# **Cluster Setup**

Materials: Two computers + One Router + Three Ethernet Cables

**Explanation**: If you have only one computer, it is also possible to establish a Hadoop Cluster using Virtual computers (e.g. installed by VirtualBox). Each computer must go through the <u>Single Node Setup</u> before this Cluster Setup. Computers are connected with

each other via a Router which assigns an IP to each computer.

Computer name	IP address	Role
Captain-CentERdata	192.168.0.2	Master
Sailor01-CentERdat a	192.168.0.10	Slave

# Rename computers

• Open a terminal (Ctrl+Alt+T):

Explanation: As shown in the above figure, the computer name is "ktyang-2009" and the user name is "ktyang"

• Type in the following commands in the terminal:

```
sudo gedit /etc/hostname
sudo gedit /etc/hosts
```

Explanation: "gedit" is a command to edit a file (you can also use command: "vi", or "vim"). "sudo" means run the command with the security privileges of the superuser. Here, two files (i.e. "/etc/hostname" and "/etc/hosts") are open for editing.

• In the two files, replace any instances of the existing computer name (e.g. "ktyang-2009") with a new one (e.g. "Captain-CentERdata"). When complete run:

```
reboot # restart the computer activate the changes.
```

• After restarting the computer, open a terminal (Ctrl+Alt+T) to check whether the computer name has been changed:

```
ktyang@Captain-CentERdata: ~
ktyang@Captain-CentERdata: ~$
ComputerName has been changed!
```

Repeating the above steps to rename the other computer as "Sailor01-CentERdata".

# Connect computers

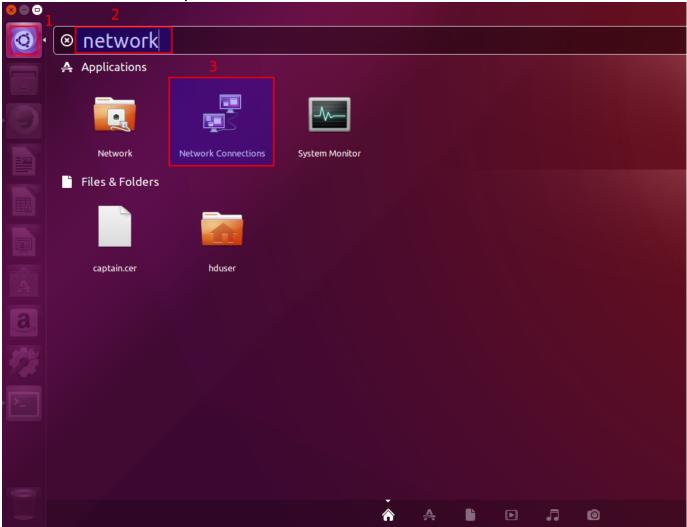
• Connecting computers with a router:

03/14/2017 1/13



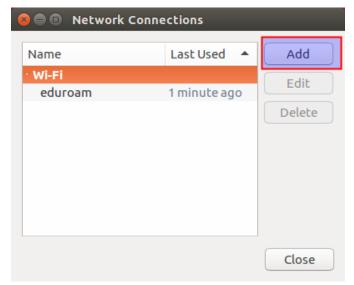
Normally, a Router contains one Internet port, and several Ethernet ports. The Internet port is connected with the Internet port on the wall. Each Ethernet port is connected with a computer.

• Set the IP address for each computer



Click the "Search you computer" => Type in "network" => Click the "Network Connections".

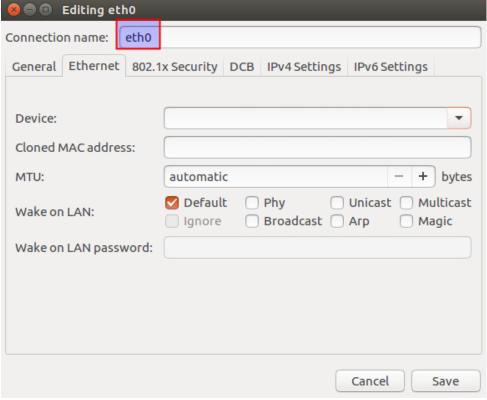
03/14/2017 2/13



Click the "Add" button.

