Single Node Setup

Materials: A computer with Linux installed + Access to the Internet
Explanation: Hadoop supports Linux better than Windows, therefore install or use a machine that has Linux installed on it (i.e. Ubuntu 16.04 64-bit) on the computer (go to Ubuntu installation tutorial). The computer must have access to the internet because we need to install some software online.

Pre-installation

The pre-installation phrase includes: creating a new user account, installing two software (JAVA and SSH).
Explanation: It is recommended to create a separate user account for Hadoop in order to isolate Hadoop file system from the Linux file system. Java is the main prerequisite for Hadoop. SSH is required when doing operations (e.g. copying files) across different computers of a Hadoop cluster.

1. Creating a new user account
   - Open a terminal (Ctrl+Alt+T) and type in the following commands:

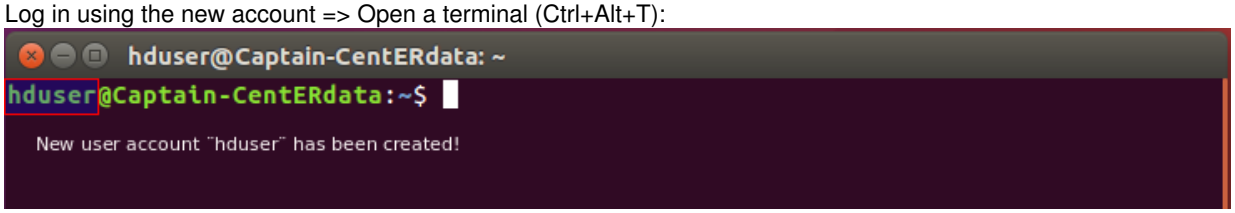
```
sudo useradd -m hduser -s /bin/bash
# Purpose: Creat a new user account named ¨hduser¨.
# ¨sudo¨ means run the command with the security privileges of the superuser.
# ¨-m¨ means creating the user's home directory if it does not exist.
# ¨-s /bin/bash¨ means using ¨/bin/bash¨ as the user's login shell.

sudo passwd hduser # Set password for the new account.

sudo adduser hduser sudo # Add hduser as a superuser

sudo reboot # Restart the computer
```

   Note: You can use other names for this new account (not necessarily use ¨hduser¨). However keep in mind that, in a Hadoop cluster, the new user account on each computer must have the same name (e.g. ¨hduser¨).
   - Log in using the new account => Open a terminal (Ctrl+Alt+T):

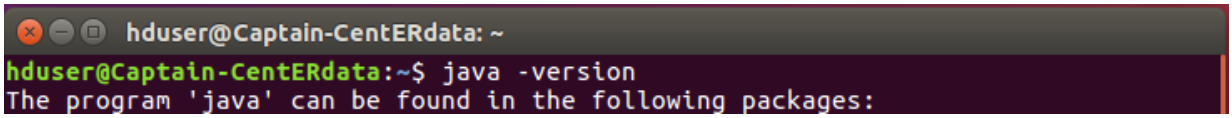


   - Update the resource of software using the following command:

```
sudo apt-get update
```

   Note: Ubuntu uses ¨apt-get¨ command to install software (e.g. JAVA, SSH) online. The online resource of software must be updated, otherwise, some software may not be found.
2. Installing JAVA
   - The following command checks whether whether JAVA has already been installed:

```
java -version
```



   The above figure shows that JAVA has not been installed.

- Install JAVA using the following commands:

```
sudo apt-get install openjdk-8-jre openjdk-8-jdk
# Install Java Runtime Environment (JRE) and Java Development Kit (JDK)
```

Note: This command may not be successful because ¨openjdk-8-jre and openjdk-8-jdk may have no installation candidate¨. In this case, please check the homepage of openjdk and install the latest version.

