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 Single Node Setup 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-CentERdata	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:**

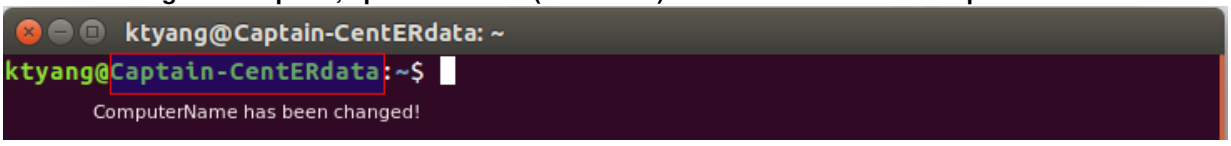
```
sudo gedit /etc/hostname
sudo gedit /etc/hosts
```

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:**



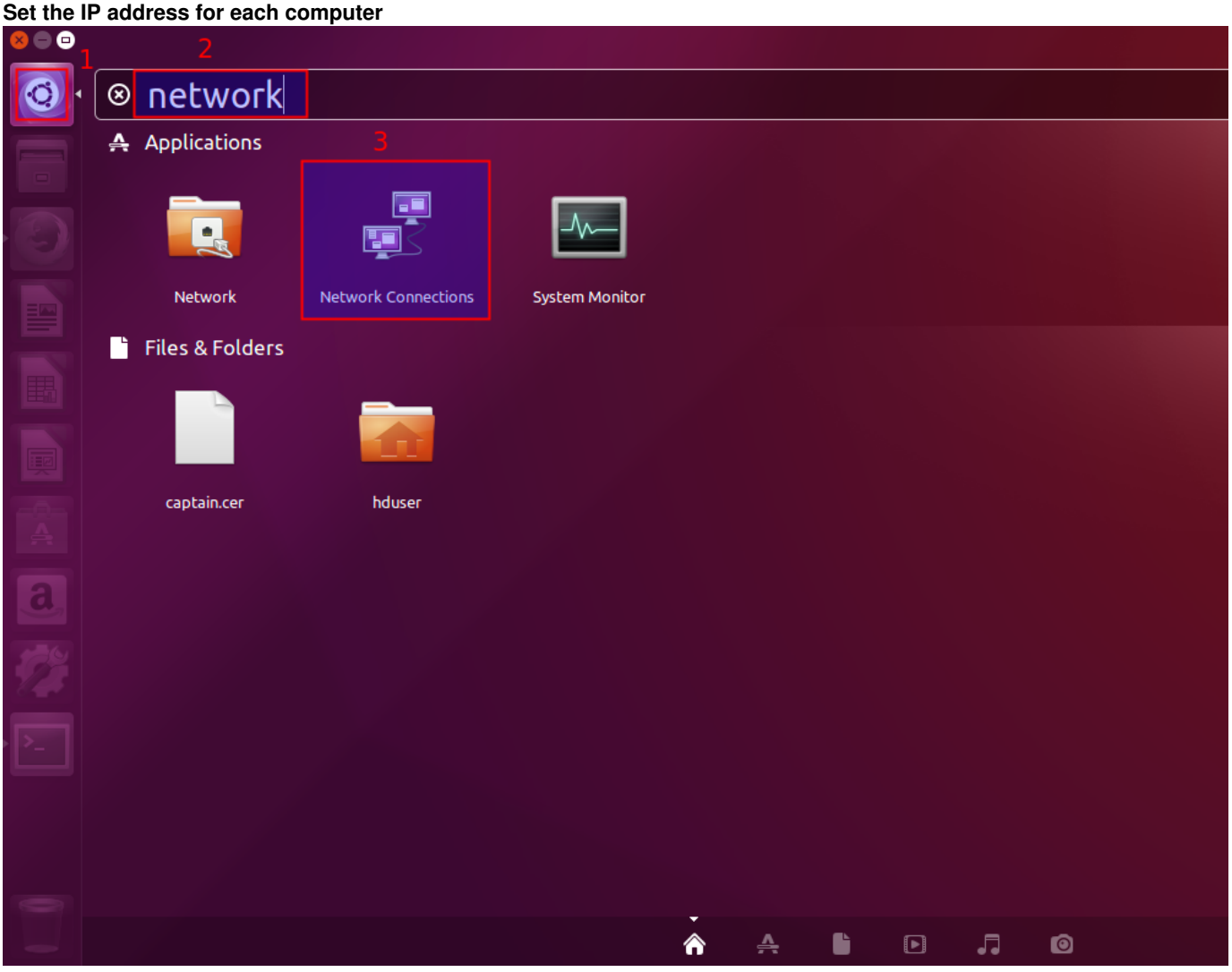
Repeating the above steps to rename the other computer as ¨Sailor01-CentERdata¨.