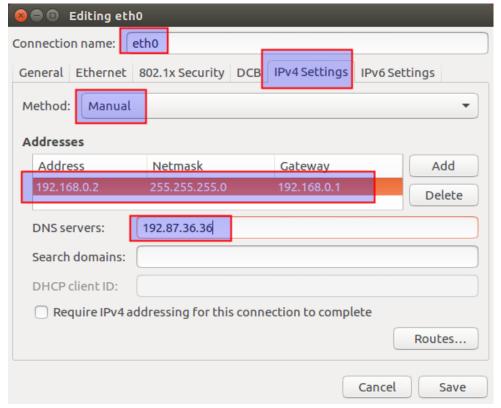


Select "Ethernet".



Create a network named "eth0".

03/14/2017 3/13



Manually set the IP address.

Restart the network:

sudo service network-manager restart

# Check the ip configuration:

ifconfig

It can be seen that the IP address of the Captain-CentERdata is set successfully. Repeat the above steps to set the IP address of the Sailor01-CentERdata. Testing on the Captain-CentERdata

```
ping 192.168.0.10 -c 4
# ping is the command to test the connection of two IP addresses.
# 192.168.0.10 is the IP address of the Sailor01-CentERdata.
# "-c 4" means transmitting 4 packets, and then stops automatically (otherwise, you need to press "Ctrl+X" to stop).
```

03/14/2017 4/13

```
hduser@Captain-CentERdata:~

hduser@Captain-CentERdata:~$ ping 192.168.0.10 -c 4

PING 192.168.0.10 (192.168.0.10) 56(84) bytes of data.

64 bytes from 192.168.0.10: icmp_seq=1 ttl=64 time=1.12 ms

64 bytes from 192.168.0.10: icmp_seq=2 ttl=64 time=0.561 ms

64 bytes from 192.168.0.10: icmp_seq=3 ttl=64 time=0.604 ms

64 bytes from 192.168.0.10: icmp_seq=4 ttl=64 time=0.306 ms

--- 192.168.0.10 ping statistics ---

4 packets transmitted, 4 received, 0% packet loss, time 3001ms

rtt min/avg/max/mdev = 0.306/0.647/1.120/0.297 ms

hduser@Captain-CentERdata:~$
```

• Enable password-free SSH between two computers:

Turn off the Firewall on each computer using the following command:

sudo ufw disable

Check whether the Firewall is disabled successfully:

sudo ufw status

```
hduser@Captain-CentERdata: ~

hduser@Captain-CentERdata: ~$ sudo ufw status
[sudo] password for hduser:
Status: inactive hduser:
hduser@Captain-CentERdata:~$
```

It shows that the status of the Firewall is inactive, therefore the Firewall is disabled successfully. Mapping the IP addresses with the computer names:

On the Captain-CentERdata, edit the file "/etc/hosts" using the following command:

sudo gedit /etc/hosts

Put the following content of "/etc/hosts" as follows:

```
127.0.0.1 localhost

192.168.0.2 Captain-CentERdata
192.168.0.10 Sailor01-CentERdata
```

Copy the file "/etc/hosts" to Sailor01-CentERdata using the following command:

```
sudo scp /etc/hosts hduser@192.168.0.10:/etc/hosts
```

Test whether Captain-CentERdata can recognize Sailor01-CentERdata by running the following command:

```
ping Sailor01-CentERdata -c 4
```

Exchange keys between computers by running the following commands only on Captain-CentERdata

```
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@Captain-CentERdata
```

03/14/2017 5/13

```
ssh-copy-id -i ~/.ssh/id_rsa.pub hduser@Sailor01-CentERdata
chmod 0600 ~/.ssh/authorized_keys
exit
```

#### • Configuring the Hadoop files on Captain-CentERdata:

There are totally 4 files to be configured: core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml. All these files are in the directory of "/usr/local/hadoop/etc/hadoop".

