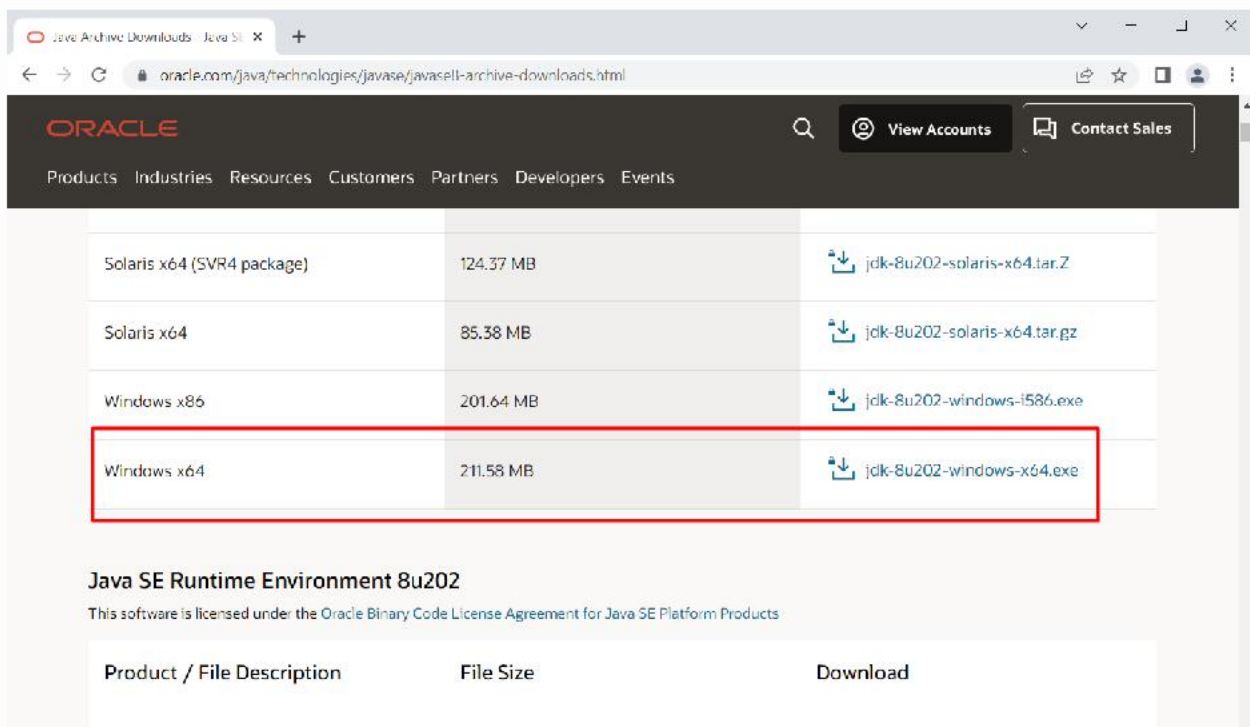


Hướng dẫn cài đặt Hadoop 3.3.0 trên Windows

Cài đặt JDK bản 1.8 (bắt buộc)

Bước 1: Tải bộ cài

- Link: <https://www.oracle.com/java/technologies/javase/javase8-archive-downloads.html>. Chọn phiên bản dành cho **Windows x64**.

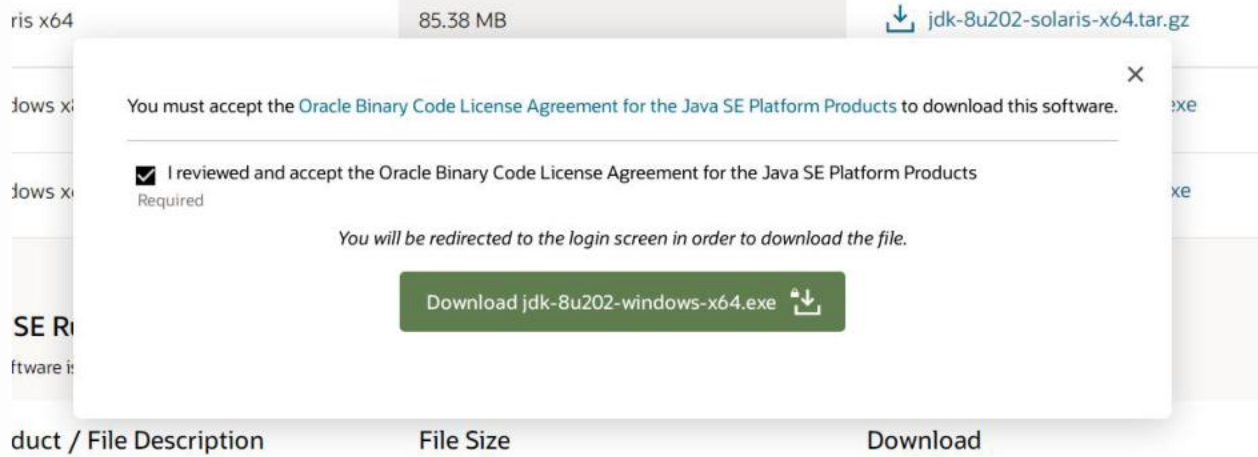


The screenshot shows the Oracle Java Archive Downloads page. The page title is "Java Archive Downloads" and the URL is "oracle.com/java/technologies/javase/javase8-archive-downloads.html". The page features a navigation bar with the Oracle logo and links for "View Accounts" and "Contact Sales". Below the navigation bar, there is a table listing various Java SE Runtime Environment (JRE) versions and their download links. The table has three columns: "Product / File Description", "File Size", and "Download". The row for "Windows x64" is highlighted with a red border, indicating the correct download link for Windows x64. Below the table, there is a section for "Java SE Runtime Environment 8u202" with a license agreement link.

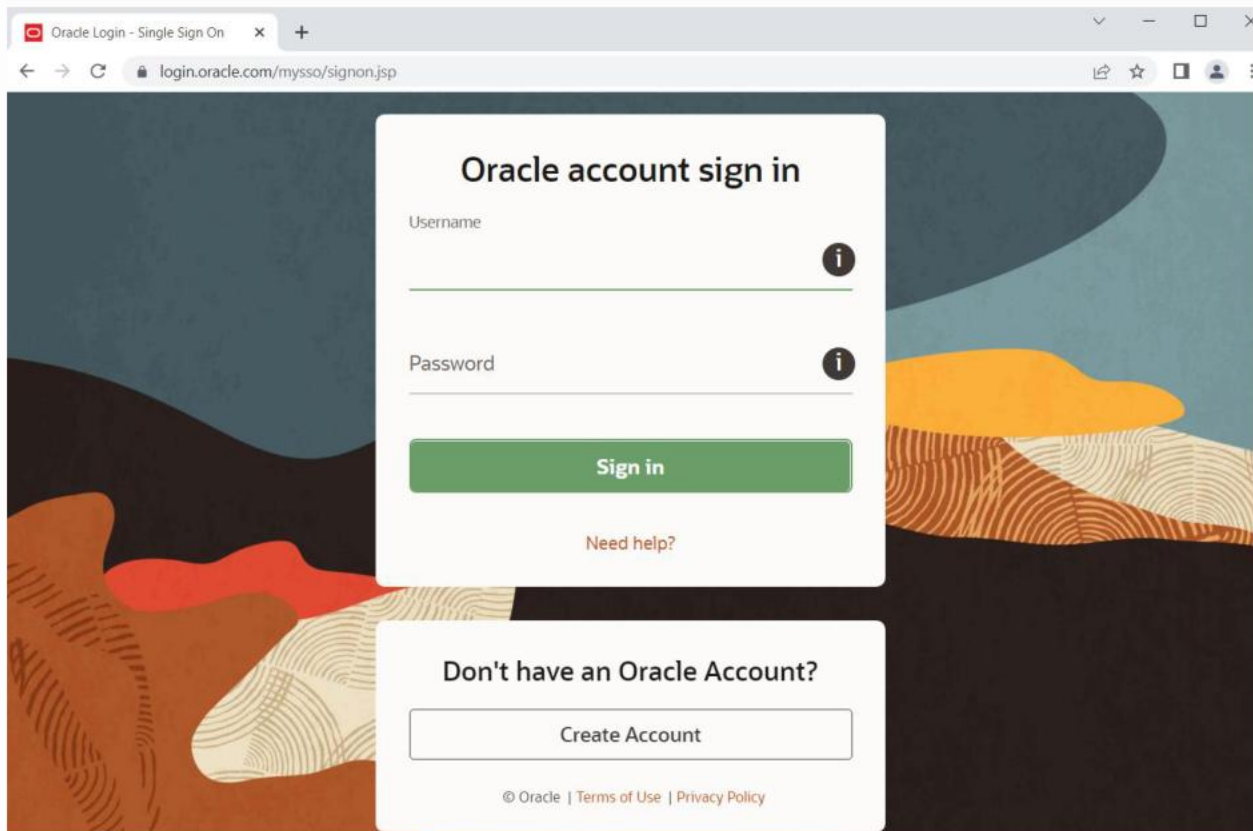
Product / File Description	File Size	Download
Solaris x64 (SVR4 package)	124.37 MB	jdk-8u202-solaris-x64.tar.Z
Solaris x64	85.38 MB	jdk-8u202-solaris-x64.tar.gz
Windows x86	201.64 MB	jdk-8u202-windows-i586.exe
Windows x64	211.58 MB	jdk-8u202-windows-x64.exe

