！！！如果不知道自己在幹嘛，全部砍掉重練最快！！！

【前置】

windows cmd 查詢空的IP arp -a

若先前有啟動先停掉hdfs(./stop-dfs.sh)和spark(./stop-all.sh)

多台開機

(先在NAT的狀態下?)修改 cd /etc/netplan/00… <init 6>

修改 /etc/hosts <第一行localhost 127...要刪掉，不然網路繞不出去>

網路改成(bridged)並指向實體網卡

ping 8.8.8.8

【clone new VM from worker 】

修改主機名稱 sudo hostnamectl set-hostname xxxxxx

確認 hostnamectl status<init 6>

【master 和 worker要交換金鑰】[SSH 公開金鑰認證：不用打密碼登入 Linux 設定教學，安全又方便](https://blog.gtwang.org/linux/linux-ssh-public-key-authentication/)

[設定 SSH 和登入公鑰生成](https://medium.com/@sleo1104/hadoop-3-2-0-%E5%AE%89%E8%A3%9D%E6%95%99%E5%AD%B8%E8%88%87%E4%BB%8B%E7%B4%B9-22aa183be33a)

##############這段不確定，看上面連結#############

workerX@ cd .ssh

workerX@ mv id\_rsa.pub id\_rsa.pub\_worker1

workerX@ cd ..

workerX@ ssh-keygen

workerX@ ls .ssh/

workerX@ scp .ssh/id\_rsa.pub master:/home/spark/.ssh/id\_rsa\_worker2.pub

workerX@ ssh master

master@ cd /.ssh

master@ ls

master@ cat id\_rsa\_worker2.pub >> authorized\_keys

master@ exit

master@ ssh master

master@ cd .ssh/

master@ scp id\_rsa.pub workerX:/home/.ssh/master.pub

master@ exit

##############用ui@偷懶複製貼上#############

【同步各台設定檔】

確認各台~對應ip是否正確 cat /etc/hosts (若有修改要scp)

master@ cd hadoop-2.10.0/etc/hadoop 改slave 改 hdfs-site.xml (worker數量)

5個檔案SCP到各台包含USER cat core.site.xml (預設HDFS放在master、預設port 8020)

master@scp -r core-site.xml hdfs-site.xml slaves yarn-site.xml mapred-site.xml.template workerX:/home/spark/hadoop-2.10.0/etc/hadoop

【如果不乾淨刪掉datdnode資料夾&重建】曾經啟動過hdfs的話，重新調整都要清掉

workerX@ cd hdfs ; rm -r datanode ; mkdir datanode

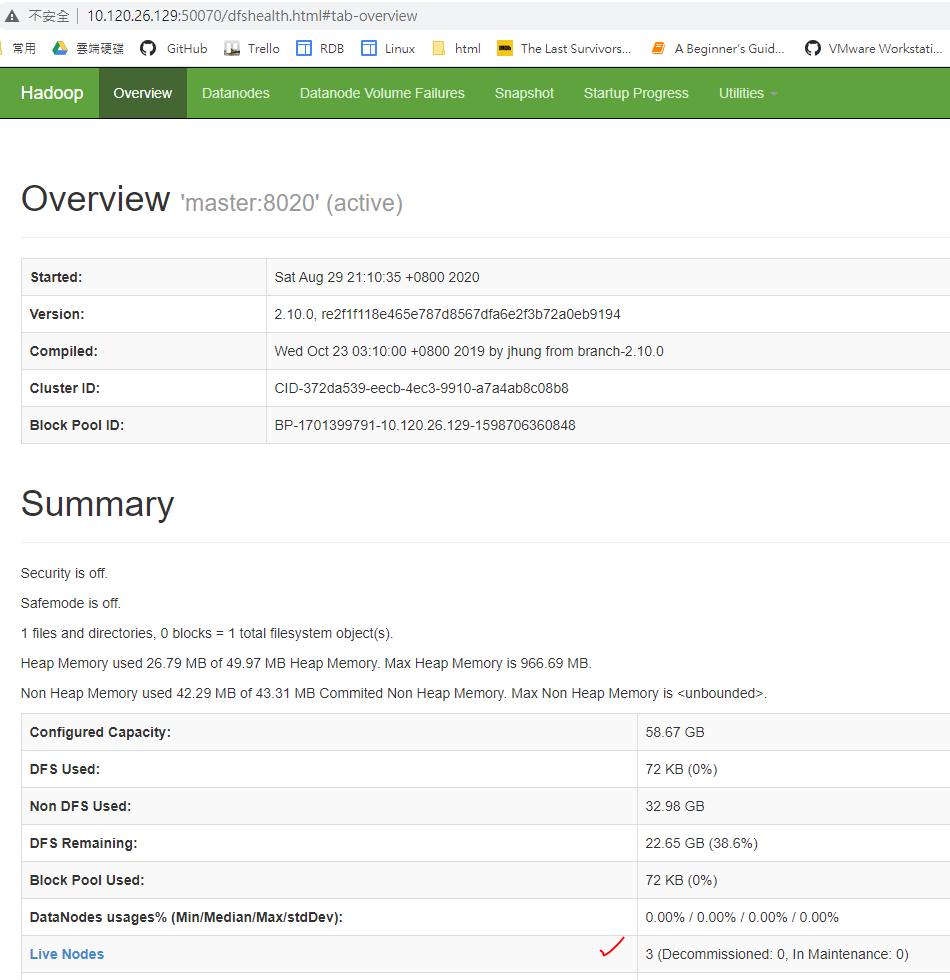
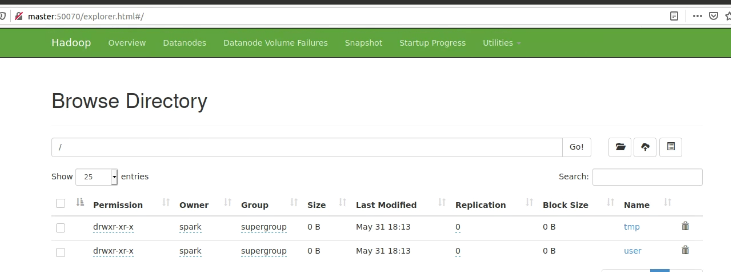
【格式化】