ls /usr/local/hadoop/etc/hadoop # Display the content

```
nduser@Captain-CentERdata: ~
hduser@Captain-CentERdata:~$ ls /usr/local/hadoop/etc/hadoop
capacity-scheduler.xml
                            httpfs-env.sh
                                                      mapred-env.sh
                            httpfs-log4j.properties
configuration.xsl
                                                      mapred-queues.xml.template
container-executor.cfg
                            httpfs-signature.secret
                                                      mapred-site.xml.template
core-site.xml
                            httpfs-site.xml
                                                      slaves
                            kms-acls.xml
hadoop-env.cmd
                                                      ssl-client.xml.example
hadoop-env.sh
                            kms-env.sh
                                                      ssl-server.xml.example
                            kms-log4j.properties
hadoop-metrics2.properties
                                                      yarn-env.cmd
hadoop-metrics.properties
                            kms-site.xml
                                                      yarn-env.sh
hadoop-policy.xml
                            log4j.properties
                                                      yarn-site.xml
hdfs-site.xml
                            mapred-env.cmd
hduser@Captain-CentERdata:~$
```

#### core-site.xml

sudo gedit /usr/local/hadoop/etc/hadoop/core-site.xml

### The content of core-site.xml is as follows:

### hdfs-site.xml

sudo gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml

# The content of **hdfs-site.xml** is as follows:

03/14/2017 6/13

#### mapred-site.xml

```
sudo cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/
local/hadoop/etc/hadoop/mapred-site.xml
sudo gedit /usr/local/hadoop/etc/hadoop/mapred-site.xml
```

#### The content of mapred-site.xml is as follows:

```
<configuration>
       property>
               <name>mapreduce.framework.name</name>
               <value>yarn</value>
       </property>
       cproperty>
               <name>mapreduce.jobhistory.address
               <value>Captain-CentERdata:10020</value>
       </property>
       cproperty>
               <name>mapreduce.jobhistory.webapp.address
               <value>Captain-CentERdata:19888
       </property>
       cproperty>
               <name>mapreduce.job.ubertask.enable</name>
               <value>true</value>
       </property>
</configuration>
```

### yarn-site.xml

```
sudo gedit /usr/local/hadoop/etc/hadoop/yarn-site.xml
```

## The content of yarn-site.xml is as follows:

• Copy the Hadoop configuration files from Captain-CentERdata to other computers (e.g. Sailor01-CentERdata):

03/14/2017 7/13

```
cd /usr/local
tar -zcf ~/hadoop.master.tar.gz hadoop # compress before copy
scp ~/hadoop.master.tar.gz hduser@Sailor01-CentERdata:~/
```

```
🤇 🖨 📵 hduser@Captain-CentERdata: /usr/local
hduser@Captain-CentERdata:~$ cd /usr/local
hduser@Captain-CentERdata:/usr/local$ tar -zcf ~/hadoop.master.tar.gz hadoop
                                                                                #
compress before copy
hduser@Captain-CentERdata:/usr/local$ ls ~
captain.cer Downloads
                                              scikit_learn_data
                                   Music
            examples.desktop
                                              Templates
                                   Pictures
Desktop
Documents
                                   Public
                                             Videos
hduser@Captain-CentERdata:/usr/local$ scp ~/hadoop.master.tar.gz hduser@Sailor01
-CentERdata:~/
hadoop.master.tar.gz
                                               100% 202MB 11.2MB/s
                                                                       00:18
hduser@Captain-CentERdata:/usr/local$
```

```
ssh Sailor01-CentERdata
sudo rm -r /usr/local/hadoop  # Delete the old Hadoop if it exists
sudo tar -zxf ~/hadoop.master.tar.gz -C /usr/local
# -z(--gzip);-x(--extract);-f(--file);-C(--directory)
sudo chown -R hduser /usr/local/hadoop # -R(--recursively)
exit
```

```
■ ■ hduser@Captain-CentERdata: ~
hduser@Captain-CentERdata:~$ ssh Sailor01-CentERdata
Welcome to Ubuntu 16.04 LTS (GNU/Linux 4.4.0-22-generic x86_64)
 * Documentation: https://help.ubuntu.com/
0 packages can be updated.
0 updates are security updates.
Last login: Mon May 30 11:14:18 2016 from 192.168.0.2
                                                             # Delete the old H
hduser@Sailor01-CentERdata:~$ sudo rm -r /usr/local/hadoop
adoop if it exists
[sudo] password for hduser:
rm: cannot remove '/usr/local/hadoop': No such file or directory
hduser@Sailor01-CentERdata:~$ ls /usr/local
bin etc games include lib man sbin share src
hduser@Sailor01-CentERdata:~$ sudo tar -zxf ~/hadoop.master.tar.gz -C /usr/local
hduser@Sailor01-CentERdata:~$ ls /usr/local
bin etc games hadoop include
                                 lib man sbin share src
hduser@Sailor01-CentERdata:~$ sudo chown -R hduser /usr/local/hadoop
hduser@Sailor01-CentERdata:~$ exit
logout
Connection to sailor01-centerdata closed.
hduser@Captain-CentERdata ~$
```