Java SE Runtime Environment 8u202
This software is licensed under the Oracle Binary Code License Agreement for Java SE Platform Products

- Tick vào checkbox và bấm nút **Download**

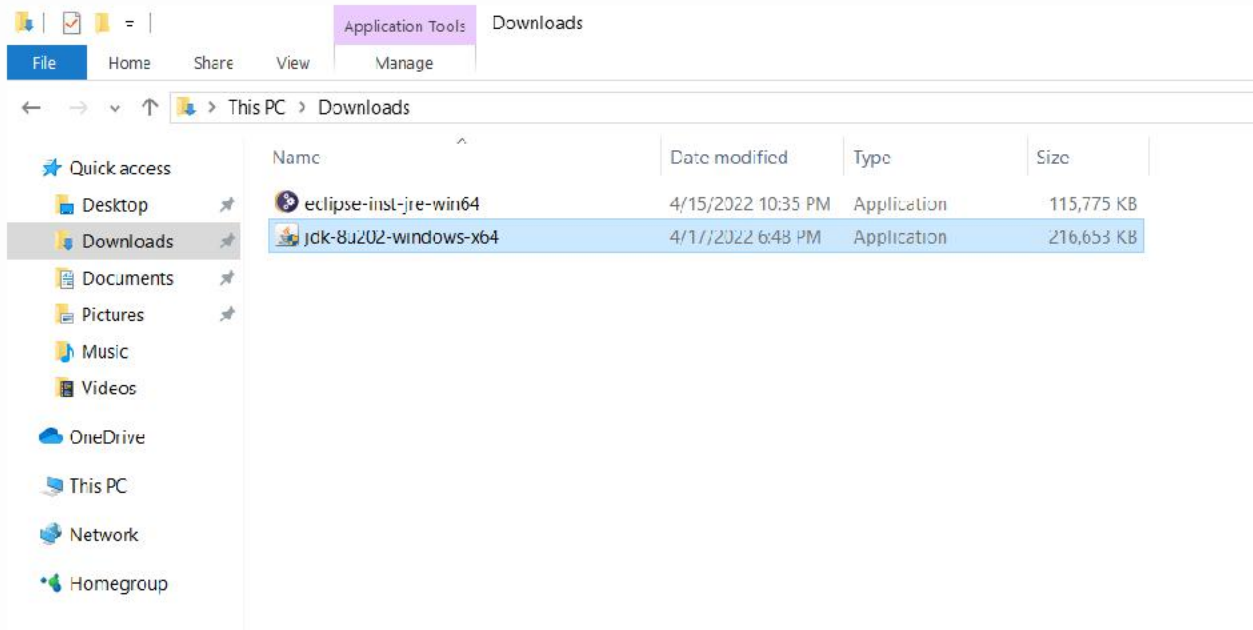


- Cần thực hiện đăng nhập tài khoản trước khi tải xuống, nếu chưa có tài khoản bấm nút **Create Account** để tạo. Sau khi đăng nhập thành công, bộ cài sẽ tự động được tải xuống.



Bước 2: Cài đặt

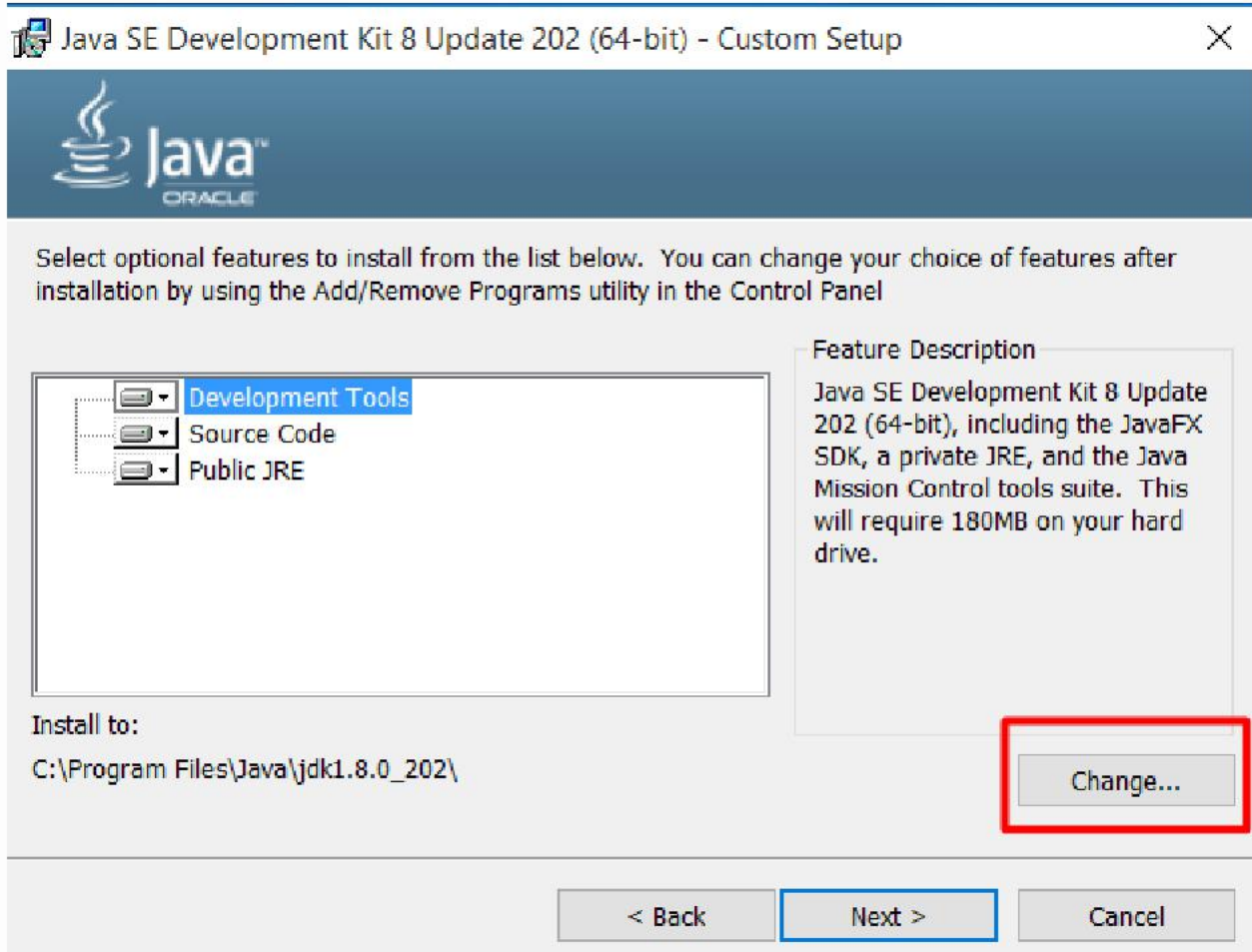
- Double click vào file cài đặt vừa tải xuống.



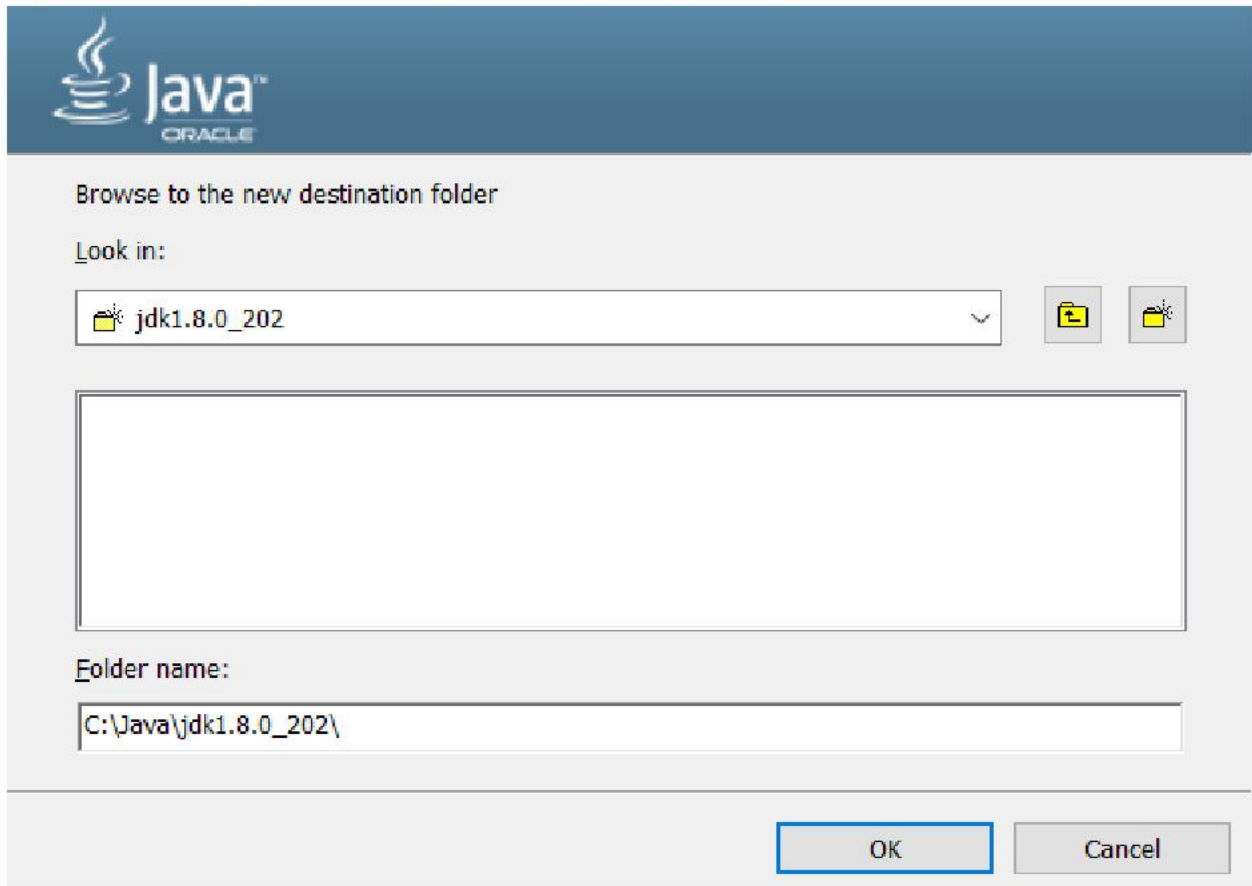
- **Bấm Next**



- Bấm **Change** để thay đổi đường dẫn lưu.



- Chọn thư mục để lưu JDK. **Lưu ý: đường dẫn không có dấu và không có khoảng trắng.**



- Tiếp tục bấm **Next**.



Select optional features to install from the list below. You can change your choice of features after installation by using the Add/Remove Programs utility in the Control Panel

- Development Tools
- Source Code
- Public JRE

Feature Description

Java SE Development Kit 8 Update 202 (64-bit), including the JavaFX SDK, a private JRE, and the Java Mission Control tools suite. This will require 180MB on your hard drive.

Install to:

C:\Program Files\Java\jdk1.8.0_202\

Change...

< Back

Next >

Cancel

Java SE Development Kit 8 Update 202 (64-bit) - Progress



Status: Updating component registration



- Bấm **OK**.



Important Information about Oracle Java SE Roadmap

Changes are coming which will impact your access to future releases of Java SE from Oracle.

Corporate users will be impacted with the April 2019 update.

These changes do not affect the version you are about to install.

For additional guidance please follow the link below.

[More information...](#)

OK

- **Bấm Next.**



Destination Folder

Click "Change" to install Java to a different folder.

Install to:
C:\Program Files\Java\jre1.8.0_202

Change...

< Back

Next >

Java Setup - Progress



- Hoàn tất cài đặt JDK 1.8, bấm **Cancel** để thoát.



Java SE Development Kit 8 Update 202 (64-bit) Successfully Installed

Click **Next Steps** to access tutorials, API documentation, developer guides, release notes and more to help you get started with the JDK.

Next Steps

Close

Bước 3: Thiết lập biến môi trường JDK

- Search từ khóa **environment** ở thanh Taskbar, chọn mục **Edit environment variables for your account**.

The image shows a Windows search interface. At the top, a search bar contains the text "environment". Below the search bar, a list of search results is displayed. The first result is "Edit the system environment variables" with a sub-label "Control panel" and a laptop icon. This result is highlighted with a red border. Below it is a "Settings" section with a sub-result "Edit environment variables for your account" and a laptop icon. A "Web" section follows, listing several search suggestions: "environment", "environmental", "environment variable", "environmental science", "environmental issues", "environment synonym", "environmental justice", and "environmental protection agency". At the bottom, there are two tabs: "My stuff" and "Web". The search bar at the bottom left contains the text "environment", which is also highlighted with a red border.

environment

Edit the system **environment** variables
Control panel

Settings

Edit **environment** variables for your account

Web

environment

environmental

environment variable

environmental science

environmental issues

environment synonym

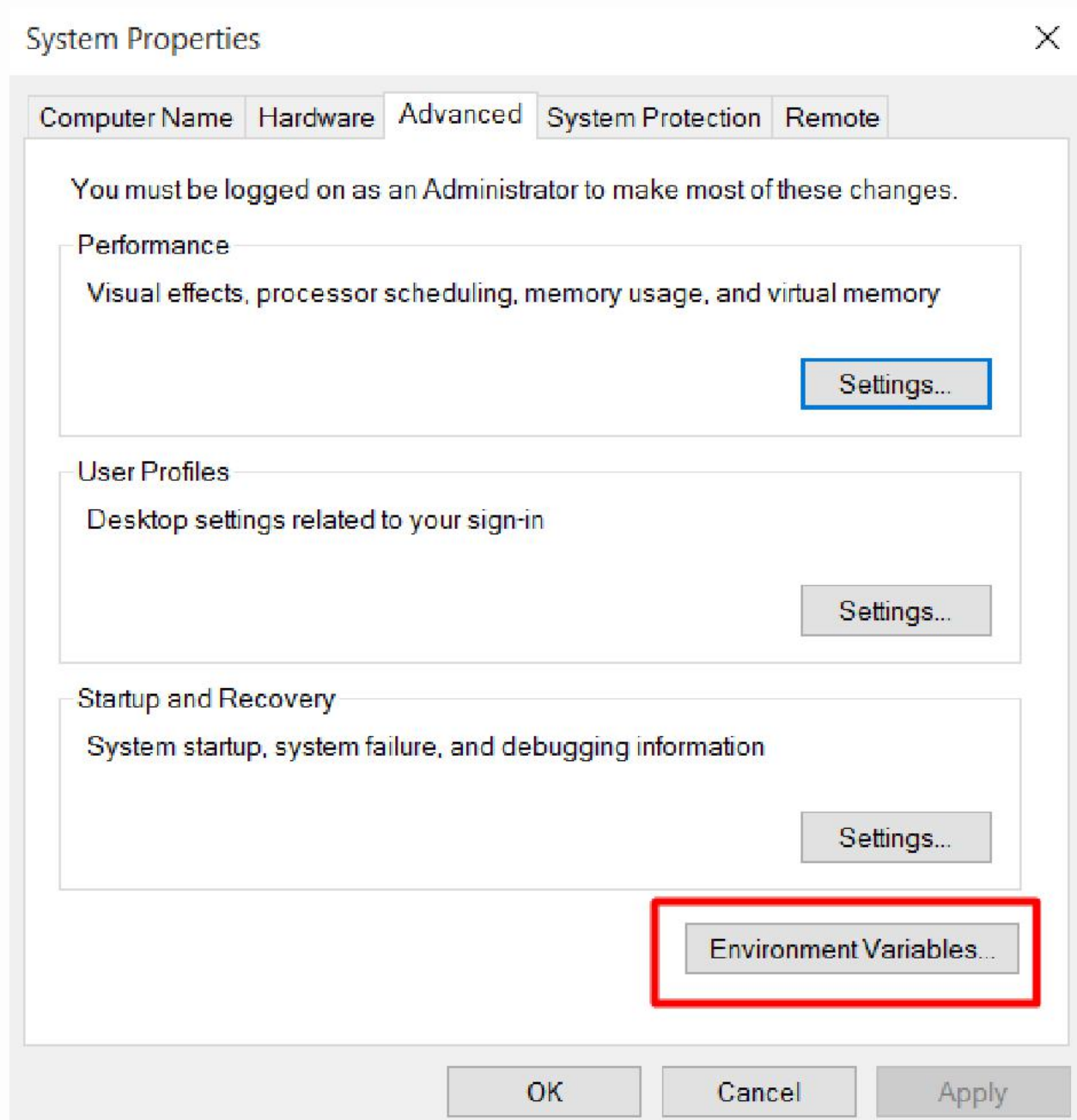
environmental justice

environmental protection agency

My stuff

Web

- Bấm **Environment Variables...**



- Bấm **New**.

Environment Variables



User variables for dachi

Variable	Value
OneDrive	C:\Users\dachi\OneDrive
TEMP	%USERPROFILE%\AppData\Local\Temp
TMP	%USERPROFILE%\AppData\Local\Temp

New... Edit... Delete

System variables

Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
NUMBER_OF_PR...	2
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\...

New... Edit... Delete

OK Cancel

- Nhập **Variable Name: JAVA_HOME** và **Variable Value** là đường dẫn lưu folder JDK **C:\Java\jdk1.8.0_202**

New User Variable ✕

Variable name:

Variable value:

- Tiếp tục làm tương tự với **System Variables**.

Environment Variables



User variables for dachi

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
OneDrive	C:\Users\dachi\OneDrive
TEMP	%USERPROFILE%\AppData\Local\Temp
TMP	%USERPROFILE%\AppData\Local\Temp

System variables

Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
NUMBER_OF_PR...	2
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\...

New System Variable



Variable name:	<input type="text" value="JAVA_HOME"/>
Variable value:	<input type="text" value="C:\Java\jdk1.8.0_202"/>
<input type="button" value="OK"/> <input type="button" value="Cancel"/>	

- Chọn **Path** và bấm **Edit** của phần **User Variables**

Environment Variables



User variables for dachi

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
OneDrive	C:\Users\dachi\OneDrive
OneDriveConsumer	C:\Users\dachi\OneDrive
Path	C:\Ruby31-x64\bin;C:\Users\dachi\AppData\Local\Microsoft\...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.RB;.RB...
TEMP	C:\Users\dachi\AppData\Local\Temp
TMP	C:\Users\dachi\AppData\Local\Temp

New...

Edit...

Delete

System variables

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
lib	C:\Program Files\SQLXML 4.0\bin\
M2_HOME	C:\Program Files\apache-maven-3.8.5
NUMBER_OF_PROCESSORS	12
OS	Windows_NT
Path	C:\Program Files (x86)\VMware\VMware Workstation\bin\;%J...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64

New...

Edit...

Delete

OK

Cancel

- Bấm **New** và nhập **%JAVA_HOME%\bin**

User variables for dachi

Variable	Value
IntelliJ IDEA Community E...	C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 20...
JAVA_HOME	C:\Java\jdk1.8.0_202
OneDrive	C:\Users\dachi\OneDrive
OneDriveConsumer	C:\Users\dachi\OneDrive
Path	C:\Ruby31-x64\bin;C:\Users\dachi\AppData\Local\Microsoft\...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.RB;.RB...
TEMP	C:\Users\dachi\AppData\Local\Temp

New...

Edit...

Delete

System variables

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
lib	C:\Program Files\SQLXML 4.0\bin\
M2_HOME	C:\Program Files\apache-maven-3.8.5
NUMBER_OF_PROCESSORS	12
OS	Windows_NT
Path	C:\Program Files (x86)\VMware\VMware Workstation\bin\;%J...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC

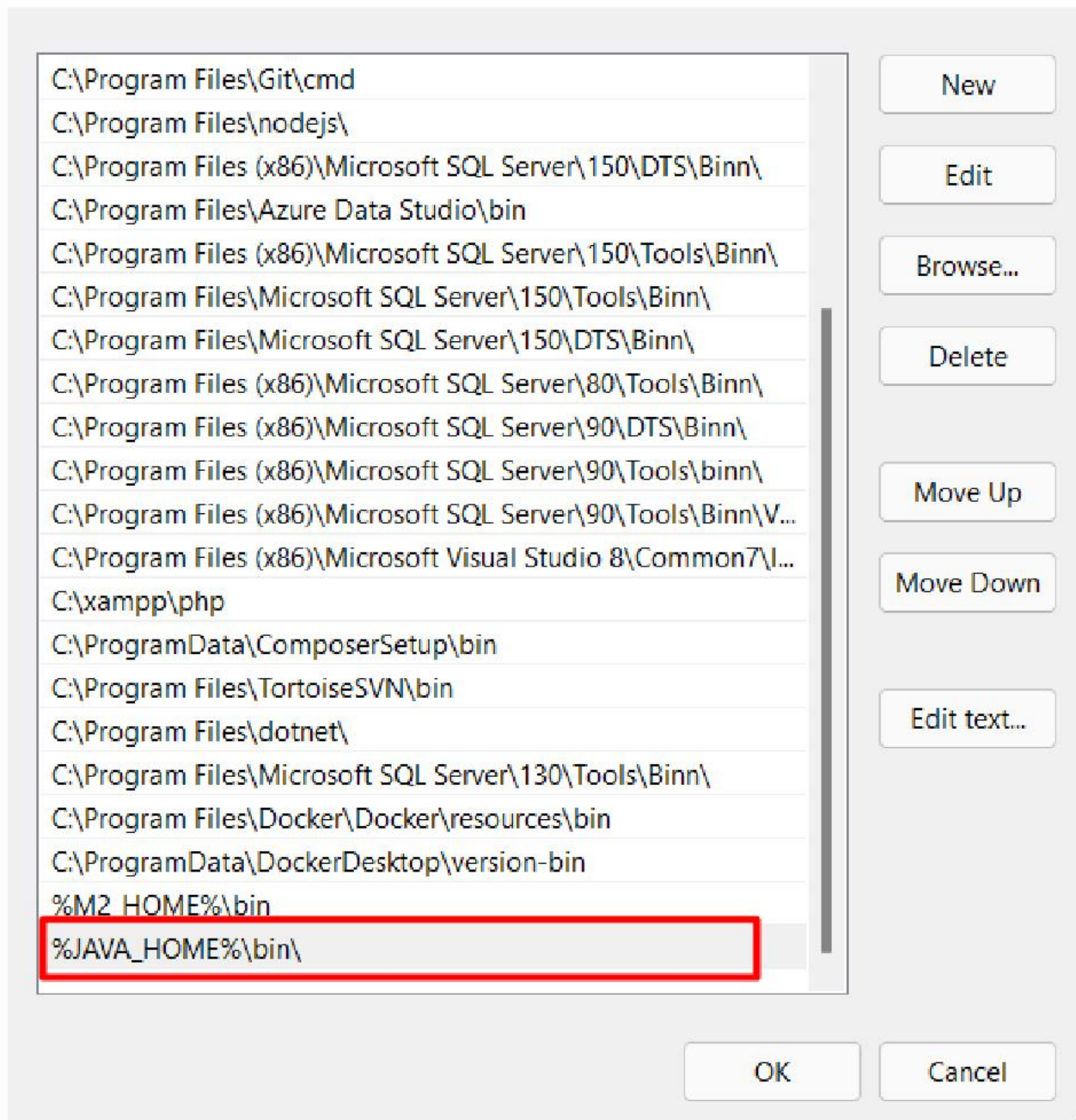
New...

Edit...

Delete

OK

Cancel



- Vào **cmd** gõ **java -version**, nếu kết quả trả về như ảnh thì quá trình thiết lập môi trường thành công.

The image shows a Windows Start menu search interface. On the left is a vertical navigation bar with icons for Home, Settings, and Recent apps. The search bar at the bottom contains the text 'cmd'. The search results are displayed in a list format, with the top result, 'Command Prompt', highlighted in a light blue box and outlined in red. Below it are several search suggestions, each preceded by a magnifying glass icon. At the bottom of the search results area, there are two tabs: 'My stuff' and 'Web', both with magnifying glass icons.