- Add the JAVA path to the environment path
  JAVA path is added by editing the file ¨~/.bashrc¨ using the following command:

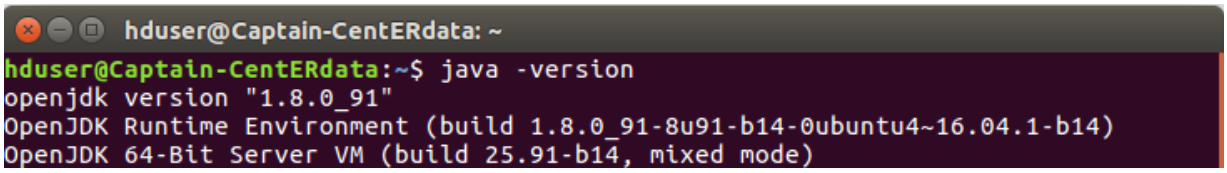
```
sudo gedit ~/.bashrc
```

Add **export JAVA_HOME=/usr** to the first line of ¨~/.bashrc¨.
Save and close the file ¨~/.bashrc¨.
Run the following command to activate the changes:

```
source ~/.bashrc
```

- Check the installation of java again:

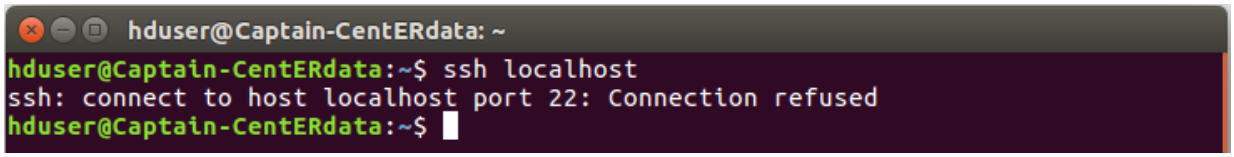
```
java -version
```



The above figure shows that JAVA has be installed successfully.

3. Installing SSH
   - The following command checks whether whether SSH has already been installed:

```
ssh localhost
```



The above figure shows that SSH has not been installed.

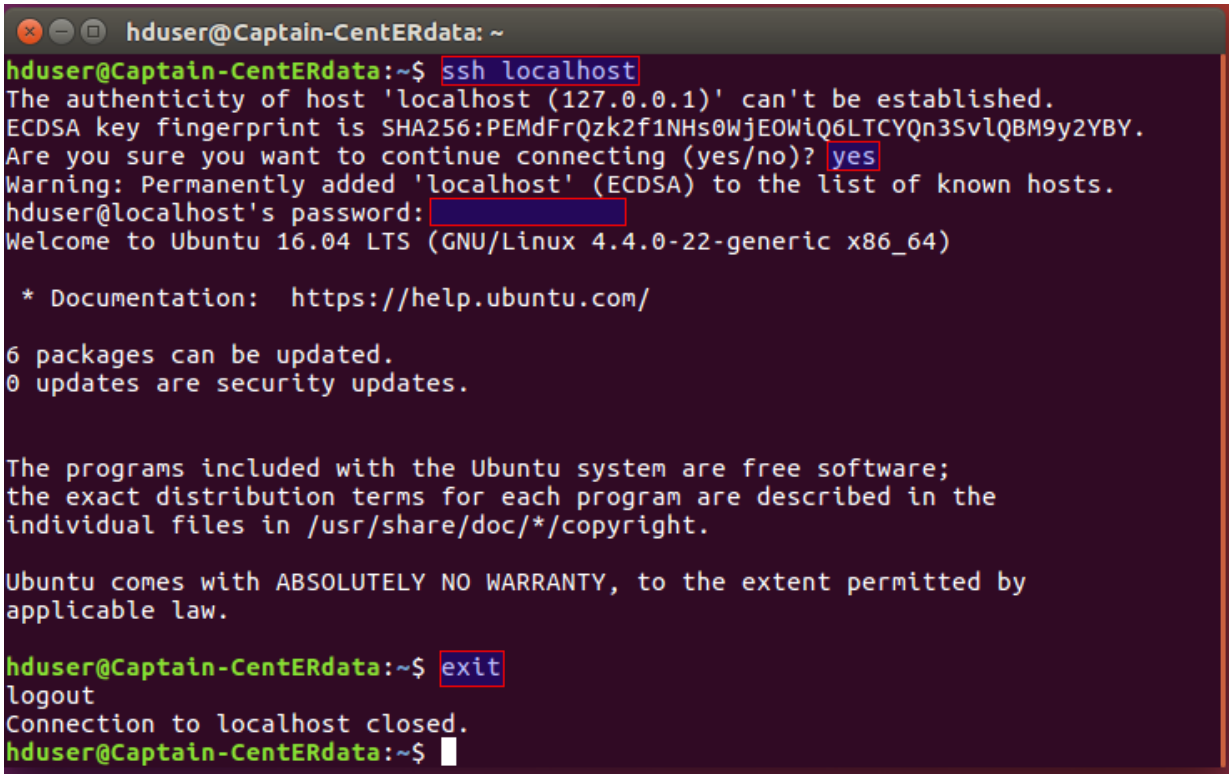
- Install SSH using the following command:

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

Note: **openssh-server** offers a very handy function **ssh-copy-id** (see Cluster Setup) to exchange SSH keys between two computers. **rsync** is a fast, versatile, remote (and local) file-copying tool. In a Hadoop cluster, there are file-copying operations between computers.

- Check the installation of SSH again:

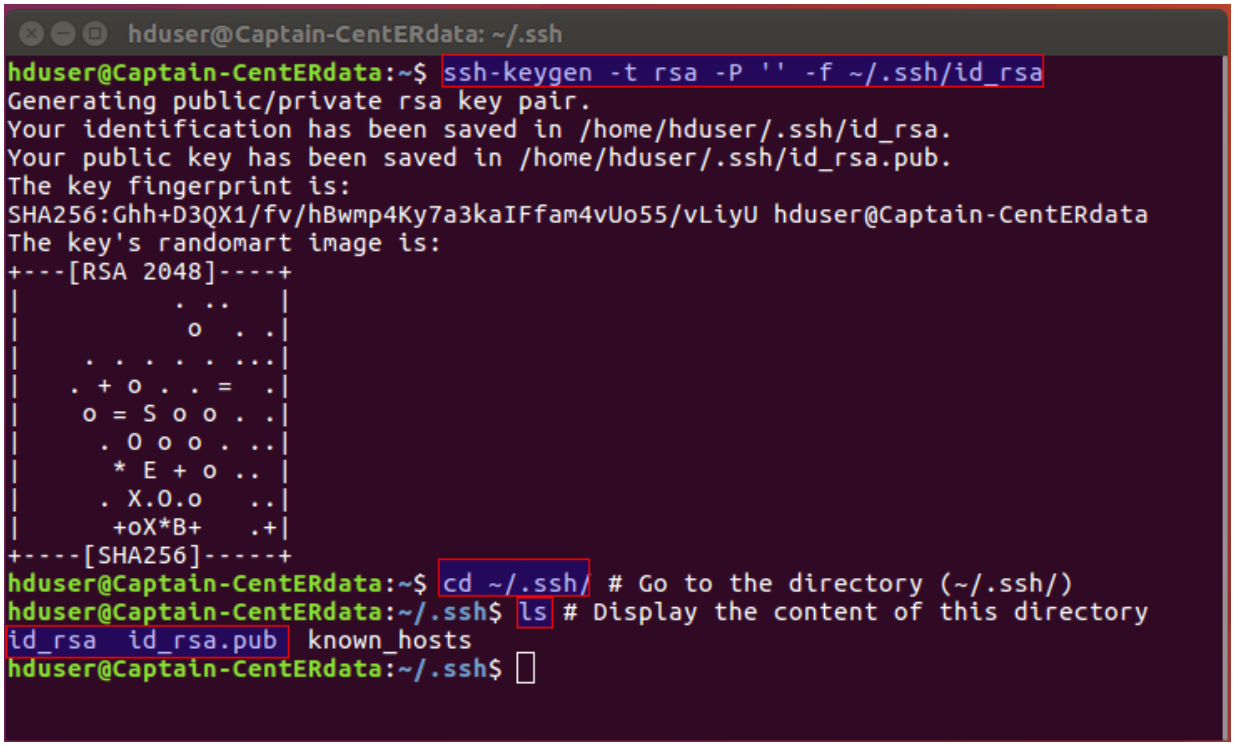
```
ssh localhost
```



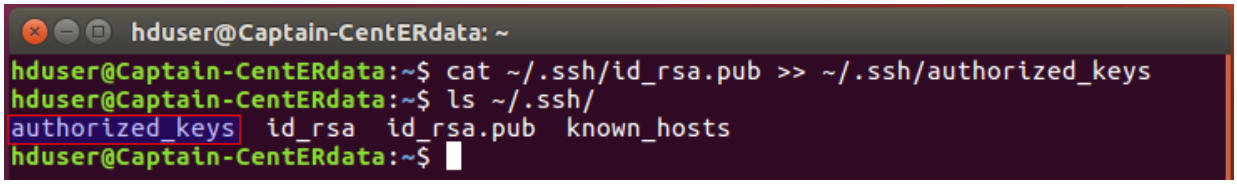
The above image shows that ssh has been installed, however, passphrase is required in order to ssh to the localhost.

- If we want to ssh to localhost without a passphrase, execute the following commands:

```
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
# (ssh-keygen) means generating keys for ssh.
# (-t rsa) sets the type of key as rsa.
# (-P '') provids '' as the passphrase.
# (-f ~/.ssh/id_rsa) specifies the filename of the key file as (~/.ssh/id_rsa).
# For more information, please refer to: http://linux.die.net/man/1/ssh-keygen
```

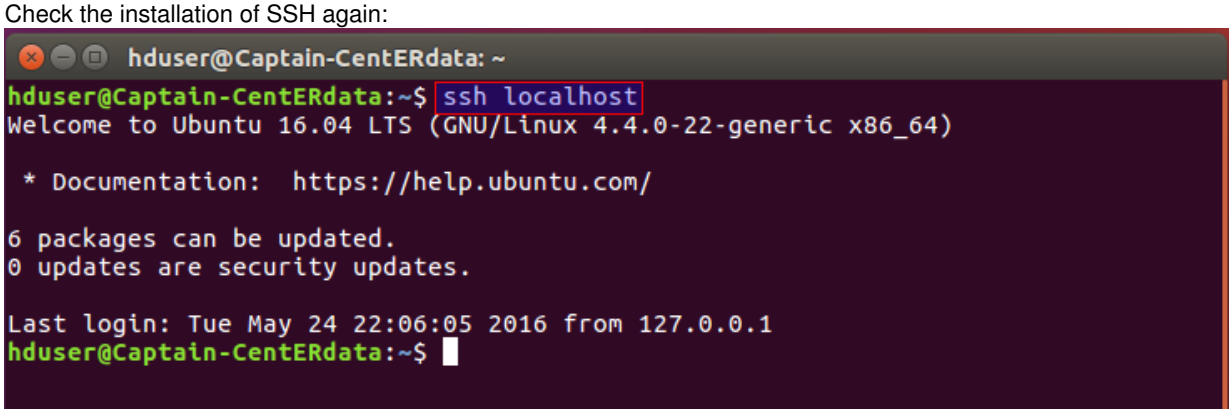



```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```




