

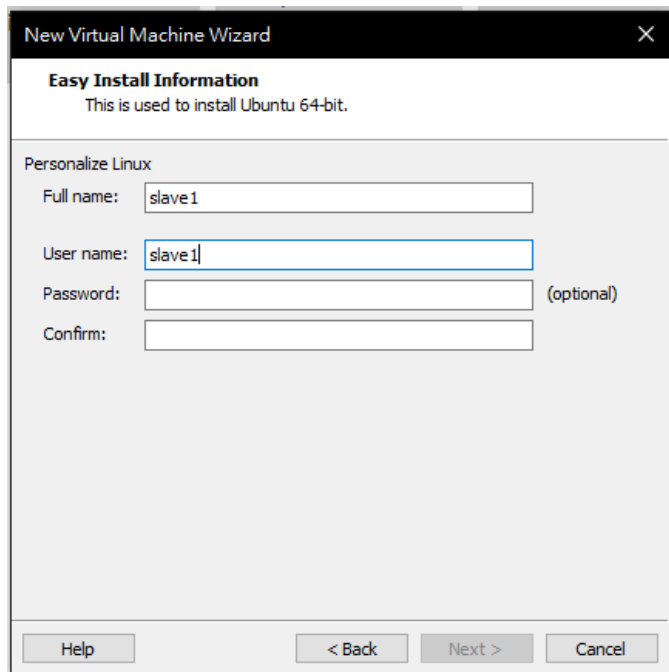
#####

#Hadoop 多節點安裝

#####

※ 如果 **gedit** 無法開啟文件，請重開 **terminal** 再試一次。

新增 slave 主機：(設置如下)



New Virtual Machine Wizard

Easy Install Information
This is used to install Ubuntu 64-bit.

Personalize Linux

Full name:

User name:

Password: (optional)

Confirm:

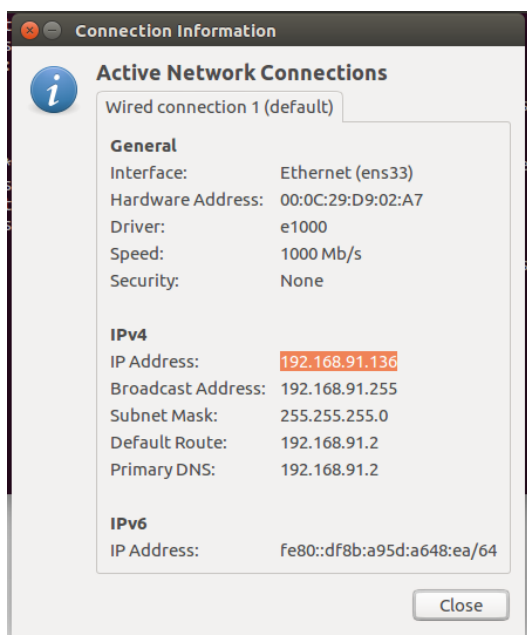
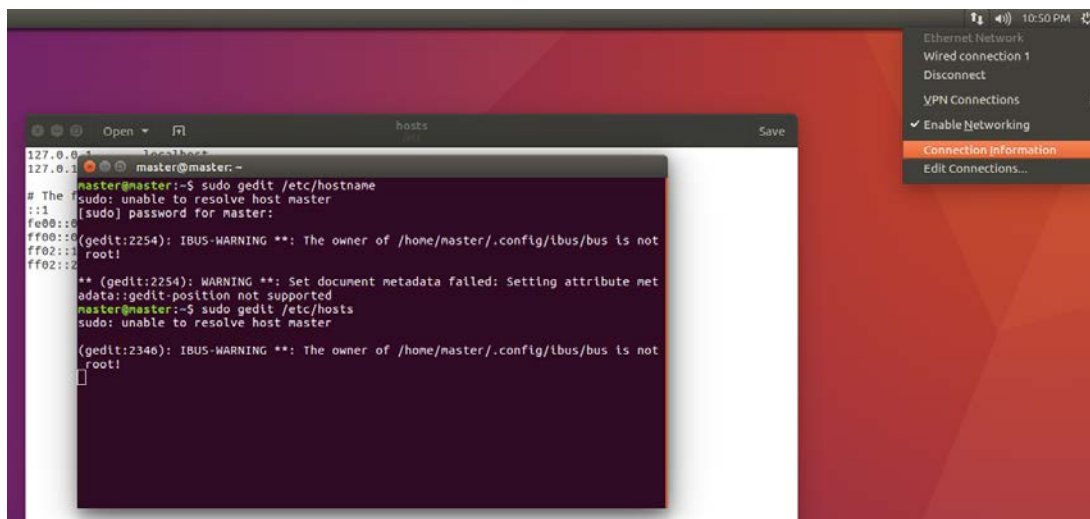
Help < Back Next > Cancel

