

# HƯỚNG DẪN CÀI ĐẶT HADOOP 3.3.0 TRÊN WINDOWS SERVER

## 1. DOWNLOAD HADOOP 3.3.0 VÀ JDK

- Truy cập vào đường dẫn sau và download Hadoop 3.3.0:

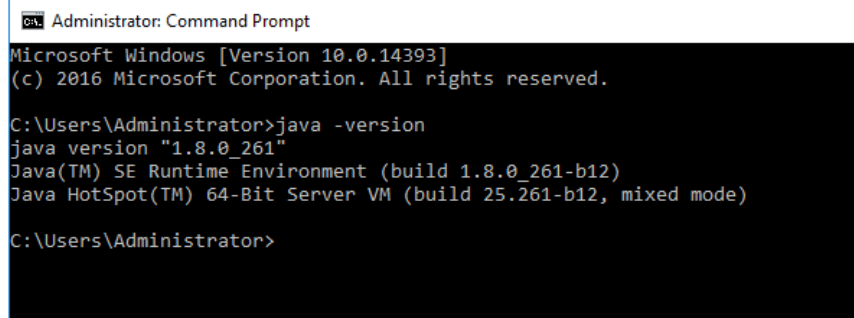
<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz>

- Kiểm tra Java đã được cài đặt chưa?

- Mở cửa sổ **cmd** (command prompt của windows), chạy lệnh sau:

```
java -version
```

- Nếu xuất hiện thông tin sau có nghĩa là máy đã có Java:



```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\Administrator>java -version
java version "1.8.0_261"
Java(TM) SE Runtime Environment (build 1.8.0_261-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.261-b12, mixed mode)

C:\Users\Administrator>
```














- Nếu máy chưa cài đặt Java, download JDK8 theo link sau:

<https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html>

- Chọn phiên bản 32 bit hoặc 64 bit tùy thuộc vào HĐH đang sử dụng:

#### Java SE Development Kit 8u261

This software is licensed under the Oracle Technology Network License Agreement for Oracle Java SE

Product / File Description	File Size	Download
Linux ARM 32 Hard Float ABI	73.4 MB	 <a href="#">jdk-8u261-linux-arm32-vfp-hflt.tar.gz</a>
Linux ARM 64 Hard Float ABI	70.3 MB	 <a href="#">jdk-8u261-linux-arm64-vfp-hflt.tar.gz</a>
Linux x86 RPM Package	121.92 MB	 <a href="#">jdk-8u261-linux-i586.rpm</a>
Linux x86 Compressed Archive	136.81 MB	 <a href="#">jdk-8u261-linux-i586.tar.gz</a>
Linux x64 RPM Package	121.53 MB	 <a href="#">jdk-8u261-linux-x64.rpm</a>
Linux x64 Compressed Archive	136.48 MB	 <a href="#">jdk-8u261-linux-x64.tar.gz</a>
macOS x64	203.94 MB	 <a href="#">jdk-8u261-macosx-x64.dmg</a>
Solaris SPARC 64-bit (SVR4 package)	125.77 MB	 <a href="#">jdk-8u261-solaris-sparcv9.tar.Z</a>
Solaris SPARC 64-bit	88.72 MB	 <a href="#">jdk-8u261-solaris-sparcv9.tar.gz</a>
Solaris x64 (SVR4 package)	134.23 MB	 <a href="#">jdk-8u261-solaris-x64.tar.Z</a>
Solaris x64	92.47 MB	 <a href="#">jdk-8u261-solaris-x64.tar.gz</a>
Windows x86	154.52 MB	 <a href="#">jdk-8u261-windows-i586.exe</a>
Windows x64	166.28 MB	 <a href="#">jdk-8u261-windows-x64.exe</a>