Command Prompt
Desktop app

Web

- cmd
- cmd ipconfig
- cmdb
- cmd commands
- cmdr
- cmdcd
- cmdd

My stuff Web

cmd|

Command Prompt

```
Microsoft Windows [Version 10.0.10240]
(c) 2015 Microsoft Corporation. All rights reserved.

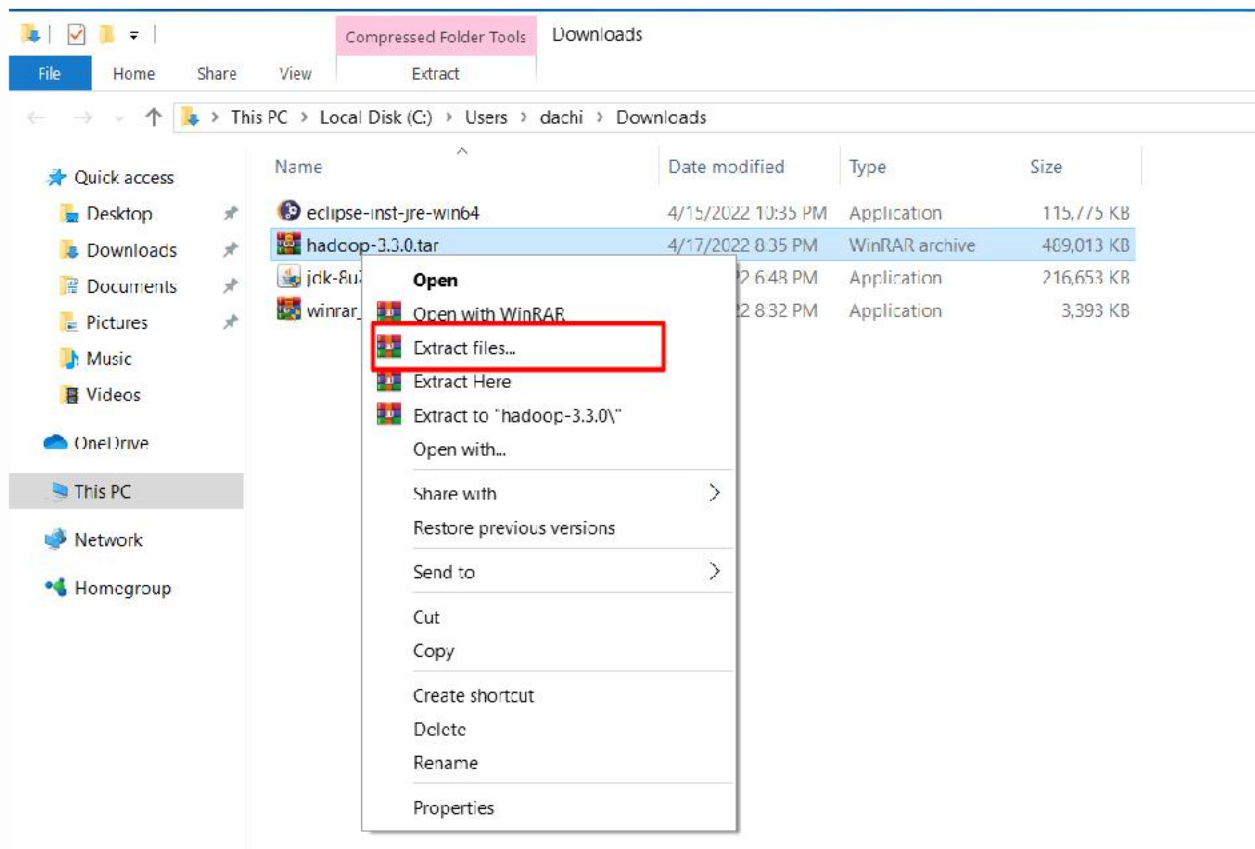
C:\Users\dachi>java -version
java version "1.8.0_202"
Java(TM) SE Runtime Environment (build 1.8.0_202-b08)
Java HotSpot(TM) 64-Bit Server VM (build 25.202-b08, mixed mode)

C:\Users\dachi>
```

Cài đặt Hadoop

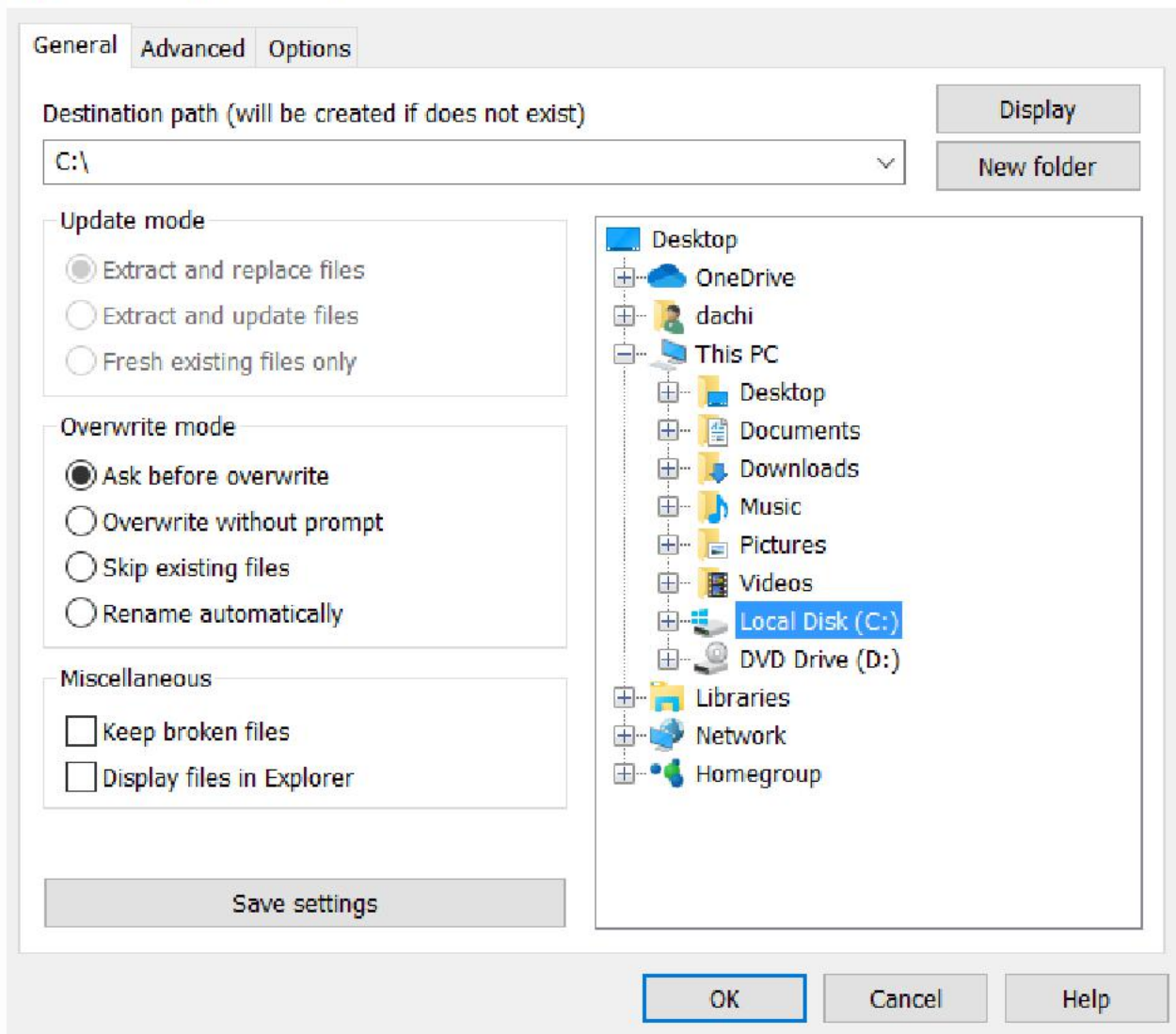
Bước 1: Tải Hadoop 3.3.0

- Truy cập link <https://archive.apache.org/dist/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz> file nén Hadoop 3.3.0 sẽ tự động được tải xuống
- Bấm chuột phải vào file **hadoop-3.3.0.taz.gz** và bấm **Extract files...**

