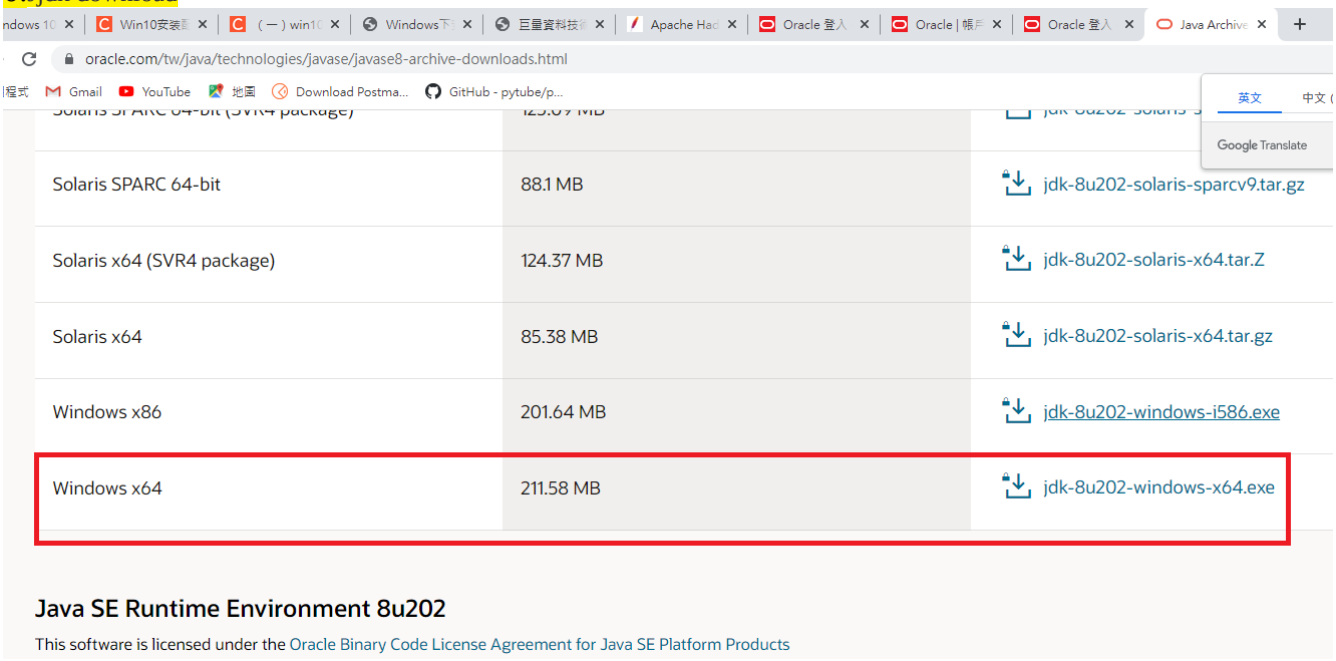
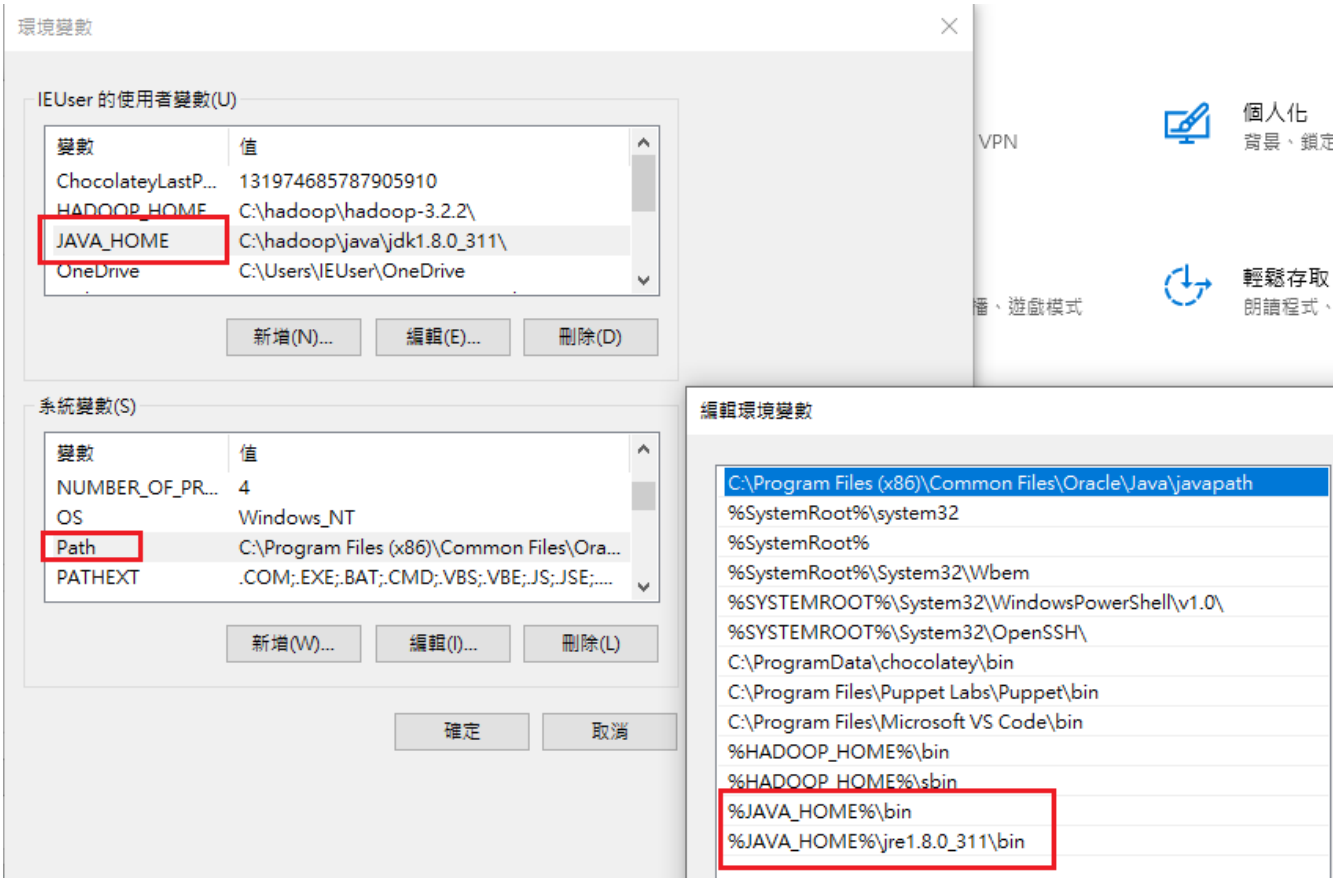
一文讀懂大數據平台 — — 寫給大數據開發初學者的話!
https://allaboutdataanalysis.medium.com/%E4%B8%80%E6%96%87%E8%AE%80%E6%87%82%E5%A4%A7%E6%95%B8%E6%93%9A%E5%B9%B3%E5%8F%B0-%E5%AF%AB%E7%B5%A6%E5%A4%A7%E6%95%B8%E6%93%9A%E9%96%8B%E7%99%BC%E5%88%9D%E5%AD%B8%E8%80%85%E7%9A%84%E8%A9%B1-d4d67c37621f

