Spark Clusters Install

☆ Install Scala (Master)

\# 安裝 Scala
```
cd /tmp; wget https://downloads.lightbend.com/scala/2.11.8/scala-
  2.11.8.tgz
tar zxvf /tmp/scala-2.11.8.tgz
mv /tmp/scala-2.11.8 /opt
```

\# 建立軟連結
```
ln -s /opt/scala-2.11.8 /opt/scala
```

\# 變更環境變數
```
vi /etc/profile
export SCALA_HOME=/opt/scala
export PATH=$SCALA_HOME/bin:$PATH
```

\# 重啟
```
source /etc/profile
```

\# 查看 Scala 版本，確認是否安裝成功
```
scala -version
```

\# 複製檔案至 Slaver1, Slaver2
```
scp -rp /opt/scala-2.11.8/ root@slaver1:/opt/scala-2.11.8
scp -rp /opt/scala-2.11.8/ root@slaver2:/opt/scala-2.11.8
scp -rp /etc/profile root@slaver1:/etc/profile
scp -rp /etc/profile root@slaver2:/etc/profile
```

\# 建立軟連結 (在 Slaver1, Slaver2 做)
```
ln -s /opt/scala-2.11.8 /opt/scala
```

☆ Install spark-2.3.1-bin-hadoop2.7 (Master)

安裝 spark-2.3.1-bin-hadoop2.7
```
cd /tmp; wget https://archive.apache.org/dist/spark/spark-
  2.3.1/spark-2.3.1-bin-hadoop2.7.tgz
tar zxvf /tmp/spark-2.3.1-bin-hadoop2.7.tgz
mv /tmp/spark-2.3.1-bin-hadoop2.7 /opt
```

變更環境變數
```
vi /etc/profile
export SPARK_HOME=/opt/spark-2.3.1-bin-hadoop2.7
export PATH=$SPARK_HOME/bin:$PATH
export PYSPARK_PYTHON=python3
export PYSPARK_DRIVER_PYTHON=ipython3
```

重啟
source /etc/profile

建立 Slaver 資訊
/opt/spark-2.3.1-bin-hadoop2.7/conf/slaves
加入
master
slaver1
slaver2

複製檔案至 Saver1, Slaver2
```
scp -rp /opt/spark-2.3.1-bin-hadoop2.7/ root@slaver1:/opt/spark-
  2.3.1-bin-hadoop2.7
scp -rp /opt/spark-2.3.1-bin-hadoop2.7/ root@slaver2:/opt/spark-
  2.3.1-bin-hadoop2.7
scp -rp /etc/profile root@slaver1:/etc/profile
scp -rp /etc/profile root@slaver2:/etc/profile
```

查看 Spark 版本，確認是否安裝成功
which spark-submit

☆ Other settings (Every server)

建立軟連結 (!!!在每台電腦做!!!)
```
ln -s /usr/bin/python3 /usr/bin/python
ln -s /usr/bin/pip3 /usr/bin/pip
```

☆ Start Spark Clusters

\# Start the spark cluster
```
cd /opt/spark-2.3.1-bin-hadoop2.7/sbin
./start-all.sh
```

☆ Reboot

\# 啟動 Zookeeper, Journalnode (!!!在每台電腦做!!!)
```
/opt/zookeeper/bin/zkServer.sh start
hadoop-daemon.sh start journalnode
```

\# 啟動 HDFS (在 Master 做)
```
start-all.sh
```

\# 啟動 Spark
```
cd /opt/spark-2.3.1-bin-hadoop2.7/sbin
./start-all.sh
```

☆ Success