

Lee Brunovsky

CSCI 6910 Cloud Computing and Security

CH4 *Hadoop & MapReduce*, **CH5** *Cloud Application Design*: Lab Exercises

HW3

Due: July 11, 2021

Submitted: July 09, 2021

Ch4Q1:

- Create 2 EC2 instances. I used t2.micro over m1.small since it was free tier.

Instance ID	Instance state	Inst...	S...	A...	Availa...	Public IPv4 DNS	Public IPv4 ...					Key n...	Launch time
i-0939d782b61c62...	Running	t2.micro	2/	+	us-east-2c	ec2-3-134-101-237.us-east-2.compute.amazonaws.com	3.134.101.237	-	-	d...	L...	GSW, HW3	2021/07/03 1
i-06832f5675331268f	Running	t2.micro	2/	+	us-east-2c	ec2-3-17-28-61.us-east-2.compute.amazonaws.com	3.17.28.61	-	-	d...	L...	GSW, HW3	2021/07/03 1

- SSH to to public DNS of both instances

ssh -i "GSWHW3.pem" ubuntu@ec2-3-134-101-237.us-east-2.compute.amazonaws.com

```
c:\Users\leebr\Desktop>ssh -i "GSWHW3.pem" ubuntu@ec2-3-134-101-237.us-east-2.compute.amazonaws.com
The authenticity of host 'ec2-3-134-101-237.us-east-2.compute.amazonaws.com (3.134.101.237)' can't be established.
ECDSA key fingerprint is SHA256:ddjHrERQEX6UPBvxegk1jZUKD9m7oZWpgSczbZKzN0c.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-3-134-101-237.us-east-2.compute.amazonaws.com,3.134.101.237' (ECDSA) to the list of know
n hosts.
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.4.0-1045-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Sat Jul  3 05:43:51 UTC 2021

System load:  0.0          Processes:           100
Usage of /:   16.4% of 7.69GB Users logged in:     0
Memory usage: 22%         IPv4 address for eth0: 172.31.41.162
Swap usage:   0%
```

ssh -i "GSWHW3.pem" ubuntu@ec2-3-17-28-61.us-east-2.compute.amazonaws.com

```
C:\Users\leebr>cd c:\users\leebr\desktop
c:\Users\leebr\Desktop>ssh -i "GSWHW3.pem" ubuntu@ec2-3-17-28-61.us-east-2.compute.amazonaws.com
The authenticity of host 'ec2-3-17-28-61.us-east-2.compute.amazonaws.com (3.17.28.61)' can't be established.
ECDSA key fingerprint is SHA256:Z/xeu6T+qSvKuwbSorQFic5UHNm+NTai/gWHPVo5JVQ.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-3-17-28-61.us-east-2.compute.amazonaws.com,3.17.28.61' (ECDSA) to the list of known host
s.
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.4.0-1045-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Sat Jul  3 05:43:54 UTC 2021

System load:  0.0          Processes:           100
Usage of /:   16.4% of 7.69GB Users logged in:     0
Memory usage: 23%         IPv4 address for eth0: 172.31.42.235
Swap usage:   0%
```

-Install Java 6+ on each instance

sudo apt-get update | sudo apt install openjdk-8-jre-headless

```
ubuntu@ip-172-31-41-162:~$ sudo apt-get update
Hit:1 http://us-east-2.ec2.archive.ubuntu.com/ubuntu focal InRelease

ubuntu@ip-172-31-42-235:~$ sudo apt-get update
Hit:1 http://us-east-2.ec2.archive.ubuntu.com/ubuntu focal InRelease

ubuntu@ip-172-31-41-162:~$ sudo apt install openjdk-8-jre-headless
Reading package lists... Done

ubuntu@ip-172-31-42-235:~$ sudo apt install openjdk-8-jre-headless
Reading package lists... Done

ubuntu@ip-172-31-41-162:~$ java -version
openjdk version "1.8.0_292"
OpenJDK Runtime Environment (build 1.8.0_292-8u292-b10-0ubuntu1~20.04-b10)
OpenJDK 64-Bit Server VM (build 25.292-b10, mixed mode)

ubuntu@ip-172-31-42-235:~$ java -version
openjdk version "1.8.0_292"
OpenJDK Runtime Environment (build 1.8.0_292-8u292-b10-0ubuntu1~20.04-b10)
OpenJDK 64-Bit Server VM (build 25.292-b10, mixed mode)
```

-Install and configure Hadoop on each instance

wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz

```
ubuntu@ip-172-31-41-162:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
--2021-07-03 05:58:06-- https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
Resolving archive.apache.org (archive.apache.org)... 138.201.131.134, 2a01:4f8:172:2ec5::2
Connecting to archive.apache.org (archive.apache.org)|138.201.131.134|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 214092195 (204M) [application/x-gzip]
Saving to: 'hadoop-2.7.3.tar.gz'

hadoop-2.7.3.tar.gz      100%[=====>] 204.17M  14.5MB/s   in 15s
2021-07-03 05:58:21 (13.9 MB/s) - 'hadoop-2.7.3.tar.gz' saved [214092195/214092195]
```

```
ubuntu@ip-172-31-42-235:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
--2021-07-03 05:57:56-- https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
Resolving archive.apache.org (archive.apache.org)... 138.201.131.134, 2a01:4f8:172:2ec5::2
Connecting to archive.apache.org (archive.apache.org)|138.201.131.134|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 214092195 (204M) [application/x-gzip]
Saving to: 'hadoop-2.7.3.tar.gz'

hadoop-2.7.3.tar.gz      100%[=====>] 204.17M  14.7MB/s   in 15s
2021-07-03 05:58:11 (13.8 MB/s) - 'hadoop-2.7.3.tar.gz' saved [214092195/214092195]
```

tar xvzf hadoop-2.7.3.tar.gz

(End of long pkg: Concatenated for brevity)

```
hadoop-2.7.3/LICENSE.txt  hadoop-2.7.3/LICENSE.txt
hadoop-2.7.3/NOTICE.txt   hadoop-2.7.3/NOTICE.txt
hadoop-2.7.3/README.txt   hadoop-2.7.3/README.txt
ubuntu@ip-172-31-41-162:~$  ubuntu@ip-172-31-42-235:~$
```

Change the host name of the node & reboot instances

```
ubuntu@ip-172-31-41-162:~$ sudo nano /etc/hostname  ubuntu@ip-172-31-41-162: ~
ubuntu@ip-172-31-42-235:~$ sudo nano /etc/hostname  GNU nano 4.8
master
```

Add private IPv4 addressees for master and slave nodes.

```
ubuntu@master:~$ sudo nano /etc/hosts  ubuntu@master: ~
ubuntu@slave:~$ sudo nano /etc/hosts  GNU nano 4.8
127.0.0.1 localhost
172.31.41.162 master
172.31.42.235 slave
```

Rename hadoop-2.7.3 to hadoop for simplicity

```
ubuntu@master:~$ mv hadoop-2.7.3 hadoop  ubuntu@slave:~$ mv hadoop-2.7.3 hadoop
```

Update JAVA_HOME path in hadoop-env.sh

```
ubuntu@master:~/hadoop/etc/hadoop$ nano hadoop-env.sh  ubuntu@slave:~/hadoop/etc/hadoop$ nano hadoop-env.sh
ubuntu@master: ~/hadoop/etc/hadoop  hadoop-env.sh  ubuntu@slave: ~/hadoop/etc/hadoop  hadoop-env.sh
# The java implementation to use.
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
```

Create and copy public key to allow authentication-less SSH

```
ubuntu@master:/home$ ssh-keygen -t rsa -f /home/ubuntu/.ssh/id_rsa
Generating public/private rsa key pair.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/ubuntu/.ssh/id_rsa
Your public key has been saved in /home/ubuntu/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:6PIHpUtJqzPPuqG/csEgVpj94UYTcztkhKT/fo1kkQM ubuntu@master
The key's randomart image is:
+---[RSA 3072]-----+
|  =.+..          |
|  o..+Eo .       |
|  ..+.+.o.       |
|  ..o .+o=o      |
|  ..+.o.=S       |
|  +.*o           |
|  o=+oo          |
|  ..*=..o..      |
|  .+o=.         |
+---[SHA256]-----+
```

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

```
ubuntu@master:/home$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
ubuntu@master:/home$ cat ~/.ssh/authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDAQDQ67ncw/7oPxTjY22qSSX7FklspQPmLwsljrJB/Bfn5oIXcP05KGR37nBkyAoiofG26+IaHSvk02RcSs
85moCo7zWqJPLUaWxN2A5sXdkzb+f5JFx6as9qqVEPTQ3qJBtELkoazacjy4MPxdET3ZWchYzgU0fzFCVubV4Jk2NPRkQp4s1XQtffffjreqSEpOmyFWhBKYL
cjh3uhaZ6YNmudSvu5RogARVTFVROHmDcnB0b0JzMjxBbX8tvQ1tV/dINKGpW8sB0WP3+AsB1WAQ09BovcIp2Wv1aaLQvuD1817MPNBK4WJXrmTVRxj6Lbk
Rr/LPn7sBiWnhwWcjzL1 GSW, HW3
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDAQDGMUYW4QMcrT5478j5ESTnh52/RvQ6xel4CdigepqNB1W8cFQhft1zRp/zB8jxc5LNR32v93BT78k+gkde
upWq6sUVrwWapm47/soFOaZvvhTCHXqlwZ9RypLHDE2yTgmw4zv/n9d0GDcatSDY8lJGs2GD+Lh7kYCJEIr6HspJldtGL/c74y2K/3MPJGP921RZn9jDpInn
ihZQQQwoAv8w8ICV4PTpQADQCnI61s0b2ywYacrb0E+LjrmTUZJj1/bx+YrRy9WiWRJEQHTJnA9QEek4ntdxSF10QczvP1+/ptLL2vWRg/oCNxLiBdZuMoeD
OEw31Fa4vt1IwXoxJkEebVtJJ4msG8btPCR6iGZZAiE14NuVopGnys1qjpSDpf41ypN1I5z9Ak0nm8vFeVm9r90y0Qg+wczRJgJzxEGHvIt65IZnUy3V0r8p
lkZiofekLkSTbk0xssIPka4A8LYcsD/iFi8HwnxKlazRLK7+TrbzKi67RPsrkKhKn0ee5zk= ubuntu@master
```

