

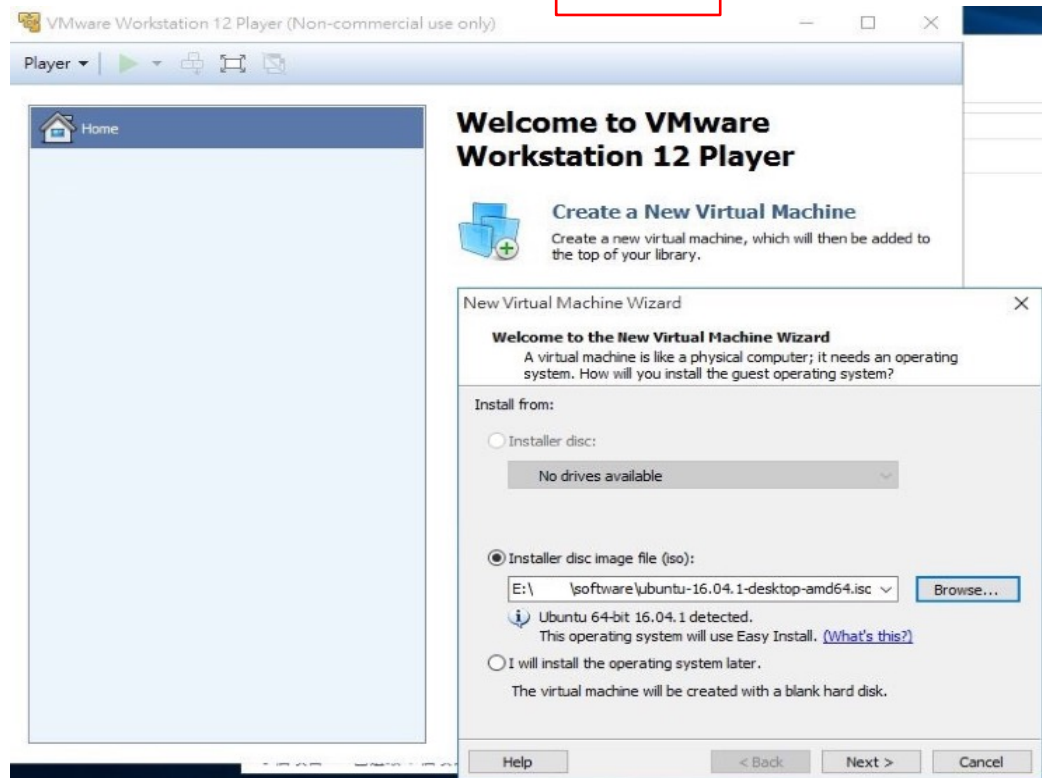
Hadoop Single Node Cluster Setup

1. install Ubuntu

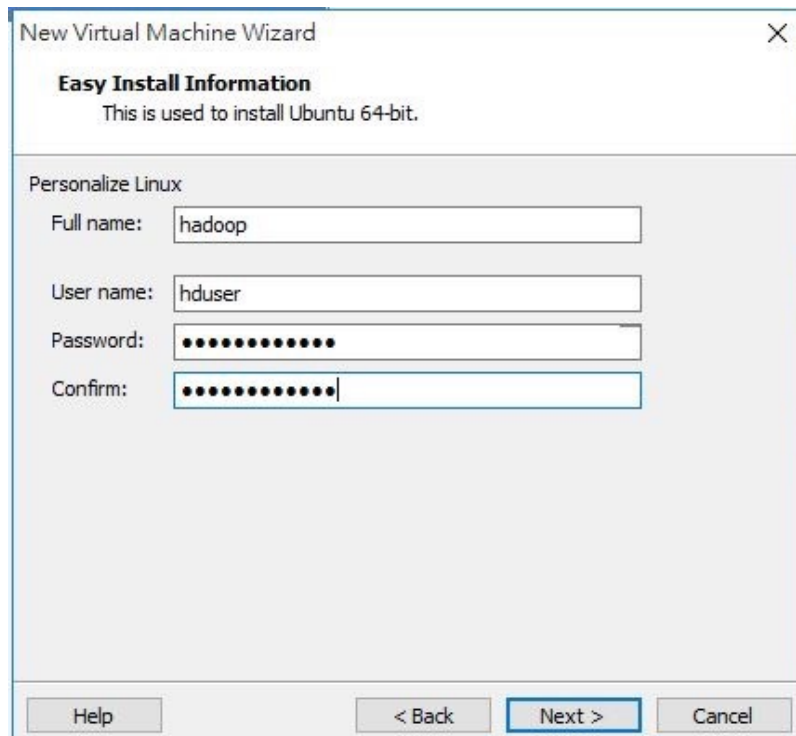
- 下載 Ubuntu，選擇最新的64位元桌面版本



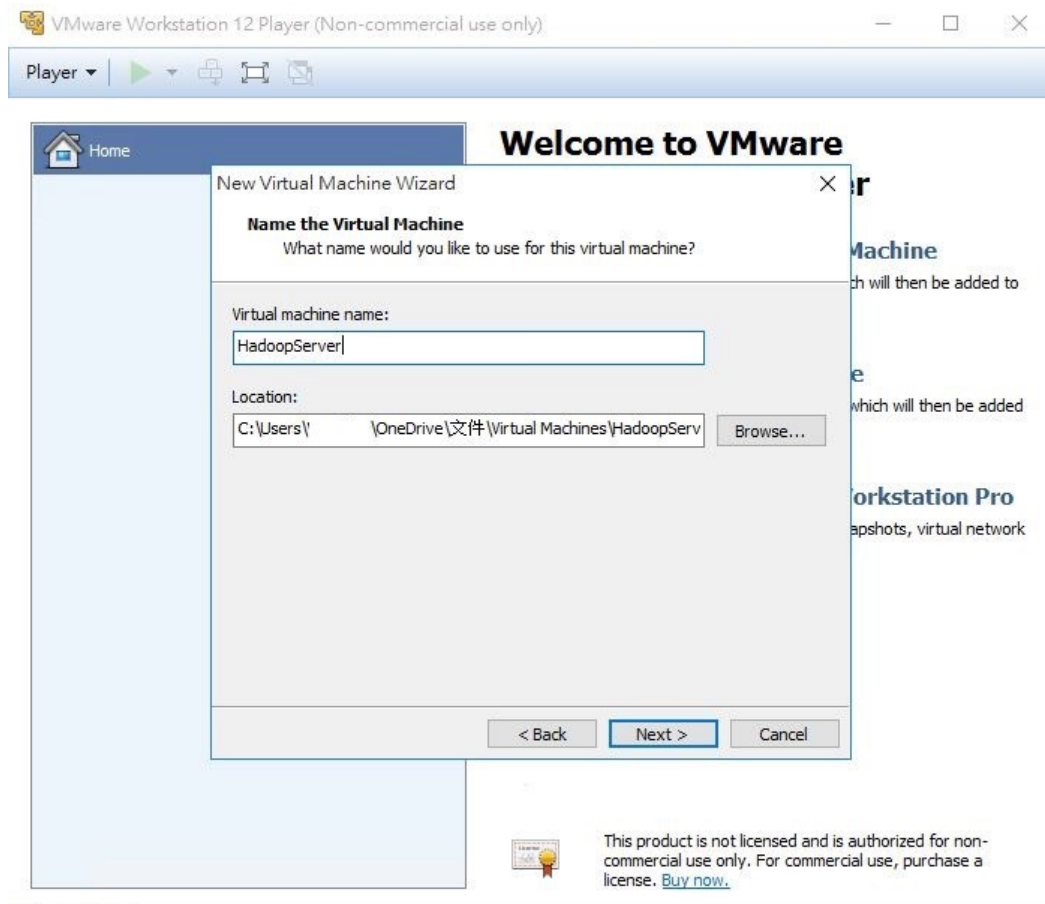
- create a new VM



- 輸入管理員帳密→hduser / hduser



- 輸入伺服器名稱



- 一個 VM 分配 20GB

New Virtual Machine Wizard

Specify Disk Capacity

How large do you want this disk to be?

The virtual machine's hard disk is stored as one or more files on the host computer's physical disk. These file(s) start small and become larger as you add applications, files, and data to your virtual machine.

Maximum disk size (GB):

Recommended size for Ubuntu 64-bit: 20 GB

☐ Store virtual disk as a single file

☒ Split virtual disk into multiple files

Splitting the disk makes it easier to move the virtual machine to another computer but may reduce performance with very large disks.

Help

< Back

Next >

Cancel

New Virtual Machine Wizard

Ready to Create Virtual Machine

Click Finish to create the virtual machine and start installing Ubuntu 64-bit and then VMware Tools.

The virtual machine will be created with the following settings:

Name:	hadoopServer
Location:	C:\Users\WeiShun\OneDrive\文件\Virtual Machines\...
Version:	Workstation 12.0
Operating System:	Ubuntu 64-bit
Hard Disk:	20 GB, Split
Memory:	1024 MB
Network Adapter:	NAT
Other Devices:	CD/DVD, USB Controller, Printer, Sound Card

Customize Hardware...