### • Run the Hadoop:

Run the following command on the Captain-CentERdata:

Set the slaves and masters (Only on Captain-CentERdata) slaves

sudo gedit /usr/local/hadoop/etc/hadoop/slaves

The content of slaves is as follows:

03/14/2017 8/13

#### masters

sudo gedit /usr/local/hadoop/etc/hadoop/masters

The content of masters is as follows:

Captain-CentERdata

#### Format the name node

hdfs namenode -format # This command is need to be run only once.

```
🦻 🗐 🏻 hduser@Captain-CentERdata: ~
16/05/30 12:58:56 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total
heap and retry cache entry expiry time is 600000 millis
16/05/30 12:58:56 INFO util.GSet: Computing capacity for map NameNodeRetryCache
16/05/30 12:58:56 INFO util.GSet: VM type = 64-bit
16/05/30 12:58:56 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.
1 KB
16/05/30 12:58:56 INFO util.GSet: capacity
                                                           = 2^15 = 32768 entries
16/05/30 12:58:56 INFO namenode.FSImage: Allocated new BlockPoolId: BP-817315929
-192.168.0.2-1464605936285
16/05/30 12:58:56 INFO common.Storage: Storage directory /usr/local/hadoop/tmp/d
fs/name has been successfully formatted.
16/05/30 12:58:56 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
16/05/30 12:58:56 INFO util.ExitUtil: Exiting with status 0
16/05/30 12:58:56 INFO namenode.NameNode: SHUTDOWN MSG:
SHUTDOWN_MSG: Shutting down NameNode at Captain-CentERdata/192.168.0.2
hduser@Captain-CentERdata:~$
```

### Start dfs and yarn

```
start-dfs.sh && start-yarn.sh
hadoop-daemon.sh start datanode # start datanode
mr-jobhistory-daemon.sh start historyserver
```

Use the following command to see the process:

jps

03/14/2017 9/13

```
nduser@Captain-CentERdata: ~
hduser@Captain-CentERdata:~$ start-dfs.sh
Starting namenodes on [Captain-CentERdata]
Captain-CentERdata: starting namenode, logging to /usr/local/hadoop/logs/hadoop-
hduser-namenode-Captain-CentERdata.out
Sailor01-CentERdata: starting datanode, logging to /usr/local/hadoop/logs/hadoop
-hduser-datanode-Sailor01-CentERdata.out
Starting secondary namenodes [Captain-CentERdata]
Captain-CentERdata: starting secondarynamenode, logging to /usr/local/hadoop/log
s/hadoop-hduser-secondarynamenode-Captain-CentERdata.out
hduser@Captain-CentERdata:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resource
manager-Captain-CentERdata.out
Sailor01-CentERdata: starting nodemanager, logging to /usr/local/hadoop/logs/yar
n-hduser-nodemanager-Sailor01-CentERdata.out
hduser@Captain-CentERdata:~$ mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /usr/local/hadoop/logs/mapred-hduser-historys
erver-Captain-CentERdata.out
hduser@Captain-CentERdata:~$ jps
7203 Jps
6680 SecondaryNameNode
7160 JobHistoryServer
6873 ResourceManager
6492 NameNode
hduser@Captain-CentERdata:~$
```