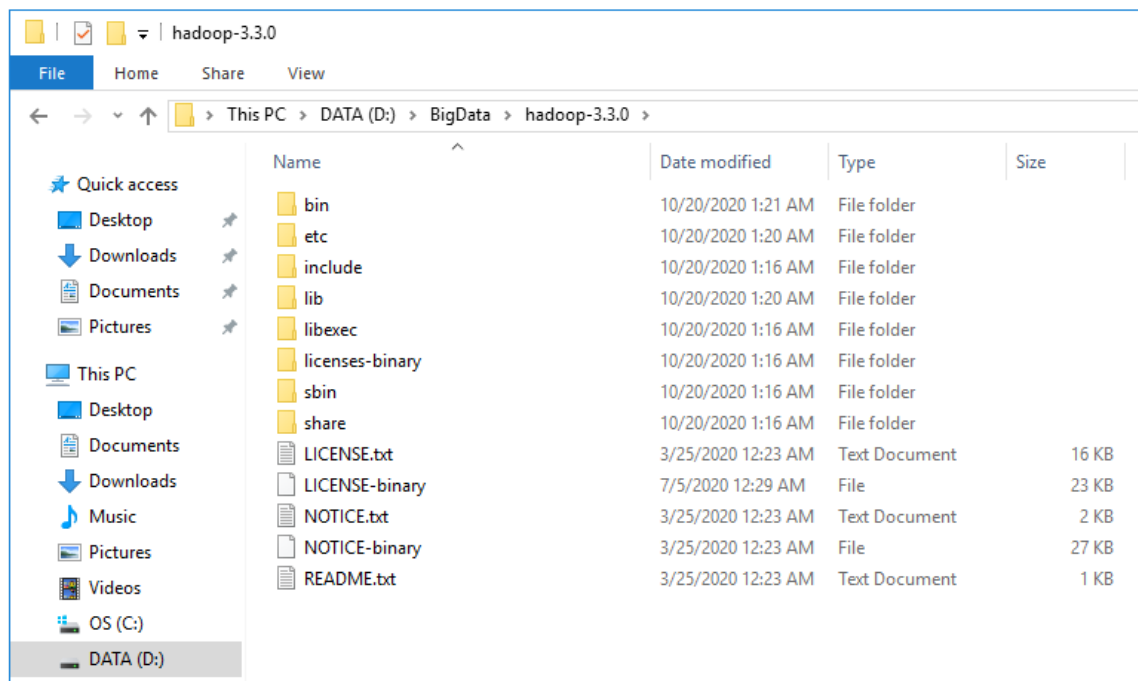
- Đường dẫn sau khi cài đặt như sau:

C:\Program Files\Java\jdk1.8.0\_261

## 2. SETUP ĐƯỜNG DẪN VÀ CÀI ĐẶT BIẾN MÔI TRƯỜNG

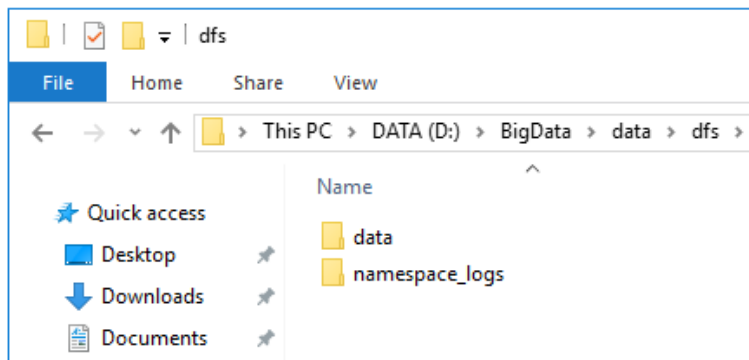
- Giải nén file ***hadoop-3.3.0.tar.gz*** đã download tại đường dẫn:

D:\BigData\hadoop-3.3.0



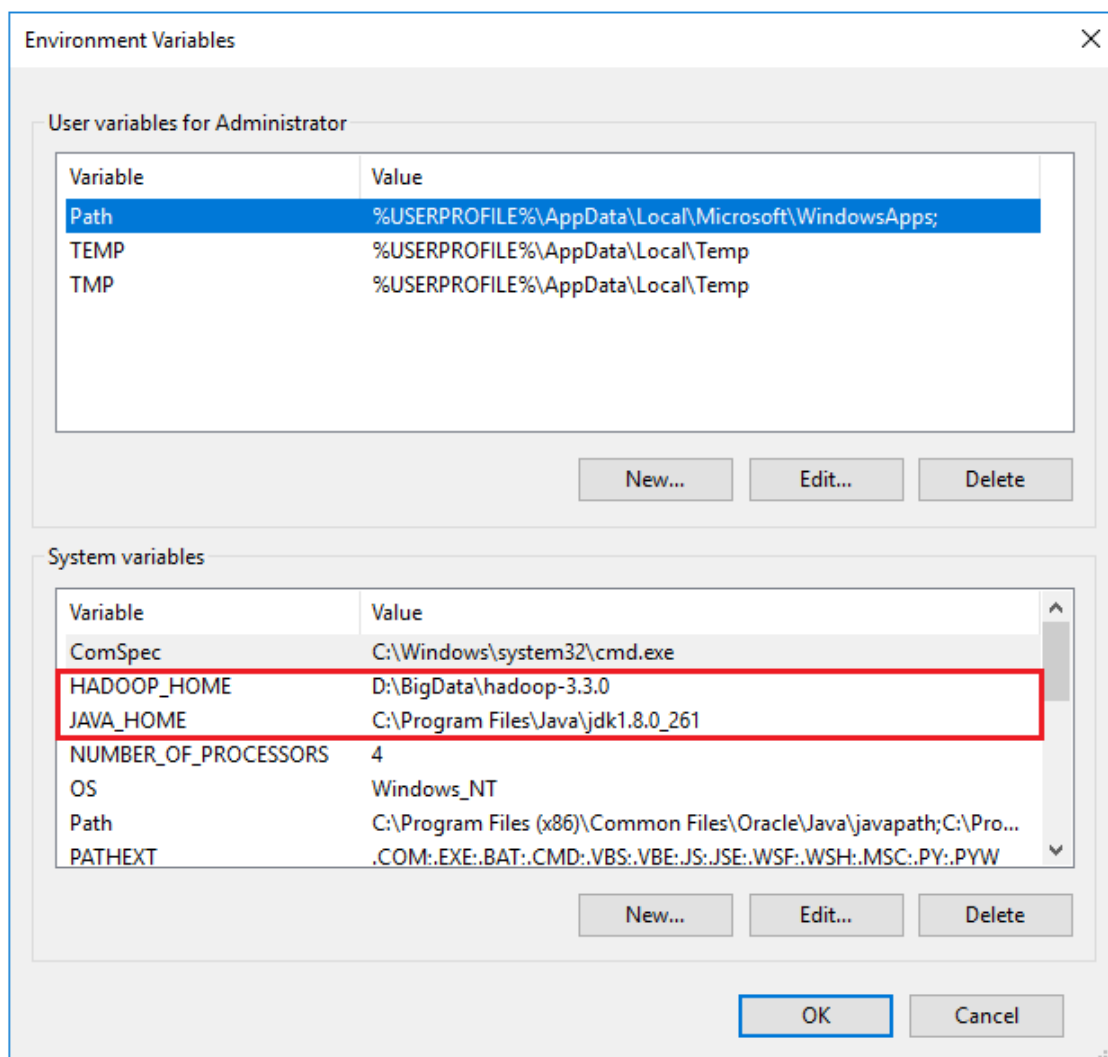
- Tạo 2 đường dẫn như sau:

D:\BigData\data\dfs\namespace\_logs  
D:\BigData\data\dfs\data



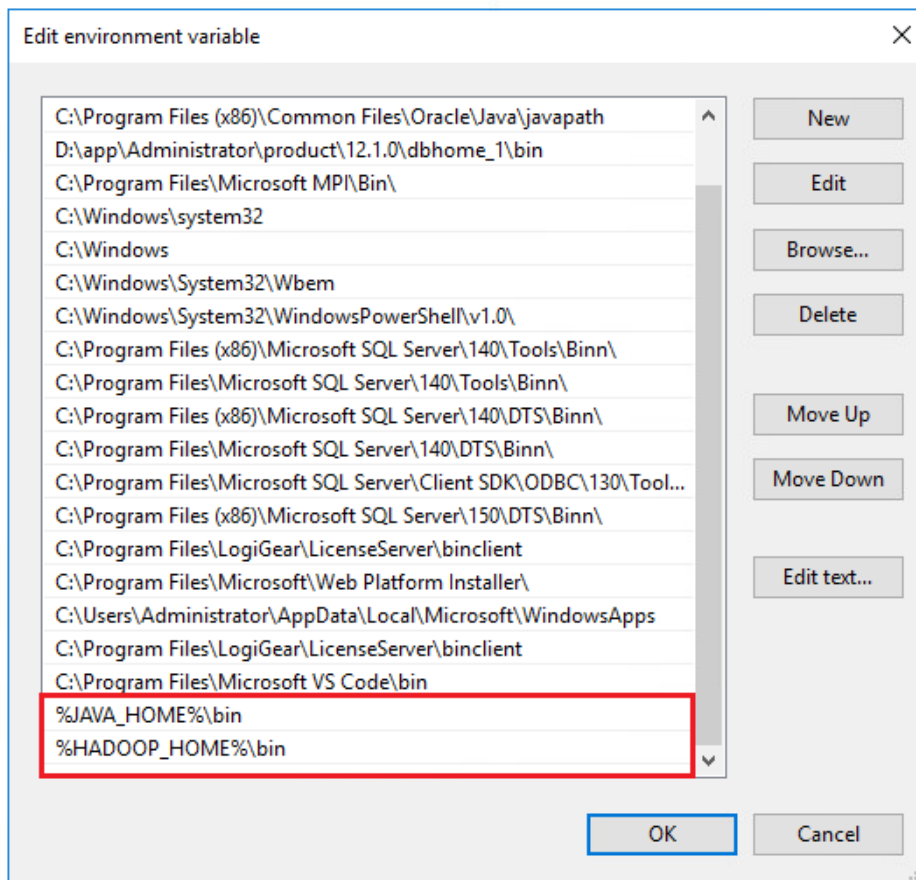
- Cài đặt biến môi trường:

HADOOP\_HOME="D:\BigData\hadoop-3.3.0"  
JAVA\_HOME="C:\Program Files\Java\jdk1.8.0\_261"



- Cập nhật thêm 2 thông tin vào biến môi trường PATH

```
%JAVA_HOME%\bin
%HADOOP_HOME%\bin
```



### 3. CÀI ĐẶT HADOOP

Trong thư mục ***hadoop-3.3.0*** đã giải nén, cần cấu hình 5 files cần thiết sau:

- ✓ *hadoop-env.cmd*
- ✓ *core-site.xml*
- ✓ *hdfs-site.xml*
- ✓ *mapred-site.xml*
- ✓ *yarn-site.xml*

Tất cả 5 files (GVHD cung cấp cho HV) đều đặt trong đường dẫn:

D:\BigData\hadoop-3.3.0\etc\hadoop


- Cấu hình ***hadoop-env.cmd***: Khai báo set JAVA\_HOME theo đường dẫn đã cài đặt JDK ở bước trên:

```
core-site.xml x hdfs-site.xml x mapred-site.xml x yarn-site.xml x hadoop-env.cmd x
1 @echo off
2 @rem Licensed to the Apache Software Foundation (ASF) under one or more
3 @rem contributor license agreements. See the NOTICE file distributed with
4 @rem this work for additional information regarding copyright ownership.
5 @rem The ASF licenses this file to You under the Apache License, Version 2.0
6 @rem (the "License"); you may not use this file except in compliance with
7 @rem the License. You may obtain a copy of the License at
8 @rem
9 @rem http://www.apache.org/licenses/LICENSE-2.0
10 @rem
11 @rem Unless required by applicable law or agreed to in writing, software
12 @rem distributed under the License is distributed on an "AS IS" BASIS,
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:\Progra~1\Java\jdk1.8.0_261
26
27 @rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
28 @rem set JSVC_HOME=%JSVC_HOME%
```

- Cấu hình ***core-site.xml***

```
core-site.xml x hdfs-site.xml x mapred-site.xml x yarn-site.xml x hadoop-env.cmd x
1 <?xml version="1.0" encoding="UTF-8"?>
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>fs.default.name</name>
22 <value>hdfs://0.0.0.0:19000</value>
23 </property>
24 </configuration>
25
```