Connect computers

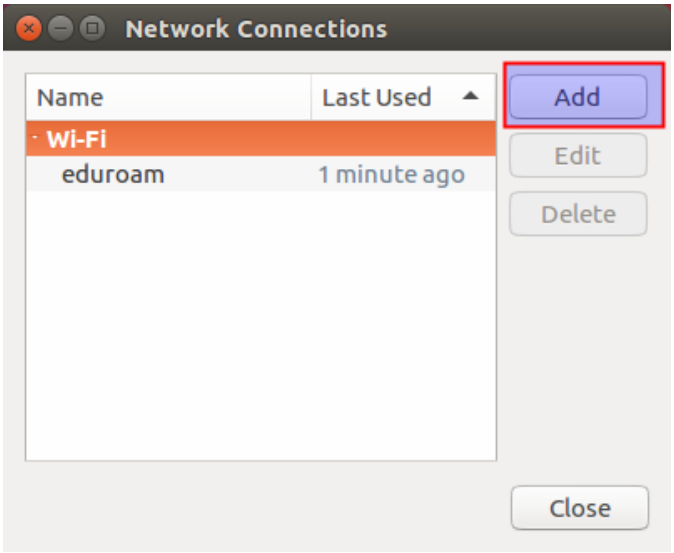
- **Connecting computers with a router:**

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.

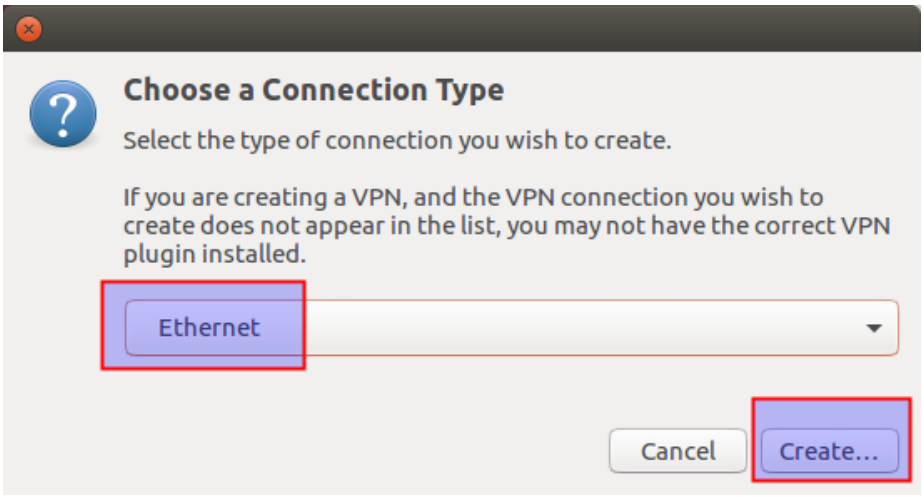
- **Set the IP address for each computer**



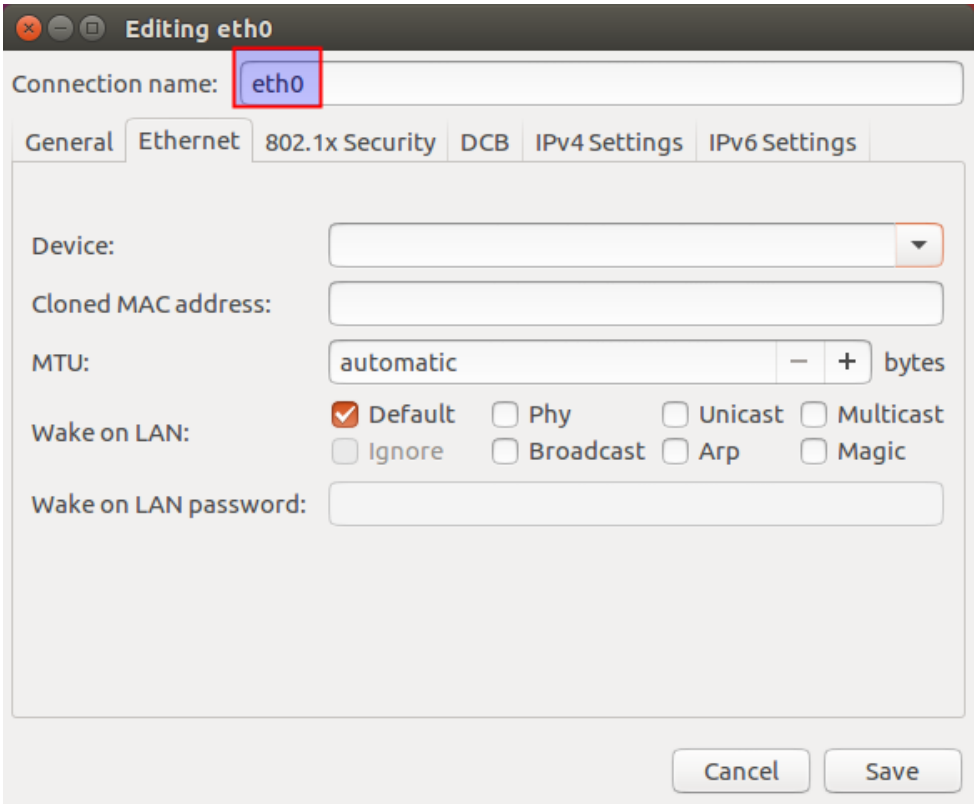
Click the ¨Search you computer¨ => Type in ¨network¨ => Click the ¨Network Connections¨.



