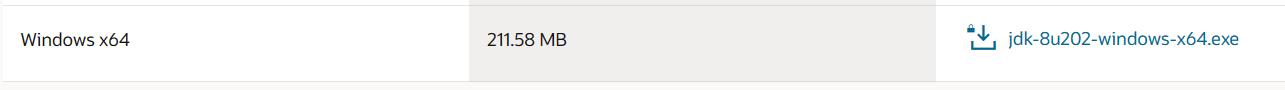
# 1. CÀI ĐẶT JDK

## 1.1 Tải JDK 1.8

Hadoop sử dụng JDK 1.8

Ta vào link sau: [Java Archive Downloads - Java SE 8 (oracle.com)](https://www.oracle.com/java/technologies/javase/javase8-archive-downloads.html)

****

Sau khi tải về ta có file



Sau đó chạy để cài đặt các gói về

A screenshot of a computer

Description automatically generated

## 1.2 Cài đặt biến môi trường JDK 1.8

Có thể tìm ở thanh công cụ “Edit the system environment variables”

**A black text on a white background

Description automatically generated**

Chọn “Environment variables”

**A screenshot of a computer program

Description automatically generated**

Dẫn đường dẫn phù hợp cho JAVA\_HOME

**A screenshot of a computer

Description automatically generated**

Và Path

**A white rectangular object with black lines

Description automatically generated**

Kiểm tra xem hệ thống đã nhận jdk chưa

**A black background with white text

Description automatically generated**

# 2. CÀI ĐẶT HADOOP

## 2.1 Tải Hadoop

Hadoop sử dụng Hadoop-3.3.6

Ta vào link sau: [Apache Hadoop](https://hadoop.apache.org/release/3.3.6.html)

Sau khi tải về và giải nén

A grey rectangular object with white text

Description automatically generated

## 2.2 Cài đặt biến môi trường Hadoop

Máy master và máy slave đều phải cài đặt chung 1 đường dẫn

Ví dụ ở dưới là “C:\hadoop-3-3-6” thì cả máy master và slave đều lưu hadoop ở địa chỉ “C:\hadoop-3-3-6”

Có thể tìm ở thanh công cụ “Edit the system environment variables”

**A black text on a white background

Description automatically generated**

Chọn “Environment variables”

**A screenshot of a computer program

Description automatically generated**

Dẫn đường dẫn phù hợp cho HADOOP

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

Và Path

****

# 3. CẤU HÌNH CÁC FILE CỦA HADOOP

## 3.1 Cấu hình địa chỉ ip

Mở command dưới quyền admin

Notepad C:\Windows\System32\Drivers\etc\hosts

A screenshot of a computer program

Description automatically generated

Điền hostname

192.168.171.148 ngocphung-master

192.168.171.233 phuthanh-slave

192.168.171.201 trantien-slave

192.168.171.179 thuyquynh-slave

192.168.56.1 tientran-slave

A screen shot of a computer

Description automatically generated

## 3.2 Tạo folder data

Tại ổ C tạo folder dfs

Trong folder dfs có 2 folder con là name và data

A screenshot of a computer

Description automatically generated

Tạo thêm folder tmp vào ổ C

A yellow folder with white text

Description automatically generated

## 3.3 core-site.xml (cấu hình trên cả master và slave)

Vào địa chỉ "C:\hadoop-3-3-6\etc\hadoop\core-site.xml"

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://ngocphung-master:9000</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/C:/tmp</value>

<description>A base for other temporary directories.</description>

</property>

</configuration>

A computer screen shot of white text

Description automatically generated

## 3.4 mapred-site.xml (cấu hình trên cả master và slave)

Vào địa chỉ "C:\hadoop-3-3-6\etc\hadoop\mapred-site.xml" “S:\Applications\Hadoop\hadoop-3-3-6\etc\hadoop”

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>ngocphung-master:9001</value>

</property>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

A computer screen shot of white text

Description automatically generated

## 3.5 hdfs-site.xml (cấu hình trên cả master và slave)

Vào địa chỉ "C:\hadoop-3-3-6\etc\hadoop\hdfs-site.xml"