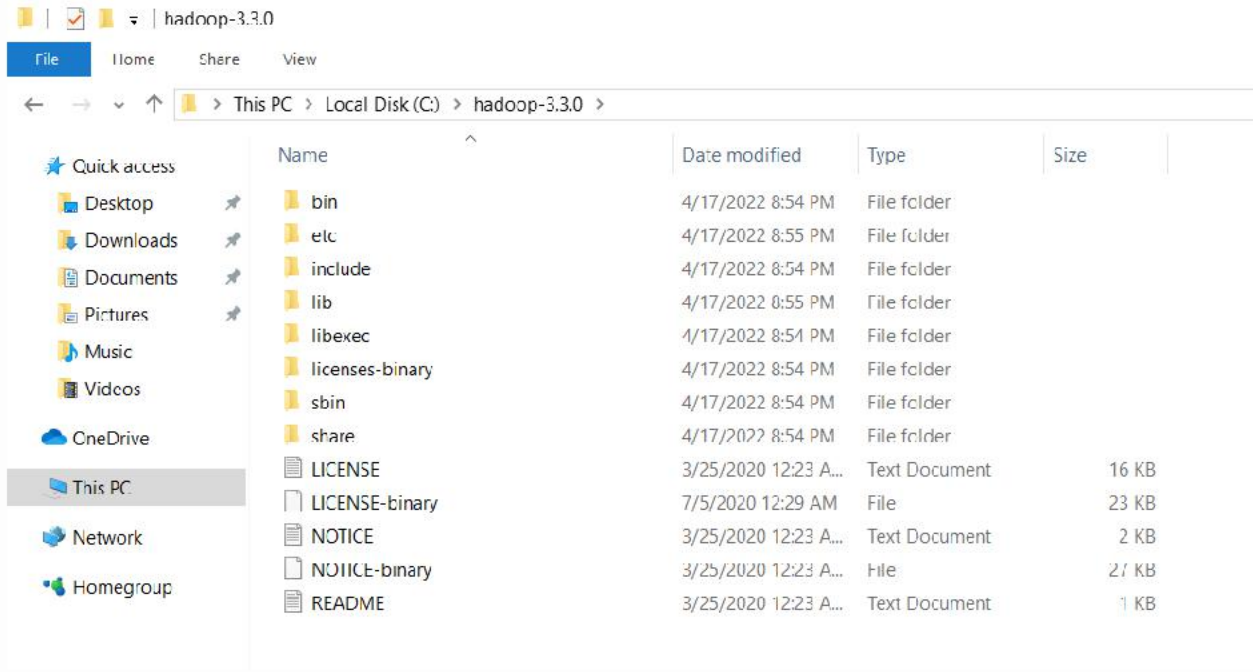


- Chọn ổ **C** và bấm **OK**

Extraction path and options



- Mở ổ C sẽ thấy thư mục **hadoop-3.3.0**



Bước 2: Thiết lập biến môi trường cho Hadoop

- Làm tương tự thiết lập biến môi trường cho JDK. Với **Variable Name: HADOOP_HOME** và **Variable Value** là đường dẫn lưu folder Hadoop **C:\hadoop-3.3.0**

Environment Variables



User variables for dachi

Variable	Value
HADOOP_HOME	C:\hadoop-3.3.0
JAVA_HOME	C:\Java\jdk1.8.0_202
OneDrive	C:\Users\dachi\OneDrive
Path	C:\Java\jdk1.8.0_202
TEMP	C:\Users\dachi\AppData\Local\Temp

New... Edit... Delete

System variables

Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
HADOOP_HOME	C:\hadoop-3.3.0
JAVA_HOME	C:\Java\jdk1.8.0_202
NUMBER_OF_PR...	2
OS	Windows_NT

New... Edit... Delete

OK Cancel

- Thêm 2 giá trị `%HADOOP_HOME%\bin` và `%HADOOP_HOME%\sbin` vào biến **Path** của phần **User Variables**

User variables for dachi

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
OneDrive	C:\Users\dachi\OneDrive
OneDriveConsumer	C:\Users\dachi\OneDrive
Path	C:\Ruby31-x64\bin;C:\Users\dachi\AppData\Local\Microsoft\...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.RB;.RB...
TEMP	C:\Users\dachi\AppData\Local\Temp
TMP	C:\Users\dachi\AppData\Local\Temp

New... Edit... Delete

System variables

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
lib	C:\Program Files\SQLXML 4.0\bin\
M2_HOME	C:\Program Files\apache-maven-3.8.5
NUMBER_OF_PROCESSORS	12
OS	Windows_NT
Path	C:\Program Files (x86)\VMware\VMware Workstation\bin\;%J...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64

New... Edit... Delete

OK Cancel

Edit environment variable



C:\Ruby31-x64\bin

%USERPROFILE%\AppData\Local\Microsoft\WindowsApps

C:\Users\dachi\AppData\Local\Programs\Microsoft VS Code\bin

C:\Users\dachi\AppData\Roaming\npm

C:\Program Files\Azure Data Studio\bin

C:\Users\dachi\AppData\Local\Programs\Azure Data Studio\bin

C:\Users\dachi\AppData\Local\Android\Sdk\platform-tools

C:\Users\dachi\Documents\flutter\bin

C:\Users\dachi\AppData\Roaming\Composer\vendor\bin

%USERPROFILE%\dotnet\tools

C:\Users\dachi\anaconda3\Library\bin

C:\Users\dachi\anaconda3\Scripts

C:\Users\dachi\anaconda3\condabin

%JAVA_HOME%\bin

%HADOOP_HOME%\bin

%HADOOP_HOME%\sbin

%IntelliJ IDEA Community Edition%

New

Edit

Browse...

Delete

Move Up

Move Down

Edit text...

OK

Cancel

- Tương tự thêm vào biến **Path** của phần **System Variables**

User variables for dachi

Variable	Value
IntelliJ IDEA Community E...	C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 20...
JAVA_HOME	C:\Java\jdk1.8.0_202
OneDrive	C:\Users\dachi\OneDrive
OneDriveConsumer	C:\Users\dachi\OneDrive
Path	C:\Ruby31-x64\bin;C:\Users\dachi\AppData\Local\Microsoft\...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.RB;.RB...
TEMP	C:\Users\dachi\AppData\Local\Temp

New...

Edit...

Delete

System variables

Variable	Value
JAVA_HOME	C:\Java\jdk1.8.0_202
lib	C:\Program Files\SQLXML 4.0\bin\
M2_HOME	C:\Program Files\apache-maven-3.8.5
NUMBER_OF_PROCESSORS	12
OS	Windows_NT
Path	C:\Program Files (x86)\VMware\VMware Workstation\bin\;%J...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC

New...

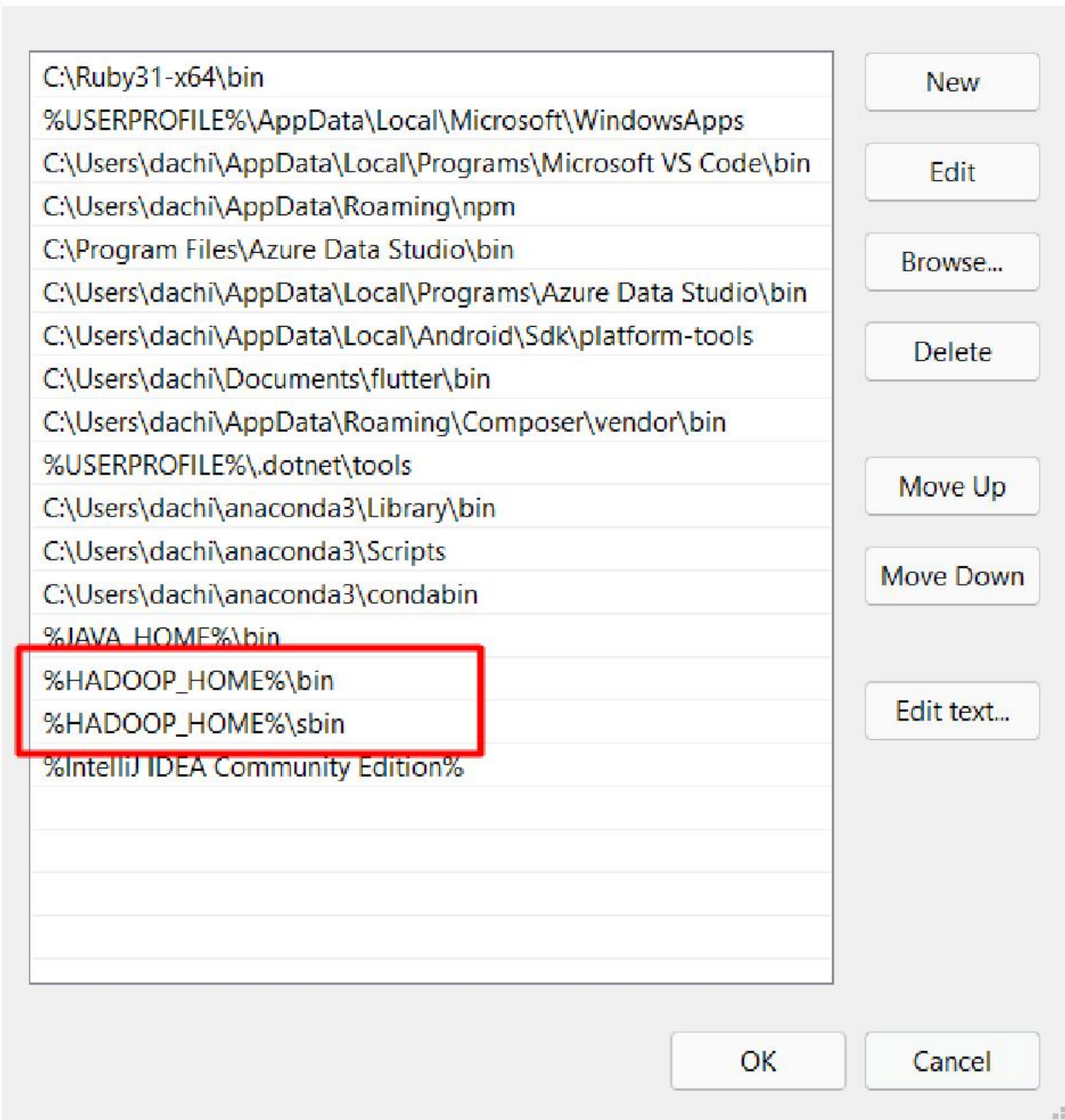
Edit...

