1. Install Homebrew and Cask

Homebrew is a free and open-source software package management system that simplifies the installation of software on Apple’s macOS operating system. Originally written by Max Howell, the package manager has gained popularity in the Ruby on Rails community and earned praise for its extensibility. Homebrew has been recommended for its ease of use as well as its integration into the command line.

$ ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"

$ brew install caskroom/cask/brew-cask

1. Install Java

$ brew update

$ brew cask install java

1. Configure SHH

In order to keep the safety of Hadoop remote administration as well as user sharing among Hadoop nodes, Hadoop requires SSH protocol. First, go to System Preferences -> Sharing, change Allow access for: **All Users**. Then open Terminal, input ssh localhost, if terminal returns Last login: Sun Jul 2 16:57:36 2017, which means that you have configured SSH Keys successfully before.

If you suffer the problem of ssh: connect to host localhost port 22: Connection refused, it happens since the remote login is closed.

$ sudo systemsetup -getremotelogin

Remote Login: off

You need to open port 22 in Mac OS X:

$ sudo systemsetup -setremotelogin on

$ ssh localhost

Last login: ...

And if you did not get the information of Last login: ..., then you need to create a new configuration:

$ ssh-keygen -t rsa

Executing the command above will generate a id\_rsa file in .ssh directory under the current user directory, after that, input the following command:

$ cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

which is used to authorize your public key to the local to prevent the passphrase request when you login, then, input ssh localhost again, if it will return Last login: ....

1. Install Hadoop

First, install Hadoop via Homebrew: brew install hadoop, it will install the hadoop under /usr/local/Cellar/hadoop. Then, you need to modify the configuration files.

Go to usr/local/Cellar/hadoop/2.8.0/libexec/etc/hadoop, then open hadoop-env.sh

export HADOOP\_OPTS="$HADOOP\_OPTS -Djava.net.preferIPv4Stack=true"

change to

export HADOOP\_OPTS="$HADOOP\_OPTS -Djava.net.preferIPv4Stack=true -Djava.security.krb5.realm= -Djava.security.krb5.kdc="

export JAVA\_HOME="/Library/Java/JavaVirtualMachines/jdk1.7.0\_79.jdk/Contents/Home"

Then configure HDFS address and port number, open core-site.xml, input following content in <configuration></configuration> tag

*<!-- Put site-specific property overrides in this file. -->*

<configuration>

<property>

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

<value>/usr/local/Cellar/hadoop/hdfs/tmp</value>

<description>A base for other temporary directories.</description>

</property>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:8020</value>

</property>

</configuration>

Configure jobtracker address and port number in map-reduce, first sudo cp mapred-site.xml.template mapred-site.xml to make a copy of mapred-site.xml, and open mapred-site.xml, add

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>localhost:8021</value>

</property>

</configuration>

Set HDFS default backup, the default value is 3, we should change to 1, open hdfs-site.xml, add

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

Before running background program, we should format the installed HDFS first, executing command hdfs namenode -format, when terminal returns a long inforamtion like:

17/07/02 16:11:05 INFO namenode.NameNode: STARTUP\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

......

17/07/02 16:11:07 INFO namenode.NameNode: SHUTDOWN\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SHUTDOWN\_MSG: Shutting down NameNode at haodemacbook-pro.local/192.168.1.4

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

It means that we finish HDFS configuration, and Hadoop is ready to launch. Besides, maybe you will get a warning

$ ... WARN util.NativeCodeLoader: Unable to load native-hadoop library **for** your platform... using builtin-java classes where applicable

It happens since you are running on 64-bit system but Hadoop native library is based on 32-bit. This is not a big issue. If it appears, you can fixed by refering this link: [here](https://stackoverflow.com/questions/19943766/hadoop-unable-to-load-native-hadoop-library-for-your-platform-warning).