JobTracker 和 TaskTracker 詳解
https://blog.csdn.net/u012599988/article/details/46860307

0.Jdk download

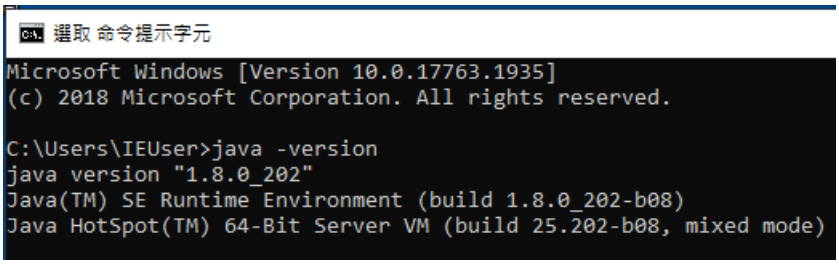


安裝路徑 C:\hadoop\java\jdk1.8.0_311 不可空白



a.
測試 java -version

https://blog.csdn.net/qq_36888550/article/details/105038790



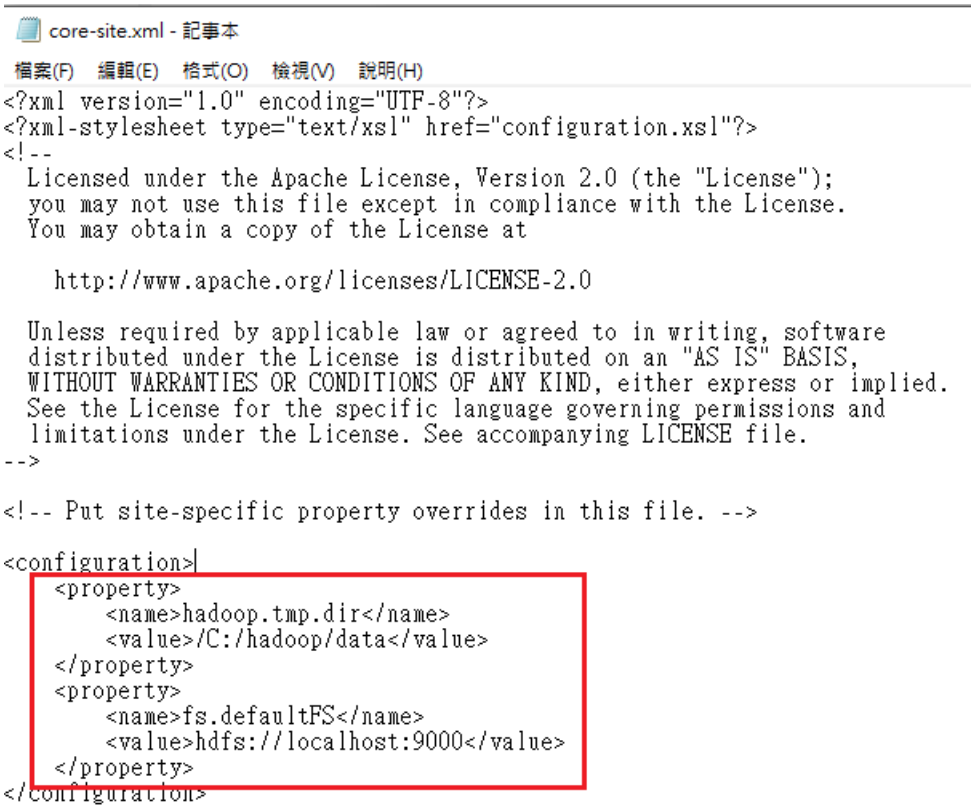
1.Hadoop 安裝 官網下載
http://hadoop.apache.org/releases.html

a. tar.gz 檔解壓縮
b. 設定環境變數
C:\hadoop 添加到 HADOOP_HOME，
并在 Path 添加 %HADOOP_HOME%\bin 和 %HADOOP_HOME%\sbin 。
WIN10 安装配置 Hadoop
https://zhuanlan.zhihu.com/p/111844817

http://debussy.im.nuu.edu.tw/sjchen/BigData/%E5%B7%A8%E9%87%8F%E8%B3%87%E6%96%99%E6%8A%80%E8%A1%93%E8%88%87%E6%87%89%E7%94%A8%E7%92%B0%E5%A2%83%E5%AE%89%E8%A3%9D%E8%AC%9B%E7%BE%A9-Hadoop.html