○ Cấu hình ***hdfs-site.xml***

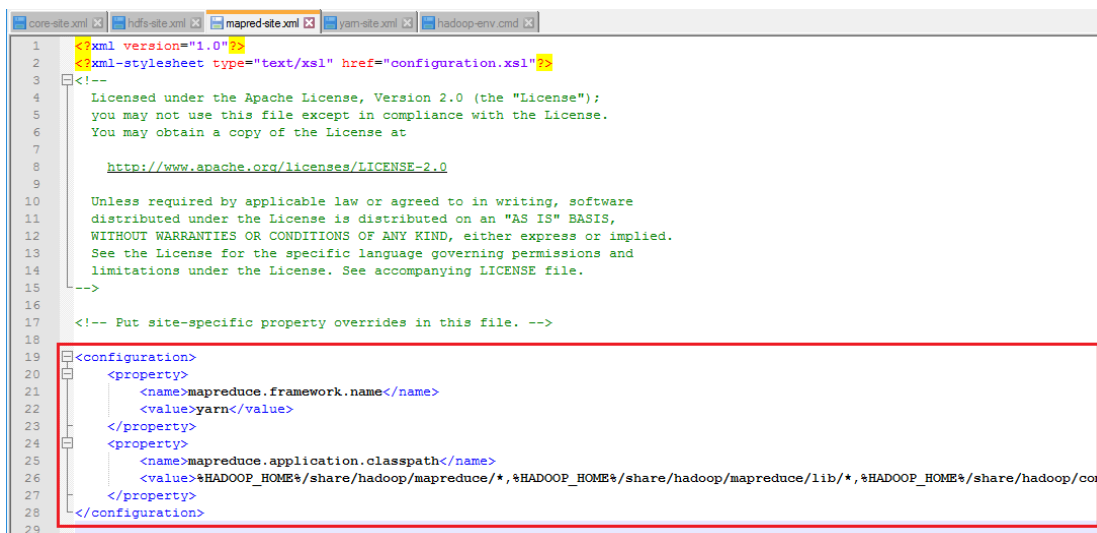


```

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
18     <!-- Put site-specific property overrides in this file. -->
19
20     <configuration>
21       <property>
22         <name>dfs.permissions</name>
23         <value>>false</value>
24       </property>
25       <property>
26         <name>dfs.replication</name>
27         <value>1</value>
28       </property>
29       <property>
30         <name>dfs.namenode.name.dir</name>
31         <value>file:///D:/BigData/data/dfs/namespace_logs</value>
32       </property>
33       <property>
34         <name>dfs.datanode.data.dir</name>
35         <value>file:///D:/BigData/data/dfs/data</value>
36       </property>
37     </configuration>

```

○ Cấu hình ***mapred-site.xml***: Nếu tìm không thấy mapred-site.xml thì tìm file mapred-site.xml.template, sau đó đổi tên lại thành mapred-site.xml

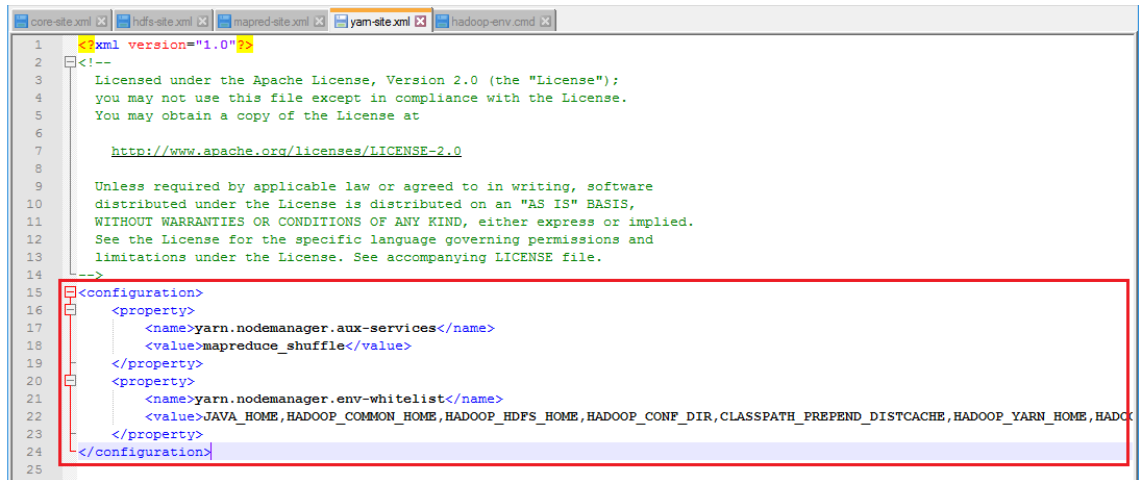


```

1      <?xml version="1.0"?>
2      <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3
4      <!--
5      Licensed under the Apache License, Version 2.0 (the "License");
6      you may not use this file except in compliance with the License.
7      You may obtain a copy of the License at
8
9      http://www.apache.org/licenses/LICENSE-2.0
10
11     Unless required by applicable law or agreed to in writing, software
12     distributed under the License is distributed on an "AS IS" BASIS,
13     WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14     See the License for the specific language governing permissions and
15     limitations under the License. See accompanying LICENSE file.
16
17     -->
18
19     <!-- Put site-specific property overrides in this file. -->
20
21     <configuration>
22       <property>
23         <name>mapreduce.framework.name</name>
24         <value>yarn</value>
25       </property>
26       <property>
27         <name>mapreduce.application.classpath</name>
28         <value>${HADOOP_HOME}/share/hadoop/mapreduce/*:${HADOOP_HOME}/share/hadoop/mapreduce/lib/*:${HADOOP_HOME}/share/hadoop/compat/*
29       </property>
30     </configuration>