1. Launch Hadoop

Go to /usr/local/Cellar/hadoop/2.8.0/sbin, execute:

$ ./start-dfs.sh *# start HDFS service*

$ ./stop-dfs.sh *# stop HDFS service*

Ternimal will return the following information:

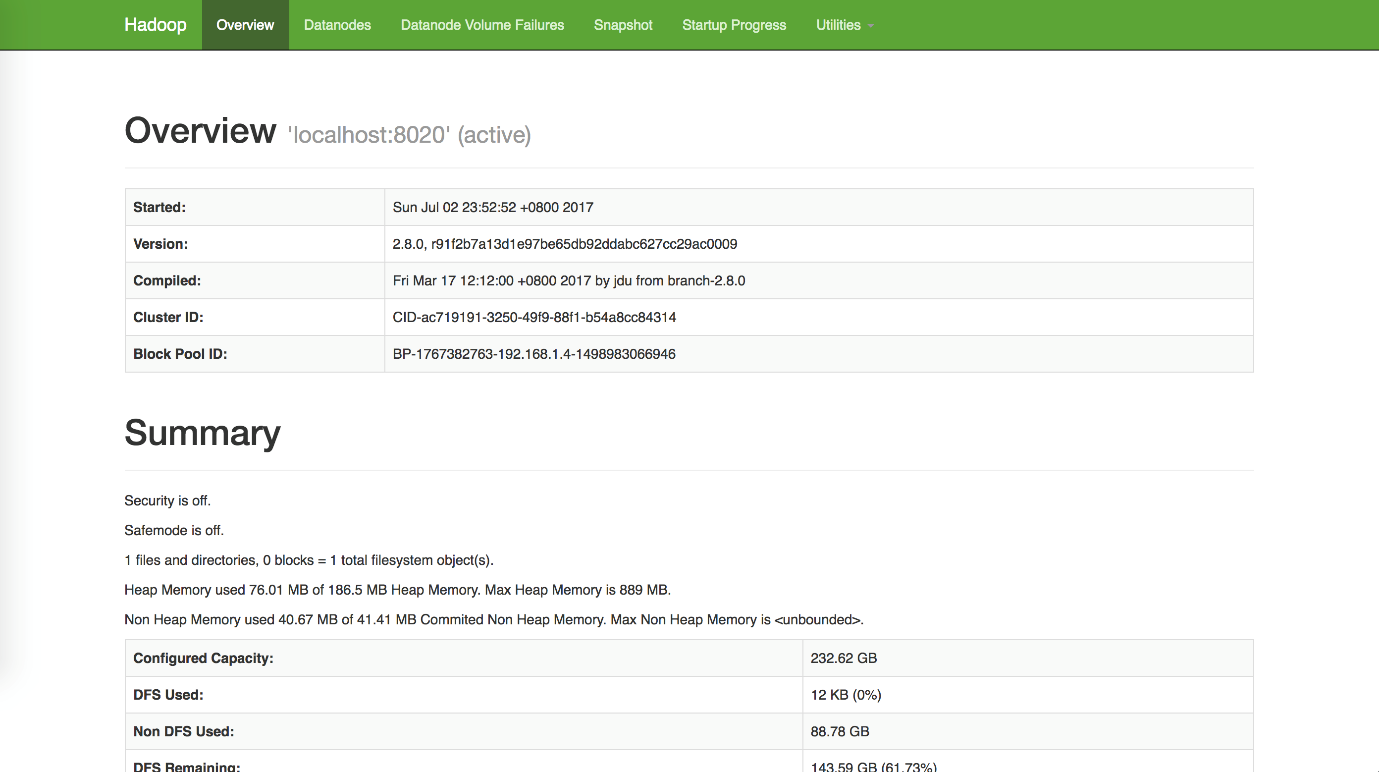
Starting namenodes on [localhost]

localhost: starting namenode, logging to /usr/local/Cellar/hadoop/2.8.0/libexec/logs/hadoop-zhanghao-namenode-HaodeMacBook-Pro.local.out

localhost: starting datanode, logging to /usr/local/Cellar/hadoop/2.8.0/libexec/logs/hadoop-zhanghao-datanode-HaodeMacBook-Pro.local.out

