Follow the steps in below link to configure Hadoop and Yarn on your system:-

<https://www.quickprogrammingtips.com/big-data/how-to-install-hadoop-on-mac-os-x-el-capitan.html>

Steps to configure passwordless login:-

As in above case munichong is a user (munichong@GrindPad)

1. In my case: Login as hduser
2. Firstly, remove the directorysudo rm -rf ~/.ssh
3. Use to re-generate /.ssh directory with default setting:
4. [hduser@localhost ~]$ ssh-keygen
5. Here we do copy and paste the content of id\_rsa.pub into authorised\_keys file created by using above command)
6. [hduser@localhost ~]$ sudo cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys
7. [hduser@localhost ~]$ chmod -R 750 ~/.ssh/authorized\_keys
8. [hduser@localhost ~]$ ssh localhost

The authenticity of host 'localhost (127.0.0.1)' can't be established. RSA key fingerprint is 04:e8:80:64:dc:71:b5:2f:c0:d9:28:86:1f:61:60:8a. Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'localhost' (RSA) to the list of known hosts. Last login: Mon Jan 4 14:31:05 2016 from localhost.localdomain

HADOOP Installation verification.

<http://localhost:9870/dfshealth.html#tab-overview>

Configuring multiple nodes in a pseudo-distributed mode:

[1) https://www.quora.com/Is-it-possible-to-have-multiple-data-nodes-in-pseudo-distributed-Hadoop](1)%20https://www.quora.com/Is-it-possible-to-have-multiple-data-nodes-in-pseudo-distributed-Hadoop)

2) <https://bigdata.wordpress.com/2010/05/27/hadoop-cookbook-4-how-to-run-multiple-data-nodes-on-one-machine/>

In order to format HDFS and start from the scratch, we make use of this command. This should only be done once during initial setup. Otherwise all your data would be formatted.

bin/hdfs namenode -format (for version > 0.21)

Important – The nodes keep on running until we close or shutdown. I ran the start-dfs on Saturday, closed terminal and logged back on Tuesday to observe that the processes were still active and Hadoop localhost shows the process was started on saturday.

3) Work on creating multiple nodes in pseudo-distributed mode.

- Next action item is to create multiple nodes using a shell script.

sysctl -n hw.ncpu --- To display the number of cores in my CPU.

Installing Hadoop in a multi-cluster Environment.

<https://examples.javacodegeeks.com/enterprise-java/apache-hadoop/apache-hadoop-wordcount-example/>

Setting-up Apache Storm on a multi-cluster environment:-

<https://medium.com/real-time-streaming/setting-up-a-single-node-apache-storm-cluster-3dda02add2e9>

Beginner’s guide to siddhi.

<https://medium.com/@niruhan/a-beginners-guide-to-siddhi-complex-event-processor-efa4bd68a71c>

<http://localhost:9390/editor> ----- Siddhi Editor.

<https://stackoverflow.com/questions/42643505/difference-between-esper-and-apache-storm>

Interesting reading, study the above link.

Apache Storm Single Node Cluster Installation --- <https://www.youtube.com/watch?v=O43QCacyKrg&list=PL9ooVrP1hQOFuaHCbwTxi6gycZl27fsKm&index=8>

Apache storm very good installation Document : <https://www.tutorialspoint.com/apache_storm/apache_storm_installation.htm>

Study about Apache storm from pluralsight. --- DONE.

Check out kafka stream. This provides Real-time event analysis.

<https://www.confluent.io/blog/introducing-kafka-streams-stream-processing-made-simple/>