<configuration>

<property>

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

<value>/C:/dfs/name</value>

</property>

<property>

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

<value>/C:/dfs/data</value>

</property>

<property>

<name>dfs.replication</name>

<value>4</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.datanode.use.datanode.hostname</name>

<value>false</value>

</property>

<property>

<name>dfs.namenode.datanode.registration.ip-hostname-check</name>

<value>false</value>

</property>

<property>

<name>dfs.namenode.http-address</name>

<value>ngocphung-master:50070</value>

<description>Your NameNode hostname for httpaccess.</description>

</property>

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>ngocphung-master:50090</value>

<description>Your Secondary NameNode hostname for httpaccess.</description>

</property>

</configuration>

**A screenshot of a computer program

Description automatically generated**

## 3.6 yarn-site.xml (cấu hình trên cả master và slave)

Vào trong địa chỉ "C:\hadoop-3-3-6\etc\hadoop\yarn-site.xml"

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

<description>Long running service which executes on Node Manager(s) and provides MapReduce Sort and Shuffle functionality.</description>

</property>

<property>

<name>yarn.log-aggregation-enable</name>

<value>true</value>

<description>Enable log aggregation so application logs are moved onto hdfs and are viewable via web ui after the application completed. The default location on hdfs is '/log' and can be changed via yarn.nodemanager.remote-app-log-dir property</description>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>ngocphung-master:8030</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>ngocphung-master:8031</value>

</property>

<property>

<name>yarn.resourcemanager.address</name>

<value>ngocphung-master:8032</value>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>ngocphung-master:8033</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>ngocphung-master:8088</value>

</property>

</configuration>

A screenshot of a computer program

Description automatically generated

## 3.7 hadoop-env.cmd (cấu hình trên cả master và slave)

Vào địa chỉ "C:\hadoop-3-3-6\etc\hadoop\hadoop-env.cmd"

Set up địa chỉ java giống vào java\_home

A black background with white text

Description automatically generated

Dán lệnh này ở cuối file

set HADOOP\_IDENT\_STRING=%USERNAME%

set HADOOP\_PREFIX=%HADOOP\_HOME%

set HADOOP\_CONF\_DIR=%HADOOP\_PREFIX%\etc\hadoop

set YARN\_CONF\_DIR=%HADOOP\_CONF\_DIR%

set PATH=%PATH%;%HADOOP\_PREFIX%\bin

A screen shot of a computer

Description automatically generated

3.8 Edit file wokers (Làm trên máy master)

Vào địa chỉ "C:\hadoop-3-3-6\etc\hadoop\workers"

Trên máy master chỉ ghi tên máy slave

A screenshot of a computer

Description automatically generated

3.9 Edit file wokers (Làm trên máy slave)

Vào địa chỉ "C:\hadoop-3-3-6\etc\hadoop\workers"

Trên máy slave chỉ ghi tên máy của mình

A screenshot of a computer

Description automatically generated

# 4. Cập nhật file bin CẬP NHẬT FILE BIN

Tải từ link này: <https://github.com/s911415/apache-hadoop-3.1.0-winutils>

Tải về và extract ra

A screenshot of a computer

Description automatically generated

Tìm folder gần với phiên bản của phiên bản mình dùng nhất

Em dùng hadoop-3.3.6 nên em sẽ dùng folder hadoop-3.3.5

A screenshot of a computer

Description automatically generated

Copy file bin và đè lên file bin của “S:\Applications\Hadoop\hadoop-3-3-6”

A screenshot of a computer

Description automatically generated

# 5. FORMAT HỆ THỐNG

## 5.1 Format

Hdfs namenode -format



## 5.2 Chép file

“S:\Applications\Hadoop\hadoop-3-3-6\share\hadoop\yarn\timelineservice”

Copy file hadoop-yarn-server-timelineservice-3.3.6

Và dán ra “S:\Applications\Hadoop\hadoop-3-3-6\share\hadoop\yarn”

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

# 6. Chạy thử

Start-all

A computer screen with white text

Description automatically generated

Namenode

A screenshot of a computer

Description automatically generated

Datanode

A screenshot of a computer

Description automatically generated

Nodemanager

A screenshot of a computer

Description automatically generated

Resourcemanager

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated