Hadoop Multi Nodes Cluster Setup

底下步驟將建立一個 Hadoop Multi Nodes Cluster，包含三個 VM

伺服器名稱	類別	HDFS	YARN
HadoopMaster	Master	NameNode	ResourceManager
HadoopSlave1	Node	DataNode	NodeManager
HadoopSlave2	Node	DataNode	NodeManager

1. 複製 single node cluster 到 hadoopslave1
   - 若下為使用VMWare 個人版的做法，若為專業版可以非常簡單的複製已存在的VM
   - 先於 VM 儲存資料夾中建立要放置三個VM檔案的資料夾

   - 將之前建好的single node cluster VM資料夾中的檔案複製到 HadoopSlave1中
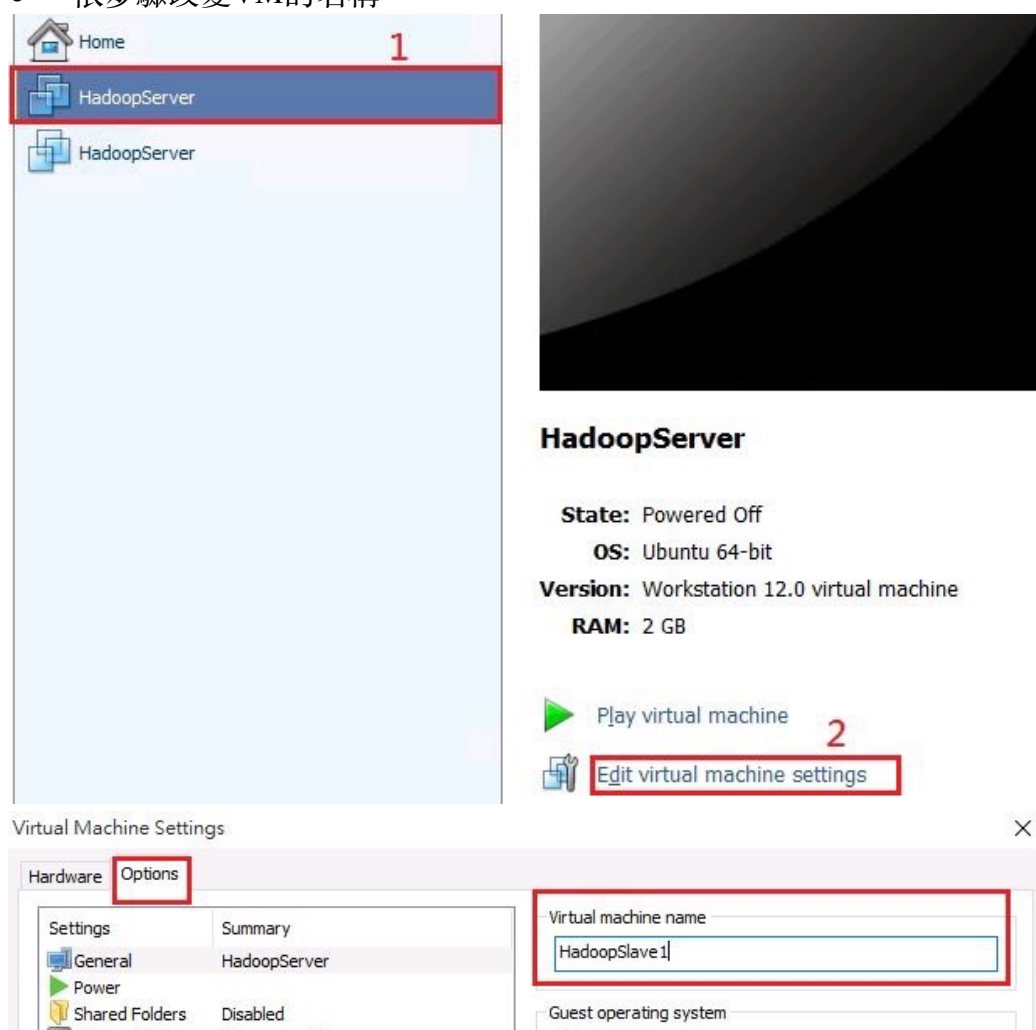