1. 修改**所有主機**名稱：master 節點設置為 master，slave 節點設置為 slave1。

`sudo gedit /etc/hostname`

設置完須重開虛擬機

2. 查詢每台主機的 IP 位址：



3. 編輯所有主機之/etc/hosts：

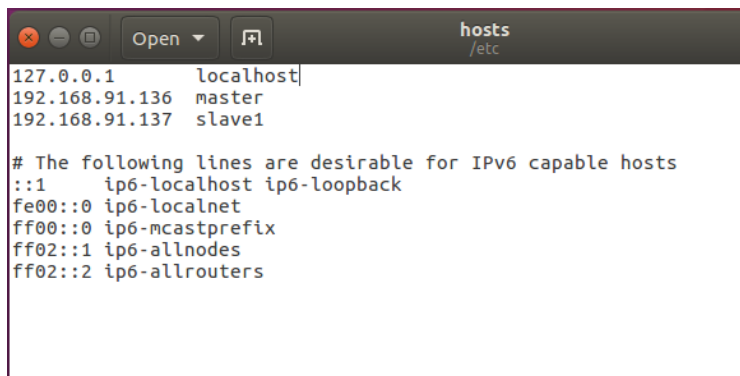
sudo gedit /etc/hosts

將 IP 位址、hostname 對應，寫入/etc/hosts

192.168.91.136 master

192.168.91.137 slave1

(IP 位址) (hostname)



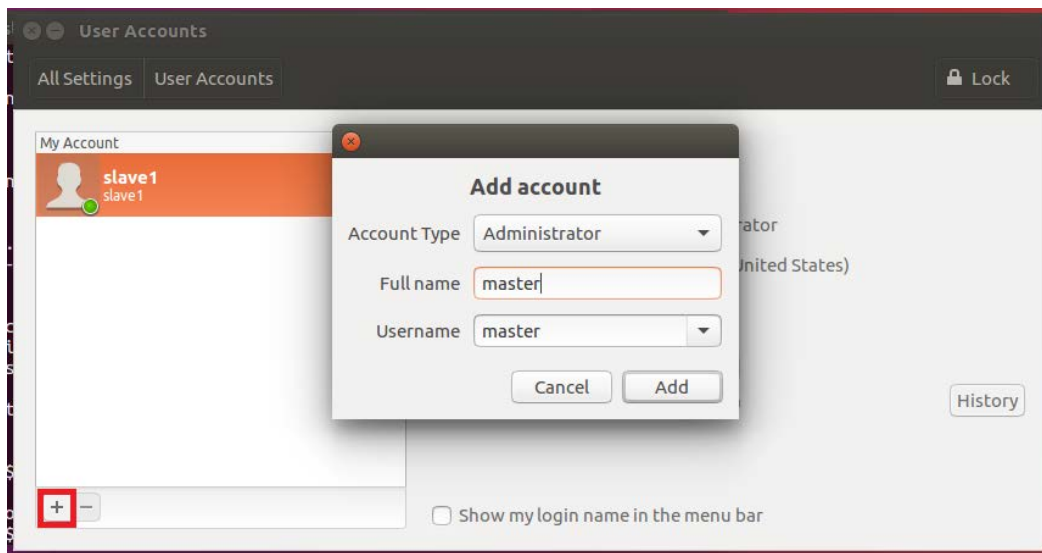
4. 在 slave 建立新使用者

System menu showing options: About This Computer, Ubuntu Help..., System Settings..., Lock (Ctrl+Alt+L), Guest Session, slave1, Log Out..., Suspend, Shut Down...