```
chmod 0600 ~/.ssh/authorized_keys
```

- Check the installation of SSH again:



The above figure shows that SSH has been installed successfully and we can ssh to the localhost without a passphrase. Up to now, we have finished the pre-installation, next we will introduce the steps of Hadoop installation.

Hadoop installation

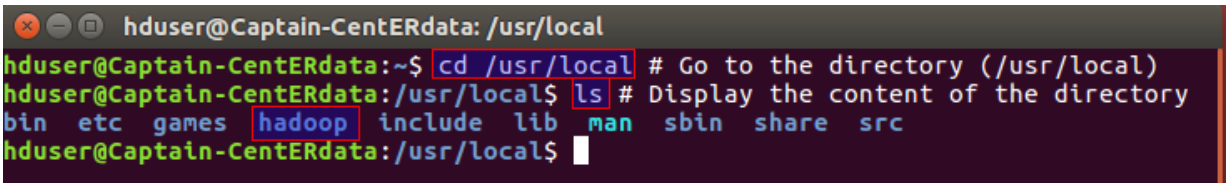
1. Download a recent stable release from one of the Apache Download Mirrors

Version	Release Date	Tarball	GPG	
2.6.4	11 February, 2016	source	signature	F75
		binary	signature	C58
2.7.2	25 January, 2016	source	signature	7D4
		binary	signature	49A
2.6.3	17 Dec, 2015	source	signature	FA0
		binary	signature	ADA

We download the Hadoop2.7.2 binary (do not download the source because it needs to be compiled before being used).
Hadoop-downloaded.png
It can be seen that the compressed file of Hadoop2.7.2 is download to the folder ¨Downloads¨.

2. Uncompress, move and rename the Hadoop file using the following commands:

```
sudo tar -zxf ~/Downloads/hadoop-2.7.2.tar.gz -C /usr/local
#-z(gzip), -x(extract), -f(file), -C(directory)
cd /usr/local/
sudo mv ./hadoop-2.7.2/ ./hadoop
sudo chown -R hduser ./hadoop
```



3. Add the Hadoop path to the environment path:

```
sudo gedit ~/.bashrc
```

Add the following line to the file ¨~/.bashrc¨.

```
export HADOOP_HOME=/usr/local/hadoop
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_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_INSTALL=$HADOOP_HOME
```

Save and close the file ¨~/.bashrc¨.
Run the following command to activate the changes:

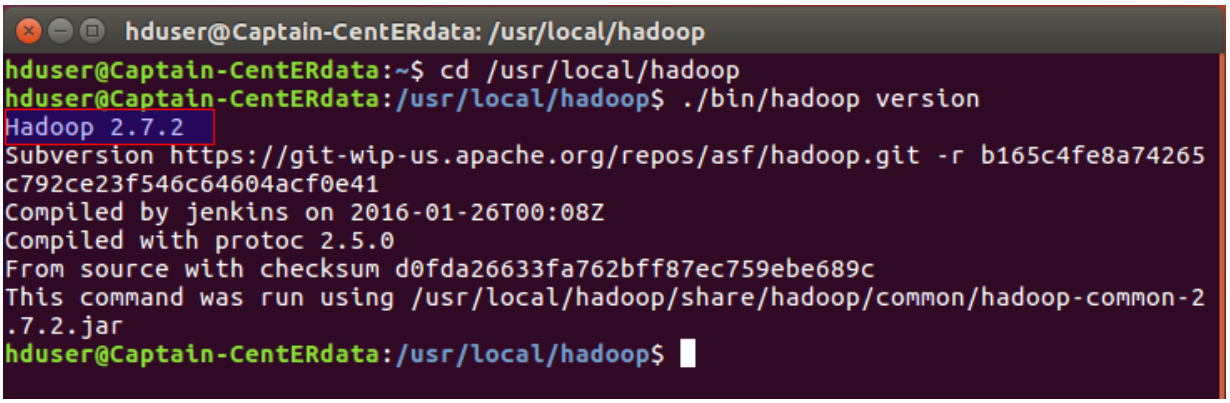
```
source ~/.bashrc
```

Up to now, the Hadoop for Signal Node Setup is finished. Next we will check whether Hadoop is installed properly, and then run some examples given by the official website.

Hadoop testing

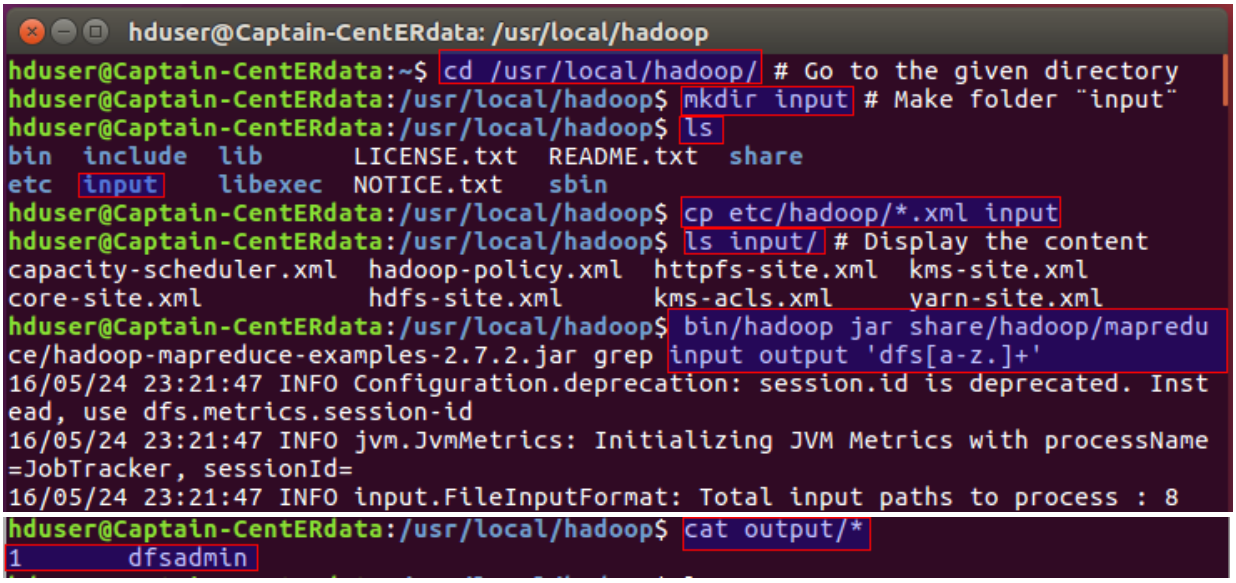
1. Check the version of Hadoop