c. 修改 core-site.xml

```
    <property>
 <name>hadoop.tmp.dir</name>
 <value>/C:/hadoop/data</value>
 </property>
 <property>
 <name>fs.default.name</name>
 <value>hdfs://localhost:9000</value>
 </property>
<property>
      <name>fs.defaultFS</name>
      <value>hdfs://192.168.111.128</value>
   </property>
```



d.修改 hdfs-site.xml

DATA 資料夾下建 namenode & datanode & snn 資料夾

```
<property>
        <name>dfs.replication</name>
        <value>1</value>
```

```
    </property>
    <property>
        <name>dfs.permissions</name>
        <value>false</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:/C:/hadoop/data/namenode</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:/C:/hadoop/data/datanode</value>
    </property>
    <property>
        <name>fs.checkpoint.dir</name>
        <value>file:/C:/hadoop/data/snn</value>
    </property>
    <property>
        <name>dfs.checkpoint.edits.dir</name>
        <value>file:/C:/hadoop/data/snn</value>
    </property>
```

hdfs-site.xml - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

```
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>/C:/hadoop/data/namenode</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>/C:/hadoop/data/datanode</value>
    </property>
</configuration>
```

e.mapred-site.xml

```
 <property>
 <name>mapreduce.framework.name</name>
 <value>yarn</value>
 </property>
 <property>
 <name>mapred.job.tracker</name>
```

```
        <value>hdfs://localhost:9001</value>
    </property>
```

hadoop 3.0.0 版本之後，各個 service 的環境變數已經不互相繼承，必須要經過 configuration 設定才行。

https://mathsigit.github.io/blog_page/2017/11/16/hole-of-submitting-mr-of-hadoop300RC0/

```
<property>
        <name>yarn.app.mapreduce.am.env</name>
        <value>HADOOP_MAPRED_HOME=/opt/hadoop</value>
    </property>
    <property>
        <name>mapreduce.map.env</name>
        <value>HADOOP_MAPRED_HOME=/opt/hadoop</value>
    </property>
    <property>
        <name>mapreduce.reduce.env</name>
        <value>HADOOP_MAPRED_HOME=/opt/hadoop</value>
    </property>
```