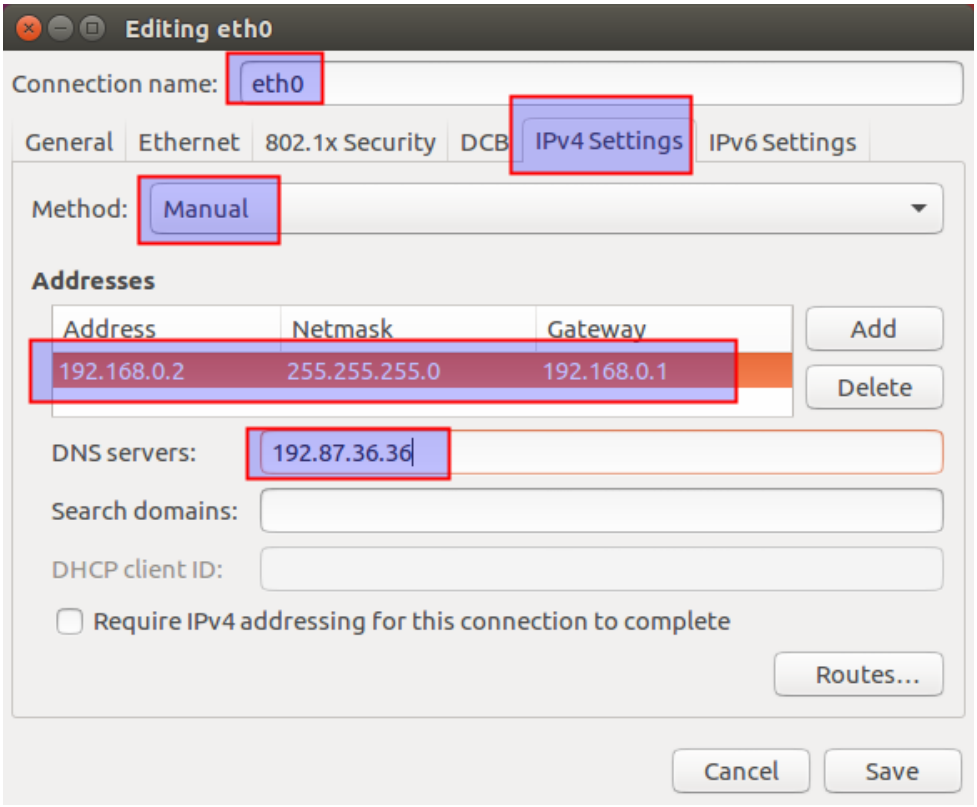
Click the ¨Add¨ button.



Select ¨Ethernet¨.



Create a network named ¨eth0¨.

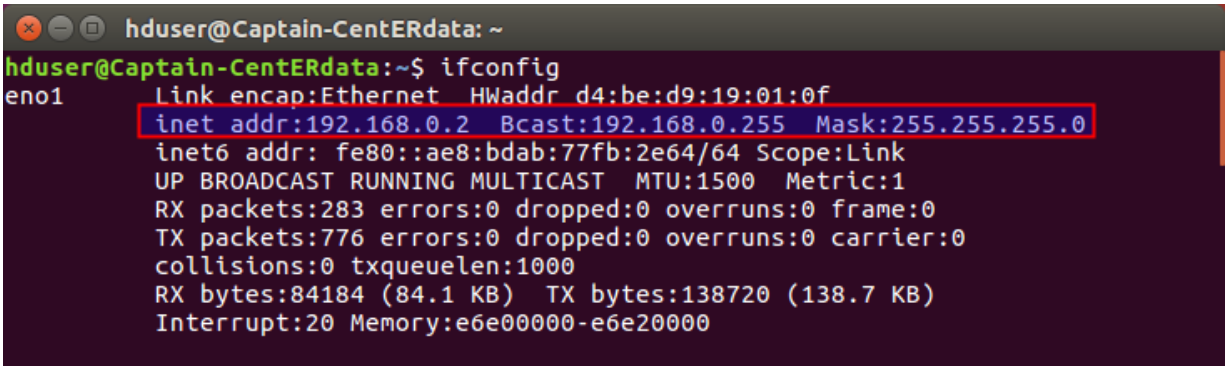


Manually set the IP address.
Restart the network:

```
sudo service network-manager restart
```

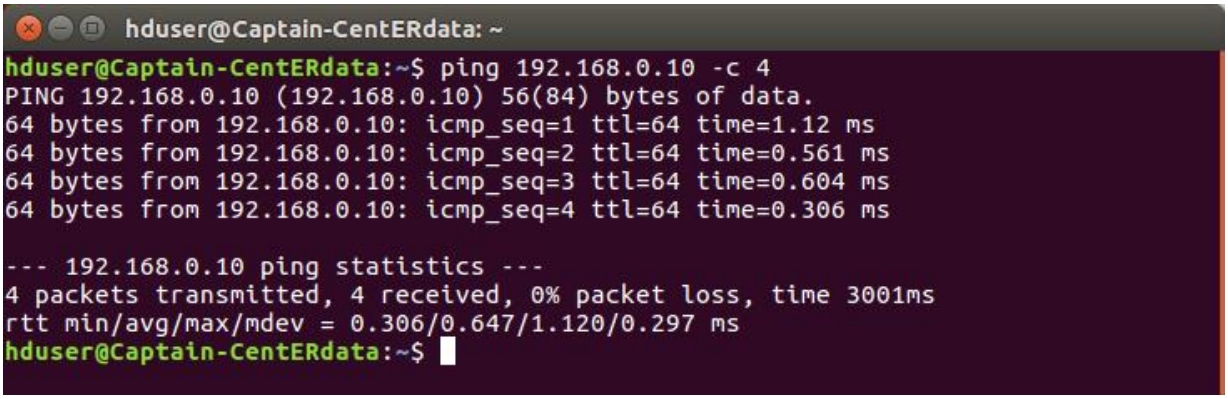
Check the ip configuration:

```
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 pr
ess ¨Ctrl+X¨ to stop).
```

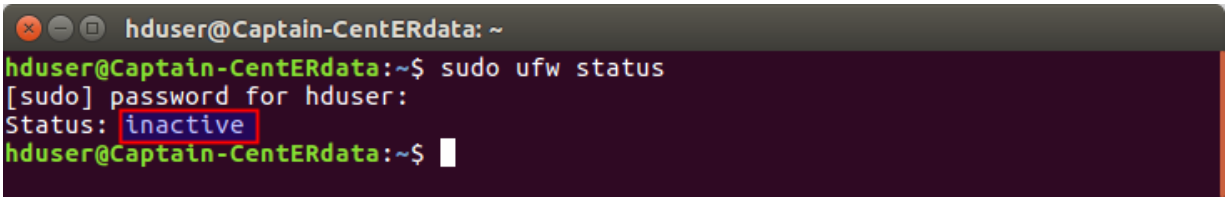


- **Enable password-free SSH between two computers:**
  Turn off the Firewall on each computer using the following command:

```
sudo ufw disable
```

Check whether the Firewall is disabled successfully:

```
sudo ufw status
```



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
```

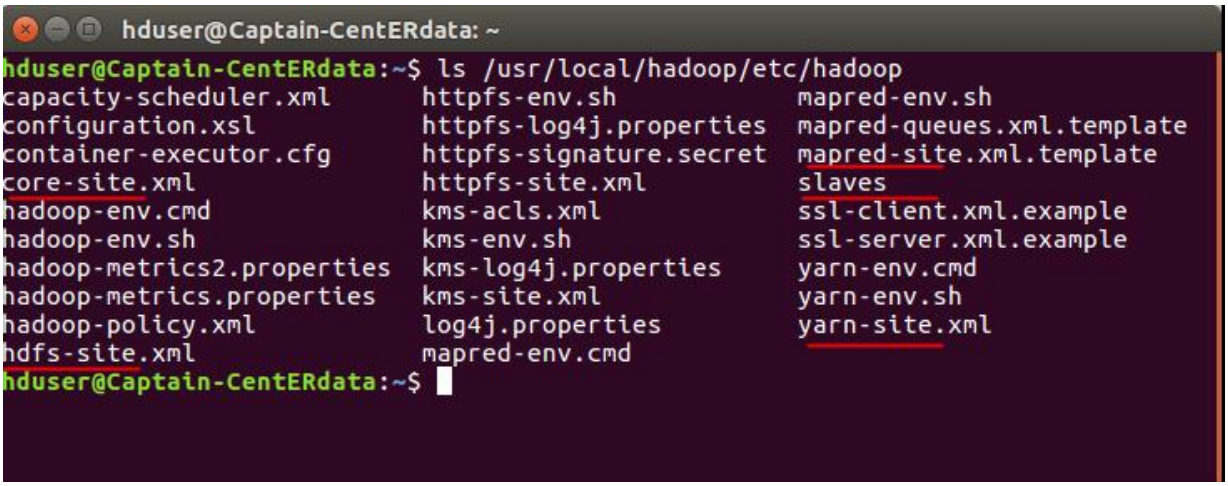
Exchange keys between computers by running the following commands only on Captain-CentERdata

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

```
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
```



core-site.xml

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

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

```
<configuration>
    <property>
            <name>fs.defaultFS</name>
            <value>hdfs://Captain-CentERdata:9000</value>
    </property>
    <property>
            <name>hadoop.tmp.dir</name>
            <value>file:/home/hduser/hdtmp</value>
            <description>Abase for other temporary directories.</description>
    </property>
</configuration>
```

hdfs-site.xml

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

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

```
<configuration>
        <property>
                <name>dfs.namenode.secondary.http-address</name>
                <value>Captain-CentERdata:50090</value>
        </property>
        <property>
                <name>dfs.replication</name>
```