Use the following command to see whether all the DataNode is running properly:

hdfs dfsadmin -report

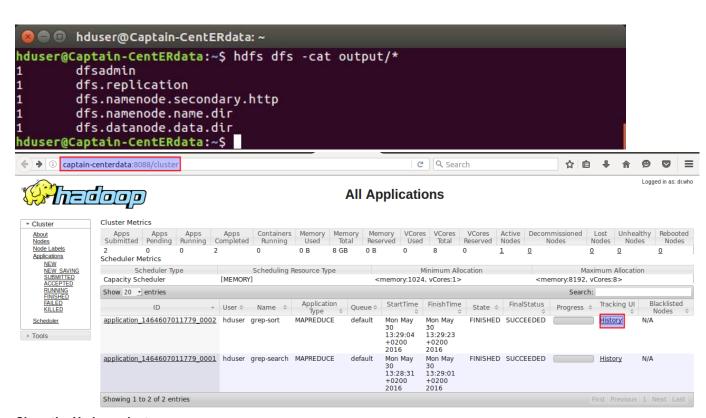
```
🤊 🗐 📵 hduser@Captain-CentERdata: ~
hduser@Captain-CentERdata:~$ hdfs dfsadmin -report
Configured Capacity: 488030355456 (454.51 GB)
Present Capacity: 457653862400 (426.22 GB)
DFS Remaining: 457653837824 (426.22 GB)
DFS Used: 24576 (24 KB)
DFS Used%: 0.00%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0
Live datanodes (1):
Name: 192.168.0.10:50010 (Sailor01-CentERdata)
Hostname: Sailor01-CentERdata
Decommission Status : Normal
Configured Capacity: 488030355456 (454.51 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 30376493056 (28.29 GB)
DFS Remaining: 457653837824 (426.22 GB)
DFS Used%: 0.00%
DFS Remaining%: 93.78%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
```

## • Examples-grep:

```
hdfs dfs -mkdir -p input
```

03/14/2017 10/13

```
hdfs dfs -put /usr/local/hadoop/etc/hadoop/*.xml input
hdfs dfs -ls input
hadoop jar /usr/
local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep input output '
dfs[a-z.]+'
hdfs dfs -cat output/*
```



# • Close the Hadoop cluster:

```
stop-yarn.sh && stop-dfs.sh
mr-jobhistory-daemon.sh stop historyserver
```

```
🤰 🗐 📵 hduser@Captain-CentERdata: ~
hduser@Captain-CentERdata:~$ stop-yarn.sh
stopping yarn daemons
stopping resourcemanager
Sailor01-CentERdata: stopping nodemanager
Sailor01-CentERdata: nodemanager did not stop gracefully after 5 seconds: killin
g with kill -9
no proxyserver to stop
hduser@Captain-CentERdata:~$ stop-dfs.sh
Stopping namenodes on [Captain-CentERdata]
Captain-CentERdata: stopping namenode
Sailor01-CentERdata: stopping datanode
Stopping secondary namenodes [Captain-CentERdata]
Captain-CentERdata: stopping secondarynamenode
hduser@Captain-CentERdata:~$ mr-jobhistory-daemon.sh stop historyserver
stopping historyserver
hduser@Captain-CentERdata:~$ jps
8883 Jps
hduser@Captain-CentERdata:~$
```

# For more examples, please go to <u>Hadoop examples</u>

# **Trouble shoot**

Problem: Name node is in safe mode. Can not delete /user/hduser/grep-temp

03/14/2017 11/13

```
hdfs dfsadmin -safemode leave # Force the namenode to leave safe mode hdfs fsck # Sort out any inconsistencies crept in the hdfs
```

Problem: sign\_and\_send\_pubkey: signing failed: agent refused operation Cause: backage **ssh** does not support well the function "ssh-copy-id" Solution: remove **ssh**, install **openssh-server**, clean the residual packages

```
sudo apt-get remove ssh # Remove ssh
sudo apt-get install openssh-server # Install openssh-server
sudo apt autoremove
# Remove packages that were automatically installed to satisfy dependencies for some package and t
hat are no more needed.
sudo apt autoclean # Clear out the local repository of retrieved package files.
sudo reboot
```

Problem: Configured Capacity: 0 (0 B) Cause: DataNote can not be activated

Solution: delete the folder "/home/hduser/hdtmp/dfs/data/" on each machine

```
stop-dfs.sh && stop-yarn.sh # Stop Hadoop # First stop the cluster
sudo rm -r -f /home/hduser/hdtmp/dfs/data/
# Delete the folder "/home/hduser/hdtmp/dfs/data/", please note that this command should run all m
achines.
start-dfs.sh && start-yarn.sh # Start Hadoop
jps # Check the nodes to see whether DataNote is activated
```

Problem: Captain-CentERdata:9000 failed on connection exception

Cause: /home/hduser/hdtmp/dfs/name in the Master machine has been destroyed.

