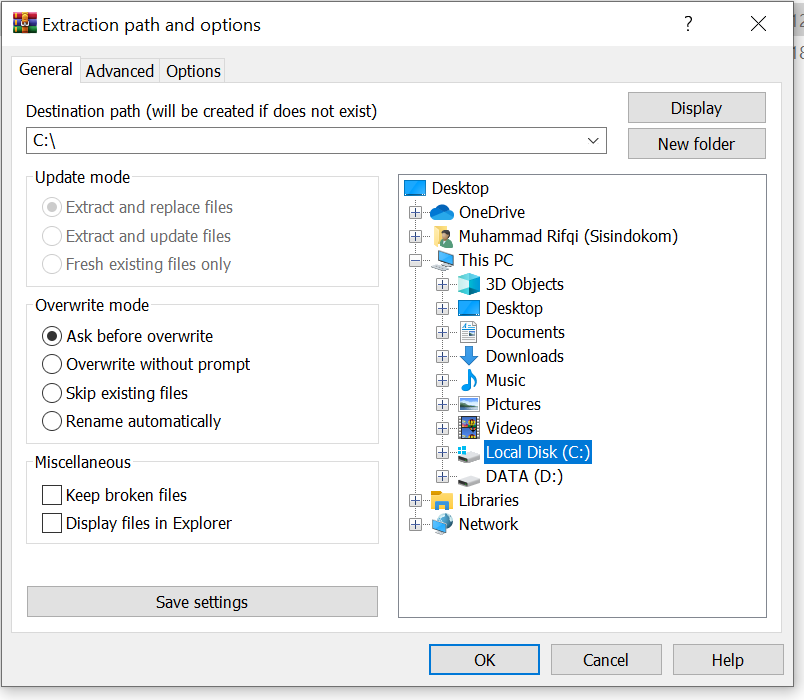
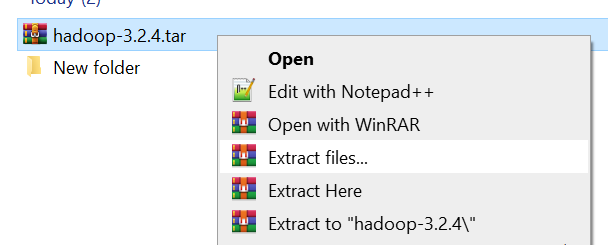
1. Tahap selanjutnya download package hadoop berikut. Di halaman resmi Hadoop <https://hadoop.apache.org/releases.html>, lalu download Hadoop version





1. Setelah Selesai Download kita akan melakukan extract file Hadoop yang telah di download di File C:

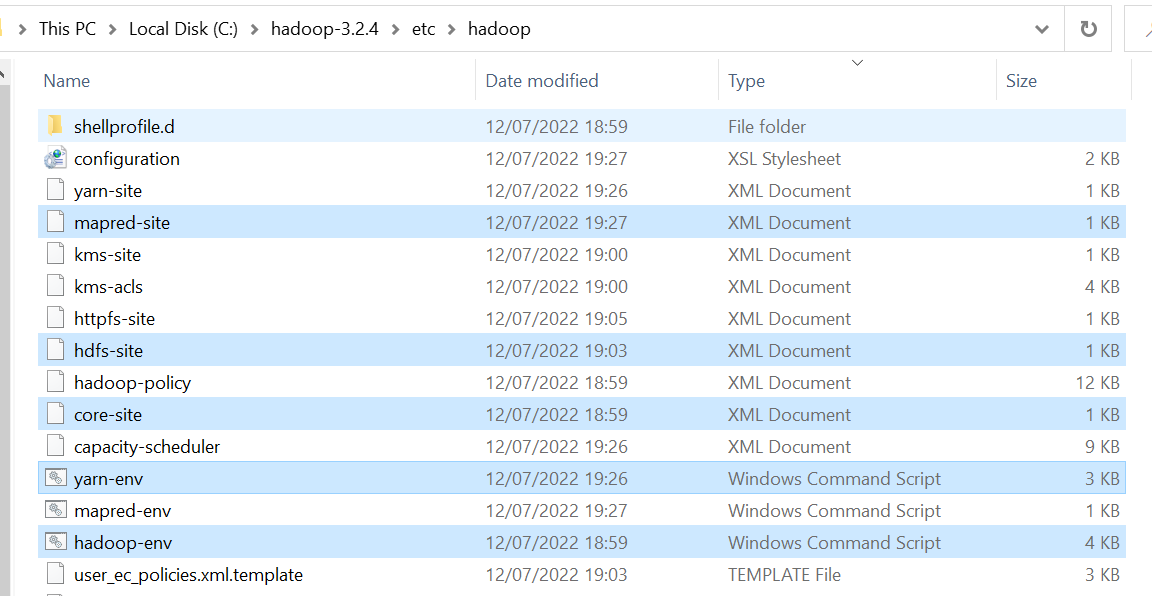
- jika ada permintaan untuk administrator maka klik yes



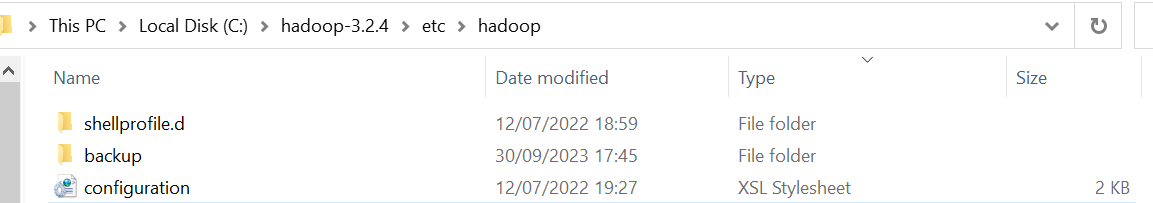
**C. Konfigurasi Hadoop**

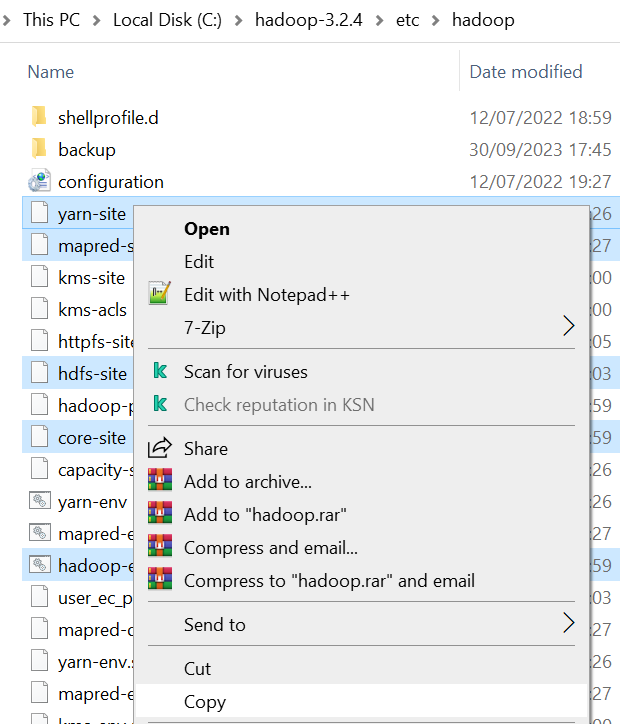
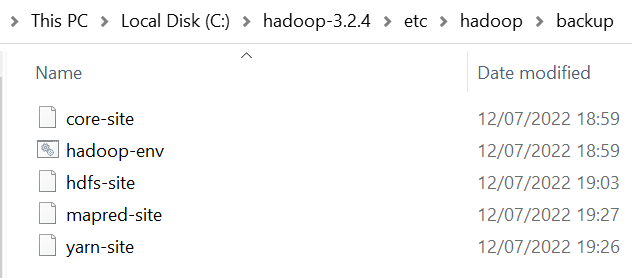
1. Setelah berhasil melakukan instalasi kita akan melakukan beberapa configurasi di “C:\hadoop-3.2.4\etc\hadoop” untuk dapat menjalankan Hadoop di. Untuk melakukan configurasi di butuhkan aplikasi text editor. Beberapa file yang akan di lakukan configurasi sebagai berikut :

* core-site.xml – (penjelasan)
* hadoop-env.cmd – (penjelasan)
* hdfs-site.xml – (penjelasan)
* mapred-site.xml – (penjelasan)
* yarn-site.xml – (penjelasan)



1. Sebelum melakukan configurasi alangkah baiknya kita melakukan backup configurasi terlebidahulu. Buat folder baru dengan nama backup lalu copy file yang akan di edit

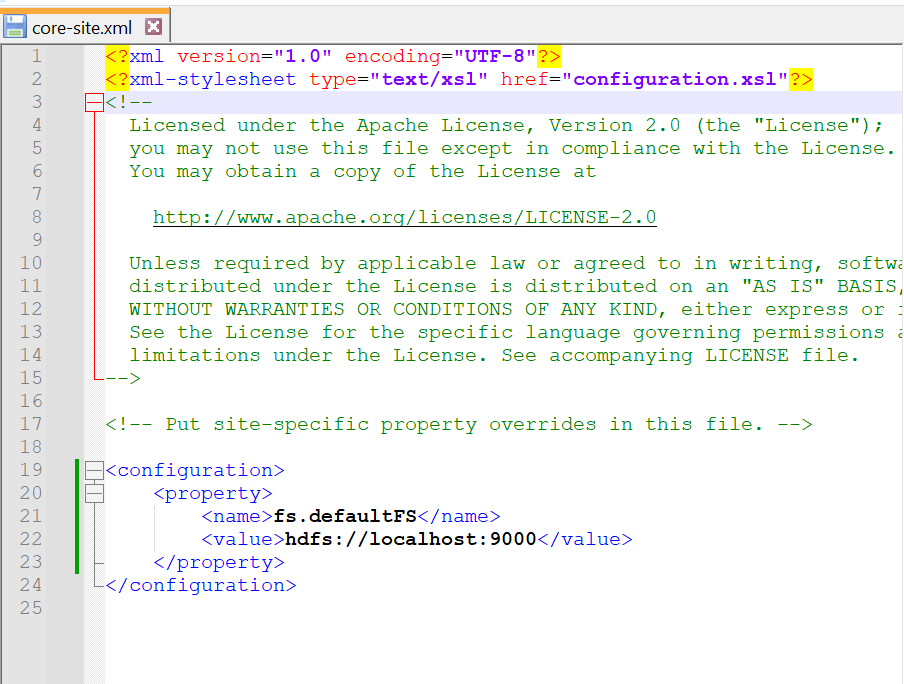


1. Configurasi core-site.xml kita akan melakukan configurasi secara simple untuk pointing dari HDFS kita dengan cara memasukan configurasi sebagai berikut:

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

Paste di dalam <configuration> tag



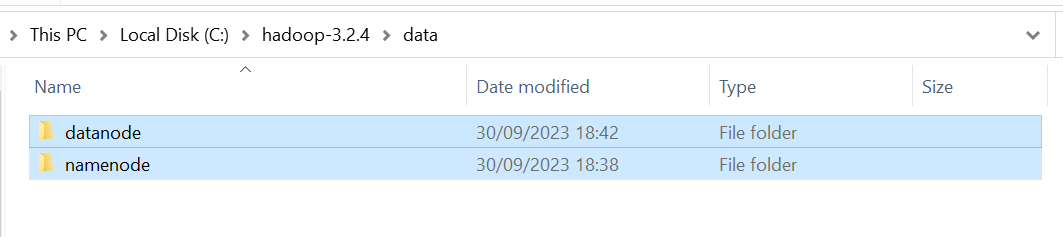
1. Configurasi hadoop-env.cmd kita akan melakukan configurasi untuk menentukan environment java yang di pakai saat berjalannya Hadoop. Masukan variable JAVA\_HOME yang sebelumnya telah di buat untuk env user

set JAVA\_HOME= C:\Java\jdk-1.8



1. Configurasi hdfs-site.xml merupakan configurasi yang penting dalam system HDFS … (penjelasan)  
   <https://mti.binus.ac.id/2018/06/28/hadoop-distributed-file-system-hdfs/>

configurasinya sebagai berikut:



<property>

<name>dfs.replication</name>

<value>1</value>

</property><property>

<name>dfs.namenode.name.dir</name>

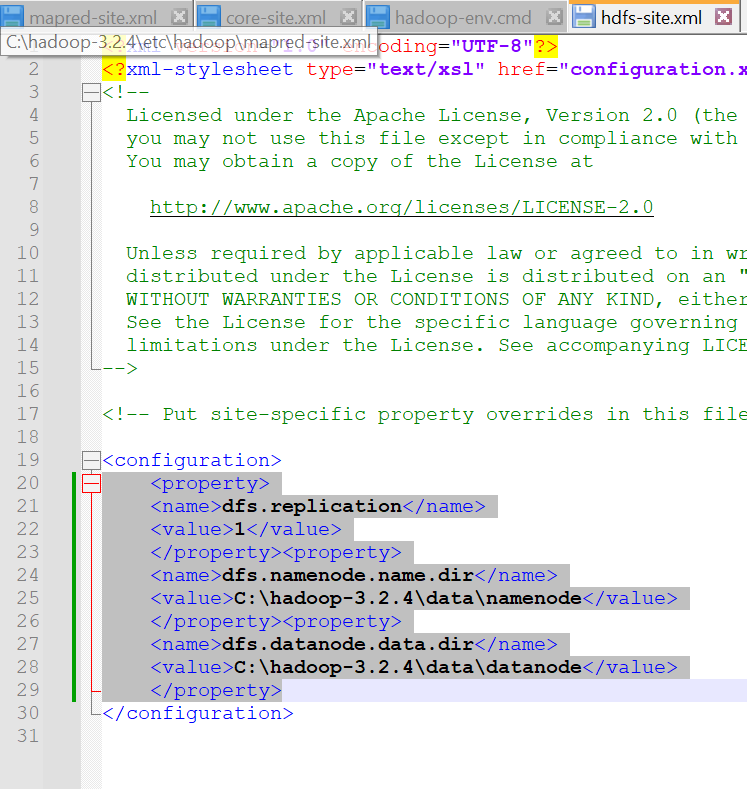
<value>C:\hadoop-3.2.4\data\namenode</value>

</property><property>

<name>dfs.datanode.data.dir</name>

<value>C:\hadoop-3.2.4\data\datanode</value>

</property>

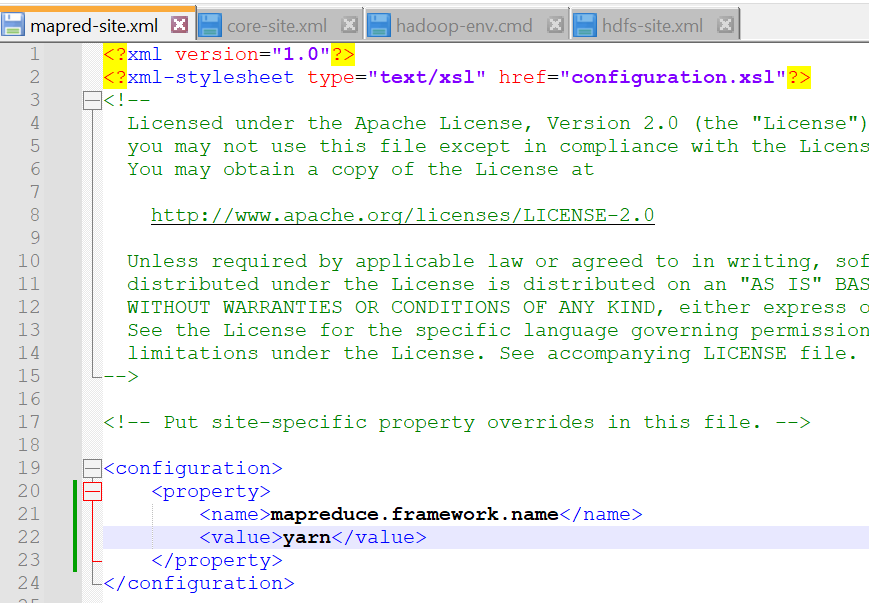


1. Configurasi .. (penjelasan)

|  |  |  |
| --- | --- | --- |
| mapreduce.framework.name | local | The runtime framework for executing MapReduce jobs. Can be one of local, classic or yarn. |

<https://hadoop.apache.org/docs/r2.7.1/hadoop-mapreduce-client/hadoop-mapreduce-client-core/mapred-default.xml>

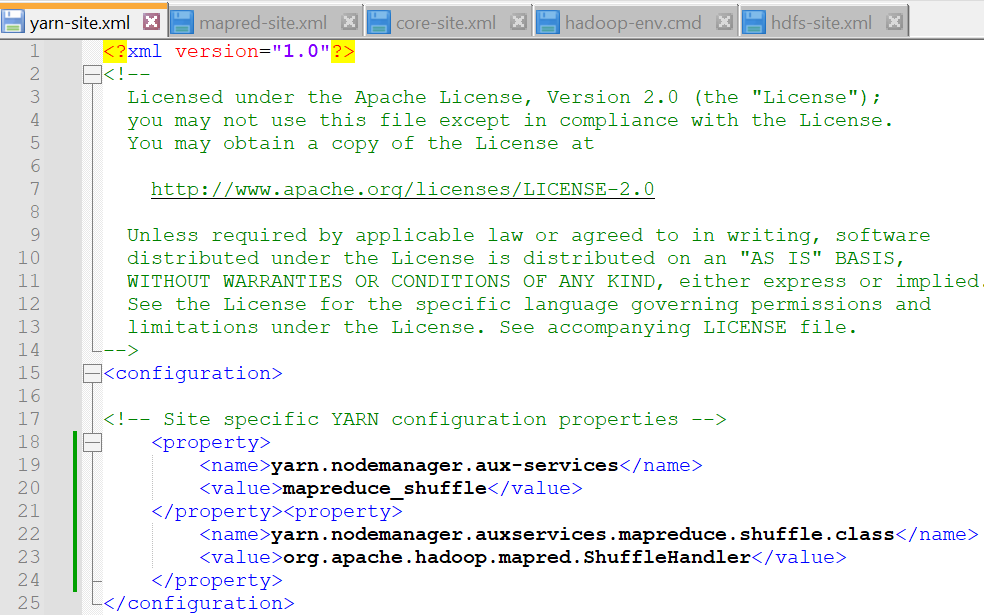
<https://www.edureka.co/blog/explaining-hadoop-configuration/>