```
                <value>1</value>
        </property>
        <property>
                <name>dfs.namenode.name.dir</name>
                <value>file:/home/hduser/hdtmp/dfs/name</value>
        </property>
        <property>
                <name>dfs.datanode.data.dir</name>
                <value>file:/home/hduser/hdtmp/dfs/data</value>
        </property>
</configuration>
```

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

yarn-site.xml

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

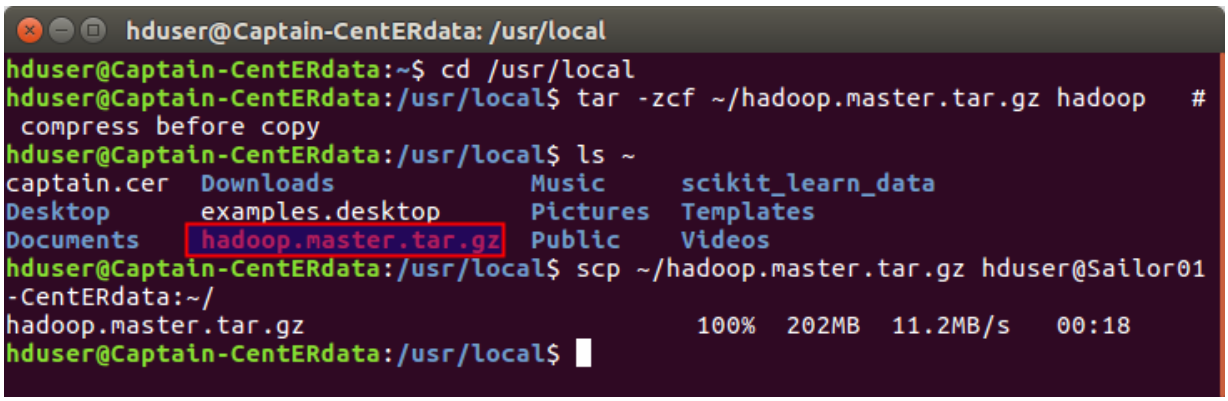
The content of **yarn-site.xml** is as follows:

```
<configuration>
        <property>
                <name>yarn.resourcemanager.hostname</name>
                <value>Captain-CentERdata</value>
        </property>
        <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
        </property>
</configuration>
```

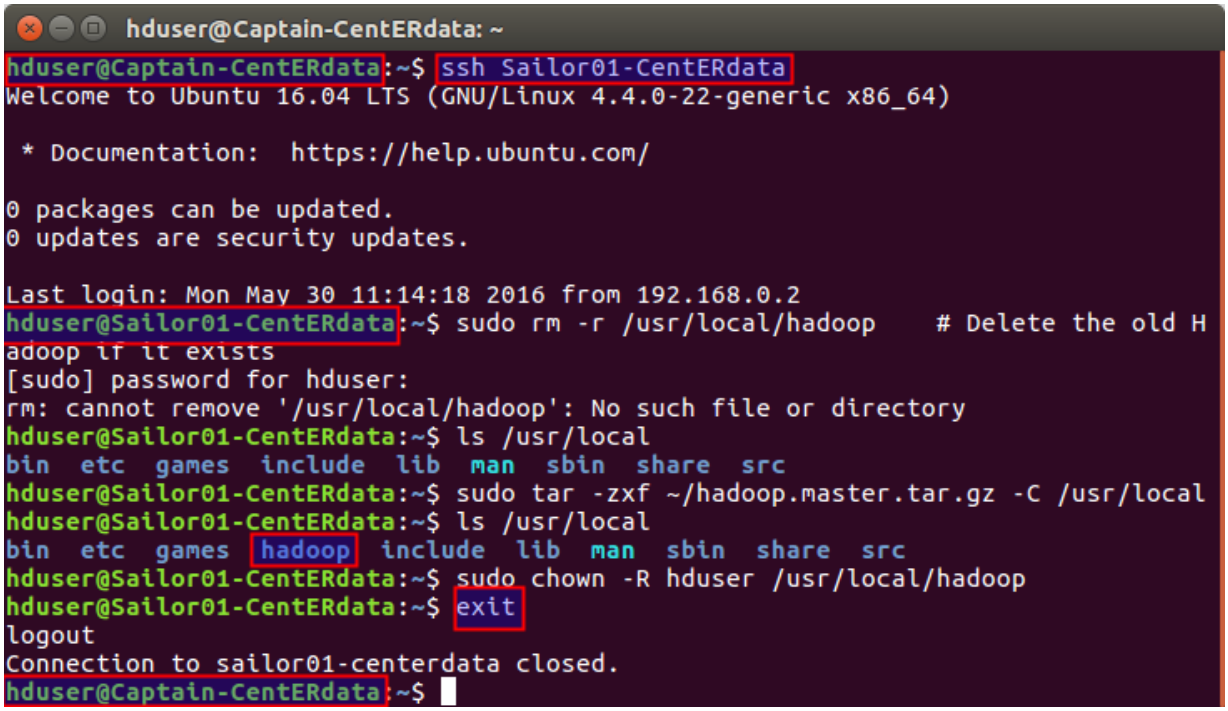
- **Copy the Hadoop configuration files from Captain-CentERdata to other computers (e.g. Sailor01-CentERdata):**

Run the following command on the Captain-CentERdata:

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




```
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
```



- **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:

```
Sailor01-CentERdata
```

masters

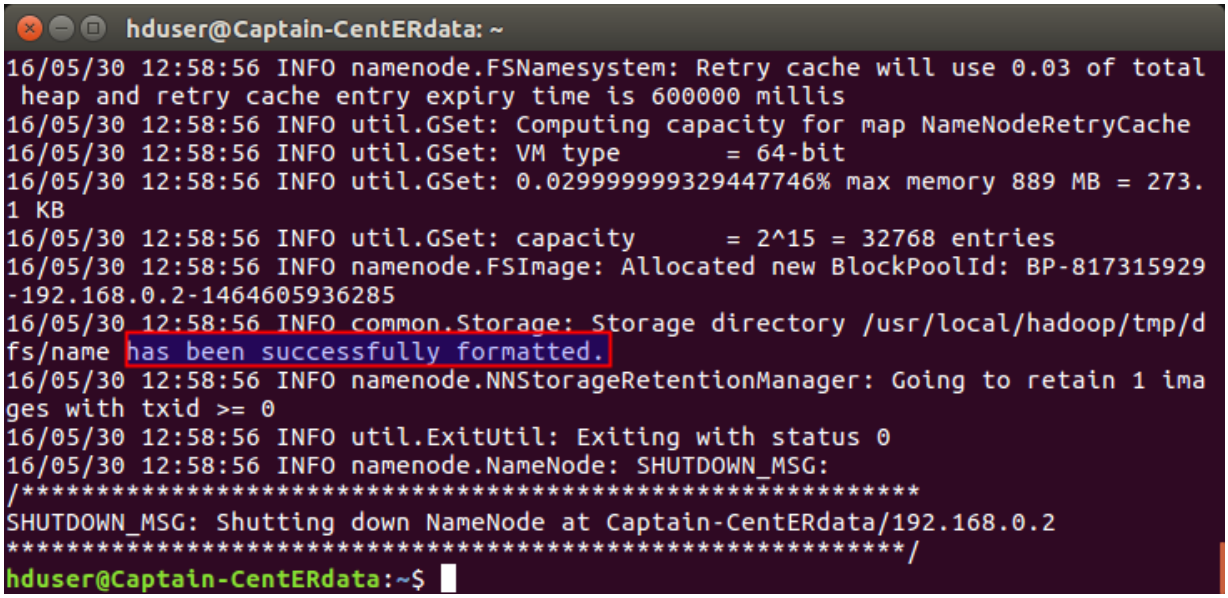
```
sudo gedit /usr/local/hadoop/etc/hadoop/masters
```

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.
```