Solution:

```
stop-dfs.sh && stop-yarn.sh # Stop Hadoop # First stop the cluster
hdfs namenode -format # Format the name node. This command is need to be run only once.
start-dfs.sh && start-yarn.sh # Start Hadoop
hdfs dfsadmin -report # Report Hadoop
```

Problem: When running "hdfs dfsadmin -report", get error "9000 failed on connection exception" Solution: on the following command on Captain-CentERdata

sudo chown -R hduser /usr/local/hadoop

Problem: datanode crashes Solution: restart datanode

hadoop-daemon.sh start datanode

03/14/2017 12/13

# Reference

- 1. https://www.tutorialspoint.com/hadoop/hadoop multi node cluster.htm
- 2. http://www.powerxing.com/install-hadoop-cluster/

3.

http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-multi-nod e-cluster/

4.

https://chawlasumit.wordpress.com/2015/03/09/install-a-multi-node-hadoop-cluster-on-ubuntu-14-04/

5.

http://fibrevillage.com/mobile/storage/628-how-to-add-a-new-datanode-to-a-run ning-hadoop-cluster

- 6. http://doctuts.readthedocs.io/en/latest/hadoop.html
- 7. https://tecadmin.net/set-up-hadoop-multi-node-cluster-on-centos-redhat/#
- 8. <a href="http://arturmkrtchyan.com/how-to-setup-multi-node-hadoop-2-yarn-cluster">http://arturmkrtchyan.com/how-to-setup-multi-node-hadoop-2-yarn-cluster</a>
  9.

http://www.quuxlabs.com/tutorials/running-hadoop-on-ubuntu-linux-multi-node-c luster/

10. http://pingax.com/install-apache-hadoop-ubuntu-cluster-setup/

## **Files**

ComputerName.png	10.8 KB	05/22/2016	Kai-Tao Yang
ComputerNameNew.png	11.7 KB	05/22/2016	Kai-Tao Yang
Computer-connection-router.jpg	121 KB	05/25/2016	Kai-Tao Yang
network-add.png	15.8 KB	05/29/2016	Kai-Tao Yang
network-eth0.png	36.2 KB	05/29/2016	Kai-Tao Yang
netwrok-captain.png	44.3 KB	05/29/2016	Kai-Tao Yang
network-captain-success.png	42.8 KB	05/29/2016	Kai-Tao Yang
network-add-ethernet.png	22.5 KB	05/29/2016	Kai-Tao Yang
network-ping-from-captain.png	53.2 KB	05/29/2016	Kai-Tao Yang
network-find-app.png	139 KB	05/29/2016	Kai-Tao Yang
firewall-inactive.png	17.7 KB	05/29/2016	Kai-Tao Yang
Hadoop-5-files.png	51.7 KB	05/30/2016	Kai-Tao Yang
Hadoop-compress-scp.png	55.1 KB	05/30/2016	Kai-Tao Yang
Hadoop-access-Sailor01.png	84.3 KB	05/30/2016	Kai-Tao Yang
Hadoop-format-successful.png	84.5 KB	05/30/2016	Kai-Tao Yang
Hadoop-start.png	91.2 KB	05/30/2016	Kai-Tao Yang
Hadoop-report.png	101 KB	05/30/2016	Kai-Tao Yang
Hadoop-stop.png	63.6 KB	05/30/2016	Kai-Tao Yang
Hadoop-output.png	26 KB	05/30/2016	Kai-Tao Yang
Hadoop-website.png	101 KB	05/30/2016	Kai-Tao Yang

03/14/2017 13/13