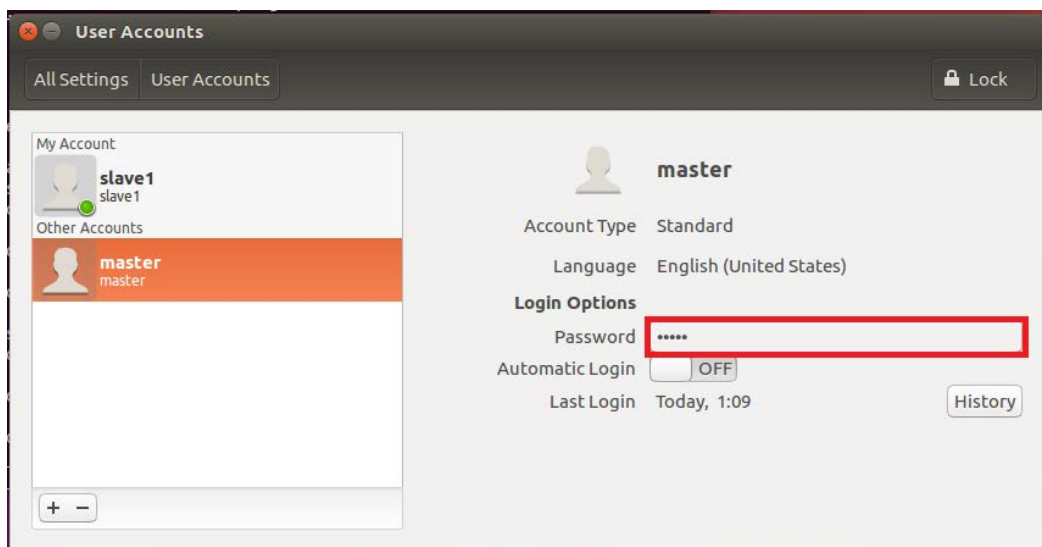
System Settings window showing categories: Personal (Appearance, Brightness & Lock, Language Support, Online Accounts, Security & Privacy, Text Entry), Hardware (Bluetooth, Color, Displays, Keyboard, Mouse & Touchpad, Network, Power, Printers, Sound, Wacom Tablet), and System (Backups, Details, Software & Updates, Time & Date, Universal Access, User Accounts).

User Accounts window showing user 'slave1' (Administrator, English (United States)). An 'Authenticate' dialog box is overlaid, stating: 'Authentication is required to change user data. An application is attempting to perform an action that requires privileges. Authentication is required to perform this action.' The dialog includes a password field and buttons for 'Cancel' and 'Authenticate'.

Name 須設為 master



設置密碼：



5. 在 slave 安裝 ssh：

```
sudo apt-get install openssh-server
```

```
sudo /etc/init.d/ssh start
```

```
sudo su master(更換使用者)
```

```
cd ~
```

```
ssh-keygen -t rsa -P ""
```

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
ssh localhost
```

```
exit
```

```

slave1@slave1:~$ ssh localhost
slave1@slave1:~$ ssh localhost
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.15.0-88-generic x86_64)

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

86 packages can be updated.
0 updates are security updates.

Last login: Tue Feb 25 05:10:28 2020 from 127.0.0.1
exit
slave1@slave1:~$ exit
logout
Connection to localhost closed.
slave1@slave1:~$

```

6. 在 master 輸入指令：

```
ssh-copy-id -i ~/.ssh/id_rsa.pub slave1
```

7. 在 master 用 ssh 登入 slave：

```
ssh slave1
```

```
exit
```

```

Connection to slave1 closed.
master@master:~/hdfs$ ssh-copy-id -i ~/.ssh/id_rsa.pub slave1
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/master/.ssh/id_rsa.pub"
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys
master@slave1's password:

Number of key(s) added: 1

Now try logging into the machine, with: "ssh 'slave1'"
and check to make sure that only the key(s) you wanted were added.

master@master:~/hdfs$ ssh slave1
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.15.0-88-generic x86_64)

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

86 packages can be updated.
0 updates are security updates.

New release '18.04.4 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Wed Feb 26 01:22:17 2020 from 192.168.91.145
master@slave1:~$ exit
logout
Connection to slave1 closed.
master@master:~/hdfs$

```

8. 在 master 下指令修改如下四個檔案，core-site.xml, mapred-site.xml, hdfs-site.xml, yarn-site.xml：

```
sudo gedit /opt/hadoop/etc/hadoop/core-site.xml
```

#core-site.xml，指定 namenode

```
<configuration>
```

```
  <property>
```

```
    <name>fs.defaultFS</name>
```

```
    <value>hdfs://master:9000</value>
```

```
  </property>
```

```
  <property>
```

```
    <name>hadoop.tmp.dir</name>
```

```
    <value>file:/home/master/tmp</value>
```

```
    <description>Abase for other temporary directories.</description>
```

```
  </property>
```

```
</configuration>
```

```
*****
```

```
sudo gedit /opt/hadoop/etc/hadoop/mapred-site.xml
```

#mapred-site.xml 指定 jobtracker

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>master:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>master:19888</value>
  </property>
  <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>
</configuration>
```

sudo gedit /opt/hadoop/etc/hadoop/yarn-site.xml

#yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>master</value>
  </property>
  <property>
    <name>yarn.nodemanager.resource.memory-mb</name>
    <value>10240</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

```

</property>
</configuration>
*****

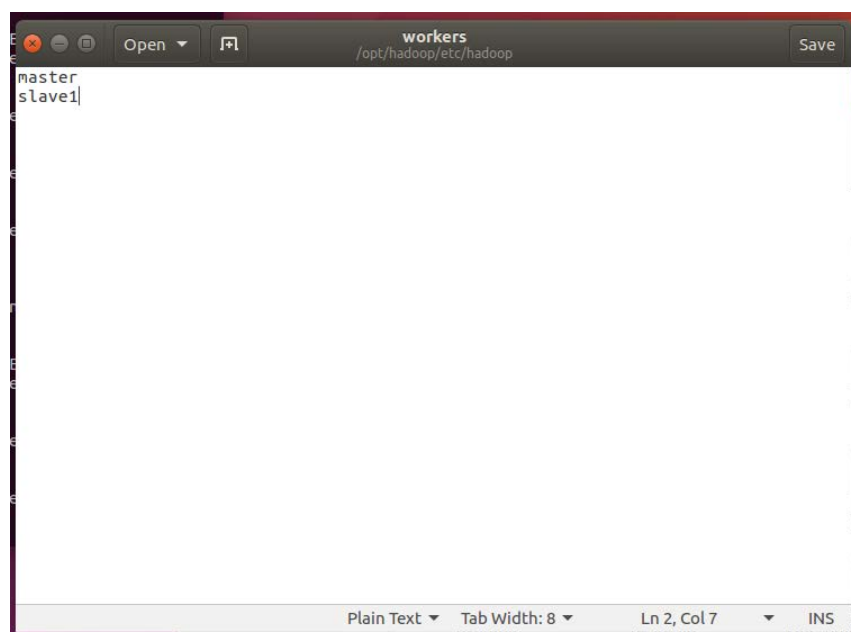
sudo gedit /opt/hadoop/etc/hadoop/hdfs-site.xml
#hdfs-site.xml
<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>master:50090</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <!--注意此處，因為有兩個 slave 節點，所以由 1 改為 2 -->
    <value>2</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/home/master/hdfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/home/master/hdfs/data</value>
  </property>
</configuration>
*****

```

9. 在 master 下指令增加節點：

```
sudo gedit /opt/hadoop/etc/hadoop/workers
```

寫入 master、slave 主機名稱



10. 在 master 下指令，將 master 的環境複製到 slave 中

cd /opt

壓縮 hadoop：

sudo tar -zcf ./hadoop.tar.gz ./hadoop

傳送給 slave 端：(@前應放 slave 的使用者名稱；@後應放 slave 的主機名稱)

scp ./hadoop.tar.gz slave1@slave1:~/Downloads

scp -r jdk slave1@slave1:~/Downloads

11. 在 slave 下指令配置 slave 的環境

su slave1

sudo gedit /etc/sudoers

在 root ALL=(ALL:ALL) ALL

下增加 master ALL=(ALL:ALL) ALL

```
# Cmnd alias specification
# User privilege specification
root    ALL=(ALL:ALL) ALL
master  ALL=(ALL:ALL) ALL
# Members of the admin group may gain root privileges
```

sudo su master (使用者變換)

sudo gedit ~/.bashrc (配置環境變數)

將以下程式碼加在檔案最下面並且儲存

export CLASS_PATH="/opt/jdk/lib"

export JAVA_HOME="/opt/jdk"

export HADOOP_HOME="/opt/hadoop"

export MAVEN_HOME="/opt/maven"

export MAHOUT_HOME="/opt/mahout"

export

PATH=\$PATH:\$JAVA_HOME/bin:\$HADOOP_HOME/bin:\$HADOOP_HOME/sbin:\$MAVEN_HOME/bin:\$MAHOUT_HOME/bin

export HADOOP_CONF_DIR="/opt/hadoop/conf"

export HADOOP_MAPRED_HOME="/opt/hadoop"

export HADOOP_COMMON_HOME="/opt/hadoop"

export HADOOP_HDFS_HOME="/opt/hadoop"

export HADOOP_LOCAL="/opt/hadoop"

export YARN_HOME="/opt/hadoop"


```

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

export CLASS_PATH="/opt/jdk/lib"
export JAVA_HOME="/opt/jdk"
export HADOOP_HOME="/opt/hadoop"
export MAVEN_HOME="/opt/maven"
export MAHOUT_HOME="/opt/mahout"
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$MAVEN_HOME/bin:$MAHOUT_HOME/
bin
export HADOOP_CONF_DIR="/opt/hadoop/conf"
export HADOOP_MAPRED_HOME="/opt/hadoop"
export HADOOP_COMMON_HOME="/opt/hadoop"
export HADOOP_HDFS_HOME="/opt/hadoop"
export HADOOP_LOCAL=""/opt/hadoop"
export YARN_HOME="/opt/hadoop"