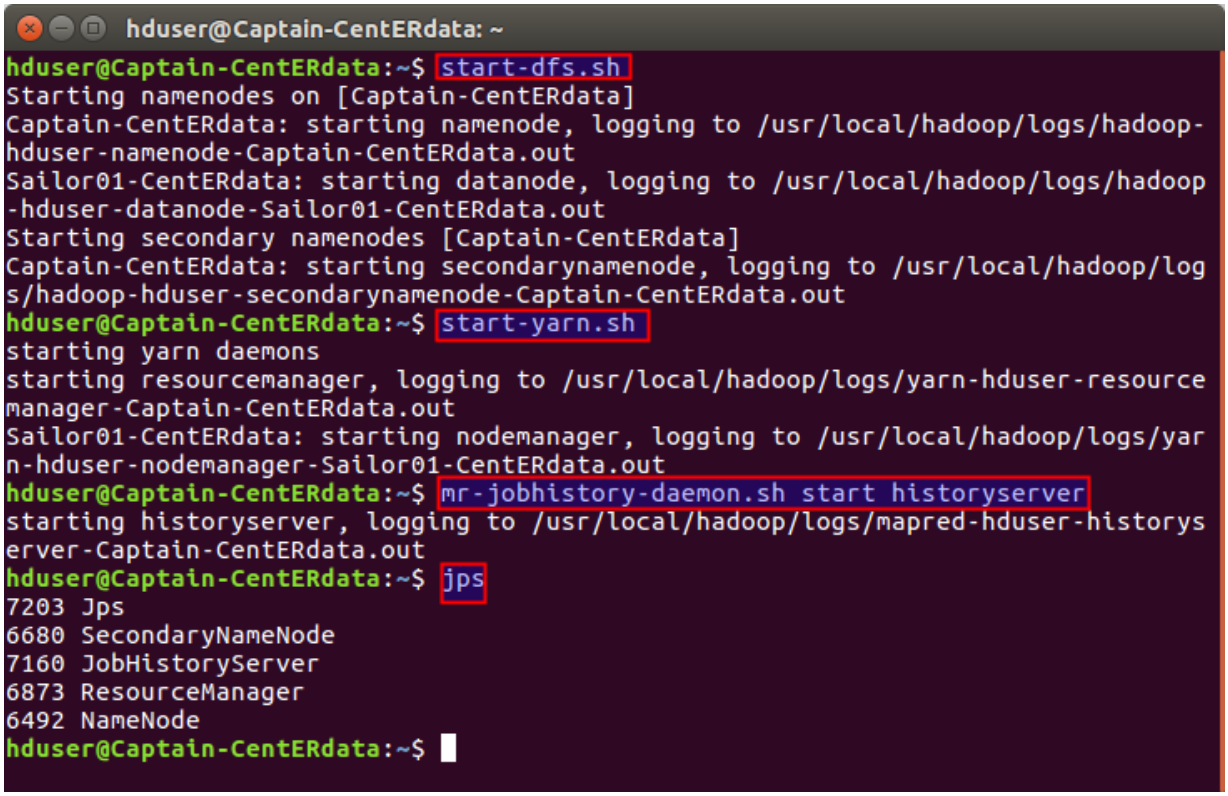


Start dfs and yarn

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

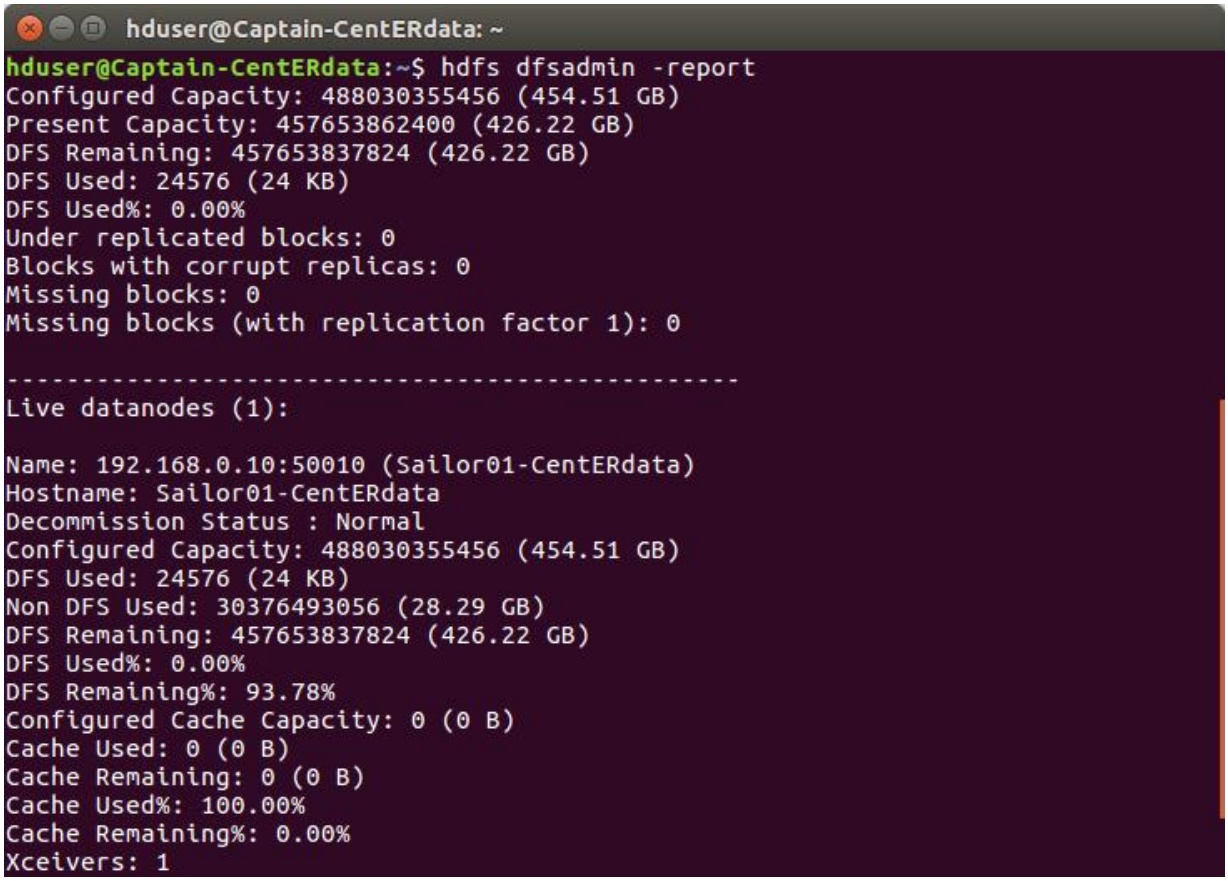
Use the following command to see the process:

```
jps
```