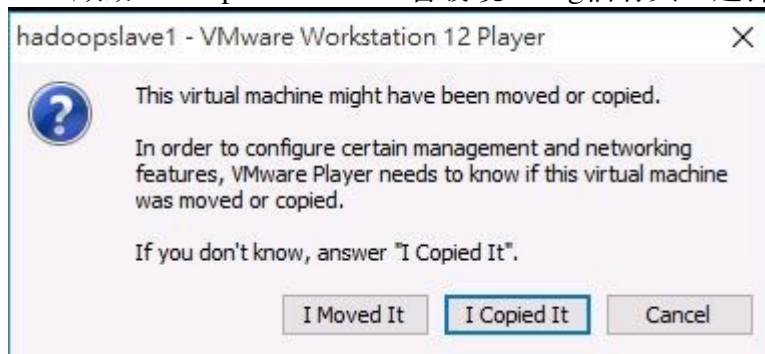   - 到VM介面，選擇 [Home]➜[Open a Virtual Machine]
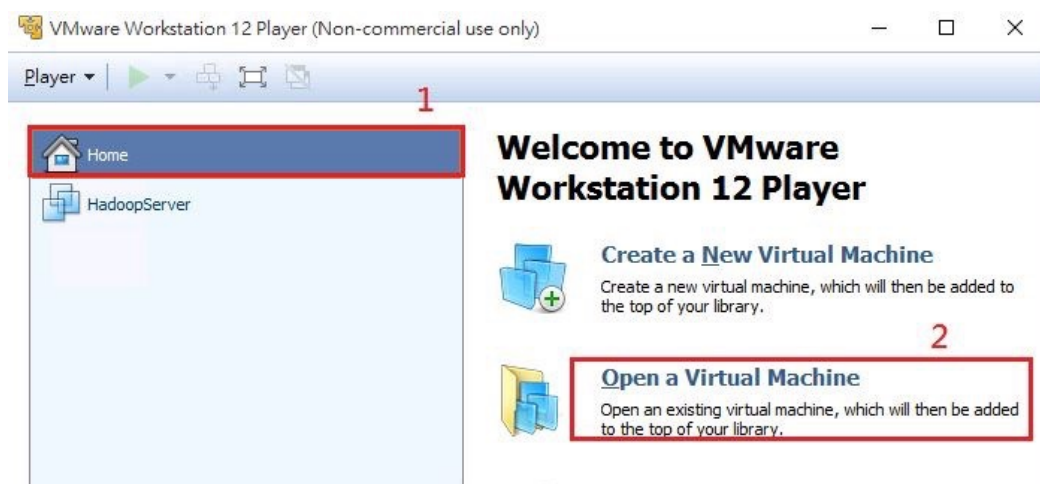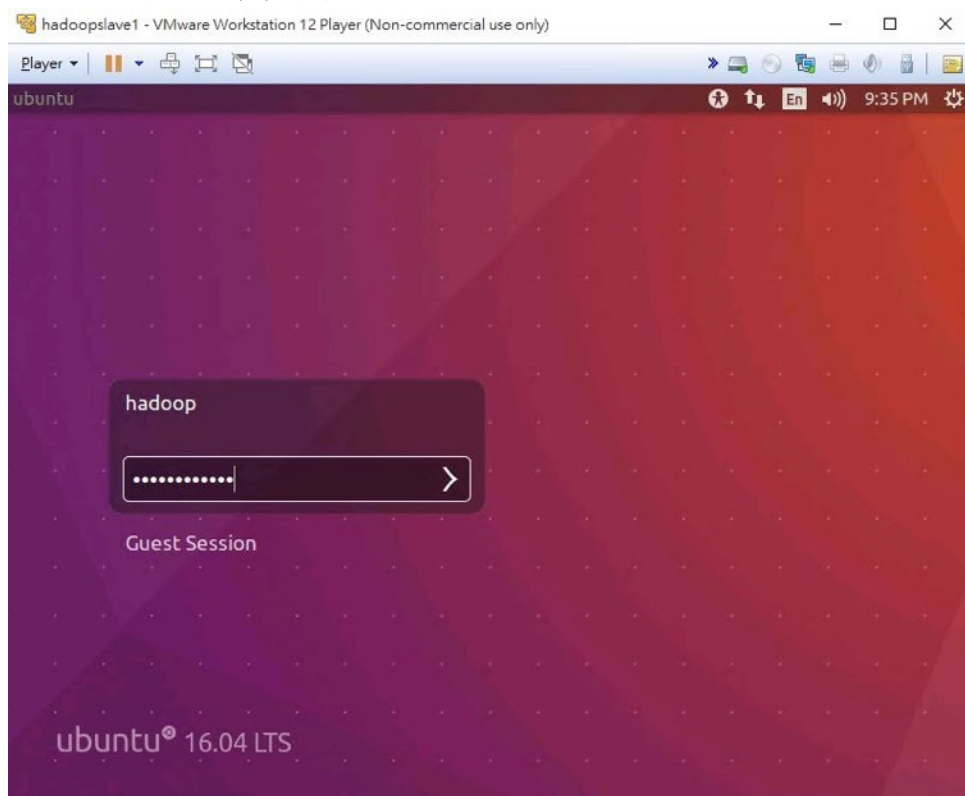
- 到HadoopSlave1資料夾選擇configuration檔