Delete

OK

Cancel

Edit environment variable



- Vào **cmd** gõ **hadoop version**, nếu kết quả trả về như ảnh thì quá trình thiết lập môi trường thành công.

```
Command Prompt
Microsoft Windows [Version 10.0.10240]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\dachi>hadoop version
Hadoop 3.3.0
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r aa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled by brahma on 2020-07-06 11:44Z
Compiled with protoc 3.7.1
From source with checksum 5dc29b802d6ccd77b262ef9d04d19c4
This command was run using /C:/hadoop-3.3.0/share/hadoop/common/hadoop-common-3.3.0.jar

C:\Users\dachi>
```

Bước 3: Cấu hình các tập tin cho Hadoop

- Truy cập thư mục **C:\hadoop-3.3.0\etc\hadoop**
- Cấu hình core-site.xml như dưới đây:

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

- Cấu hình mapred-site.xml như dưới đây:

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

- Tạo thư mục **data** trong **C:\hadoop-3.3.0**
- Tạo thư mục con **datanode** trong **C:\hadoop-3.3.0\data**
- Tạo thư mục con **namenode** trong **C:\hadoop-3.3.0\data**
- Sau đó cấu hình hdfs-site.xml như sau:

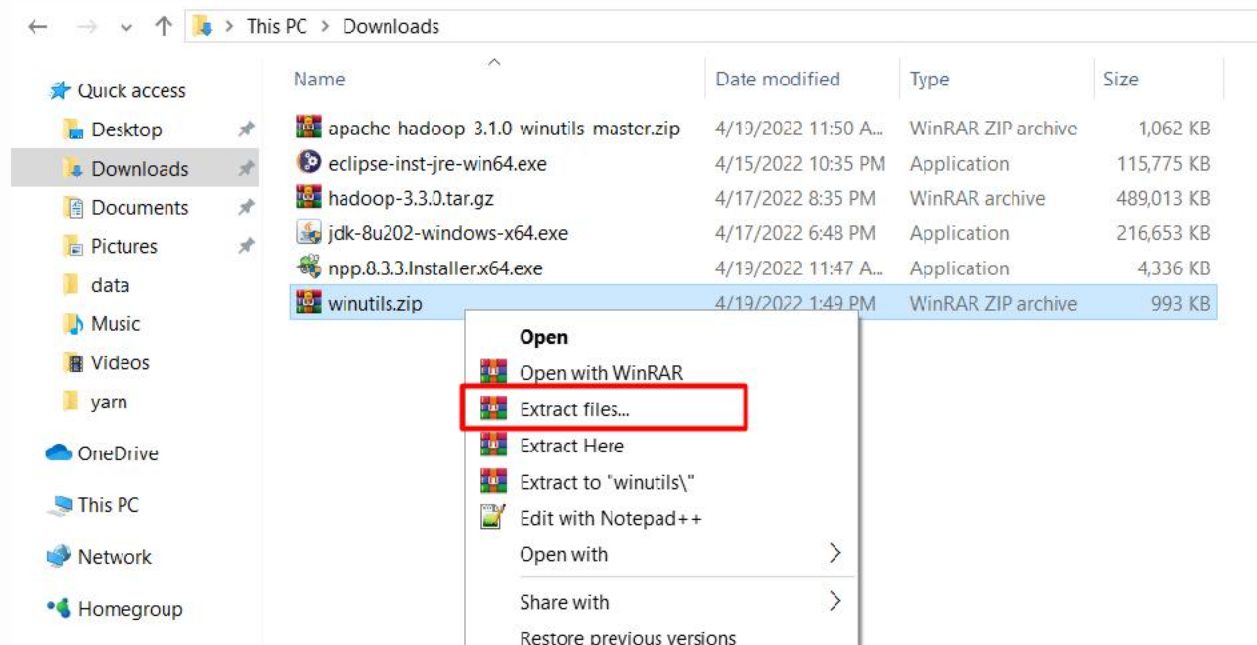
```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/hadoop-3.3.0/data/namenode</value>
  </property>
</configuration>
```

```
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/hadoop-3.3.0/data/datanode</value>
</property>
</configuration>
```

- Cấu hình yarn-site.xml như dưới đây:

```
<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>
```

- Tải [apache-hadoop-winutils](#)
- Giải nén sẽ thấy thư mục **bin** bên trong. Chép đè thư mục bin này vào thư mục **C:\hadoop-3.3.0\bin**



Extraction path and options



General | Advanced | Options

Destination path (will be created if does not exist)

C:\hadoop-3.3.0

Update mode

- Extract and replace files
- Extract and update files
- Fresh existing files only

Overwrite mode

- Ask before overwrite
- Overwrite without prompt
- Skip existing files
- Rename automatically

Miscellaneous

- Keep broken files
- Display files in Explorer

Save settings

Display

New folder

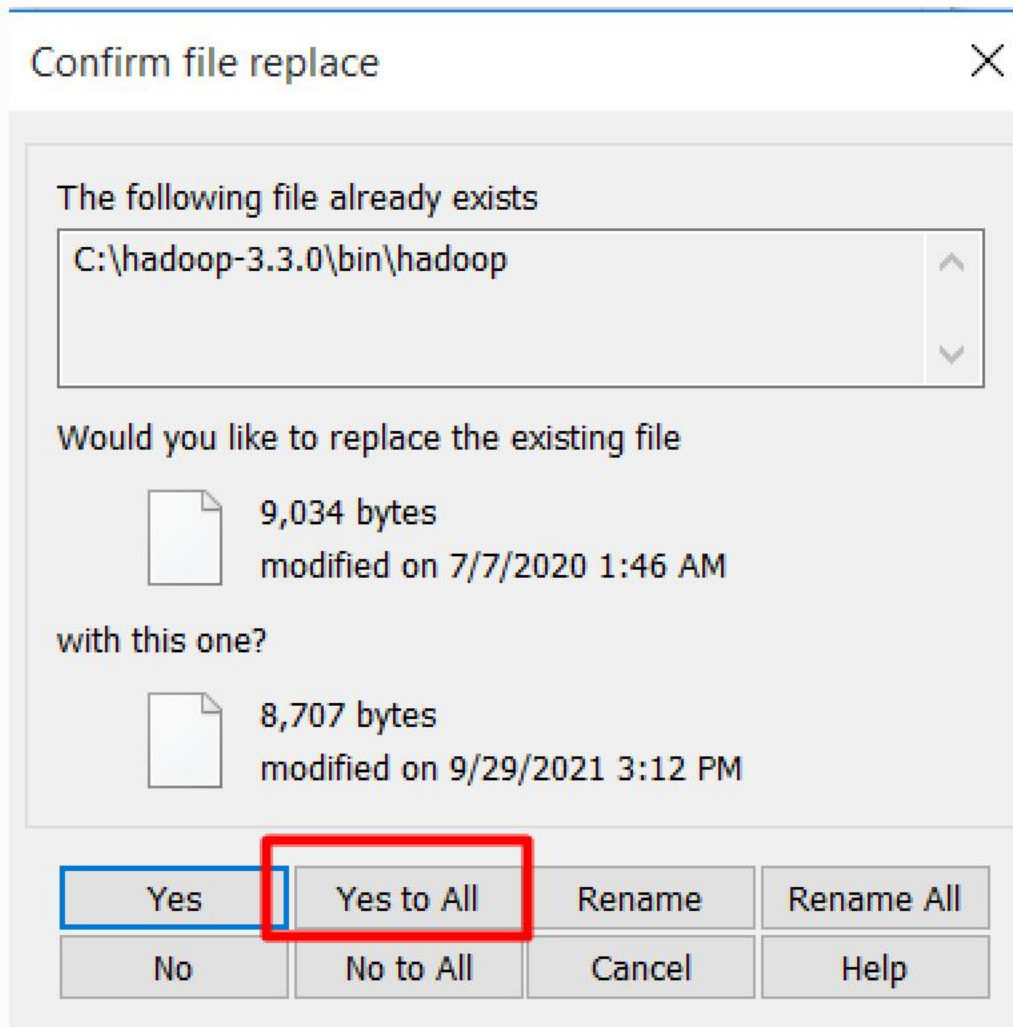
This PC

- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- Local Disk (C:)
 - \$WINDOWS.~BT
 - hadoop-3.3.0
 - Java
 - PerfLogs
 - Program Files
 - Program Files (x86)
 - ProgramData
 - tmp
 - Users
 - Windows
- DVD Drive (D:)

