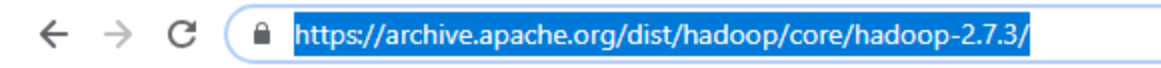


Cài đặt Hadoop trên windows








1. Chuẩn bị

- Tải file hadoop-2.7.3.tar từ link:














<https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/>



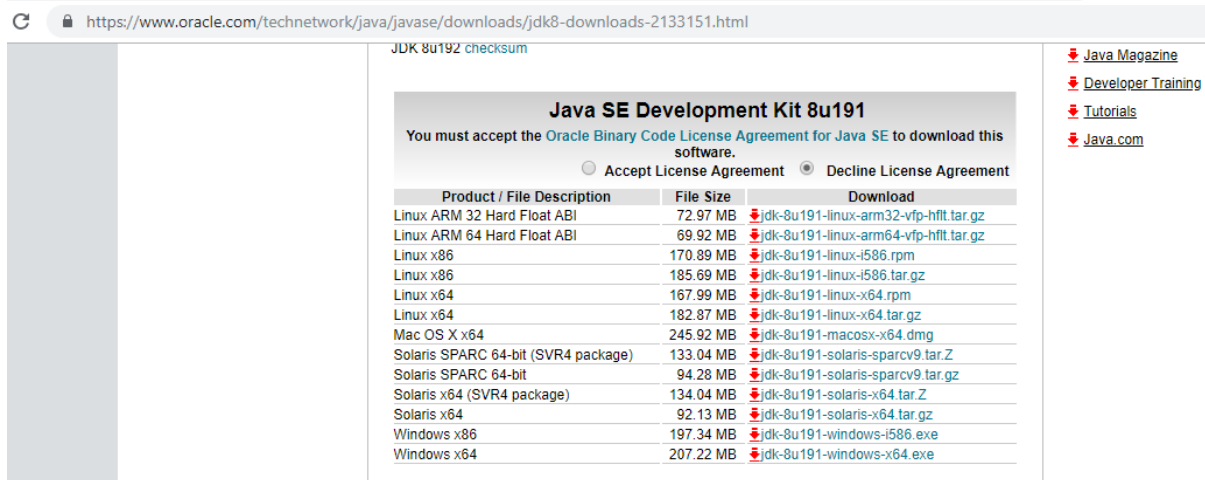
Index of /dist/hadoop/core/hadoop-2.7.3

Name	Last modified	Size	Description
 Parent Directory	-	-	
 hadoop-2.7.3-src.tar.gz	2016-08-25 19:25	17M	
 hadoop-2.7.3-src.tar.gz.asc	2016-08-25 19:25	521	
 hadoop-2.7.3-src.tar.gz.mds	2016-08-25 19:25	1.1K	
 hadoop-2.7.3.tar.gz	2016-08-25 19:25	204M	
 hadoop-2.7.3.tar.gz.asc	2016-08-25 19:25	521	
 hadoop-2.7.3.tar.gz.mds	2016-08-25 19:25	958	

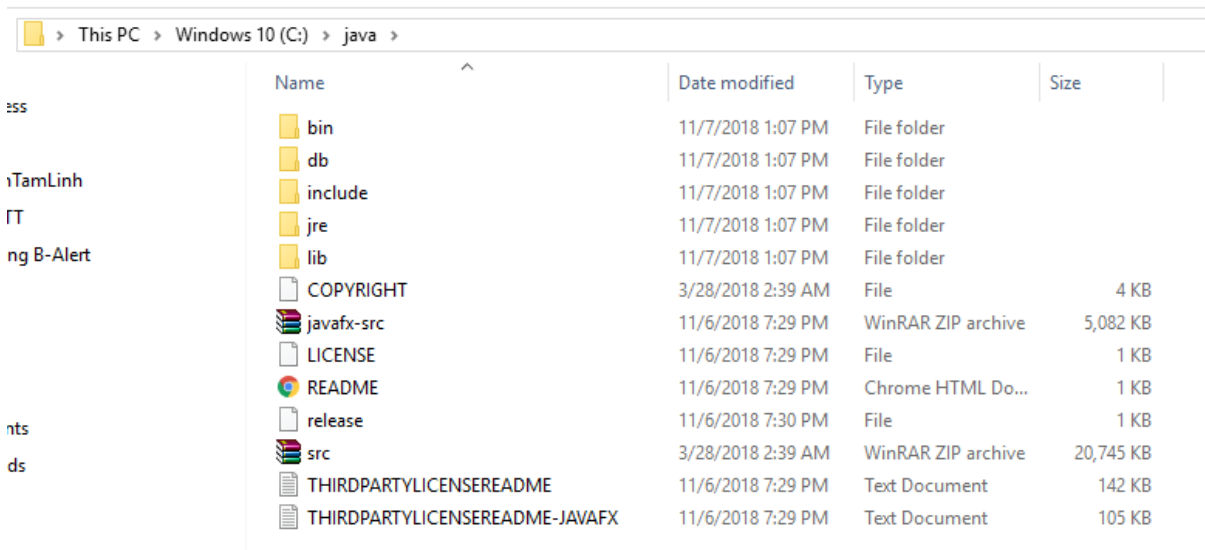
- Giải nén ra ổ C, ta được thư mục C:\hadoop-2.7.3 (chưa thư mục contrib và data):d

» This PC » Windows 10 (C:) » hadoop-2.7.3				
	Name	Date modified	Type	Size
iLinh	 bin	11/14/2018 4:35 PM	File folder	
	 contrib	11/14/2018 5:17 PM	File folder	
	 data	11/14/2018 4:21 PM	File folder	
	 etc	11/7/2018 3:07 PM	File folder	
	 include	11/7/2018 3:07 PM	File folder	
	 lib	11/7/2018 3:07 PM	File folder	
	 libexec	11/7/2018 3:07 PM	File folder	
	 logs	11/14/2018 4:49 PM	File folder	
	 sbin	11/7/2018 3:07 PM	File folder	
	 share	11/7/2018 3:08 PM	File folder	
-Alert	 LICENSE	8/17/2016 6:49 PM	Text Document	83 KB
	 NOTICE	8/17/2016 6:49 PM	Text Document	15 KB
	 README	8/17/2016 6:49 PM	Text Document	2 KB

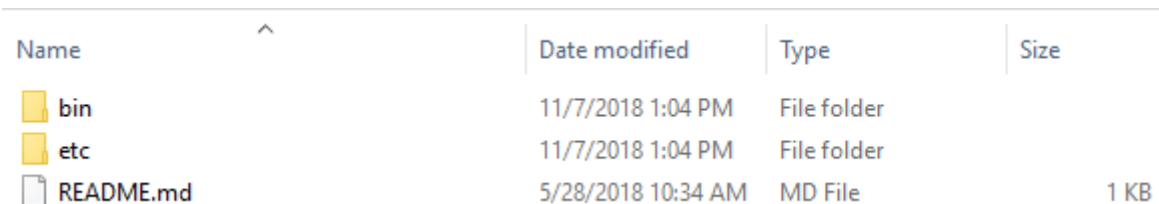
- Tải jdk, ví dụ file jdk-8u171-windows-x64 từ trang:
<https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>



