1. Configurasi hdfs-site.xml … (penjelasan)

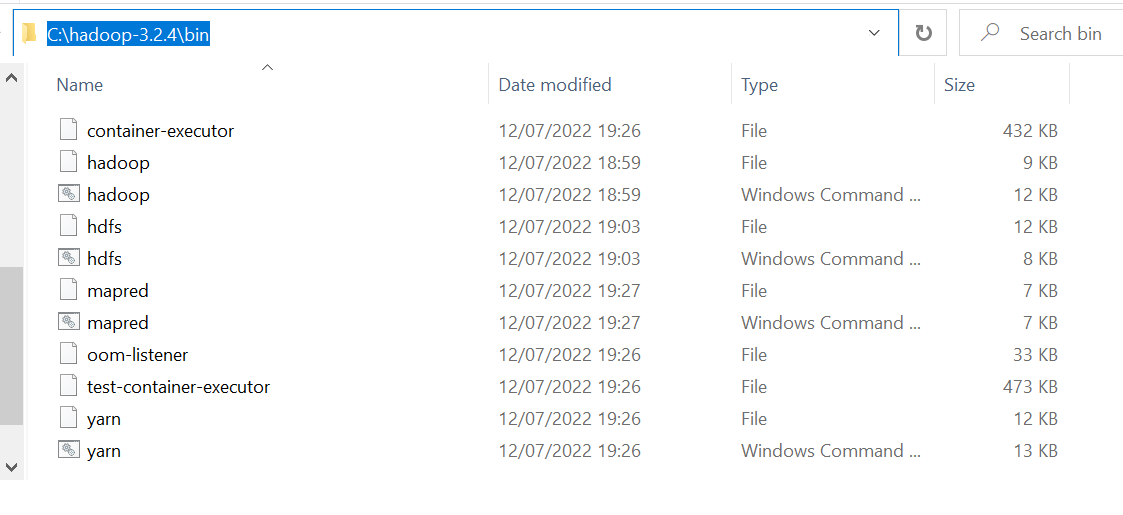
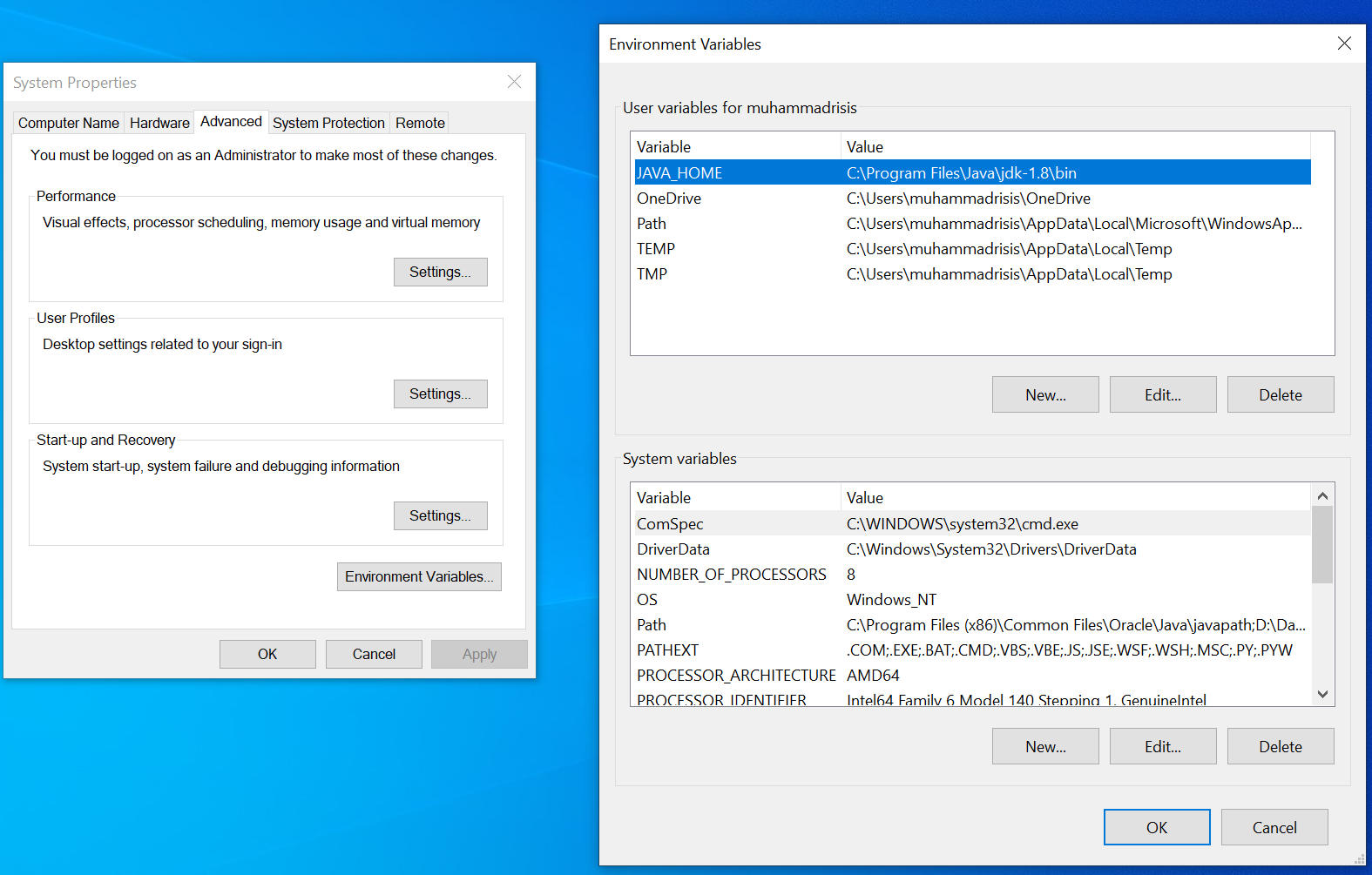
<https://hadoop.apache.org/docs/r2.7.1/hadoop-yarn/hadoop-yarn-site/YARN.html>

a

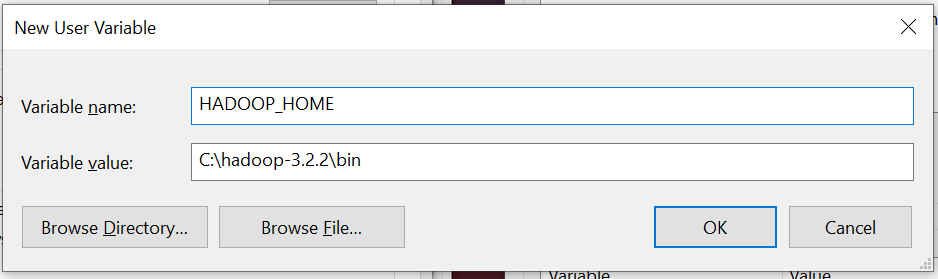


**D. Konfigurasi Environment Variable Hadoop**

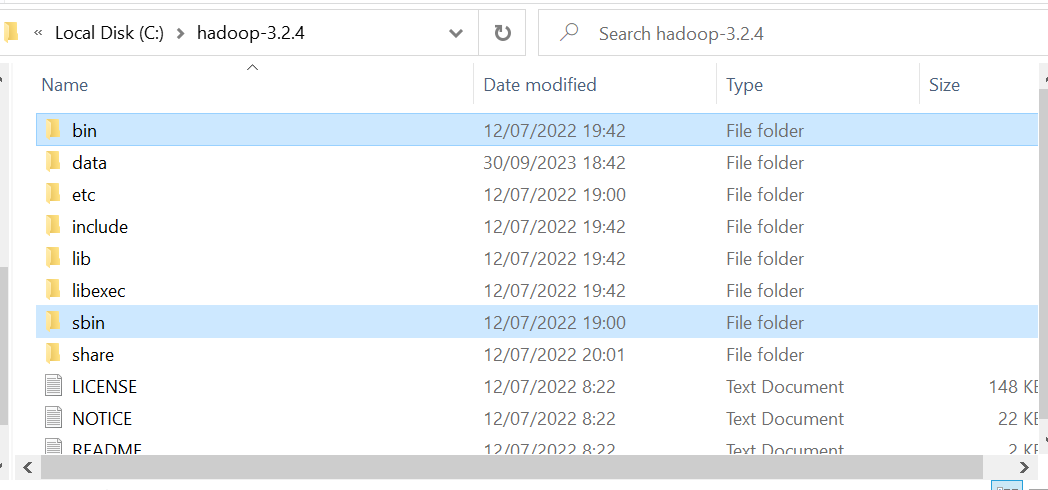
1. Selanjutnya kita akan melakukan konfigurasi Environment Variable agar dapat di kenali oleh windows. Path C:\hadoop-3.2.2\bin harus di daftarkan

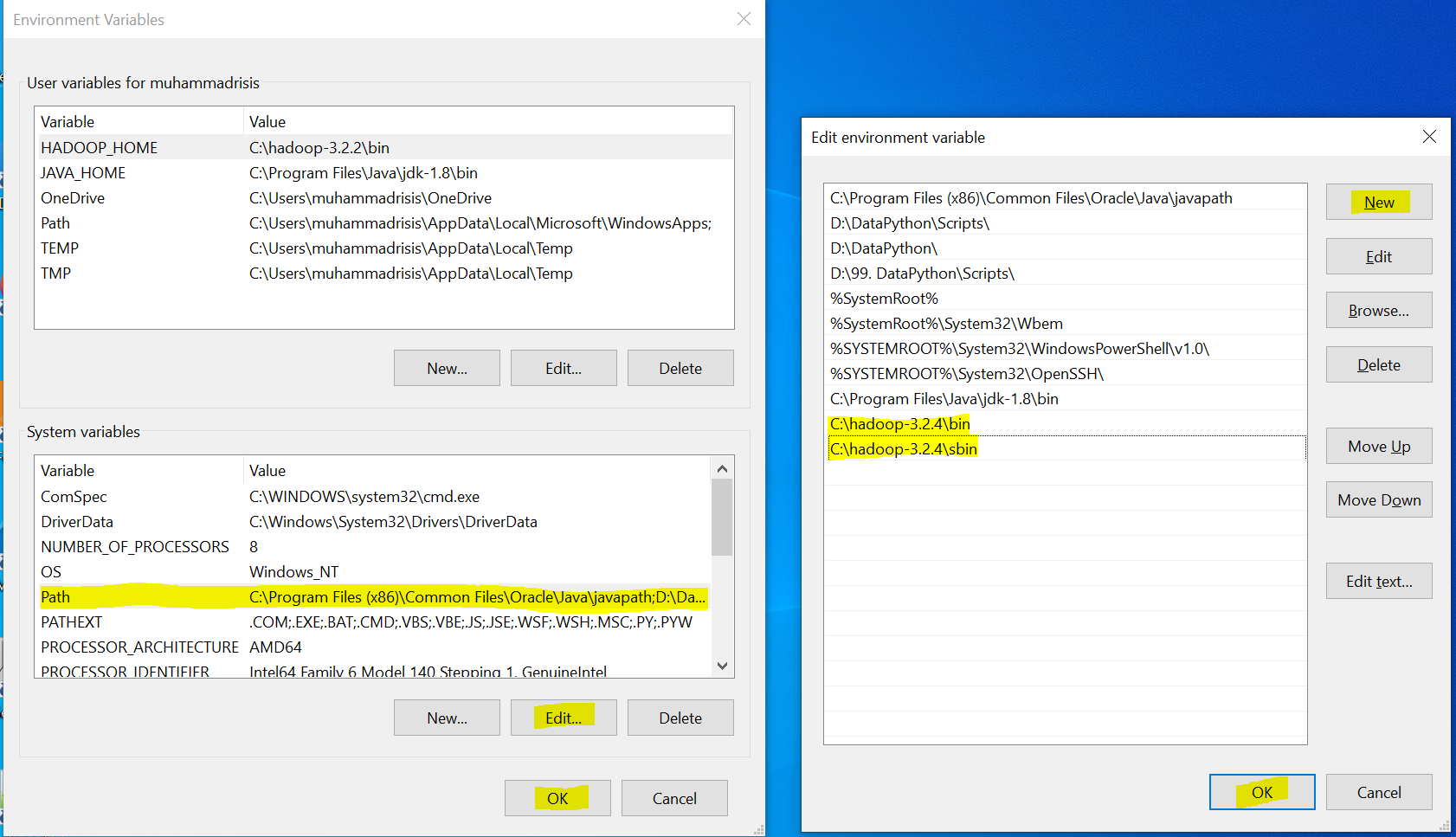
1. Selanjutnya kita masukan jalur dari binary Hadoop ke dalam user variable dengan nama HADOOP\_HOME dan valuenya C:\hadoop-3.2.2\bin



1. Daftarkan juga di variable path agar dapat terbaca oleh system. Kedua file di bawah harus di daftarkan ke dapam “path” system