Starting secondary namenodes [0.0.0.0]

It means the local service launched successfully, then open Resource Manager in browser through the link http://localhost:50070, you can see the following page

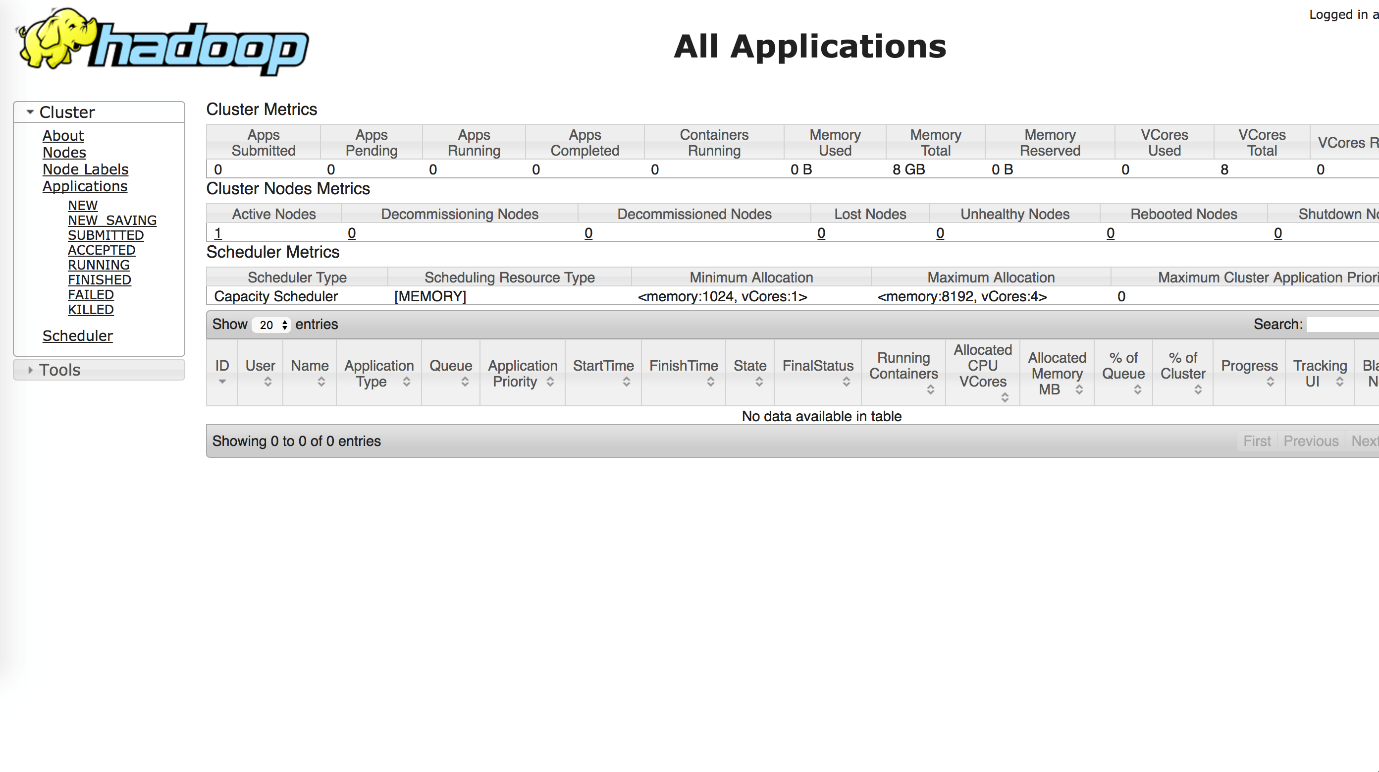


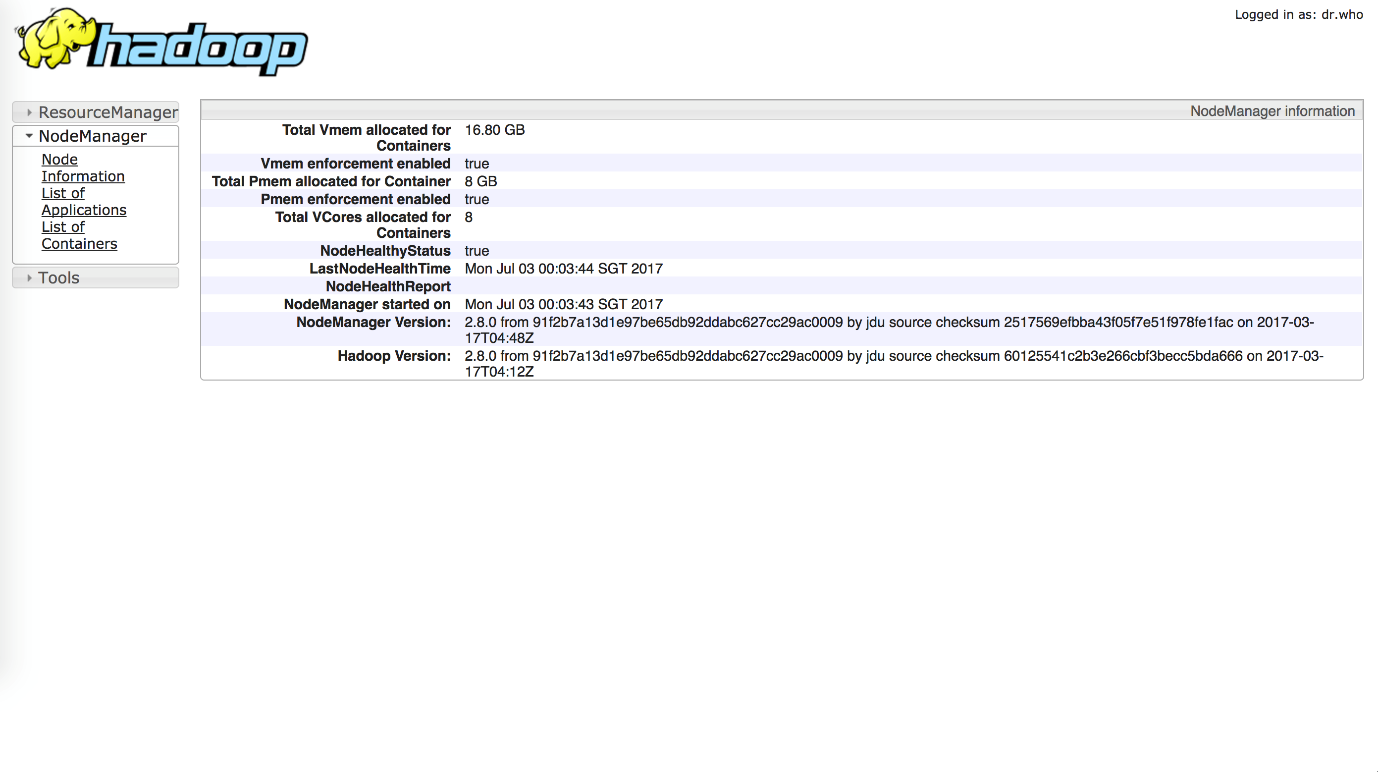
Samely, under current diretory, you can start the JobTracker through the commands:

$ ./start-yarn.sh *# start yarn, MapReduce framework*

$ ./stop-yarn.sh *# stop yarn*

Then open browser and go to the page http://localhost:8088, Specific Node Information http://localhost:8042, you will see





Simply, you can execute ./start-all.sh and ./stop-all.sh to start or close all the hadoop service. Finally, open /etc/profile and add the configuration information of Hadoop environment variables.

export HADOOP\_HOME=/usr/local/Cellar/hadoop/2.8.0

export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin

Then you can start and close Hadoop under the user directory rather than go to /usr/local/Cellar/hadoop/2.8.0/sbin every time.

Reference:

<https://isaacchanghau.github.io/post/install_hadoop_mac/>