- Cài đặt jdk, ta được thư mục jdk1.8.0_171 trong C:\Program Files\Java
- Vì Hadoop không sử dụng đường dẫn có dấu cách nên ta chép thư mục jdk1.8.0_171 ra ổ ngoài C và rename thành java
















- Tải file hadooponwindows-master.zip và giải nén, ra thư mục có nội dung:



Nội dung thư mục bin:

Name	Date modified	Type	Size
 container-executor	5/28/2018 10:34 AM	File	157 KB
 hadoop	5/28/2018 10:34 AM	File	7 KB
 hadoop	5/28/2018 10:34 AM	Windows Comma...	9 KB
 hadoop.dll	5/28/2018 10:34 AM	Application extens	84 KB
 hadoop	5/28/2018 10:34 AM	Exports Library File	17 KB
 hadoop	5/28/2018 10:34 AM	Object File Library	29 KB
 hadoop	5/28/2018 10:34 AM	Program Debug D...	475 KB
 hdfs	5/28/2018 10:34 AM	File	12 KB
 hdfs	5/28/2018 10:34 AM	Windows Comma...	7 KB
 hdfs.dll	5/28/2018 10:34 AM	Application extens	59 KB
 hdfs	5/28/2018 10:34 AM	Exports Library File	9 KB
 hdfs	5/28/2018 10:34 AM	Object File Library	15 KB
 hdfs	5/28/2018 10:34 AM	Program Debug D...	452 KB
 hdfs_static	5/28/2018 10:34 AM	Object File Library	341 KB
 hdfs_static	5/28/2018 10:34 AM	Program Debug D...	92 KB
 libwinutils	5/28/2018 10:34 AM	Object File Library	1,211 KB
 mapred	5/28/2018 10:34 AM	File	6 KB
 mapred	5/28/2018 10:34 AM	Windows Comma...	6 KB
 rcc	5/28/2018 10:34 AM	File	2 KB
 test-container-executor	5/28/2018 10:34 AM	File	200 KB
 winutils	5/28/2018 10:34 AM	Application	107 KB
 winutils	5/28/2018 10:34 AM	Program Debug D...	875 KB
 yarn	5/28/2018 10:34 AM	File	13 KB
 yarn	5/28/2018 10:34 AM	Windows Comma...	11 KB


- Chép tất cả các file trong thư mục bin vào thư mục “C:\hadoop-2.7.3\bin”
- Tải file hadoop-eclipse-plugin-2.6.0.jar về máy
- Tạo thư mục contrib trong thư mục “C:\hadoop-2.7.3”

Windows 10 (C:) > hadoop-2.7.3				
Name	^	Date modified	Type	Size
 bin		11/14/2018 4:35 PM	File folder	
 contrib		11/14/2018 5:17 PM	File folder	
 data		11/14/2018 4:21 PM	File folder	
 etc		11/7/2018 3:07 PM	File folder	
 include		11/7/2018 3:07 PM	File folder	
 lib		11/7/2018 3:07 PM	File folder	
 libexec		11/7/2018 3:07 PM	File folder	
 logs		11/14/2018 4:49 PM	File folder	
 sbin		11/7/2018 3:07 PM	File folder	
 share		11/7/2018 3:08 PM	File folder	
 LICENSE		8/17/2016 6:49 PM	Text Document	83 KB
 NOTICE		8/17/2016 6:49 PM	Text Document	15 KB
 README		8/17/2016 6:49 PM	Text Document	2 KB

- Chép file hadoop-eclipse-plugin-2.6.0.jar vào thư mục “C:\hadoop-2.7.3\contrib”

re View

This PC > Windows 10 (C:) > hadoop-2.7.3 > contrib

Name	Date modified	Type	Size
 hadoop-eclipse-plugin-2.6.0.jar	12/13/2014 6:58 AM	JAR File	30,05!

- Tải file eclipse-jee-oxygen-3a-win32-x86_64.zip về máy
- Giải nén được thư mục eclipse có nội dung như sau:

Name	Date modified	Type	Size
configuration	11/19/2018 9:07 AM	File folder	
dropins	4/5/2018 11:14 AM	File folder	
features	4/5/2018 11:14 AM	File folder	
p2	11/19/2018 9:06 AM	File folder	
plugins	11/19/2018 9:03 AM	File folder	
readme	4/5/2018 11:14 AM	File folder	
.eclipseproduct	12/22/2017 2:01 AM	ECLIPSEPRODUCT...	1 KB
artifacts	4/5/2018 11:14 AM	XML Document	292 KB
eclipse	4/5/2018 11:16 AM	Application	313 KB
eclipse	11/19/2018 9:05 AM	Configuration sett...	1 KB
eclipsec	4/5/2018 11:16 AM	Application	26 KB

- Chép file `hadoop-eclipse-plugin-2.6.0.jar` vào thư mục “`..\eclipse\plugins`”

Name	Date modified	Type	Size
hadoop-eclipse-plugin-2.6.0.jar	12/13/2014 6:58 AM	JAR File	30,059 KB
java_cup.runtime_0.10.0.v201005080400.jar	4/5/2018 11:14 AM	JAR File	17 KB
javaewah_1.1.6.v20160919-1400.jar	4/5/2018 11:14 AM	JAR File	180 KB
javax.activation_1.1.0.v201211130549.jar	4/5/2018 11:14 AM	JAR File	54 KB
javax.annotation_1.2.0.v201602091430.jar	4/5/2018 11:14 AM	JAR File	29 KB
javax.el_2.2.0.v201303151357.jar	4/5/2018 11:14 AM	JAR File	55 KB
javax.inject_1.0.0.v20091030.jar	4/5/2018 11:14 AM	JAR File	17 KB
javax.jws_2.0.0.v201005080400.jar	4/5/2018 11:14 AM	JAR File	22 KB

-

2. Cấu hình

- Vào thư mục `C:\hadoop-2.7.3\etc\hadoop`
- Mở file `core-site.xml`, cấu hình như sau:

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:19000</value>
  </property>
</configuration>
```

- Mở file `mapred-site.xml`, cấu hình như sau (trong đó Admin là tên người dùng windows đang đăng nhập, tên này không có dấu cách):

```
<configuration>
```

```

<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
<property>
  <name>mapreduce.job.user.name</name>
  <value>Admin</value>
</property>
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
<property>
  <name>yarn.apps.stagingDir</name>
  <value>/user/Admin/staging</value>
</property>
</configuration>

```

- Mở file yarn-site.xml, cấu hình như sau:

```

<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-
services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>

```

- Mở file hdfs-site.xml, cấu hình như sau:

```

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>

```

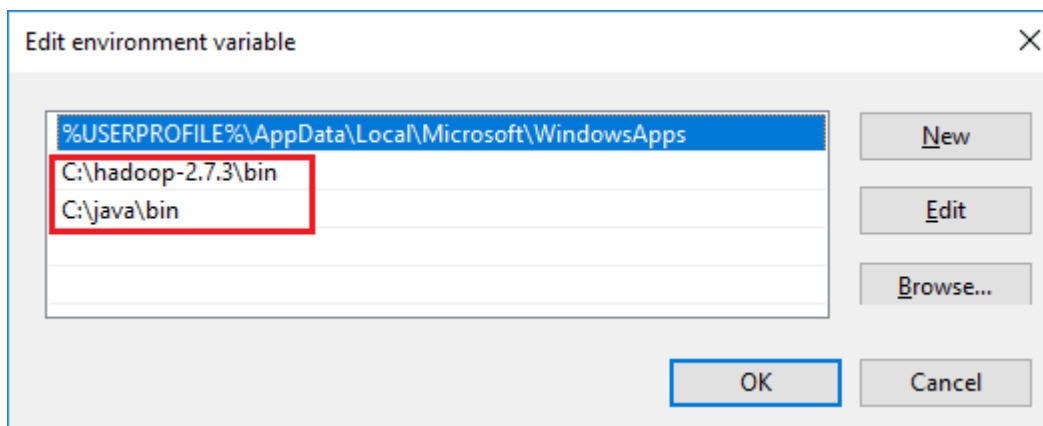
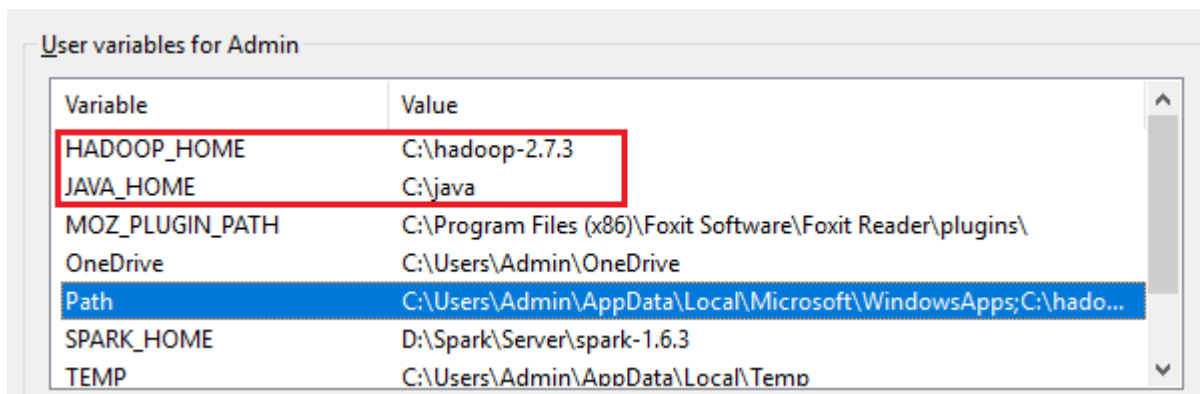
```

    <value>file:///C:/hadoop-
2.7.3/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///C:/hadoop-
2.7.3/data/datanode</value>
  </property>

```

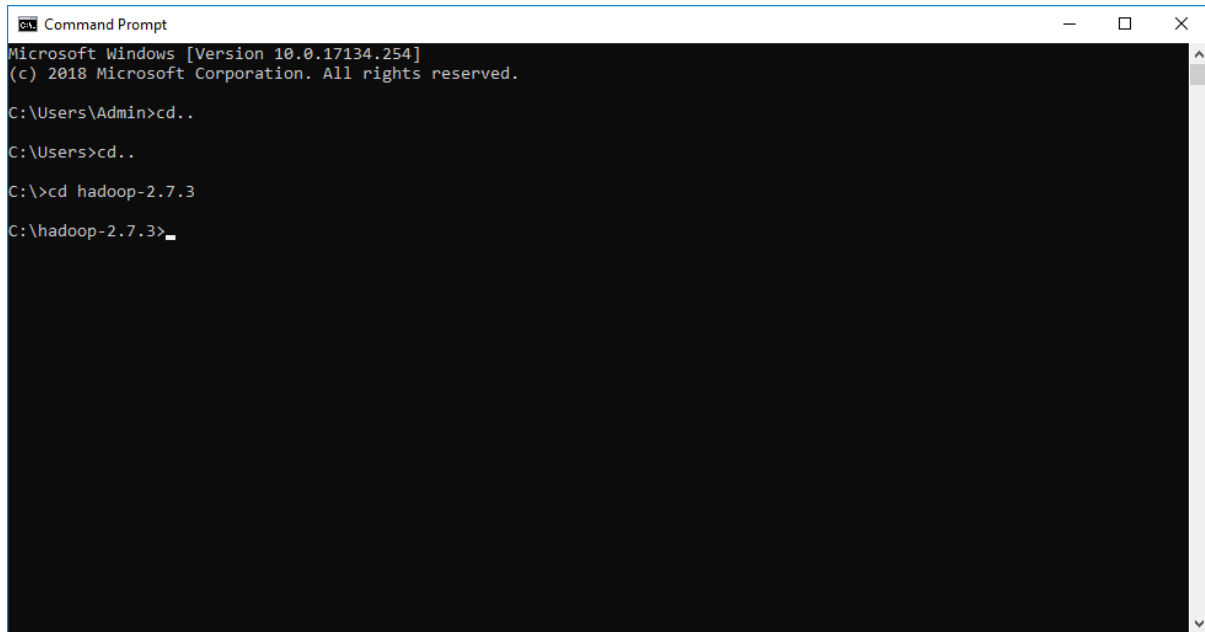
- Mở file `hadoop-env.cmd`, cấu hình như sau:

- Tìm và che dòng có nội dung: `set JAVA_HOME=%JAVA_HOME% (@rem set JAVA_HOME=%JAVA_HOME%)`
- Bổ sung dòng: `set JAVA_HOME=C:\java`
- Bổ sung 2 biến môi trường của Windows và biến Path



3. Kiểm tra việc cài đặt và cấu hình

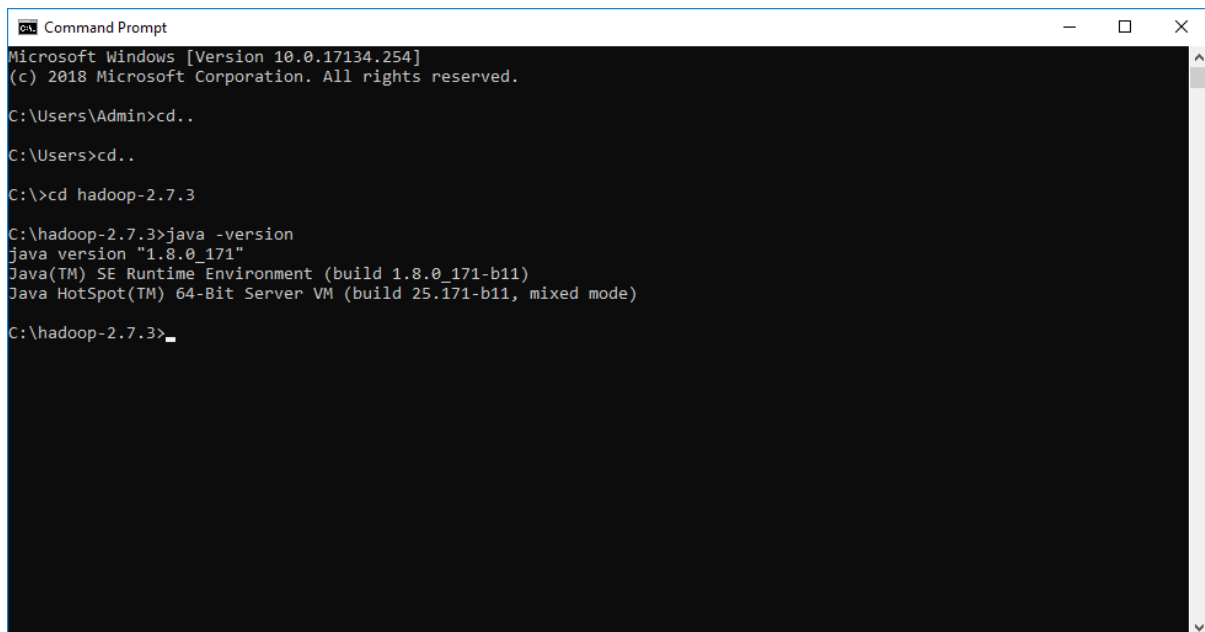
- Chạy `cmd.exe`, vào thư mục `"C:\hadoop-2.7.3"`