1. 1. Klik “Path” di system variable, 2. Klik Tombol Edit, 3. Klik New, 4. Masukan path dari folder bin Hadoop C:\hadoop-3.2.2\bin 5. Masukan path dari folder sbin Hadoop C:\hadoop-3.2.2\sbin, 6. Klik OK, 7. Klik OK



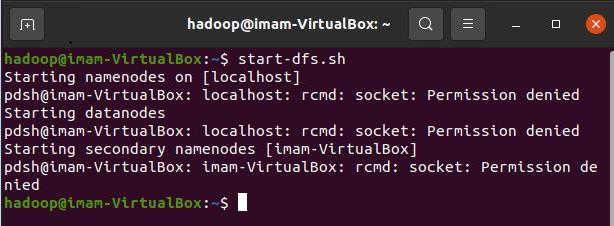
**Be careful and click on OK in all windows related to the environment variables to save the changes**

**E. Menjalankan Hadoop**

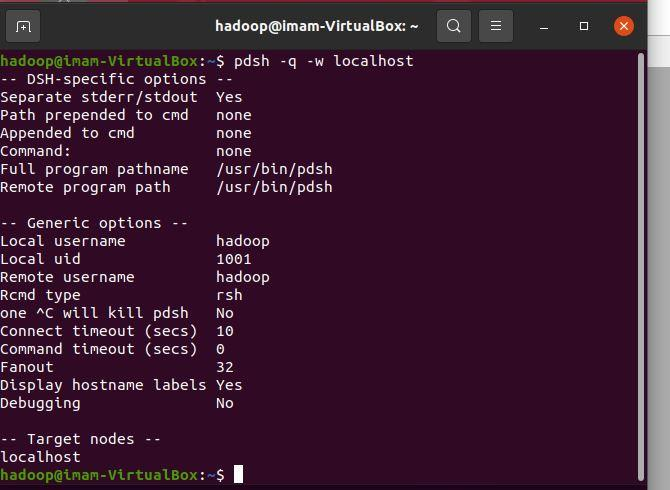
1. Tahap selanjutnya kita sudah bisa menjalankan hadoopnya dimulai dari perintah untuk masuk ke file hadoop sbin.



1. Selanjutnya menjalankan hadoop cluster atau dfs (distributed file system) service dengan menghasilkan namenodes, datanodes, dan secondary namenodes.
2. Selanjutnya menjalankan hadoop cluster atau dfs (distributed file system) service dengan menghasilkan namenodes, datanodes, dan secondary namenodes.



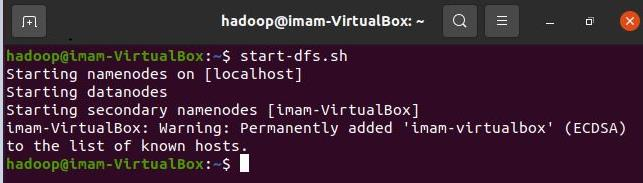
Jika mengalami hal diatas lakukan perintah berikut untuk cek tipe rcmd.



Ubah menjadi ssh dengan perintah berikut

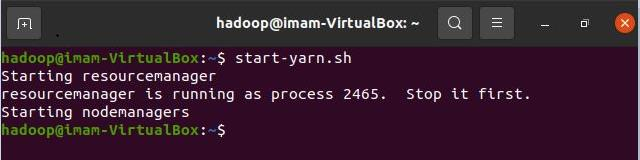


Kemudian jalankan perintah kembali

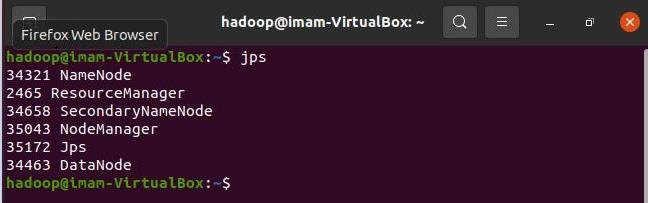


3. Selain itu jalankan juga YARN (Yet Another Resource Negotiator) service dengan menghasilkan resourcemanager dan nodemanagers.

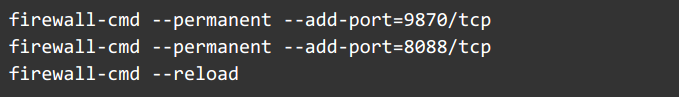


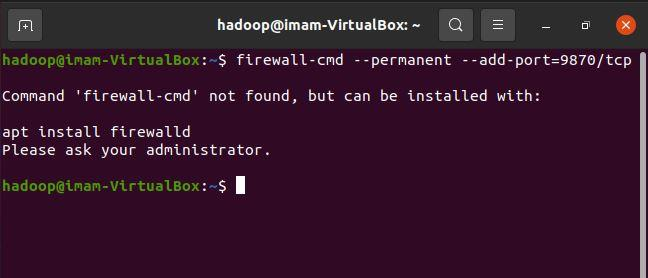


4. Setelah berhasil dijalankan, kita dapat melihat statusnya apakah sudah berhasil atau belum dengan jps (java virtual machine process status tool).

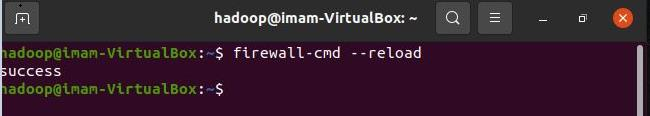
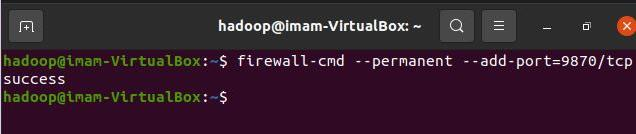
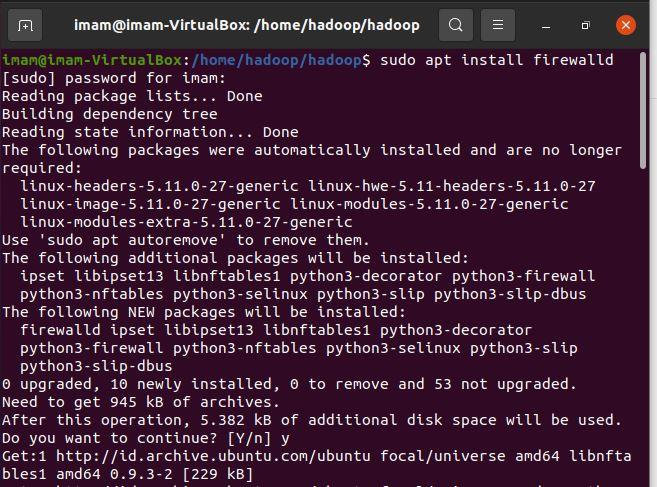
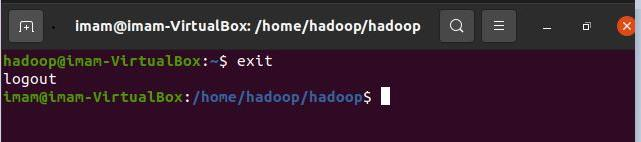


5. Untuk menjalankan hadoop kita dapat menggunakan browser dengan IP 127.0.0.1 yang sudah kita setting sebelumnya di file core-site.xml dengan port 9870 dan 8088. Namun kita masih perlu menyesuaikan portnya agar bisa diakses dengan perintah berikut.