OK

Cancel

Help



- Sau đó format lại namenode và datanode, mở command line lên, gõ 2 lệnh sau (Bước format này chỉ cần làm 1 lần):

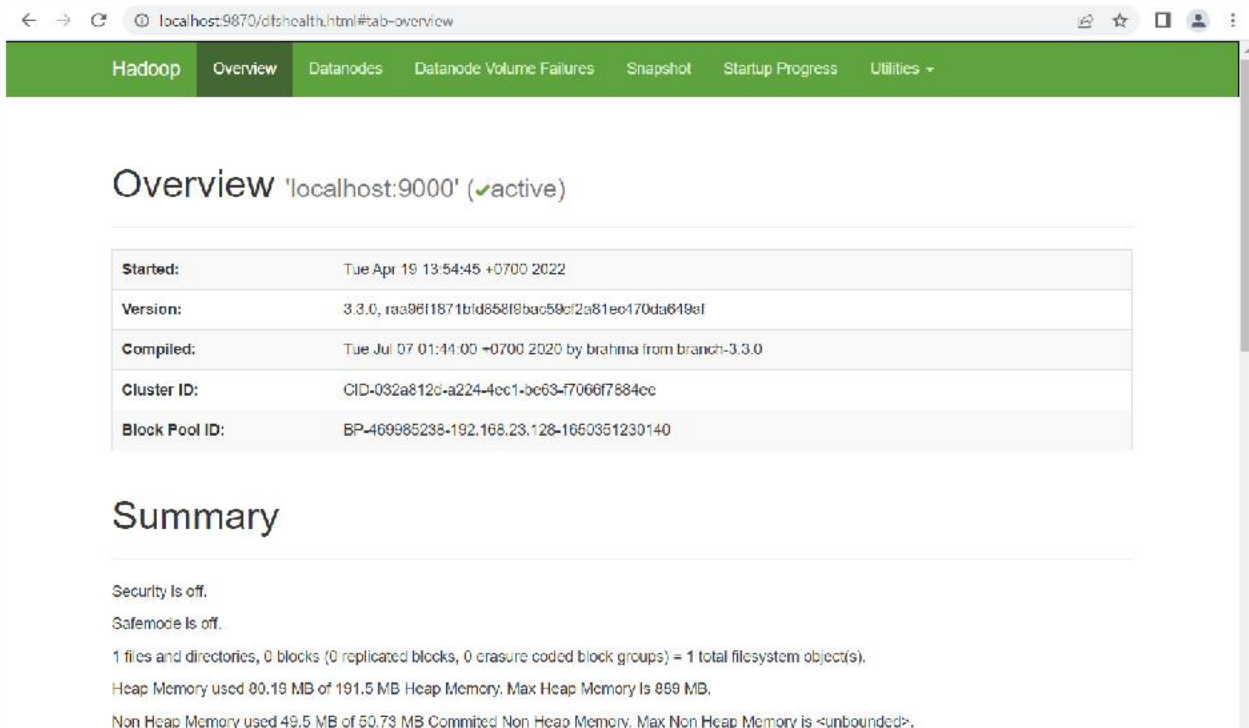
```
hdfs namenode -format  
hdfs datanode -format
```

- Tiếp theo sao chép file: **C:\hadoop-3.3.0\share\hadoop\yarn\timelineservice\hadoop-yarn-server-timelineservice-3.3.0.jar** vào **C:\hadoop-3.3.0\share\hadoop\yarn\hadoop-yarn-server-timelineservice-3.3.0.jar**
- Quá trình cài đặt Hadoop đã hoàn tất, để chạy Hadoop di chuyển tới thư mục **C:\hadoop-3.3.0\sbin**. Gõ lệnh:

```
start-all
```

- 4 ứng dụng sẽ được chạy và phải đảm bảo các ứng dụng không bị kết thúc:

- Name node web page: <http://localhost:9870/>



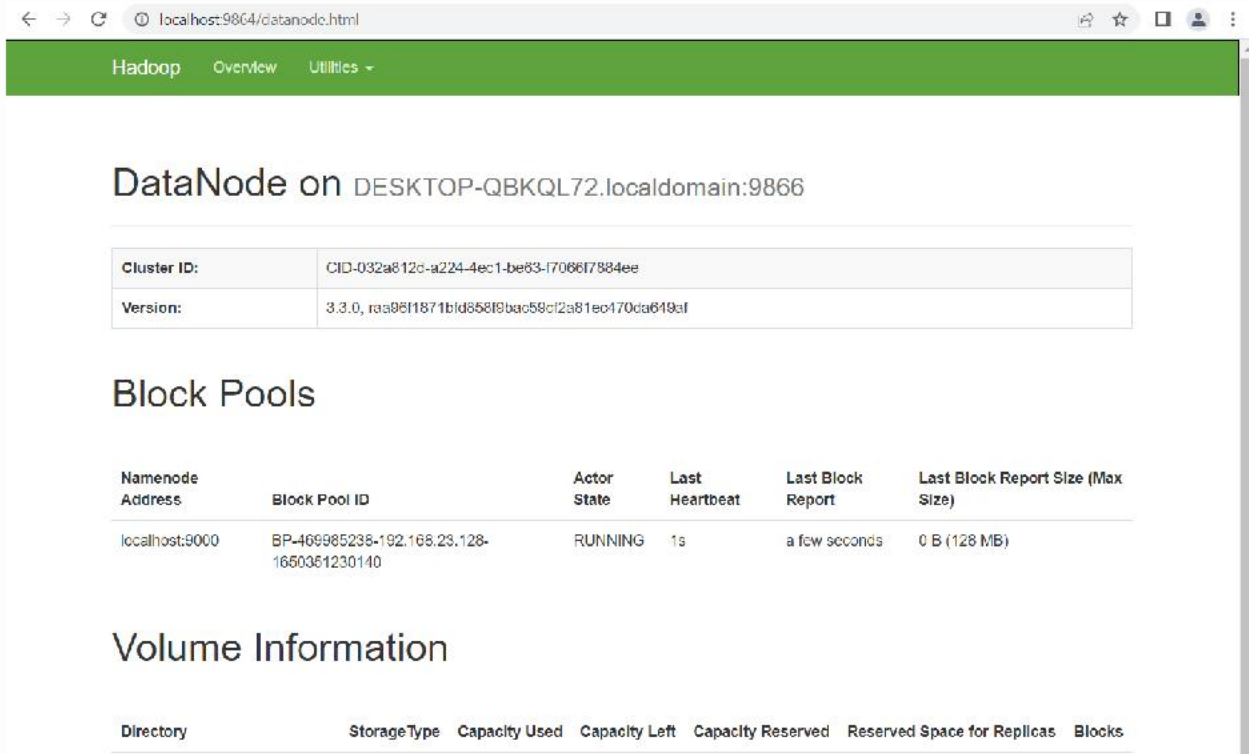
The screenshot shows the Hadoop Name Node Overview page for 'localhost:9000' (active). The page has a green navigation bar with 'Hadoop' and 'Overview' selected. Below the navigation bar, the title is 'Overview 'localhost:9000' (active)'. A table provides key information:

Started:	Tue Apr 19 13:54:45 +0700 2022
Version:	3.3.0, raa96f1871bid858f9bac58cf2a81e0470da649ef
Compiled:	Tue Jul 07 01:44:00 +0700 2020 by brahma from branch-3.3.0
Cluster ID:	CID-032a812c-a224-4ec1-be63-f7066f7884ee
Block Pool ID:	BP-469985238-192.168.23.128-1650351230140

Below the table is a 'Summary' section with the following text:

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 80.19 MB of 191.5 MB Heap Memory. Max Heap Memory Is 899 MB.
 Non Heap Memory used 49.5 MB of 50.73 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

- Data node web page: <http://localhost:9864/>



The screenshot shows the Hadoop Data Node page for 'DESKTOP-QBKQL72.localdomain:9866'. The page has a green navigation bar with 'Hadoop', 'Overview', and 'Utilities' visible. The title is 'DataNode on DESKTOP-QBKQL72.localdomain:9866'. A table provides key information:

Cluster ID:	CID-032a812c-a224-4ec1-be63-f7066f7884ee
Version:	3.3.0, raa96f1871bid858f9bac58cf2a81e0470da649ef

Below the table is a 'Block Pools' section with a table:

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-469985238-192.168.23.128-1650351230140	RUNNING	1s	a few seconds	0 B (128 MB)

Below the Block Pools table is a 'Volume Information' section with a table:

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
-----------	-------------	---------------	---------------	-------------------	-----------------------------	--------