```
Command Prompt
Microsoft Windows [Version 10.0.17134.254]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Admin>cd..
C:\Users>cd..
C:\>cd hadoop-2.7.3
C:\hadoop-2.7.3>
```

- Gõ lệnh `java -version` để kiểm tra phiên bản cài đặt java



```
Command Prompt
Microsoft Windows [Version 10.0.17134.254]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Admin>cd..
C:\Users>cd..
C:\>cd hadoop-2.7.3
C:\hadoop-2.7.3>java -version
java version "1.8.0_171"
Java(TM) SE Runtime Environment (build 1.8.0_171-b11)
Java HotSpot(TM) 64-Bit Server VM (build 25.171-b11, mixed mode)
C:\hadoop-2.7.3>
```

- Làm rỗng các thư mục `"C:\hadoop\data\namenode"` và `"C:\hadoop\data\datanode"`: Nếu không rỗng thì xóa hết thư mục con và file trong đó
- Gõ lệnh `"hdfs namenode -format"`, có thể gặp lỗi:


```
C:\Windows\System32\cmd.exe
java.lang.RuntimeException: org.xml.sax.SAXParseException; systemId: file:/C:/hadoop-2.7.3/etc/hadoop/hdfs-site.xml; lineNumber: 32; columnNumber: 1; The end-tag for element type "property" must end with a '>' delimiter.
    at org.apache.hadoop.conf.Configuration.loadResource(Configuration.java:2645)
    at org.apache.hadoop.conf.Configuration.loadResources(Configuration.java:2492)
    at org.apache.hadoop.conf.Configuration.getProps(Configuration.java:2405)
    at org.apache.hadoop.conf.Configuration.set(Configuration.java:1143)
    at org.apache.hadoop.conf.Configuration.set(Configuration.java:1115)
    at org.apache.hadoop.conf.Configuration.setBoolean(Configuration.java:1451)
    at org.apache.hadoop.util.GenericOptionsParser.processGeneralOptions(GenericOptionsParser.java:321)
    at org.apache.hadoop.util.GenericOptionsParser.parseGeneralOptions(GenericOptionsParser.java:487)
    at org.apache.hadoop.util.GenericOptionsParser.<init>(GenericOptionsParser.java:170)
    at org.apache.hadoop.util.GenericOptionsParser.<init>(GenericOptionsParser.java:153)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1422)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1550)
Caused by: org.xml.sax.SAXParseException; systemId: file:/C:/hadoop-2.7.3/etc/hadoop/hdfs-site.xml; lineNumber: 32; columnNumber: 1; The end-tag for element type "property" must end with a '>' delimiter.
    at org.apache.xerces.parsers.DOMParser.parse(Unknown Source)
    at org.apache.xerces.jaxp.DocumentBuilderImpl.parse(Unknown Source)
    at javax.xml.parsers.DocumentBuilder.parse(DocumentBuilder.java:150)
    at org.apache.hadoop.conf.Configuration.parse(Configuration.java:2480)
    at org.apache.hadoop.conf.Configuration.parse(Configuration.java:2468)
    at org.apache.hadoop.conf.Configuration.loadResource(Configuration.java:2539)
    ... 11 more
18/05/28 16:00:25 INFO util.ExitUtil: Exiting with status 1
18/05/28 16:00:25 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at rum-pc/192.168.0.2
*****/
C:\hadoop-2.7.3>
```

- Kiểm tra thông tin trong hình chữ nhật dài màu đỏ, trường hợp trên cần kiểm tra lại file “hdfs-site.xml” xem lỗi ký tự ‘>’ như thế nào
- Nếu không lỗi hoặc đã sửa lỗi thì nội dung trong hình chữ ngắn màu đỏ là “Exiting with status 0”

```
Command Prompt
18/05/28 16:01:32 INFO util.GSet: VM type = 64-bit
18/05/28 16:01:32 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
18/05/28 16:01:32 INFO util.GSet: capacity = 2^18 = 262144 entries
18/05/28 16:01:32 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
18/05/28 16:01:32 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
18/05/28 16:01:32 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
18/05/28 16:01:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
18/05/28 16:01:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
18/05/28 16:01:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
18/05/28 16:01:32 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
18/05/28 16:01:32 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
18/05/28 16:01:32 INFO util.GSet: Computing capacity for map NameNodeRetryCache
18/05/28 16:01:32 INFO util.GSet: VM type = 64-bit
18/05/28 16:01:32 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
18/05/28 16:01:32 INFO util.GSet: capacity = 2^15 = 32768 entries
18/05/28 16:01:32 INFO namenode.FSImage: Allocated new BlockPoolId: BP-437396816-192.168.0.2-1527498092168
18/05/28 16:01:32 INFO common.Storage: Storage directory C:\hadoop-2.7.3\data\namenode has been successfully formatted.
18/05/28 16:01:32 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop-2.7.3\data\namenode\current\fsimage.ckpt_000000000000000000 using no compression
18/05/28 16:01:32 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop-2.7.3\data\namenode\current\fsimage.ckpt_000000000000000000 of size 352 bytes saved in 0 seconds.
18/05/28 16:01:32 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
18/05/28 16:01:32 INFO util.ExitUtil: Exiting with status 0
18/05/28 16:01:32 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at rum-pc/192.168.0.2
*****/
C:\Users\RumIT>
```

- Nếu gõ lệnh trên lần 2, có thể có thông báo (vì đã format rồi, gõ Y Enter)

```
Command Prompt - hdfs namenode -format
18/11/17 14:17:11 INFO namenode.FSNamesystem: HA Enabled: false
18/11/17 14:17:11 INFO namenode.FSNamesystem: Append Enabled: true
18/11/17 14:17:13 INFO util.GSet: Computing capacity for map INodeMap
18/11/17 14:17:13 INFO util.GSet: VM type = 64-bit
18/11/17 14:17:13 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
18/11/17 14:17:13 INFO util.GSet: capacity = 2^20 = 1048576 entries
18/11/17 14:17:13 INFO namenode.FSDirectory: ACLs enabled? false
18/11/17 14:17:13 INFO namenode.FSDirectory: XAttrs enabled? true
18/11/17 14:17:13 INFO namenode.FSDirectory: Maximum size of an xattr: 16384
18/11/17 14:17:13 INFO namenode.NameNode: Caching file names occurring more than 10 times
18/11/17 14:17:13 INFO util.GSet: Computing capacity for map cachedBlocks
18/11/17 14:17:13 INFO util.GSet: VM type = 64-bit
18/11/17 14:17:13 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
18/11/17 14:17:13 INFO util.GSet: capacity = 2^18 = 262144 entries
18/11/17 14:17:13 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
18/11/17 14:17:13 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
18/11/17 14:17:13 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
18/11/17 14:17:13 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
18/11/17 14:17:13 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
18/11/17 14:17:13 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
18/11/17 14:17:13 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
18/11/17 14:17:13 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
18/11/17 14:17:13 INFO util.GSet: Computing capacity for map NameNodeRetryCache
18/11/17 14:17:13 INFO util.GSet: VM type = 64-bit
18/11/17 14:17:13 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
18/11/17 14:17:13 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory C:\hadoop-2.7.3\data\namenode ? (Y or N) Y
```

4. Chạy Hadoop

- Vào thư mục “C:\hadoop-2.7.3\sbin”, gõ lệnh “start-all.cmd” và enter