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

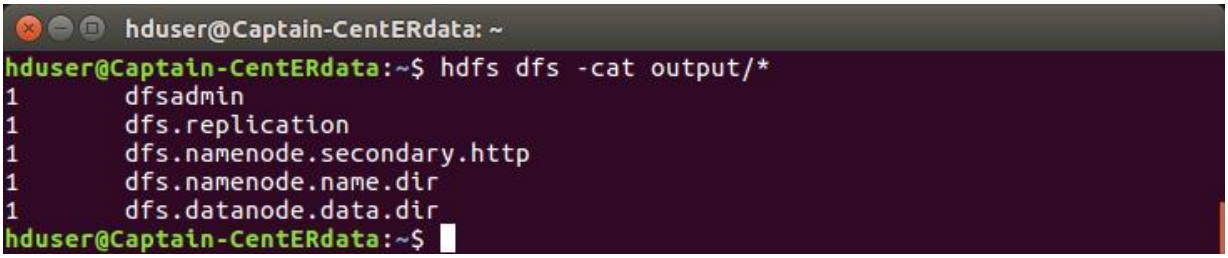
```
hdfs dfsadmin -report
```

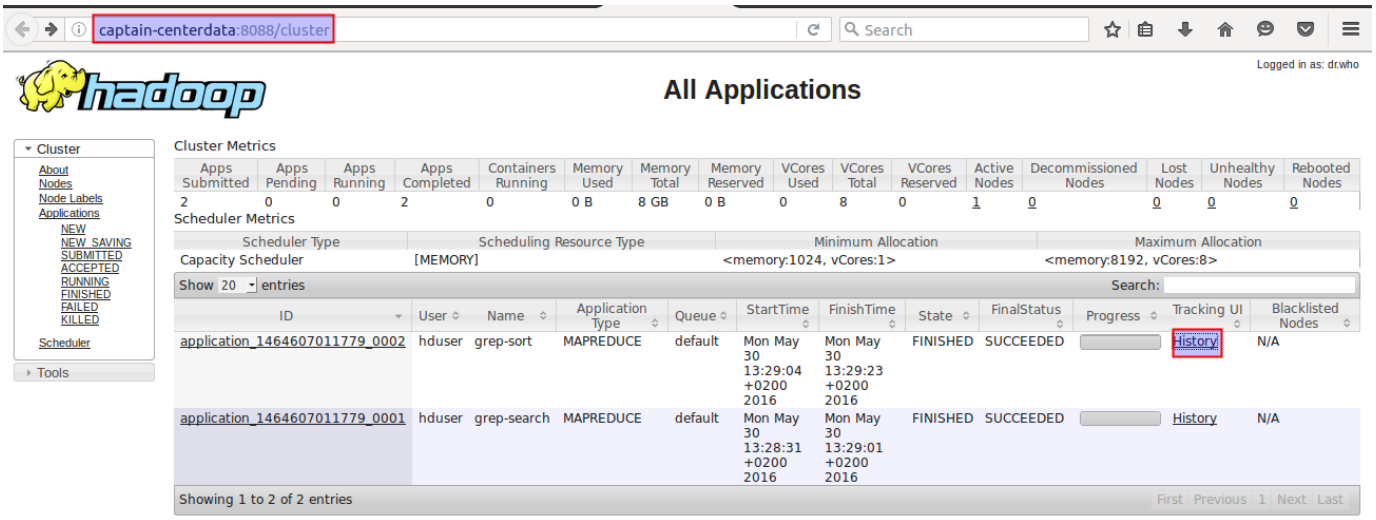


- **Examples-grep:**

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

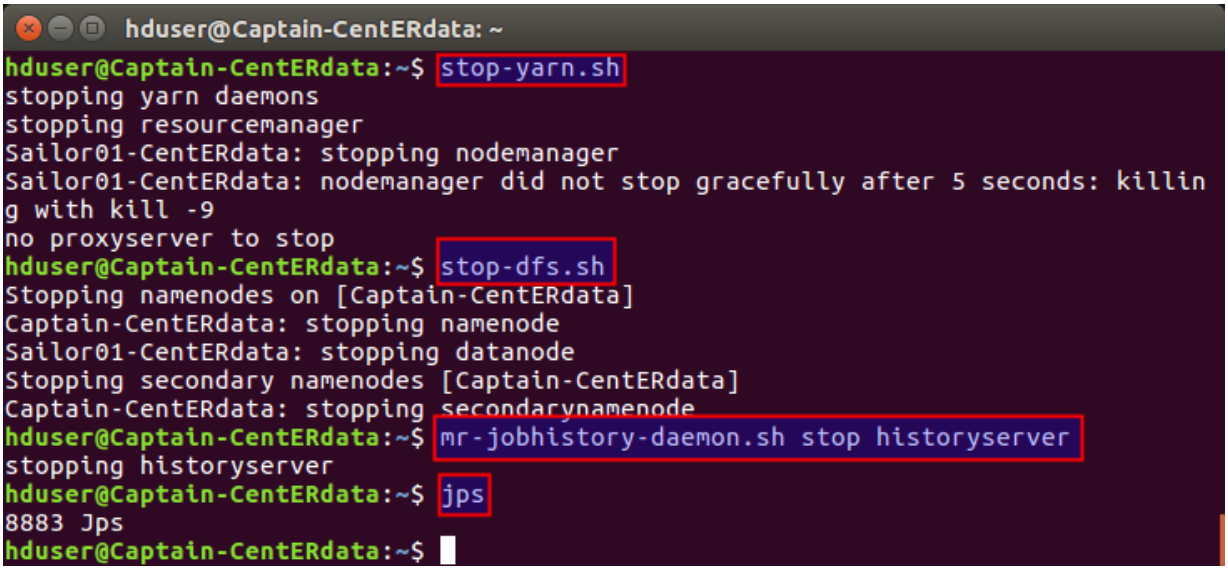
```
hdfs dfs -put /usr/local/hadoop/etc/hadoop/*.xml input
hdfs dfs -ls input
hadoop jar /usr/
local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep input output '
dfs[a-z.]+'
hdfs dfs -cat output/*
```





- **Close the Hadoop cluster:**

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



For more examples, please go to Hadoop examples

Trouble shoot

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

```
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
```

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-node-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-running-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. http://arturmkrtchyan.com/how-to-setup-multi-node-hadoop-2-yarn-cluster
9. http://www.quuxlabs.com/tutorials/running-hadoop-on-ubuntu-linux-multi-node-cluster/
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