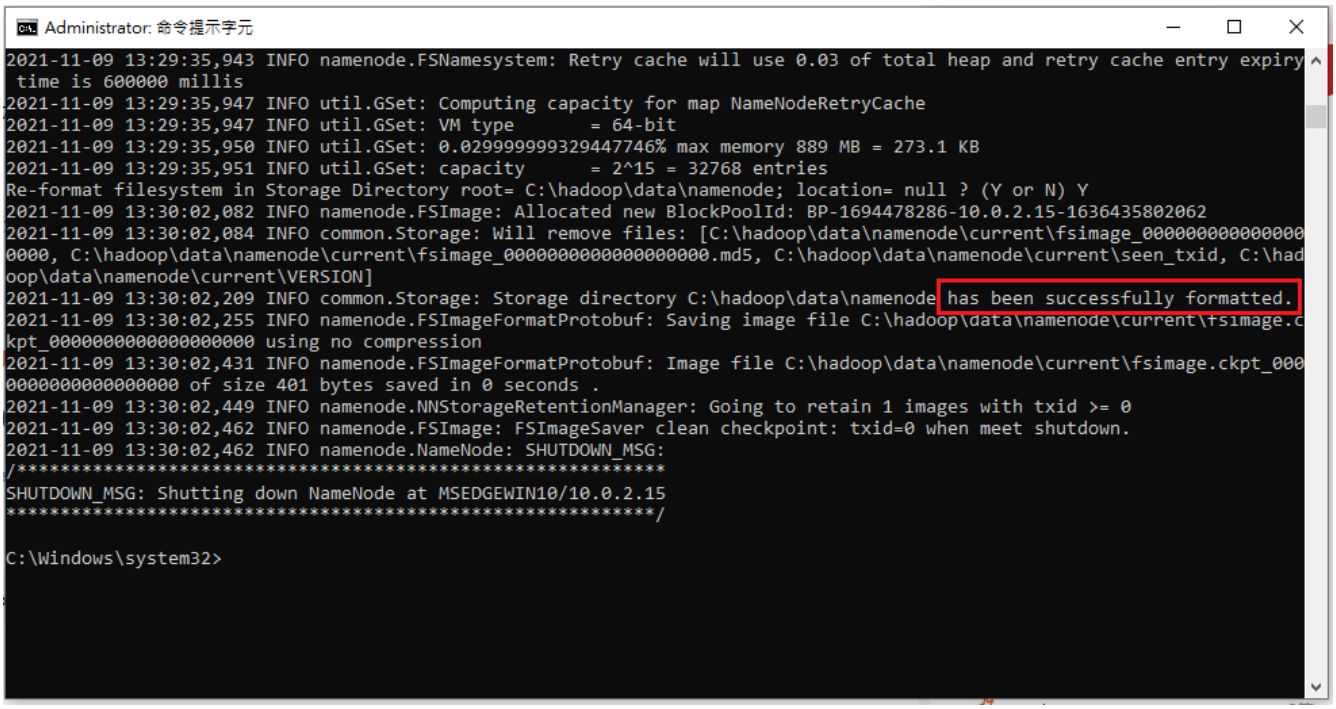
f.yarn-site.xml

```
<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.hahoop.mapred.ShuffleHandler</value>
 </property>
 <property>
 <name>yarn.nodemanager.local-dirs</name>
 <value>/C:/hadoop/hadoop-3.2.2/tmp</value>
 </property>
 <property>
 <name>yarn.resourcemanager.address</name>
 <value>localhost:8032</value>
 </property>
 <property>
 <name>yarn.resourcemanager.scheduler.address</name>
 <value>localhost:8030</value>
 </property>
 <property>
 <name>yarn.resourcemanager.resourc-tracker.address</name>
 <value>localhost:8031</value>
 </property>
 <property>
 <name>yarn.resourcemanager.admin.address</name>
 <value>localhost:8033</value>
 </property>
 <property>
 <name>yarn.resourcemanager.webapp.address</name>
 <value>localhost:8088</value>
<property>
        <name>yarn.nodemanager.resource.memory-mb</name>
        <value>106496</value>
    </property>
    <property>
```

```
    <name>yarn.scheduler.minimum-allocation-mb</name>
    <value>2048</value>
</property>
<property>
    <name>yarn.scheduler.maximum-allocation-mb</name>
    <value>106496</value>
</property>
<property>
    <name>yarn.app.mapreduce.am.resource.mb</name>
    <value>4096</value>
</property>
<property>
    <name>yarn.app.mapreduce.am.command-opts</name>
    <value>-Xmx3276m</value>
</property>
```

```
    </property>
```

g.hdfs namenode -format



h.cmd 中输入 **start-all.cmd** 启动，输入 **stop-all.cmd** 关闭

win10 搭建 Hadoop3.2.2 和 HBase2.3.5 单机版
https://blog.csdn.net/liu320yj/article/details/117823724
成功版