```
SHUTDOWN_MSG: Shutting down NameNode at tutrung/172.22.176.1
*****/

C:\hadoop-2.7.3>cd sbin
C:\hadoop-2.7.3\sbin>start-all.cmd
```

```
Command Prompt
18/11/17 14:24:15 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
18/11/17 14:24:15 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
18/11/17 14:24:15 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
18/11/17 14:24:15 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
18/11/17 14:24:15 INFO util.GSet: Computing capacity for map NameNodeRetryCache
18/11/17 14:24:15 INFO util.GSet: VM type = 64-bit
18/11/17 14:24:15 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
18/11/17 14:24:15 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory C:\hadoop-2.7.3\data\namenode ? (Y or N) Y
18/11/17 14:24:23 INFO namenode.FSImage: Allocated new BlockPoolId: BP-87471494-172.22.176.1-1542493463152
18/11/17 14:24:23 INFO common.Storage: Storage directory C:\hadoop-2.7.3\data\namenode has been successfully formatted.
18/11/17 14:24:23 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop-2.7.3\data\namenode\current\fsimage.ckpt_000000000000000000 using no compression
18/11/17 14:24:23 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop-2.7.3\data\namenode\current\fsimage.ckpt_000000000000000000 of size 352 bytes saved in 0 seconds.
18/11/17 14:24:24 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
18/11/17 14:24:24 INFO util.ExitUtil: Exiting with status 0
18/11/17 14:24:24 INFO namenode.NameNode: SHUTDOWN_MSG:
*****/
SHUTDOWN_MSG: Shutting down NameNode at tutrung/172.22.176.1
*****/

C:\hadoop-2.7.3>cd sbin
C:\hadoop-2.7.3\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\hadoop-2.7.3\sbin>
```

- Một loạt cửa sổ cmd hiện ra chứng tỏ server đang chạy, nếu tắt đồng nghĩa với stop server

```
Apache Hadoop Distribution - hadoop namenode
    at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.startCommonServices(FSNamesystem.java:1058)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.startCommonServices(NameNode.java:679)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:665)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:812)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:796)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1493)
    at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1559)
18/11/17 14:33:27 WARN common.Util: Path C:\hadoop-2.7.3\data\namenode should be specified as a URI in configuration files. Please update hdfs configuration.
18/11/17 14:33:27 INFO namenode.LeaseManager: Number of blocks under construction: 0
18/11/17 14:33:27 INFO namenode.LeaseManager: Number of blocks under construction: 0
18/11/17 14:33:27 INFO namenode.FSNamesystem: initializing replication queues
18/11/17 14:33:27 INFO hdfs.StateChange: STATE* Leaving safe mode after 12 secs
18/11/17 14:33:27 INFO hdfs.StateChange: STATE* Network topology has 0 racks and 0 datanodes
18/11/17 14:33:27 INFO hdfs.StateChange: STATE* UnderReplicatedBlocks has 0 blocks
18/11/17 14:33:27 INFO blockmanagement.DatanodeDescriptor: Number of failed storage changes from 0 to 0
18/11/17 14:33:27 INFO blockmanagement.BlockManager: Total number of blocks = 0
18/11/17 14:33:27 INFO blockmanagement.BlockManager: Number of invalid blocks = 0
18/11/17 14:33:27 INFO blockmanagement.BlockManager: Number of under-replicated blocks = 0
18/11/17 14:33:27 INFO blockmanagement.BlockManager: Number of over-replicated blocks = 0
18/11/17 14:33:27 INFO blockmanagement.BlockManager: Number of blocks being written = 0
18/11/17 14:33:27 INFO hdfs.StateChange: STATE* Replication Queue initialization scan for invalid, over- and under-replicated blocks completed in 71 msec
18/11/17 14:33:27 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:19000
18/11/17 14:33:27 INFO namenode.FSNamesystem: Starting services required for active state
18/11/17 14:33:27 INFO ipc.Server: IPC Server Responder: starting
18/11/17 14:33:27 INFO ipc.Server: IPC Server listener on 19000: starting
18/11/17 14:33:27 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds

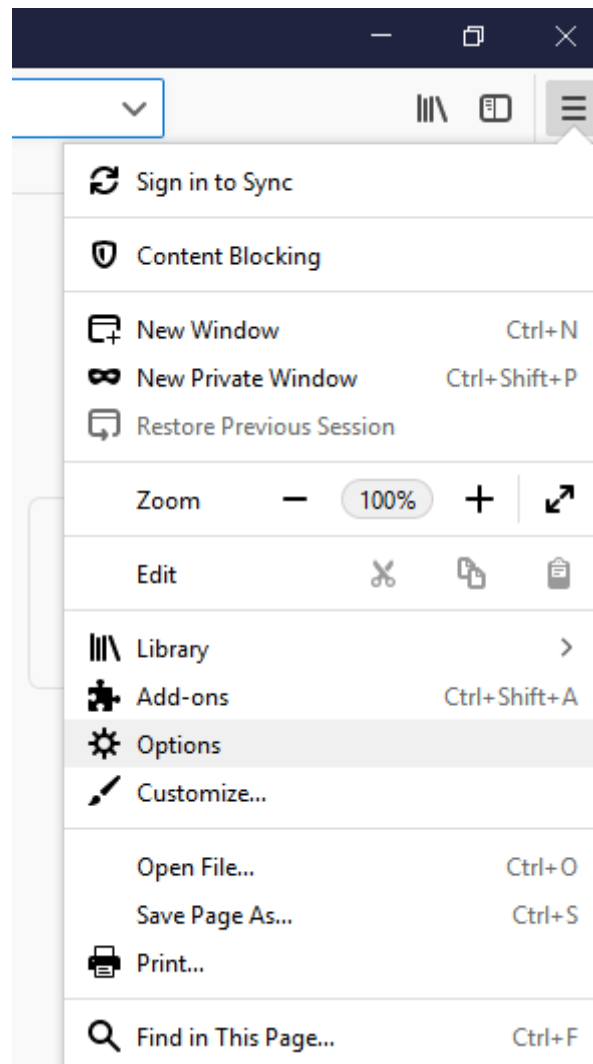
Apache Hadoop Distribution
    at org.apache.hadoop.hdfs.server.datanode.DataStorage.loadDataStorage(DataStorage.java:416)
    at org.apache.hadoop.hdfs.server.datanode.DataStorage.addStorageLocations(DataStorage.java:395)
    at org.apache.hadoop.hdfs.server.datanode.DataStorage.recoverTransitionRead(DataStorage.java:573)
    at org.apache.hadoop.hdfs.server.datanode.DataNode.initStorage(DataNode.java:1362)
    at org.apache.hadoop.hdfs.server.datanode.DataNode.initBlockPool(DataNode.java:1327)
    at org.apache.hadoop.hdfs.server.datanode.BPOfferService.verifyAndSetNamespaceInfo(BPOfferService.java:317)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.connectToNNAndHandshake(BPServiceActor.java:223)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.run(BPServiceActor.java:802)
    at java.lang.Thread.run(Thread.java:748)
18/11/17 14:33:30 FATAL datanode.DataNode: Initialization failed for Block pool <registering> (Datanode Uuid unassigned) service to localhost/127.0.0.1:19000. Exiting.
java.io.IOException: All specified directories are failed to load.
    at org.apache.hadoop.hdfs.server.datanode.DataStorage.recoverTransitionRead(DataStorage.java:574)
    at org.apache.hadoop.hdfs.server.datanode.DataNode.initStorage(DataNode.java:1362)
    at org.apache.hadoop.hdfs.server.datanode.DataNode.initBlockPool(DataNode.java:1327)
    at org.apache.hadoop.hdfs.server.datanode.BPOfferService.verifyAndSetNamespaceInfo(BPOfferService.java:317)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.connectToNNAndHandshake(BPServiceActor.java:223)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.run(BPServiceActor.java:802)
    at java.lang.Thread.run(Thread.java:748)
18/11/17 14:33:30 WARN datanode.DataNode: Ending block pool service for: Block pool <registering> (Datanode Uuid unassigned) service to localhost/127.0.0.1:19000
18/11/17 14:33:30 INFO datanode.DataNode: Removed Block pool <registering> (Datanode Uuid unassigned)
18/11/17 14:33:32 WARN datanode.DataNode: Exiting Datanode
18/11/17 14:33:32 INFO util.ExitUtil: Exiting with status 0
18/11/17 14:33:32 INFO datanode.DataNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down DataNode at tutrung/172.22.176.1
*****/
C:\hadoop-2.7.3\sbin>
```



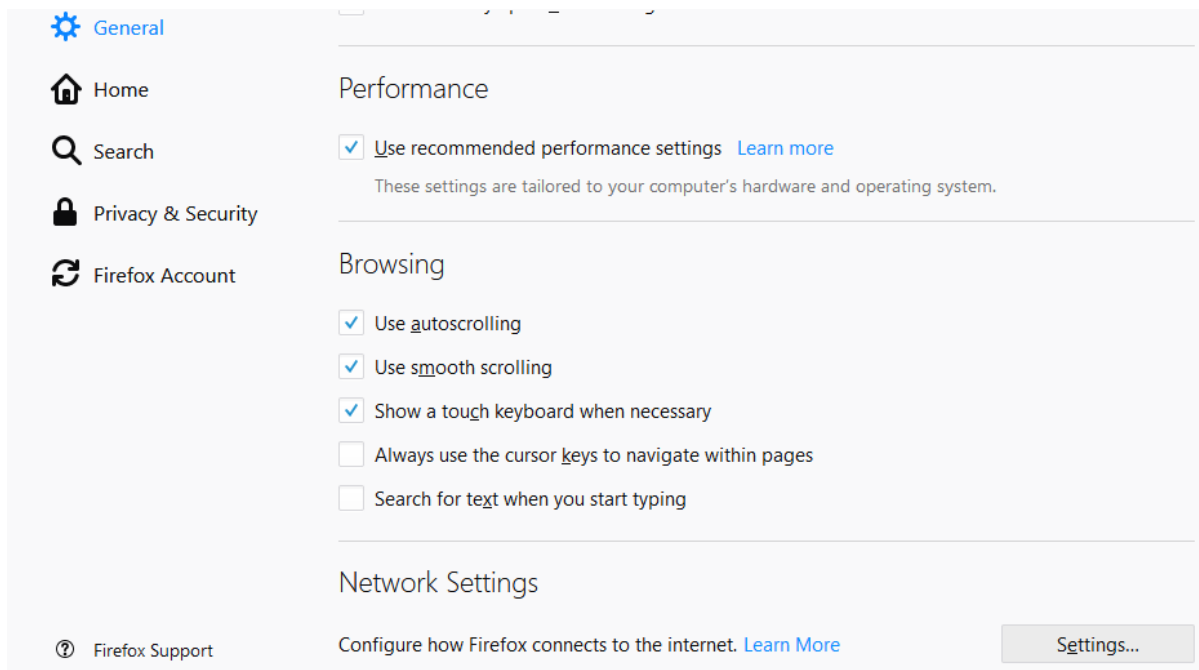
```
Apache Hadoop Distribution - yarn  resourcemanager
Nov 17, 2018 2:33:14 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.resourcemanager.webapp.RMWebServices as a root resource class
Nov 17, 2018 2:33:14 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
Nov 17, 2018 2:33:14 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.9 09/02/2011 11:17 AM'
Nov 17, 2018 2:33:14 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Nov 17, 2018 2:33:21 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Nov 17, 2018 2:33:26 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.RMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
18/11/17 14:33:26 INFO mortbay.log: Started HttpServer2$SelectChannelConnectorWithSafeStartup@0.0.0.0:8088
18/11/17 14:33:26 INFO webapp.WebApps: Web app cluster started at 8088
18/11/17 14:33:27 INFO ipc.CallQueueManager: Using callQueue class java.util.concurrent.LinkedBlockingQueue
18/11/17 14:33:27 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceManagerAdministrationProtocolPB to the server
18/11/17 14:33:27 INFO ipc.Server: IPC Server Responder: starting
18/11/17 14:33:27 INFO ipc.Server: IPC Server listener on 8033: starting
18/11/17 14:33:27 INFO ipc.Server: Starting Socket Reader #1 for port 8033
18/11/17 14:33:28 INFO util.RackResolver: Resolved tutrung.mshome.net to /default-rack
18/11/17 14:33:28 INFO resourcemanager.ResourceTrackerService: NodeManager from node tutrung.mshome.net(cmPort: 30016 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId tutrung.mshome.net:30016
18/11/17 14:33:28 INFO rmnode.RMNodeImpl: tutrung.mshome.net:30016 Node Transitioned from NEW to RUNNING
18/11/17 14:33:28 INFO capacity.CapacityScheduler: Added node tutrung.mshome.net:30016 clusterResource: <memory:8192, vCores:8>
```

```
Apache Hadoop Distribution - yarn  nodemanager
Nov 17, 2018 2:33:17 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices as a root resource class
Nov 17, 2018 2:33:17 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
Nov 17, 2018 2:33:17 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Nov 17, 2018 2:33:17 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.9 09/02/2011 11:17 AM'
Nov 17, 2018 2:33:17 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Nov 17, 2018 2:33:21 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Nov 17, 2018 2:33:26 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
18/11/17 14:33:26 INFO mortbay.log: Started HttpServer2$SelectChannelConnectorWithSafeStartup@0.0.0.0:8042
18/11/17 14:33:26 INFO webapp.WebApps: Web app node started at 8042
18/11/17 14:33:26 INFO client.RMProxy: Connecting to Resource Manager at /0.0.0.0:8031
18/11/17 14:33:27 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []
18/11/17 14:33:27 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers :[]
18/11/17 14:33:28 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 1810650918
18/11/17 14:33:28 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id -29592561
18/11/17 14:33:28 INFO nodemanager.NodeStatusUpdaterImpl: Registered with Resource Manager as tutrung.mshome.net:30016 with total resource of <memory:8192, vCores:8>
18/11/17 14:33:28 INFO nodemanager.NodeStatusUpdaterImpl: Notifying ContainerManager to unblock new container-requests
```

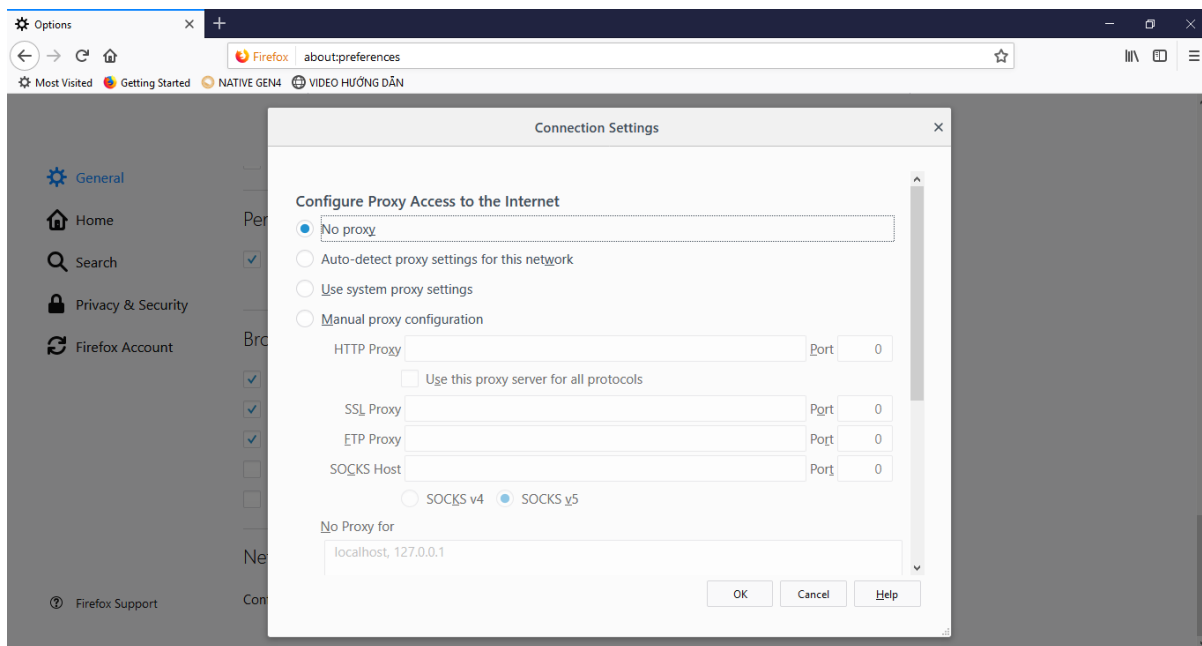
- Mở trình duyệt fileFox, nhấn settings->options



- Kéo thanh cuộn xuống dưới, chọn settings



- Chọn “No proxy”



- Kiểm tra server: gõ “localhost:50070” trên trình duyệt, nếu chạy ok ta được:

Overview 'localhost:19000' (active)

Started:	Sat Nov 17 14:33:14 PST 2018
Version:	2.7.3, rbaa91f7c6bc9cb92be5982de4719c1c8af91ccff
Compiled:	2016-08-18T01:41Z by root from branch-2.7.3
Cluster ID:	CID-b9ddfe8f-34eb-4512-a3b7-8c4d22138f6b
Block Pool ID:	BP-87471494-172.22.176.1-1542493463152

Summary

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
tutrung:50010 (127.0.0.1:50010)	1	In Service	82.88 GB	136 B	41.91 GB	40.96 GB	0	136 B (0%)	0	2.7.3

Decommissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction
------	--------------	-------------------------	------------------------------	--

- Nếu có dòng trong khung chữ nhật màu đỏ trong mục “Datanode Information” thì cài thành công. Nếu ko thì chưa nhận thư mục namenode, datanode. Hãy kiểm tra lại thông báo lỗi tương tự như sau (tại một trong các màn hình cmd):

```
2016-12-01 16:20:43,943 WARN common.Storage
(DataStorage.java:loadDataStorage(449)) - Failed to add storage directory
[DISK]file:/grid/0/hadoop/hdfs/data/
java.io.IOException: Incompatible clusterIDs in /grid/0/hadoop/hdfs/data:
namenode clusterID = CID-6c775e94-4cfb-4366-9a8b-c0f359a2b5d3; datanode
clusterID = CID-abfdf47f-98af-4924-a376-33b001206f7b
```

```
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.doTransition(DataStorage.java:799)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.loadStorageDirectory(DataStorage.java:322)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.loadDataStorage(DataStorage.java:438)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.addStorageLocations(DataStorage.java:417)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.recoverTransitionRead(DataStorage.java:595)
    at
org.apache.hadoop.hdfs.server.datanode.DataNode.initStorage(DataNode.java:1483)
    at
org.apache.hadoop.hdfs.server.datanode.DataNode.initBlockPool(DataNode.java:1448)
    at
org.apache.hadoop.hdfs.server.datanode.BPOfferService.verifyAndSetNamespaceInfo(BPOfferService.java:319)
    at
org.apache.hadoop.hdfs.server.datanode.BPServiceActor.connectToNNAndHandshake(BPServiceActor.java:267)
    at
org.apache.hadoop.hdfs.server.datanode.BPServiceActor.run(BPServiceActor.java:740)
    at java.lang.Thread.run(Thread.java:745)
2016-12-01 16:20:43,946 INFO common.Storage (Storage.java:tryLock(774)) - Lock on /grid/1/hadoop/hdfs/data/in_use.lock acquired by nodename 18164@MYDATANODE
2016-12-01 16:20:43,946 WARN common.Storage (DataStorage.java:loadDataStorage(449)) - Failed to add storage directory [DISK]file:/grid/1/hadoop/hdfs/data/
java.io.IOException: Incompatible clusterIDs in /grid/1/hadoop/hdfs/data:
namenode clusterID = CID-6c775e94-4cfb-4366-9a8b-c0f359a2b5d3; datanode
clusterID = CID-abfdf47f-98af-4924-a376-33b001206f7b
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.doTransition(DataStorage.java:799)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.loadStorageDirectory(DataStorage.java:322)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.loadDataStorage(DataStorage.java:438)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.addStorageLocations(DataStorage.java:417)
    at
org.apache.hadoop.hdfs.server.datanode.DataStorage.recoverTransitionRead(DataStorage.java:595)
    at
org.apache.hadoop.hdfs.server.datanode.DataNode.initStorage(DataNode.java:1483)
```














```

        at
org.apache.hadoop.hdfs.server.datanode.DataNode.initBlockPool(DataNode.java:144
8)
        at
org.apache.hadoop.hdfs.server.datanode.BPOfferService.verifyAndSetNamespaceInfo
(BPOfferService.java:319)
        at
org.apache.hadoop.hdfs.server.datanode.BPServiceActor.connectToNNAndHandshake(B
PServiceActor.java:267)
        at
org.apache.hadoop.hdfs.server.datanode.BPServiceActor.run(BPServiceActor.java:7
40)
-