- 依步驟改變VM的名稱

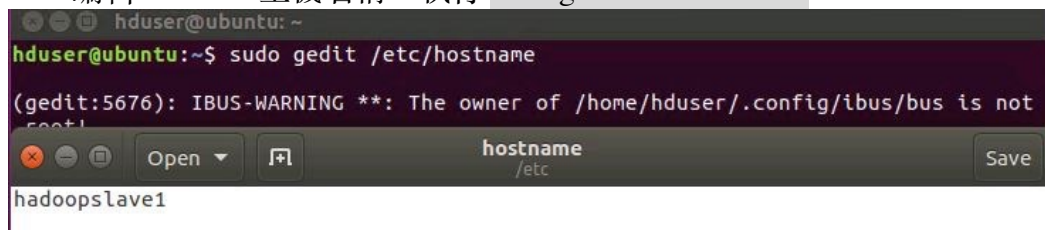- 啟動HadoopSlave1，VM會發現config檔有異，選擇 I Copied It

- 即可順利複製完成

2. 設定HadoopSlave1
   - 啟動HadoopSlave1 虛擬機
   - 編輯hostname主機名稱，執行 sudo gedit /etc/hostname

   - 編輯core-site.xml，執行 sudo gedit /usr/local/hadoop/etc/hadoop/core-site.xml，指定NameNode位置，將原本設定 localhost 處改為 hadoopmaster

```
hduser@ubuntu: ~
hduser@ubuntu:~$ sudo gedit /usr/local/hadoop/etc/hadoop/core-site.xml

                                    *core-site.xml
 Open  ▼    ⊞         /usr/local/hadoop/etc/hadoop                    Save

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

<configuration>
    <property>
        <name>fs.default.name</name>
        <value>hdfs://hadoopmaster:9000</value>
    </property>
</configuration>
```

- 編輯yarn-site.xml，執行 sudo gedit /usr/local /hadoop/etc/hadoop/yarn-site.xml，設定 RresourceManager的屬性，輸入以下指令

```
<property>
     <name>yarn.resourcemanager.resource-tracker.address</name>
     <value>hadoopmaster:8025</value>
</property>
<property>
    <name>yarn.resourcemanager.scheduler.address</name>
    <value>hadoopmaster:8030</value>
</property>
     <property>
        <name>yarn.resourcemanager.address</name>
        <value>hadoopmaster:8050</value>
     </property>
```



```
    <name>yarn.resourcemanager.address</name>
    <value>hadoopmaster:8050</value>
</property>
</configuration>
```

- 編輯mapred-site.xml，執行
  sudo gedit /usr/local/hadoop/etc/hadoop/mapred-site.xml，設定監控
  MapReduce的JobTracker工作分配狀況，將將原本設定 localhost 處改為
  hadoopmaster

```
hduser@ubuntu: ~
hduser@ubuntu:~$ sudo gedit /usr/local/hadoop/etc/hadoop/mapred-site.xml
```

```
*mapred-site.xml
/usr/local/hadoop/etc/hadoop
```

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

<configuration>
    <property>
        <name>mapred.job.tracker</name>
        <value>hadoopmaster:54311</value>
    </property>
</configuration>
```

- 編輯hdfs-site.xml，執行 sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml，設定 HDFS 分散式檔案組態


```
hduser@ubuntu: ~
hduser@ubuntu:~$ sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

```
hdfs-site.xml
/usr/local/hadoop/etc/hadoop
```