master@ cd ~/hdfs ; hadoop namenode -format

【啟動hdfs】

master@ cd ~/hadoop-2.10.0/sbin ; ./stop-dfs.sh ; ./start-dfs.sh

瀏覽器確認 master ip:50070 (HDFS是否有啟動-NN/DN)

自己架要建這些資料夾~還有對應user的

master@ hadoop fs -mkdir /tmp

master@ hadoop fs -mkdir /user

master@ hadoop fs -mkdir /user/spark

【設定spark】

master@ spark-2.4.5-bin-hadoop2.7/conf

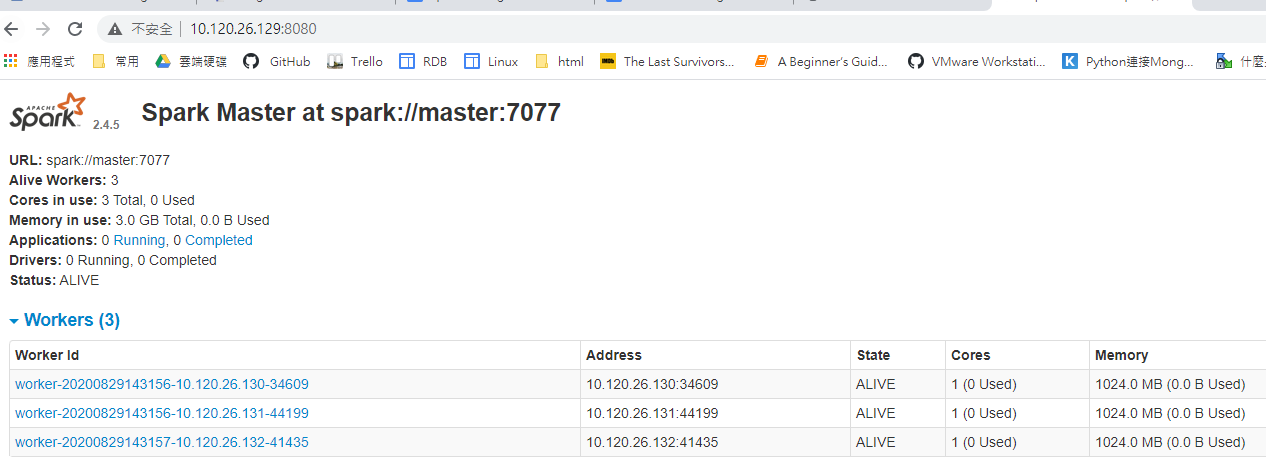
master@ sudo vi slaves

master@ scp -r slaves ui:/home/spark/spark-2.4.5-bin-hadoop2.7/conf/

master@ cd ..

master@ cd sbin

master@ ./stop-all.sh ; ./start-all.sh



【回到ui跑程式】

ui@ hadoop fs -mkdir /user/spark/spark101

ui@ hadoop fs -put ~/Desktop/spark101/movies spark101 上傳要跑的檔案

傳錯了刪資料夾 hadoop fs -rmr spark101/movies/movies

#################文件內容###########################

/etc/netplan/00-installer-config.yaml

# This is the network config written by 'subiquity'

network:

ethernets:

ens33:

addresses:

- 10.120.26.xxx/16

gateway4: 10.120.0.254

nameservers:

addresses:

- 8.8.8.8

version: 2

/etc/hosts

10.120.26.xxx ui

10.120.26.xxx master

10.120.26.xxx worker1

10.120.26.xxx worker2

10.120.26.xxx worker3

master@ ~/hadoop-2.10.0/etc/hadoop/slaves

worker1

worker2

worker3

master@ ~/hadoop-2.10.0/etc/hadoop/hdfs-site.xml

<?xml version="1.0" encoding="UTF-8"?>

<?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.namenode.name.dir</name>

<value>file:/home/spark/hdfs/namenode</value>

</property>

<property>

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

<value>file:/home/spark/hdfs/datanode</value>

</property>

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

</configuration>

#########其他補充############

spark@master cd ~/spark….../conf

vim (slaves) (spark-env.sh)

(預設core和記憶體都給很大，若要修改) vim spark-defaults.comf.temp…...

scp -r 有修改的檔案~ (ui)(workerX)

hdfs-site.xml.bak>>>大神組還有一段關於帳號權限的參數，若登入帳號都是SPARK不需

自己架要建這些資料夾

spark@worker1 mkdir hdfs

cd hdfs/

mkdir datanode

spark@master mkdir hdfs

cd hdfs/

mkdir namenode