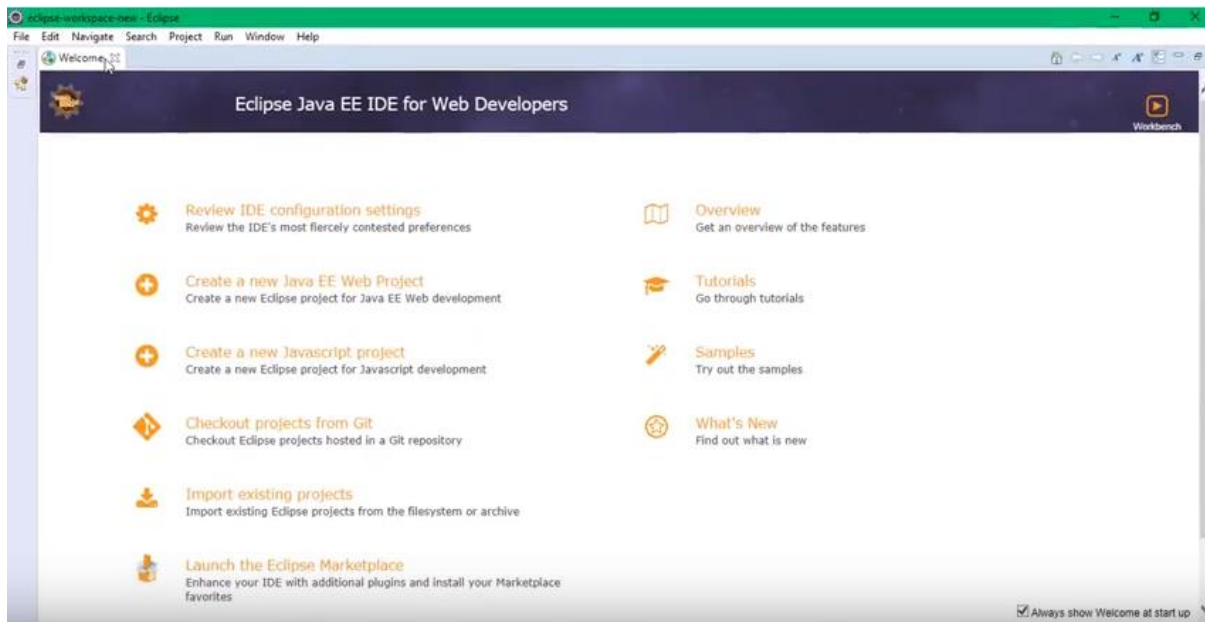
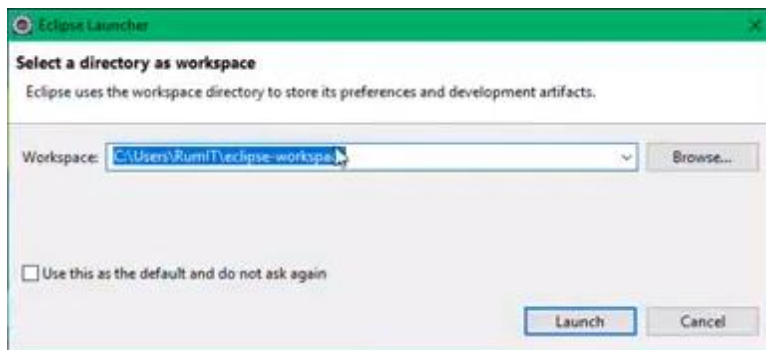
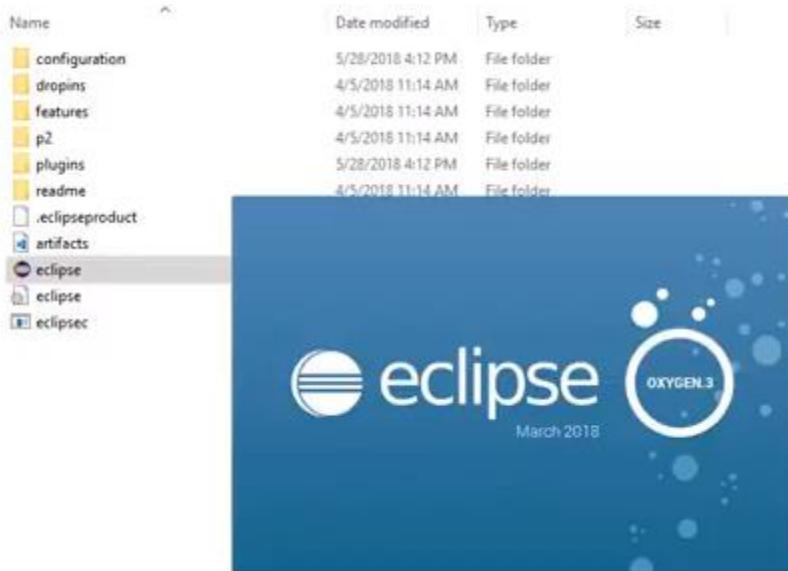
```

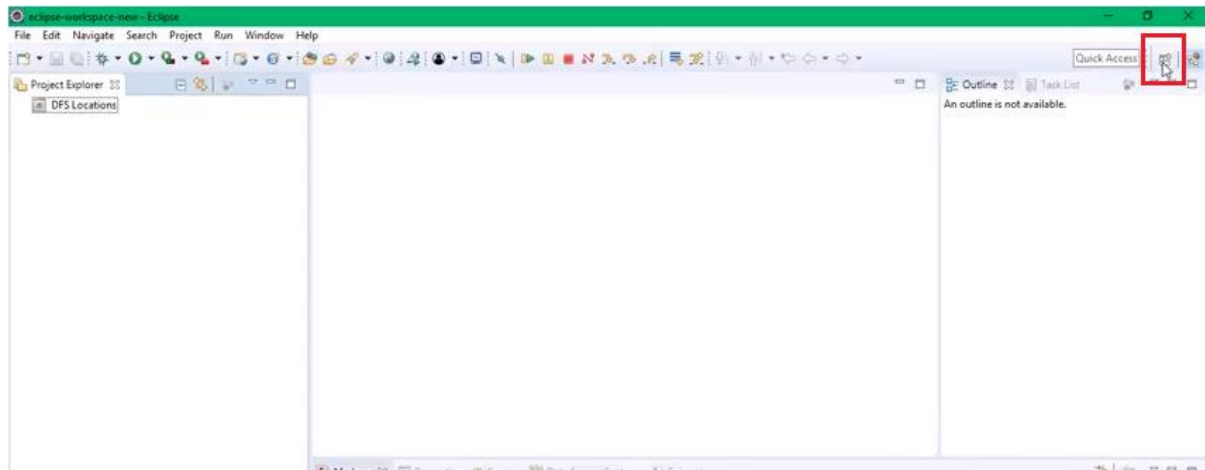
5. Chạy Eclipse

- Vào thư mục eclipse

Name	Date modified	Type	Size
 configuration	11/19/2018 9:07 AM	File folder	
 dropins	4/5/2018 11:14 AM	File folder	
 features	4/5/2018 11:14 AM	File folder	
 p2	11/19/2018 9:06 AM	File folder	
 plugins	11/19/2018 9:03 AM	File folder	
 readme	4/5/2018 11:14 AM	File folder	
 .eclipseproduct	12/22/2017 2:01 AM	ECLIPSEPRODUCT...	1 KB
 artifacts	4/5/2018 11:14 AM	XML Document	292 KB
 eclipse	4/5/2018 11:16 AM	Application	313 KB
 eclipse	11/19/2018 9:05 AM	Configuration sett...	1 KB
 eclipsec	4/5/2018 11:16 AM	Application	26 KB

- Nháy đúp file eclipse để khởi động eclipse, đợi để nhấn nút Launch, đợi đến khi màn hình chính hiện ra, close khung





- Nhấn nút trong vùng chữ nhật màu đỏ để mở hộp thoại “Open Perspective”, chọn “Map/Reduce”, nhấn Open