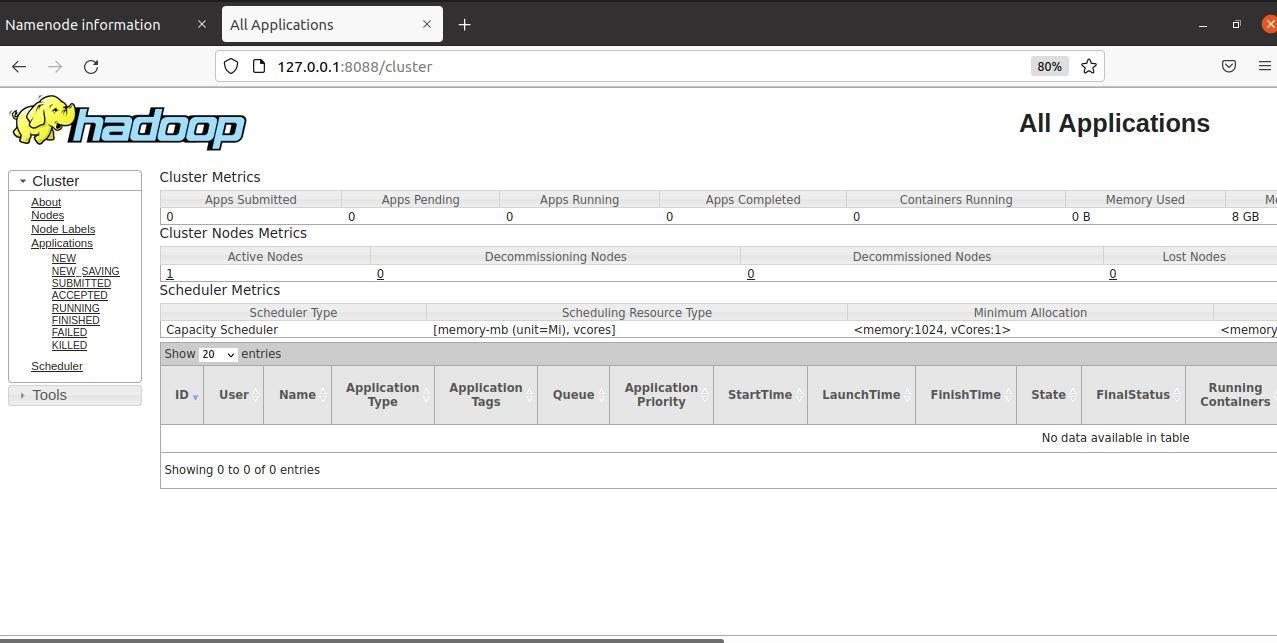
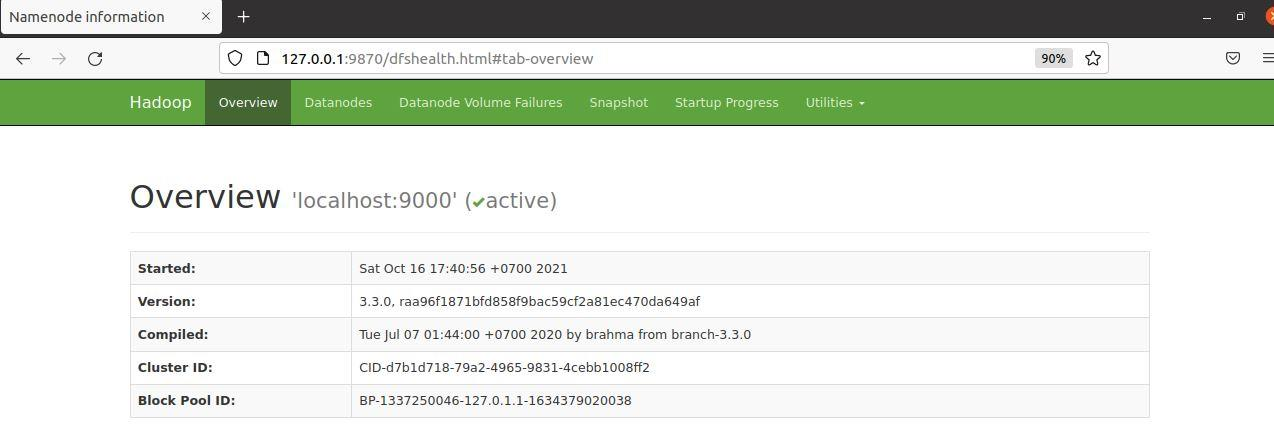
Jika tidak berhasil dijalankan dan menampilkan hasil berikut, maka perlu melakukan instalasi firewalld terlebih dahulu menggunakan superuser.

Setelah sudah berhasil diinstal maka coba kembali perintah di atas.