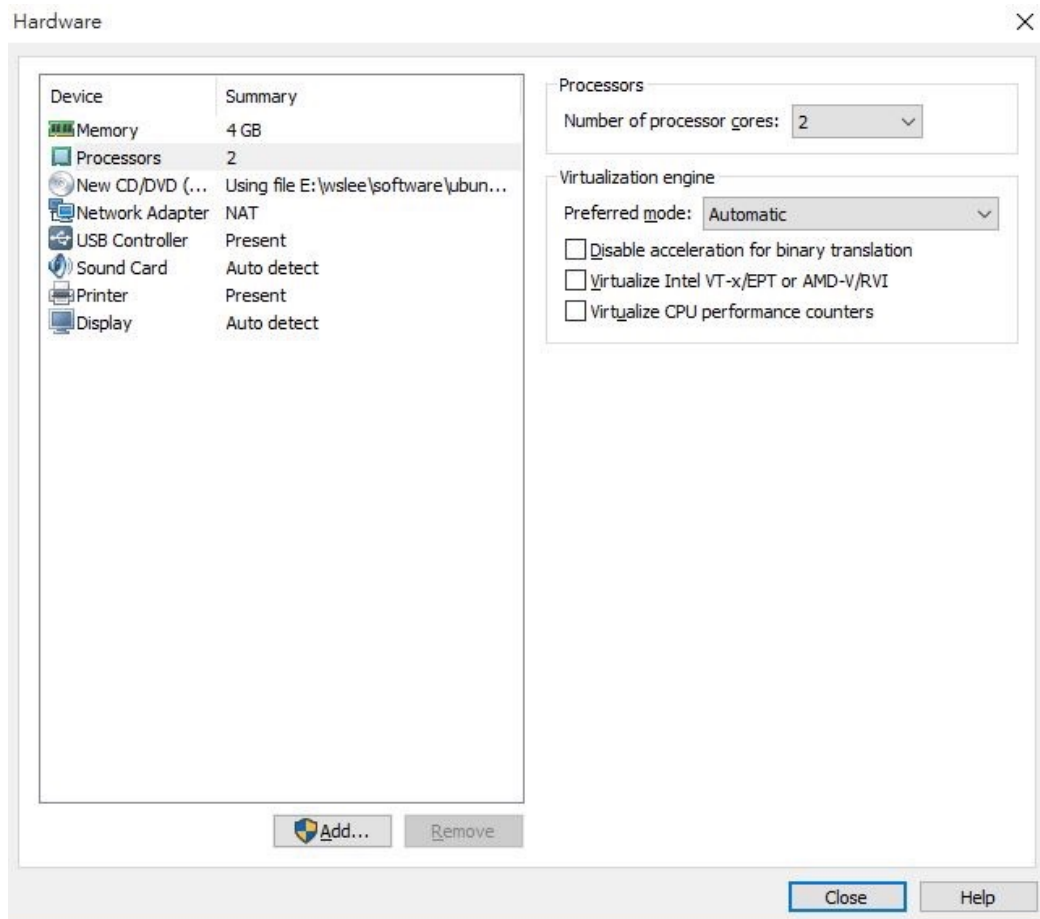
☒ Power on this virtual machine after creation