```

source ~/.bashrc

sudo su slave1 (使用者變換)

在 slave 下指令配置 java：

cd ~

cd Downloads

sudo mv jdk /opt/jdk (移到/opt 目錄下)

在 slave 下指令配置 hadoop：

cd ~

cd Downloads

tar zxvf hadoop.tar.gz (解壓縮)

sudo mv hadoop /opt/hadoop (移到/opt 目錄下)

sudo chmod -R 777 /opt/hadoop/logs (更改權限)

sudo su master (使用者變換)

source /opt/hadoop/etc/hadoop/hadoop-env.sh (讓配置生效)

java -version (測試安裝正常)

hadoop version (測試安裝正常)

```

master@slave1:/home/slave1$ source /opt/hadoop/etc/hadoop/hadoop-env.sh
master@slave1:/home/slave1$ java -version
java version "1.8.0_241"
Java(TM) SE Runtime Environment (build 1.8.0_241-b07)
Java HotSpot(TM) 64-Bit Server VM (build 25.241-b07, mixed mode)
master@slave1:/home/slave1$ hadoop version
Hadoop 3.2.1
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r b3cbbb4
67e22ea829b3808f4b7b01d07e0bf3842
Compiled by rohithsharmaks on 2019-09-10T15:56Z
Compiled with protoc 2.5.0
From source with checksum 776eaf9eee9c0ffc370bcbcb1888737
This command was run using /opt/hadoop/share/hadoop/common/hadoop-common-3.2.1.j
ar
master@slave1:/home/slave1$

```

12. 在 master 下指令啟用 hadoop (要先刪除 home/hdfs 資料夾內的 name、data)

source /opt/hadoop/etc/hadoop/hadoop-env.sh

hdfs namenode -format

source /opt/hadoop/etc/hadoop/hadoop-env.sh

start-all.sh

完成畫面：

master 主機：

```
master@master: ~  
WARNING: Stopping all Apache Hadoop daemons as master in 10 seconds.  
WARNING: Use CTRL-C to abort.  
Stopping namenodes on [master]  
Stopping datanodes  
Stopping secondary namenodes [master]  
Stopping nodemanagers  
Stopping resourcemanager  
master@master:~$ start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as master in 10 seconds.  
WARNING: This is not a recommended production deployment configuration.  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [master]  
Starting datanodes  
Starting secondary namenodes [master]  
Starting resourcemanager  
Starting nodemanagers  
master@master:~$ jps  
11872 SecondaryNameNode  
12257 NodeManager  
12117 ResourceManager  
11486 NameNode  
11631 DataNode  
12623 Jps  
master@master:~$
```

slave1 主機：

```
master@slave1: ~  
master@slave1:~$ jps  
7955 DataNode  
52430 Jps  
master@slave1:~$ jps  
53040 Jps  
52900 NodeManager  
52747 DataNode  
master@slave1:~$
```


試試以下 2 個網址有無正常運作，<http://master:8088>

All Applications - Mozilla Firefox

All Applications × Namenode information × +

← → ↺ 🏠

🔍 master:8088/cluster



All App

▼ Cluster

[About](#)
[Nodes](#)
[Node Labels](#)
[Applications](#)
[NEW](#)
[NEW_SAVING](#)
[SUBMITTED](#)
[ACCEPTED](#)
[RUNNING](#)
[FINISHED](#)
[FAILED](#)
[KILLED](#)
[Scheduler](#)

► Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	M
0	0	0	0	0	0 I

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocat
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>

Show 20 ▾ entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalSta
▼	◇	◇	◇	◇	◇	◇	◇	◇	◇	

No data a

Showing 0 to 0 of 0 entries

<http://master:9870>

Namenode information - Mozilla Firefox

All Applications × Namenode information × +

← → ↺ 🏠

🔍 master:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Util

Overview 'master:9000' (active)

Started:	Tue Feb 25 06:35:07 -0800 2020
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 08:56:00 -0700 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-17fa0a09-3b61-4b4c-b31f-251f2daf3fce
Block Pool ID:	BP-1318696061-192.168.91.136-1582639258292

Summary