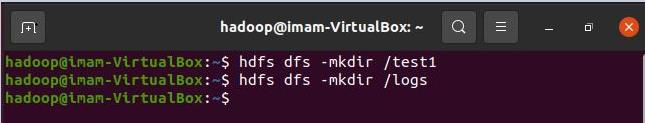


6. Terakhir buka browser dan ketikkan URL berikut

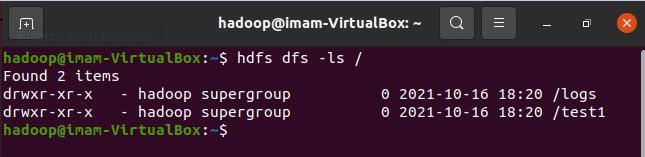
**E. Testing Hadoop**



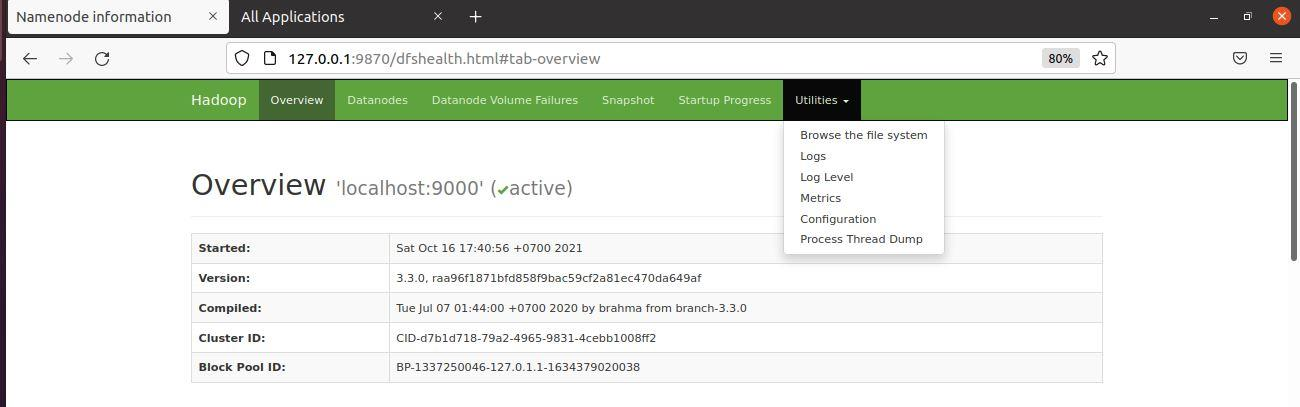
1. Tahap terakhir kita akan coba melakukan testing bagaimana cara menggunakannya, kita buat terlebih dahulu contoh direktori.

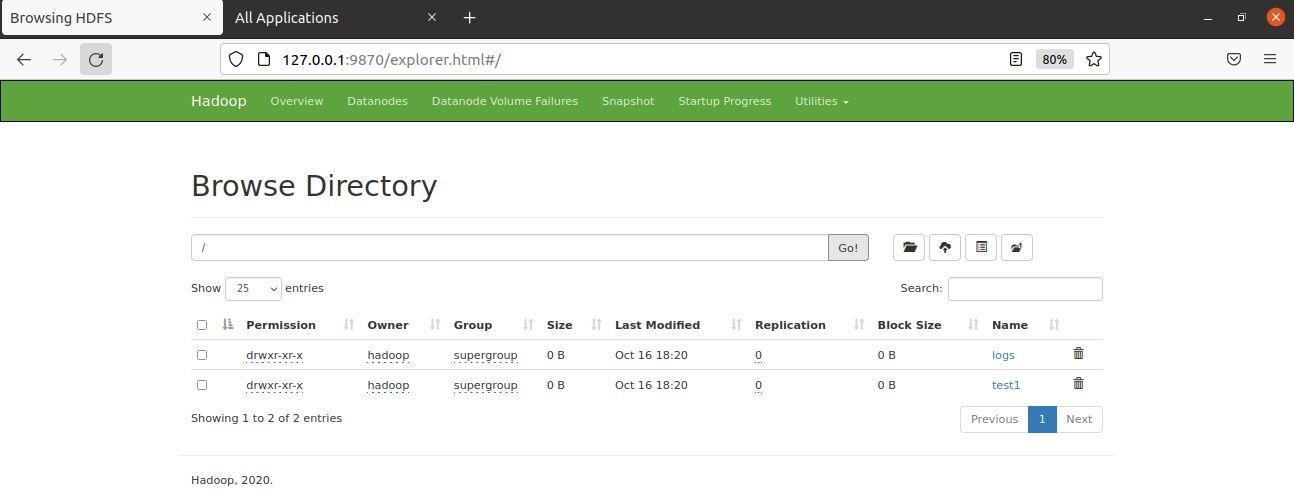


2. Perlu diingat direktori yang dibuat terletak di hdfs bukan di / (root) dan kita bisa memeriksanya dengan perintah berikut



Atau kalian bisa melihatnya langsung di browser dengan memilih menu Utilities kemudian pilih Browser the file system.



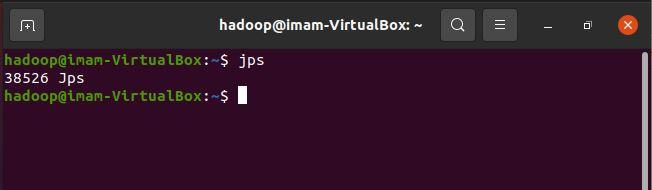
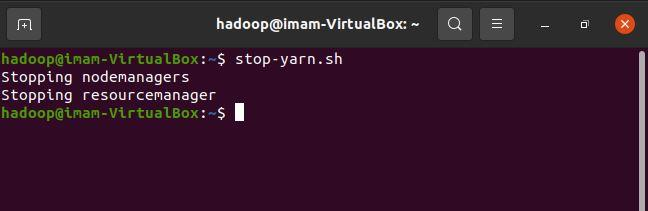
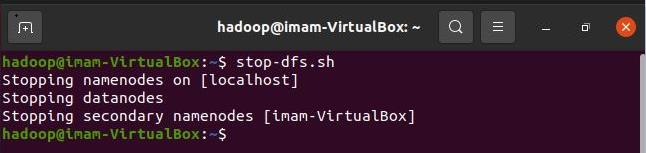
Maka akan terlihat 2 direktori file system sebelumnya dan untuk selanjutnya dapat kalian lakukan proses big data namun tidak akan dibahas pada modul ini.

**F. Mematikan Hadoop**

1. Jika kalian ingin mematikan service hadoopnya kalian dapat mengetikkan perintah berikut.



2. Jika kalian ingin menyalakan hadoop service lagi dapat juga dilakukan dengan perintah berikut, cara sebelumnya agar memastikan tidak ada yang error di setiap bagian.





**Referensi**

* <https://medium.com/skyshidigital/teknologi-big-data-dengan-hadoop-d8a2e93791a8>
* <https://tecadmin.net/install-hadoop-on-ubuntu-20-04/>
* Modul Praktikum 1: Instalasi Hadoop, Mata Kuliah Big Data 20202