C:\Windows\system32>

瀏覽器連結
http://localhost:8088/
http://localhost:9870 hdfs 3.板以上

jps 查詢那些有啟動

firewall-cmd -state

hadoop fs -ls -R /
hadoop fs -mkdir -p /User/hadoop
hdfs dfs -rm -r /User/hadoop

hadoop fs -put C:\Users\IEUser\Downloads\test.txt /user/hadoop/test/test_copy.txt
hadoop fs -copyFromLocal C:\Users\IEUser\Downloads\test.txt /user/hadoop/test/test_copy.txt

hdfs dfs -rm -r /Users/IEUser/hadoop/dir1

hadoop fs -mkdir -p /user/hadoop/dir1

hadoop fs -put /hadoop/hadoop-3.2.2/README.txt /user/hadoop/readme_test.txt

hdfs dfs -put /usr/local/hadoop/File_15007.csv /user/hadoop/readme_test.txt

hadoop fs -copyFromLocal -f C:/hadoop/hadoop-3.2.2/README.txt /User/hadoop/test/readme_test.txt

hadoop fs -put C:\Users\IEUser\Downloads\File_15007.csv /User/hadoop/test/File_15007_copy.csv

---

指令

分散式檔案系統

```
$ hdfs
```

```
$ hdfs dfs
$ hdfs dfs -help put
```

管理者指令

HDFS 目前的狀態

```
$ hdfs dfsadmin -report
```

使用者指令

建立一個 /user/hadoop 目錄路徑

```
$ hdfs dfs -mkdir -p /user/hadoop
$ hdfs dfs -
  put /haddop3/hadoop-3.2.2/README.txt /user/hadoop/readme_test.txt
```

hdfs dfs -ls -R /

-cat：用以將檔案文件的內容輸出到標準輸出裝置

```
$ hdfs dfs -cat /user/hadoop/readme_test.txt
```

-cp：用來將 HDFS 下的某個檔案文件或目錄，複製到 HDFS 的另外一個目錄下。
hdfs dfs -cp /user/hadoop/readme_test.txt /input　#移動文件到/input 目錄下

-mv：用來將 HDFS 下的某個檔案文件或目錄，移動到 HDFS 的另外一個目錄下，不可跨越檔案系統。
將 HDFS 的目錄 /input 搬移到 HDFS 下的另一個目錄/user/hadoop 下，且該目錄更名為 input02。

```
$ hdfs dfs -mv /input /user/hadoop/input02
```

-rm：刪除 HDFS 上的檔案文件。若要刪除目錄，或是刪除某檔案連同其上層目錄時，則需加上操作參數-r。

```
$ hdfs dfs -rm /user/hadoop/readme_test.txt   #刪除文件
$ hdfs dfs -rm -r /user/hadoop/input          #刪除目錄
```

```
hdfs dfs -rm -r /user
```

hdfs fsck 指令以**回報 HDFS 相關檔案架構的健康狀況**

```
$ hdfs fsck /tmp
```

hadoop 運行第一個實例 wordcount

C:\Windows\system32>hadoop fs -mkdir -p /input/wordcount

C:\Windows\system32>hdfs dfs -mkdir -p /input/wordcount

C:\Windows\system32>hdfs dfs -ls -R /
drwxr-xr-x   - IEUser supergroup   0 2021-11-26 14:45 /input
drwxr-xr-x   - IEUser supergroup   0 2021-11-26 14:45 /input/wordcount

C:\Windows\system32>hdfs dfs -put /Users/IEUser/hadoop/file2.txt /input/wordcount/file2.txt

C:\Windows\system32>hdfs dfs -mkdir -p /output

C:\Windows\system32>
C:\Windows\system32>hdfs dfs -ls -R /
drwxr-xr-x   - IEUser supergroup   0 2021-11-26 14:45 /input
drwxr-xr-x   - IEUser supergroup   0 2021-11-26 14:50 /input/wordcount
-rw-r--r--   1 IEUser supergroup   79 2021-11-26 14:50 /input/wordcount/file2.txt
drwxr-xr-x   - IEUser supergroup   0 2021-11-26 14:53 /output

hadoop jar /haddop3/hadoop-3.2.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar wordcount /input/wordcount/file2.txt /output/wordcount

yarn jar /haddop3/hadoop-3.2.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar wordcount /input/wordcount/file2.txt /output/wordcount

hadoop dfs -cat /output/wordcount/*

hadoop fs -rm -r /output/wordcount