```

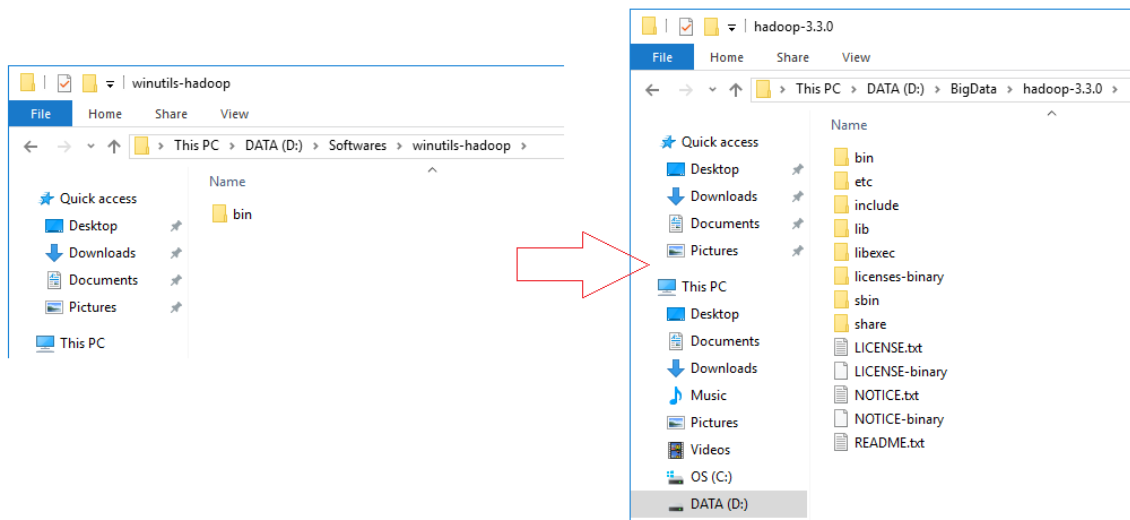
- Cấu hình **yarn-site.xml**



```
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.env-whitelist</name>
22 <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_NAMENODE_HOME,HADOOP_MAPRED_HOME</value>
23 </property>
24 </configuration>
25
```

## 4. FORMAT NAMENODE

- Giải nén tập tin **winutils-hadoop.zip** (GVHD cung cấp), sau đó chép thư mục **bin** trong thư mục vừa giải nén chép đè vào thư mục **bin** trong đường dẫn **D:\BigData\hadoop-3.3.0**:



- Mở cửa sổ cmd, gõ lệnh:

```
hadoop namenode -format
```



```
C:\Users\Administrator>hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
2020-10-20 01:35:49,250 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = WIN-5T3EUFBA83R/192.168.111.113
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.3.0
STARTUP_MSG: classpath = D:\BigData\hadoop-3.3.0\etc\hadoop;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\accessors-smart-1.2.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\asm-5.0.4.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\audience-annotations-0.5.0.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\checker-qual-2.5.2.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\commons-cli-1.2.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\commons-codec-1.11.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\commons-compress-1.19.jar;D:\BigData\hadoop-3.3.0\share\hadoop\common\lib\commons-io-2.5.jar;D:\BigData\hadoop-3.3.0\share\hadoop\com
```

- Chờ quá trình format hoàn thành.