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

<configuration>
    <property>
        <name>dfs.replication</name>
        <value>2</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value> file:/usr/local/hadoop/hadoop_data/hdfs/datanode</value>
    </property>
</configuration>
```

- 將 HadoopSlave1 關機

Shut Down

Goodbye, hadoop. Are you sure you want to close all programs and shut down the computer?
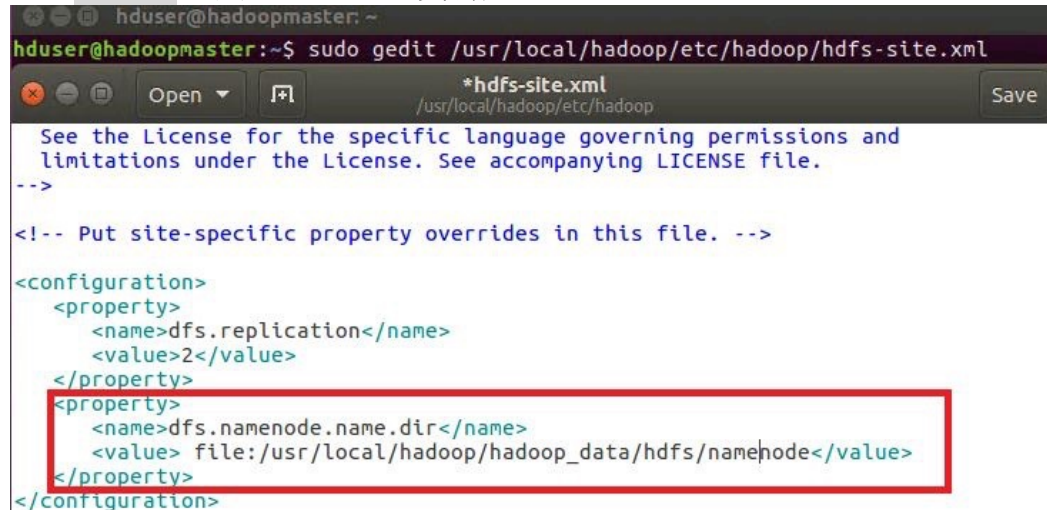
Shut Down

1. 複製HadoopSlave1的虛擬虛設定檔，來產生HadoopMaster、HadoopSlave2 等兩台虛擬機，修改 /etc/hostname 內的伺服器名稱分別為hadoopmaster和
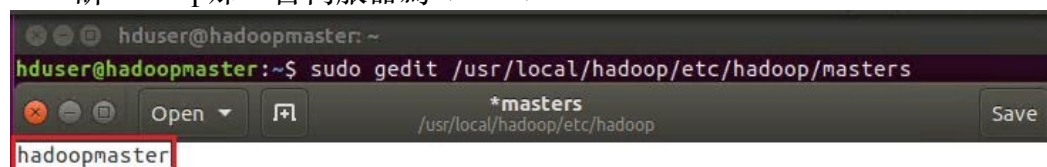
hadoopslave2，設定完之後，虛擬機需重開。

2. 設定HadoopMaster伺服器
   ● 編輯hostname主機名稱，執行 sudo gedit /etc/hostname，檔案內填入 **hadoopmaster**
   ● 重開機HadoopMaster
   ● 編輯 hdfs-site.xml，執行 sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml，設定NameNode資料夾
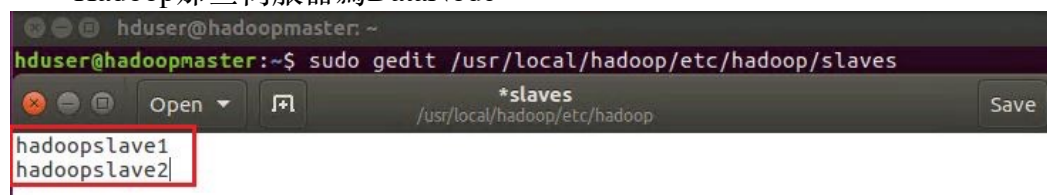


```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml

Open ▼        *hdfs-site.xml                    Save
              /usr/local/hadoop/etc/hadoop

  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>2</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value> file:/usr/local/hadoop/hadoop_data/hdfs/namenode</value>
    </property>
</configuration>
```

   ● 編輯 masters，執行 sudo gedit /usr/local /hadoop/etc/hadoop/masters，告訴Hadoop那一台伺服器為NameNode


```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ sudo gedit /usr/local/hadoop/etc/hadoop/masters

Open ▼        *masters                    Save
              /usr/local/hadoop/etc/hadoop
hadoopmaster
```

   ● 編輯 slaves，執行 sudo gedit /usr/local /hadoop/etc/hadoop/slaves，告訴Hadoop那些伺服器為DataNode


```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ sudo gedit /usr/local/hadoop/etc/hadoop/slaves

Open ▼        *slaves                    Save
              /usr/local/hadoop/etc/hadoop
hadoopslave1
hadoopslave2
```

3. 設定各節點主機名稱和相應的IP位址
   ● 啟動三台虛擬機

- 分別至各虛擬機中，執行ifconfig，取得伺服器IP

- 至各虛擬機中，編輯/etc/hosts，執行sudo gedit /etc/hosts，將節點主稱和ＩＰ位址填入

4. Master連線至slave1、slave2建立HDFS目錄
   - 切換至HadoopMaster伺服器
   - 由HadoopMaster經過SSH連至HadoopSlave1，執行 ssh hadoopslave1，注意連線成功後會如圖紅線標示

```
hduser@hadoopslave1: ~
hduser@hadoopmaster:~$ ssh hadoopslave1
The authenticity of host 'hadoopslave1 (192.168.59.134)' can't be established.
ECDSA key fingerprint is SHA256:l5HfVz2GKon2xpmavQSLRqfvPdxuogiqaF/Xjx5XV3E.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'hadoopslave1,192.168.59.134' (ECDSA) to the list of
known hosts.
hduser@hadoopslave1's password:
Welcome to Ubuntu 16.04.1 LTS (GNU/Linux 4.4.0-31-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

282 packages can be updated.
115 updates are security updates.

hduser@hadoopslave1:~$
```

- 執行
    - i.   移除hdfs目錄，sudo rm –rf /usr/local/hadoop/hadoop_data/hdfs
    - ii.   建立DataNode儲存目錄，mkdir -p /usr/local/hadoop/hadoop_data/ hdfs/datanode
    - iii. 修改目錄擁有者，sudo chown -R hduser:hduser /usr/local/hadoop/
    - iv.   中斷連線

```
hduser@hadoopmaster: ~
hduser@hadoopslave1:~$ sudo rm -rf /usr/local/hadoop/hadoop_data/hdfs
[sudo] password for hduser:
hduser@hadoopslave1:~$ sudo mkdir -p /usr/local/hadoop/hadoop_data/hdfs/datanode
hduser@hadoopslave1:~$ sudo chown hduser:hduser -R /usr/local/hadoop
hduser@hadoopslave1:~$ exit
logout
Connection to hadoopslave1 closed.
hduser@hadoopmaster:~$
```

- 重複上述步驟，設定hadoopslave2
- 建立NameNode HDFS目錄
    - v.   首先刪除HDFS目錄，執行 sudo rm –rf /usr/local/hadoop/ hadoop_data/hdfs
    - vi.   建立NameNode目錄，執行 mkdir -p /usr/local/hadoop/hadoop_data/ hdfs/namenode
    - vii. 修改目錄擁有者，sudo chown -R hduser:hduser /usr/local/hadoop/

```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ sudo rm -rf /usr/local/hadoop/hadoop_data/hdfs
hduser@hadoopmaster:~$ mkdir -p /usr/local/hadoop/hadoop_data/hdfs/namenode
hduser@hadoopmaster:~$ sudo chown -R hduser:hduser /usr/local/hadoop
hduser@hadoopmaster:~$
```

- 格式化HDFS目錄，執行hadoop namenode -format

```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

16/11/27 20:10:13 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = hadoopmaster/192.168.59.132
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.7.3
STARTUP_MSG:   classpath = /home/hduser/hadoop-2.7.3/etc/hadoop:/home/hduser/had
oop-2.7.3/share/hadoop/common/lib/asm-3.2.jar:/home/hduser/hadoop-2.7.3/share/ha
doop/common/lib/jsp-api-2.1.jar:/home/hduser/hadoop-2.7.3/share/hadoop/common/li
b/gson-2.2.4.jar:/home/hduser/hadoop-2.7.3/share/hadoop/common/lib/jetty-6.1.26.
jar:/home/hduser/hadoop-2.7.3/share/hadoop/common/lib/jersey-server-1.9.jar:/hom
e/hduser/hadoop-2.7.3/share/hadoop/common/lib/activation-1.1.jar:/home/hduser/ha
doop-2.7.3/share/hadoop/common/lib/commons-io-2.4.jar:/home/hduser/hadoop-2.7.3/
share/hadoop/common/lib/api-util-1.0.0-M20.jar:/home/hduser/hadoop-2.7.3/share/h
adoop/common/lib/commons-logging-1.1.3.jar:/home/hduser/hadoop-2.7.3/share/hadoo
p/common/lib/hadoop-annotations-2.7.3.jar:/home/hduser/hadoop-2.7.3/share/hadoop
```

5. 啟動Hadoop Multi Nodes Cluster
   - 切換至HadoopMaster
   - 執行 start-all.sh


```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [hadoopmaster]
The authenticity of host 'hadoopmaster (192.168.59.137)' can't be established.
ECDSA key fingerprint is SHA256:l5HfVz2GKon2xpmavQSLRqfvPdxuogiqaF/Xjx5XV3E.
Are you sure you want to continue connecting (yes/no)? yes
hadoopmaster: Warning: Permanently added 'hadoopmaster,192.168.59.137' (ECDSA) t
o the list of known hosts.
hadoopmaster: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser
-namenode-hadoopmaster.out
hadoopslave2: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser
-datanode-hadoopslave2.out
hadoopslave1: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser
-datanode-hadoopslave1.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resource
manager-hadoopmaster.out
hadoopslave1: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduse
r-nodemanager-hadoopslave1.out
hadoopslave2: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduse
r-nodemanager-hadoopslave2.out
```

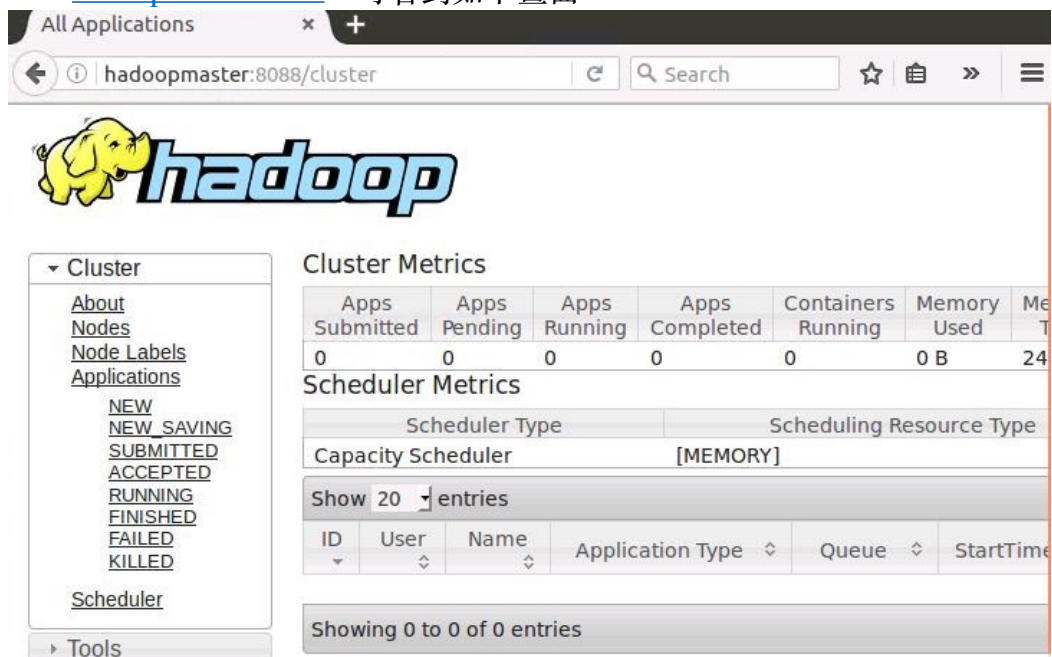   - 查看 HadoopMaster所執行的行程，輸入 jps


```
hduser@hadoopmaster: ~
hduser@hadoopmaster:~$ jps
33605 ResourceManager
33240 NameNode
33883 Jps
33454 SecondaryNameNode
hduser@hadoopmaster:~$
```

   - 查看HadoopSlave1、HadoopSlave2的行程，輸入 jps

- 開啟Hadoop ResourceManager Web介面，開啟瀏覽器，輸入http://hadoopmaster:8088，可看到如下畫面

- 查看已執行的節點，於瀏覽器➔Nodes，可看到目前有三個節點

- 開啟NameNode HDFS Web介面，於瀏覽器，輸入http://hadoopmaster:50070

- 查看DataNodes

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Faile Volu
hadoopslave2:50010 (192.168.59.135:50010)	1	In Service	17.59 GB	24 KB	5.6 GB	11.99 GB	0	24 KB (0%)	0
hadoopslave1:50010 (192.168.59.134:50010)	1	In Service	17.59 GB	24 KB	5.6 GB	11.99 GB	0	24 KB (0%)	0

6. 欲結束Hadoop，執行stop-all.sh