```
hadoop version
```



2. Run the given example ¨grep¨

```
cd /usr/local/hadoop/ # Go to the given directory
mkdir input # Make folder ¨input¨
ls
cp etc/hadoop/*.xml input
ls input/ # Display the content
bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar grep input output '
dfs[a-z.]+'
cat output/*
```



Files

AccountCreated.png	12.2 KB	05/22/2016	Kai-Tao Yang
JAVA-not-installed.png	15.2 KB	05/22/2016	Kai-Tao Yang
JAVA-installed.png	24.3 KB	05/22/2016	Kai-Tao Yang
JAVA-home-path.png	20.4 KB	05/22/2016	Kai-Tao Yang
SSH-not-installed.png	16.5 KB	05/22/2016	Kai-Tao Yang
SSH-installed-need-password.png	77.5 KB	05/23/2016	Kai-Tao Yang
Hadoop-miroor.png	19.4 KB	05/24/2016	Kai-Tao Yang
SSH-keygen.png	66.2 KB	05/24/2016	Kai-Tao Yang
SSH-authorized-keys.png	24.2 KB	05/24/2016	Kai-Tao Yang
SSH-installed.png	32.3 KB	05/24/2016	Kai-Tao Yang
Hadoop-renamed.png	29.1 KB	05/24/2016	Kai-Tao Yang
Hadoop-version.png	49.6 KB	05/24/2016	Kai-Tao Yang
Hadoop-grep-runing.png	91.1 KB	05/24/2016	Kai-Tao Yang
Hadoop-grep-result.png	7.56 KB	05/24/2016	Kai-Tao Yang
JAVA-path-checking.png	24.5 KB	05/29/2016	Kai-Tao Yang
Hadoop-environment-path.png	24.9 KB	05/30/2016	Kai-Tao Yang