## 5. KHỞI ĐỘNG HADOOP

- Truy cập đến đường dẫn **C:*|BigData|hadoop-3.3.0|sbin*** và gõ lệnh:

```
start-all.cmd
```

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

D:\BigData\hadoop-3.3.0\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\BigData\hadoop-3.3.0\sbin>
```

- Xuất hiện 4 hộp thoại command sau là dịch vụ đã chạy thành công:

```

Apache Hadoop Distribution - hadoop namenode
2020-10-20 01:38:32,153 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@720bf653{node/,file:///C:/Users/Administrator/AppData/Local/Temp/jetty-0_0_0-8042-_-any-8692132725154160344.dir/webapp/,AVAILABLE}{jar:file:/D:/BigData/hadoop-3.3.0/share/hadoop/yarn/hadoop-yarn-common-3.3.0.jar!/webapps/node}
2020-10-20 01:38:32,293 INFO server.AbstractConnector: Started ServerConnector@33617539{HTTP/1.1,[http/1.1]}{0.0.0.0:8042}
2020-10-20 01:38:32,293 INFO server.Server: Started @12161ms
2020-10-20 01:38:32,313 INFO webapp.WebApps: Web app node started at 8042
2020-10-20 01:38:32,318 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : WIN-5T3EUFBA83R:50237
2020-10-20 01:38:32,322 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2020-10-20 01:38:32,339 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2020-10-20 01:38:32,551 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []
2020-10-20 01:38:32,575 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers: []
2020-10-20 01:38:33,574 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 1475544316
2020-10-20 01:38:33,575 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id 1123922066
2020-10-20 01:38:33,575 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as WIN-5T3EUFBA83R:50237 with total resource of <memory:8192, vCores:8>

```

## 6. TRUY CẬP HADOOP SERVER BẰNG GIAO DIỆN GUI

- Từ máy client, chúng ta có thể truy cập đến máy Hadoop Server bằng trình duyệt theo cú pháp sau:

`http://IP:9870`

**IP:** IP của máy Hadoop Server.

**Port 9870:** Là port mặc định của để truy cập đến giao diện GUI của Hadoop Server.

Tham khảo các port khác tại link:

<https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-hdfs/hdfs-default.xml>

- Giao diện truy cập thành công sẽ có giao diện như sau:

## Overview '0.0.0.0:19000' (✓active)

Started:	Tue Oct 20 01:08:44 +0700 2020
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Tue Jul 07 01:44:00 +0700 2020 by brahma from branch-3.3.0
Cluster ID:	CID-7769a97b-9426-4348-955c-25183dc8da02
Block Pool ID:	BP-1151799433-192.168.111.113-1603130686368

## 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 111.14 MB of 237.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 60.84 MB of 62.61 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	60 GB
Configured Remote Capacity:	0 B
DFS Used:	324 B (0%)
Non DFS Used:	2.83 GB
DFS Remaining:	57.17 GB (95.28%)
Block Pool Used:	324 B (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)

- Truy cập đến trang upload file:

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse the file system

Logs

Log Level

Metrics

Configuration

Process Thread Dump

## Browse Directory

/

Show 25 entries

Search:

Permission Owner Group Size Last Modified Replication Block Size Name

No data available in table

Showing 0 to 0 of 0 entries

Previous Next

Hadoop, 2020.

- Truy cập đến file từ tại Hadoop Server theo cú pháp:

hdfs://**IP:PORT**

**IP.** IP của máy Hadoop Server.

**PORT:** Port đã cấu hình trên file *core-site.xml*. Cũng có thể xem tại giao diện GUI tại trang chủ



Overview '0.0.0.0:19000' (✓active)

Started:	Tue Oct 20 01:08:44 +0700 2020
Version:	3.3.0, raa96f1871bfd858f9bac59c2a81ec470da649af
Compiled:	Tue Jul 07 01:44:00 +0700 2020 by brahma from branch-3.3.0
Cluster ID:	CID-7769a97b-9426-4348-955c-25183dc8da02
Block Pool ID:	BP-1151799433-192.168.111.113-1603130686368

\_\_\_ HẾT \_\_\_