< Back

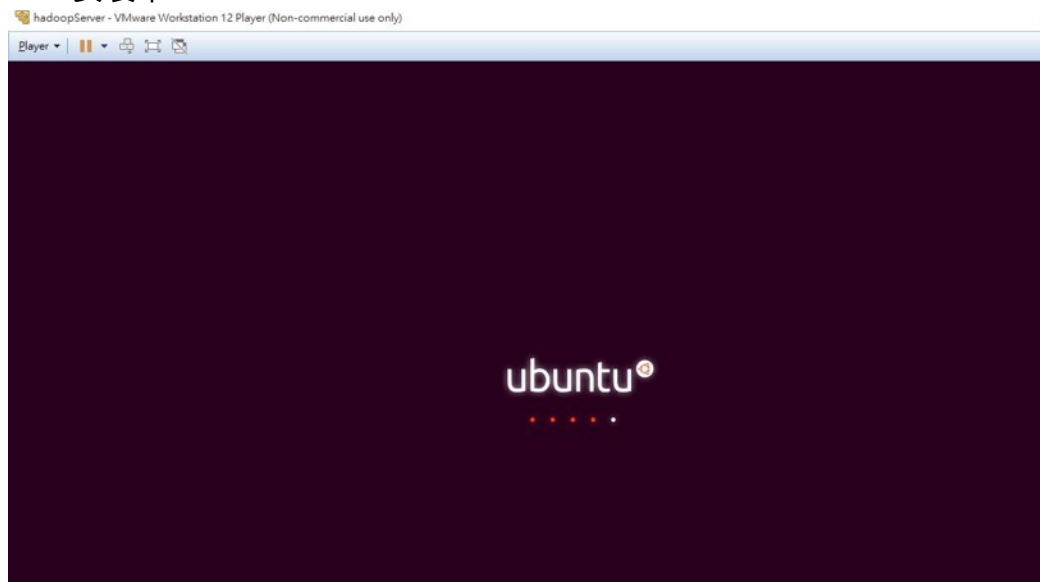
Finish

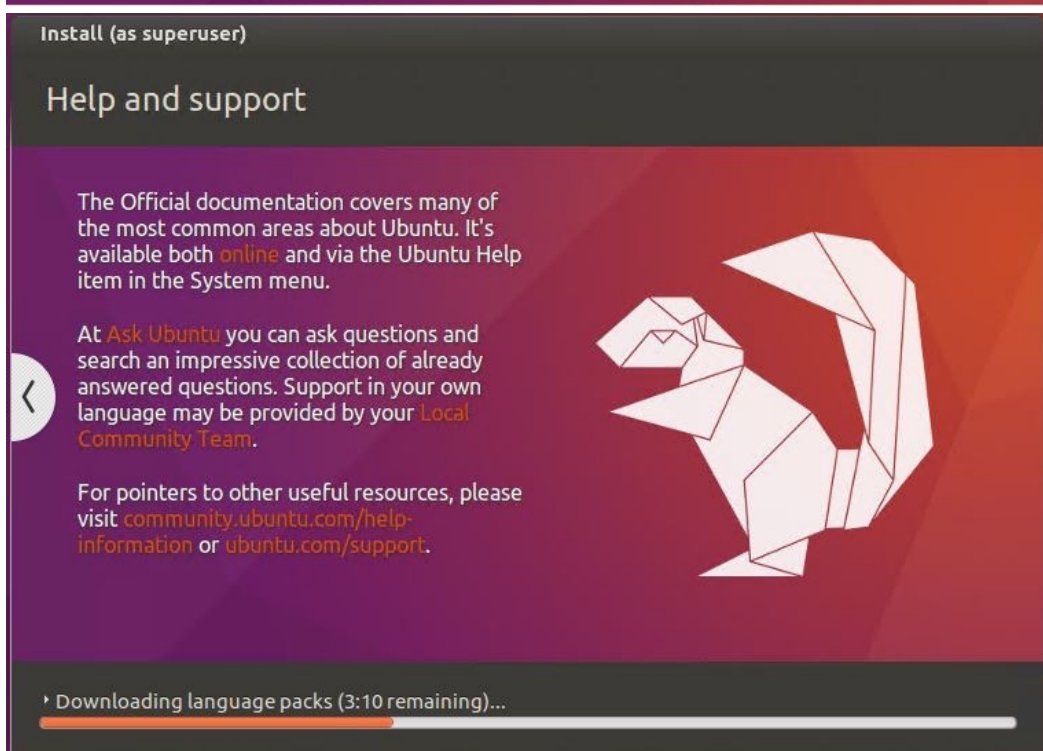
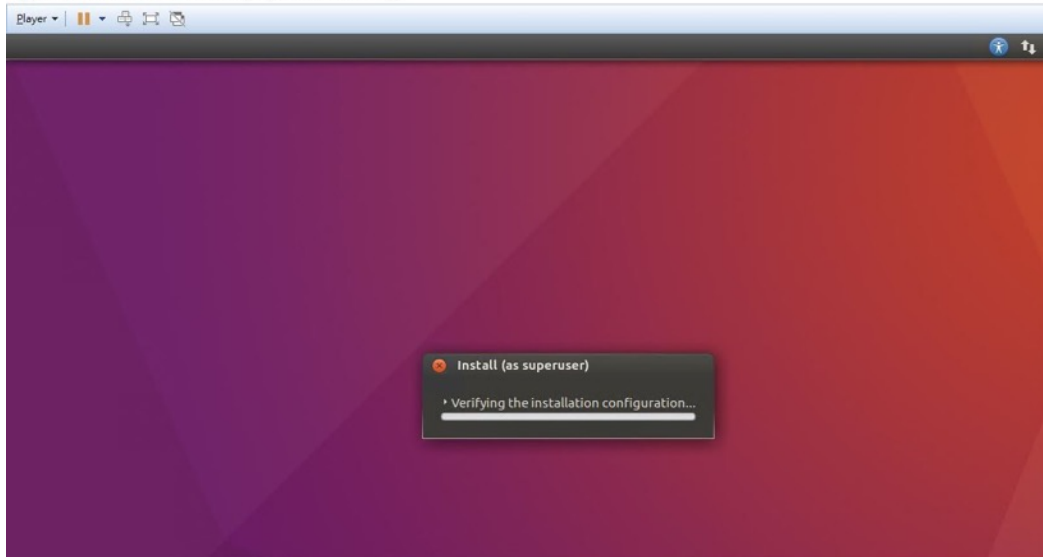
Cancel

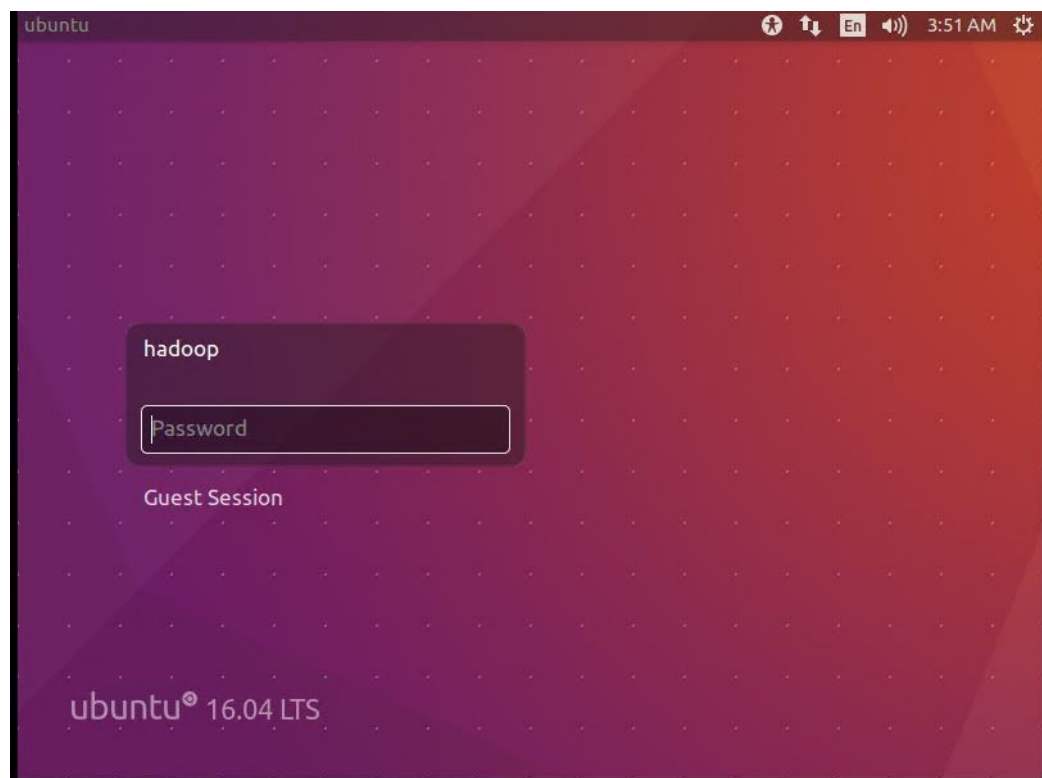
- 選擇 2 processors 和 4GB RAM



- 安裝中~







2. install Java JDK

- 開啟一個terminal，輸入 `sudo apt-get install default-jdk`，讓 apt-get 去抓最新的JDK


```
hduser@ubuntu: ~  
hduser@ubuntu:~$ sudo apt-get install default-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
  ca-certificates-java default-jdk-headless default-jre default-jre-headless  
  fonts-dejavu-extra java-common libbonobo2-0 libbonobo2-common libgif7  
  libgnome-2-0 libgnome2-common libgnomevfs2-0 libgnomevfs2-common libice-dev  
  liborbit-2-0 libpthread-stubs0-dev libsm-dev libx11-dev libx11-doc  
  libxau-dev libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk  
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless  
  x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools  
  xtrans-dev
```

- 安裝完畢，可試著輸入 `java -version` 和 `javac -version` 檢查

```
hduser@ubuntu:~$ java -version  
openjdk version "1.8.0_111"  
OpenJDK Runtime Environment (build 1.8.0_111-8u111-b14-2ubuntu0.16.04.2-b14)  
OpenJDK 64-Bit Server VM (build 25.111-b14, mixed mode)  
hduser@ubuntu:~$ javac -version  
javac 1.8.0_111
```

- 輸入 `sudo update-alternatives --display java` 取得JDK的安裝路徑

```
hduser@ubuntu: ~  
hduser@ubuntu:~$ sudo update-alternatives --display java  
java - auto mode  
link best version is /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java  
link currently points to /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java  
link java is /usr/bin/java  
slave java.1.gz is /usr/share/man/man1/java.1.gz  
/usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java - priority 1081  
slave java.1.gz: /usr/lib/jvm/java-8-openjdk-amd64/jre/man/man1/java.1.gz  
hduser@ubuntu:~$
```

3. Install SSH

- 之後連線到其他伺服器時可重複輸入密碼
- 輸入 `sudo apt-get install ssh` 安裝 SSH

```
hduser@ubuntu: ~  
hduser@ubuntu:~$ sudo apt-get install ssh  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
  ncurses-term openssh-client openssh-server openssh-sftp-server ssh-import-id  
Suggested packages:  
  ssh-askpass libpam-ssh keychain monkeysphere rssh molly-guard  
The following NEW packages will be installed:  
  ncurses-term openssh-server openssh-sftp-server ssh ssh-import-id  
The following packages will be upgraded:  
  openssh-client  
1 upgraded, 5 newly installed, 0 to remove and 280 not upgraded.  
Need to get 643 kB/1,230 kB of archives.  
After this operation, 5,244 kB of additional disk space will be used.  
Do you want to continue? [Y/n]
```

- 安裝 rsync，輸入 `sudo apt-get install rsync`

```
hduser@ubuntu: ~  
hduser@ubuntu:~$ sudo apt-get install rsync  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
rsync is already the newest version (3.1.1-3ubuntu1).  
rsync set to manually installed.  
0 upgraded, 0 newly installed, 0 to remove and 280 not upgraded.  
hduser@ubuntu:~$
```

- 輸入 `ssh-keygen -t rsa -P ''` 產生 public/private key pair

```
hduser@ubuntu:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):
Created directory '/home/hduser/.ssh'.
Your identification has been saved in /home/hduser/.ssh/id_rsa.
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:QgwqWxtL0boiz0soDZQ8rAkLg91iTBpp5oX+/SAwU58 hduser@ubuntu
The key's randomart image is:
+---[RSA 2048]-----+
|..+. |
|=%.*+ |
|&*X B + |
|=%.= E |
|= * o . S |
| + + o . |
|= + . o |
|o= . |
| +. |
+---[SHA256]-----+
```

- 查看產生的SSH key，輸入 `ll ~/.ssh`

```
hduser@ubuntu:~$ ll ~/.ssh
total 16
drwx----- 2 hduser hduser 4096 Dec  3 11:54 ./
drwxr-xr-x 16 hduser hduser 4096 Dec  3 11:54 ../
-rw----- 1 hduser hduser  672 Dec  3 11:54 id_dsa
-rw-r--r-- 1 hduser hduser  603 Dec  3 11:54 id_dsa.pub
hduser@ubuntu:~$
```

- 輸入 `cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys`
設定 key

```
hduser@ubuntu:~$ cat $HOME/.ssh/id_rsa.pub
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQCAQCq3BD/8BABDIu8Etksi05gp4XlGk0az7rXTp9iLKp8
DclJBBwC09L7pcPwnIjkLX0hZ0xBZdPzPjLgYgWzrKKiFNCEFTllWw128pDF0Tfwd+YwZtXfTLfH0LSx
mTKI/VT0CiJgsbbeuiJFAVA0wtudHM+PRndkvDatyLBb0+i6FjWaVap0ayUtpY3DHL1SK/pTNrr7oZoH
g3T2YwuHLN010+7aD55jsr7849lFPSc0NnGPkLeGz1iUNUFLwnrOpG4XOCiSaVl5cahtsSEb3+VLMqZC
RFyde13mMQ3widCAYe2DG0irJ+XmDkeBYLGiWN6Wi+tnL+NupUmv0LvrZN6v hduser@ubuntu
hduser@ubuntu:~$ cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

- 啟動SSH `ssh localhost`

```
hduser@ubuntu:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:wN0apnZoV21KYAfdRRk/+r0WU9BVJJCjH91g/oYvk04.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
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

236 packages can be updated.
82 updates are security updates.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
```

4. Install Hadoop

- 連線至 Hadoop ([網址](#))

Apache Hadoop Rel... x Ubuntu Start Page x +

hadoop.apache.org/releases.html

convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-256.

Version	Release Date	Tarball	GPG	SHA-256
3.0.0-alpha1	03 September, 2016	source	signature	checksum file
		binary	signature	checksum file
2.7.3	25 August, 2016	source	signature	227785DC 6E3E6EF8..
		binary	signature	D489DF38 08244B90..
2.6.5	08 October, 2016	source	signature	3A843F18 73D9951A..
		binary	signature	001AD18D 4B6D0FE5..
2.5.2	19 Nov, 2014	source	signature	139EF872 09C5637E..
		binary	signature	0BDB4850 A3825208..

To verify Hadoop releases using GPG:

- 下載最新版 Hadoop-2.7.3.tar.gz，輸入 `wget http://apache.stu.edu.tw/hadoop/common/hadoop-2.7.3.tar.gz`

Apache Download ... x Ubuntu Start Page x +

www.apache.org/dyn/closer.cgi/hadoop,

The Apache Way

Contribute

ASF Sponsors

We suggest the following mirror site for your download:

<http://apache.stu.edu.tw/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz>

Other mirror sites are suggested below. Please use the backup mirrors only to download PGP and MD5 signatures to verify your downloads or if no other mirrors are working.

```

hduser@ubuntu: ~
hduser@ubuntu:~$ wget http://apache.stu.edu.tw/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
--2016-12-03 15:22:50-- http://apache.stu.edu.tw/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
Resolving apache.stu.edu.tw (apache.stu.edu.tw)... 120.119.118.1, 2001:e10:c41:eee::1
Connecting to apache.stu.edu.tw (apache.stu.edu.tw)|120.119.118.1|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 214092195 (204M) [application/x-gzip]
Saving to: 'hadoop-2.7.3.tar.gz'

hadoop-2.7.3.tar.gz  0%[
] 967.11K  263KB/s  eta 14m 3s

```

- 解壓縮 Hadoop-2.7.3.tar.gz 在 Terminal 中輸入 `sudo tar -zxvf Hadoop-2.7.3.tar.gz`

```
hduser@ubuntu:~$ sudo tar -zxvf hadoop-2.7.3.tar.gz
```

- 預設安裝路徑為 /usr/local，輸入 `sudo mv hadoop-2.7.3 /usr/local/hadoop`
- 查看hadoop安裝目錄，輸入 `ll /usr/local/hadoop`

```
hduser@ubuntu: ~  
hduser@ubuntu:~$ sudo mv hadoop-2.7.3 /usr/local/hadoop  
hduser@ubuntu:~$ ll /usr/local/hadoop  
total 140  
drwxr-xr-x  9 root root  4096 Aug 18 09:49 ./  
drwxr-xr-x 11 root root  4096 Dec  3 15:36 ../  
drwxr-xr-x  2 root root  4096 Aug 18 09:49 bin/  
drwxr-xr-x  3 root root  4096 Aug 18 09:49 etc/  
drwxr-xr-x  2 root root  4096 Aug 18 09:49 include/  
drwxr-xr-x  3 root root  4096 Aug 18 09:49 lib/  
drwxr-xr-x  2 root root  4096 Aug 18 09:49 libexec/  
-rw-r--r--  1 root root 84854 Aug 18 09:49 LICENSE.txt  
-rw-r--r--  1 root root 14978 Aug 18 09:49 NOTICE.txt  
-rw-r--r--  1 root root  1366 Aug 18 09:49 README.txt  
drwxr-xr-x  2 root root  4096 Aug 18 09:49 sbin/  
drwxr-xr-x  4 root root  4096 Aug 18 09:49 share/  
hduser@ubuntu:~$
```

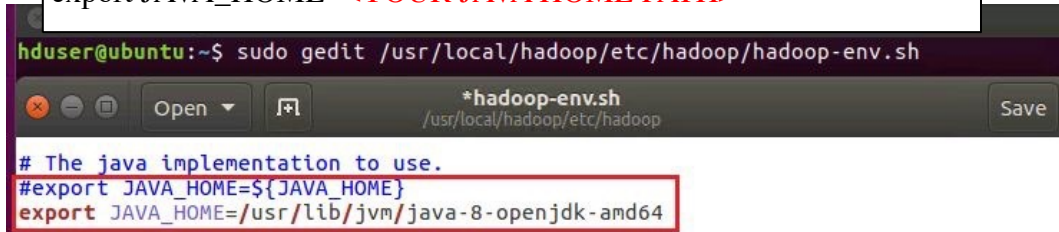
- 設定環境變數，編輯 .bashrc 輸入 `sudo gedit ~/.bashrc`，在檔案的最後加入下列字串(注意JAVA_HOME為本機的路徑)

```
hduser@ubuntu:~$ sudo gedit ~/.bashrc  
[sudo] password for hduser:  
  
hduser@ubuntu: ~  
hduser@ubuntu:~$ sudo gedit ~/.bashrc  
*.bashrc  
Save  
#Hadoop Variables  
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64  
export HADOOP_HOME=/usr/local/hadoop  
export PATH=$PATH:$HADOOP_HOME/bin  
export PATH=$PATH:$HADOOP_HOME/sbin  
export HADOOP_MAPRED_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native  
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"  
export JAVA_LIBRARY_PATH=$HADOOP_HOME/lib/native:$JAVA_LIBRARY_PATH  
#Hadoop Variables
```

```
#Hadoop Variables  
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64  
export HADOOP_HOME=/usr/local/hadoop  
export PATH=$PATH:$HADOOP_HOME/bin  
export PATH=$PATH:$HADOOP_HOME/sbin  
export HADOOP_MAPRED_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native  
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"  
export JAVA_LIBRARY_PATH=$HADOOP_HOME/lib/native:$JAVA_LIBRARY_PATH  
#Hadoop Variables
```

- 接下來編輯 Hadoop-env.sh，輸入 `sudo gedit /usr/local/hadoop/etc/hadoop/hadoop-env.sh`
- 在檔案中輸入下列字串(注意JAVA_HOME為本機的路徑)

```
export JAVA_HOME=<YOUR JAVA HOME PATH>
```

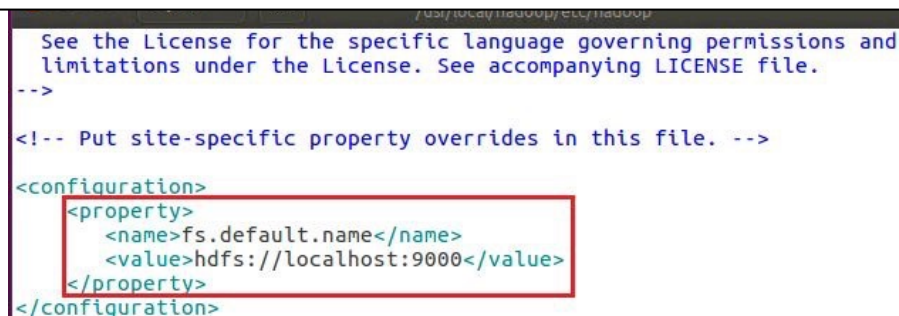


```
hduser@ubuntu:~$ sudo gedit /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

```
# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

- 編輯core-site.xml，輸入 `sudo gedit /usr/local/hadoop/etc/hadoop/core-site.xml`
- 在 xml 檔中，<configuration> tag 間輸入下列字串

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



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

- 編輯yarn-site.xml，輸入 `sudo gedit /usr/local/hadoop/etc/hadoop/yarn-site.xml`
- 在 xml 檔中，<configuration> tag 間輸入下列字串

```
hduser@ubuntu: ~  
<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.hadoop.mapred.ShuffleHandler</value>  
</property>  
  
<property>  
  <name>mapreduce.framework.name</name>  
  <value>yarn</value>  
</property>  
  
*mapred-site.xml  
/usr/local/hadoop/etc/hadoop  
Save  
  
<!-- Put site-specific property overrides in this file. -->  
<configuration>  
  <property>  
    <name>mapreduce.framework.name</name>  
    <value>yarn</value>  
  </property>  
</configuration>
```

- 編輯hdfs-site.xml，輸入 `sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml`
- 在 xml 檔中，<configuration> tag 間輸入下列字串


```
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value> file:/usr/local/hadoop/hadoop_data/hdfs/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value> file:/usr/local/hadoop/hadoop_data/hdfs/datanode</value>
</property>
```

v. `sudo chown hduser:hduser -R /usr/local/hadoop`

```
hduser@ubuntu: ~
hduser@ubuntu:~$ sudo mkdir -p /usr/local/hadoop/hadoop_data/hdfs/namenode
hduser@ubuntu:~$ sudo mkdir -p /usr/local/hadoop/hadoop_data/hdfs/datanode
hduser@ubuntu:~$ sudo chown hduser:hduser -R /usr/local/hadoop/
hduser@ubuntu:~$
```

- 格式化 Hadoop file system，輸入 Hadoop namenode -format

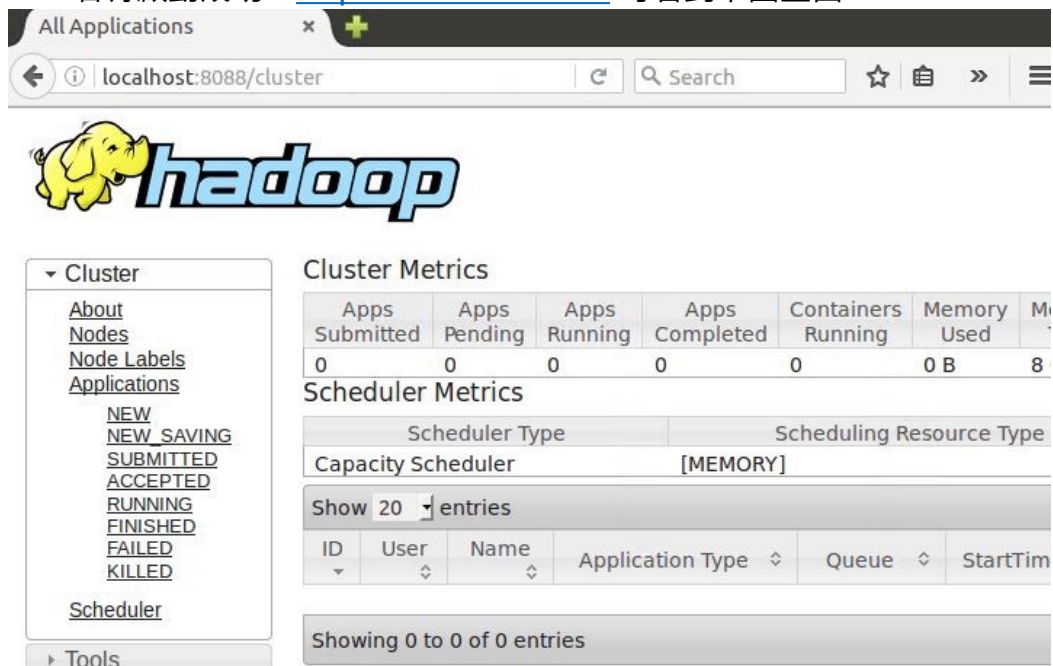
```
hduser@ubuntu: ~
hduser@ubuntu:~$ hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

16/12/03 17:10:41 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = ubuntu/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.7.3
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/
hadoop/common/lib/hadoop-auth-2.7.3.jar:/usr/local/hadoop/share/hadoop/common/li
b/jsch-0.1.42.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-logging-1.1.
53 bytes saved in 0 seconds.
16/12/03 17:10:47 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
16/12/03 17:10:47 INFO util.ExitUtil: Exiting with status 0
16/12/03 17:10:47 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at ubuntu/127.0.1.1
*****/
```

- 輸入 `start-all.sh` 啟動 Hadoop

```
hduser@ubuntu: ~  
hduser@ubuntu:~$ start-all.sh  
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh  
Starting namenodes on [localhost]  
hduser@localhost's password:  
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-na  
menode-ubuntu.out  
hduser@localhost's password:  
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-da  
tanode-ubuntu.out  
Starting secondary namenodes [0.0.0.0]  
hduser@0.0.0.0's password:  
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hd  
user-secondarynamenode-ubuntu.out  
starting yarn daemons  
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resource  
manager-ubuntu.out  
hduser@localhost's password:  
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-n  
odemanager-ubuntu.out  
hduser@ubuntu:~$
```

- 若有啟動成功，<http://localhost:8088> 可看到下圖畫面



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	...
0	0	0	0	0	0 B	8...

Scheduler Metrics

Scheduler Type		Scheduling Resource Type	
Capacity Scheduler		[MEMORY]	

Showing 0 to 0 of 0 entries

- <http://localhost:50070> 檢視 datanode 或 file system

Namenode information

+

←

localhost:50070/dfshealth.html#tab-overvie

↻

Search

☆

📄

»

☰

Startup Progress

Utilities

Overview 'localhost:54310' (active)

Started:	Sun Nov 13 16:52:16 CST 2016
Version:	2.7.3, rbaa91f7c6bc9cb92be5982de4719c1c8af91ccff
Compiled:	2016-08-18T01:41Z by root from branch-2.7.3
Cluster ID:	CID-fe287edb-4de7-45f2-afa6-9190189edf6f
Block Pool ID:	BP-2047880570-127.0.1.1-1479026925454