Install and configure hadoop on Mac

Configure ssh

Enable remote login and add current user to allow list.

System Preferences -> Sharing -> Remote Login

É	Finder	File	Edit	View	Go	Window	Help
Abou	it This Ma	ic					
Syste	em Prefer	ences.					
App S	Store		6	updates			
Rece	nt Items			>			
Force	e Quit			C # D			
Sleep Resta Shut	o art Down						
Lock	Screen			^ # Q			
Log	Dut staso	ph					

General	Desktop & Screen Saver	Dock & Menu Bar	Mission Control	Siri	Q Spotlight	Language & Region	Notifications & Focus
(Q) Internet Accounts	Passwords	Users & Groups	Accessibility	Screen Time	Extensions	Security & Privacy	
Software Update	Network	Bluetooth	Sound	Touch ID	Keyboard	Trackpad	Mouse
Displays	Printers & Scanners	Battery	Date & Time	Sharing	Time Machine	Startup Disk	
Java							

Add current user to allow access list

••• < > s	Q Search	
Computer Name: Computers stasophdet	的 MacBook Pro on your local network can access your computer at: MacBook-Pro.local	Edit
On Service Screen Sharing File Sharing Media Sharing Printer Sharing Remote Login Remote Management	 Remote Login: On To log in to this computer remotely, type "ssh state Allow full disk access for remote users Allow access for: All users Only these users: 	soph@10.248.138.47".
 Remote Apple Events Bluetooth Sharing Internet Sharing Content Caching AirPlay Receiver 	Administrators	

```
ssh-keygen -t rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

#verify ssh
ssh localhost

Install java

Confirm you have the correct version of java (version 8 or 11)

you can download java11 from the link below or install java 8 or java 11 on your own

jdk download link

https://aka.ms/download-jdk/microsoft-jdk-11.0.15-macOS-x64.tar.gz

unzip and install jdk

```
#unzip jdk
tar xvf microsoft-jdk-11.0.15-macOS-x64.tar.gz
#remove downloaded file
rm microsoft-jdk-11.0.15-macOS-x64.tar.gz
#make directories
sudo mkdir -p /Library/Java/JavaVirtualMachines
#move jdk to JavaVirtualMachines
sudo mv jdk-11.0.15+10 /Library/Java/JavaVirtualMachines
```

Open the file zshrc (bashrc if you use linux)

open -a TextEdit ~/.zshrc

Then add the following lines to the **end** of this file (Note that you can find java home path by using /usr/libexec/java_home in terminal window):

export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk-11.0.15+10/Contents/Home

Save the file, and then run the following command to take these configurations into effect:

source ~/.zshrc

Download and configure Hadoop

1. Download Hadoop and Configure HADOOP_HOME

```
mkdir ~/workdir
cd ~/workdir
#Download the Hadoop package, you can also use wget or other tools to download
the file
curl -0 https://dlcdn.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz
#unpack the package
tar xvf hadoop-3.3.2.tar.gz
```

2. Configure HADOOP_HOME

Open the file zshrc (bashrc if you use linux)

open -a TextEdit ~/.zshrc

Then add the following lines to the **end** of this file (Note that you can find java home path by using /usr/libexec/java_home in terminal window):

export HADOOP_HOME=~/workdir/hadoop-3.3.2
export HADOOP_CONF_DIR=\$HADOOP_HOME/etc/hadoop
export PATH=\$HADOOP_HOME/bin:\$HADOOP_HOME/sbin:\$PATH
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk-11.0.15+10/Contents/Home

Save the file, and then run the following command to take these configurations into effect:

source ~/.zshrc

3.Configure HDFS

hadoop-env.sh

open the hadoop environment file, hadoop-env.sh, using text edit (or use vim, emacs or nano):

open -a TextEdit \$HADOOP_CONF_DIR/hadoop-env.sh

add the following to the end of this file

export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk-11.0.15+10/Contents/Home

core-site.xml

open the HDFS core configuration file

open -a TextEdit \$HADOOP_CONF_DIR/core-site.xml

Note that it is in xml format, and every configuration should be put in between <configuration> and </configuration>. You need to add the following lines:

Please change username to your own username (You can use whoami or 1s /Users to check your username)

• hdfs-site.xml

Open hdfs-site.xml

open -a TextEdit \$HADOOP_CONF_DIR/hdfs-site.xm]

```
<property>
    <name>dfs.replication</name>
    <value>1</value>
</property>
    <name>dfs.namenode.name.dir</name>
    <value>file://${hadoop.tmp.dir}/dfs/name</value>
</property>
    <name>dfs.datanode.data.dir</name>
    <value>file://${hadoop.tmp.dir}/dfs/data</value>
</property>
```

mapred-site.xml

Open mapred-site.xml

open -a TextEdit \$HADOOP_CONF_DIR/mapred-site.xml

```
<property>
<name>mapreduce.map.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_MAPRED_HOME</value>
</property>
<property>
<name>mapreduce.reduce.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_MAPRED_HOME</value>
```

</property>

• yarn-site.xml

open -a TextEdit \$HADOOP_CONF_DIR/yarn-site.xm]

```
<property>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
```

4. Start HDFS

```
# Work in the Hadoop home folder
cd $HADOOP_HOME
#Format the NameNode
hdfs namenode -format
#Start HDFS
start-dfs.sh
#Use the command "jps" to see whether Hadoop has been started successfully. You
should see "SecondaryNameNode", "NameNode", "Jps" and "DataNode"
jps
```

You can browse the web interface for the information of NameNode and DataNode at: <u>http://loca</u><u>lhost:9870</u>.

5. Start YARN

start-yarn.sh

Try jps again, you will see "NodeManager" and "ResourceManager", and these are the main daemons of YARN.

Browse the web interface (for supervision and debugging) for the ResourceManager at: <u>http://loca</u><u>lhost:8088/</u>.

6. Using HDFS and running MapReduce in the pseudo-distributed mode

Please change username to your own username (You can use whoami or 1s /Users to show your username)

```
hdfs dfs -mkdir -p /user/username
```

See the rest in file "Lab 1".