Copy Master public key to slave

```
ubuntu@slave: ~/.ssh
GNU nano 4.8 authorized keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDAQDQ67ncw/7oPxTjY22qSSX7FklspQPmLwsljrJB/Bfn5oIXcP05KGR37nBkyAoiofG26+IaHSvk02RcSs
85moCo7zWqJPLUaWxN2A5sXdkzb+f5JFx6as9qqVEPTQ3qJBtELkoazacjy4MPxdET3ZWchYzgU0fzFCVubV4Jk2NPRkQp4s1XQtffffjreqSEpOmyFWh
BKYLcjh3uhaZ6YNmudSvu5RogARVTFVROHmDcnB0b0JzMjxBbX8tvQ1tV/dINKGpW8sB0WP3+AsB1WAQ09BovcIp2Wv1aaLQvuD1817MPNBK4WJXrmTVRxj6Lbk
Rr/LPn7sBiWnhwWcjzL1 GSW, HW3
= ubuntu@master_
```

```
ubuntu@slave:~/.ssh$ cat authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDAQDQ67ncw/7oPxTjY22qSSX7FklspQPmLwsljrJB/Bfn5oIXcP05KGR37nBkyAoiofG26+IaHSvk02Rc
S85moCo7zWqJPLUaWxN2A5sXdkzb+f5JFx6as9qqVEPTQ3qJBtELkoazacjy4MPxdET3ZWchYzgU0fzFCVubV4Jk2NPRkQp4s1XQtffffjreqSEpOmyFWh
BKYLcjh3uhaZ6YNmudSvu5RogARVTFVROHmDcnB0b0JzMjxBbX8tvQ1tV/dINKGpW8sB0WP3+AsB1WAQ09BovcIp2Wv1aaLQvuD1817MPNBK4WJXrmTVRxj6Lbk
Rr/LPn7sBiWnhwWcjzL1 GSW, HW3
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDAQDGMUYW4QMcrT5478j5ESTnh52/RvQ6xel4CdigepqNB1W8cFQhft1zRp/zB8jxc5LNR32v93BT78k+gkde
upWq6sUVrwWapm47/soFOaZvvhTCHXqlwZ9RypLHDE2yTgmw4zv/n9d0GDcatSDY8lJGs2GD+Lh7kYCJEIr6HspJldtGL/c74y2K/3MPJGP921RZn9jDpInn
ihZQQQwoAv8w8ICV4PTpQADQCnI61s0b2ywYacrb0E+LjrmTUZJj1/bx+YrRy9WiWRJEQHTJnA9QEek4ntdxSF10QczvP1+/ptLL2vWRg/oCNxLiBdZuMoeD
OEw31Fa4vt1IwXoxJkEebVtJJ4msG8btPCR6iGZZAiE14NuVopGnys1qjpSDpf41ypN1I5z9Ak0nm8vFeVm9r90y0Qg+wczRJgJzxEGHvIt65IZnUy3V0r8p
lkZiofekLkSTbk0xssIPka4A8LYcsD/iFi8HwnxKlazRLK7+TrbzKi67RPsrkKhKn0ee5zk= ubuntu@master
```

Save host keys for each node (repeat same steps on slave)

```
ubuntu@master:~$ ssh master
The authenticity of host 'master (172.31.41.162)' can't be established.
ECDSA key fingerprint is SHA256:ddjHrERQEX6UPBvxegk1jZUKD9m7oZWpg5czbZKzN0c.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'master,172.31.41.162' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.4.0-1045-aws x86_64)
```

```
ubuntu@master:~$ ssh slave
The authenticity of host 'slave (172.31.42.235)' can't be established.
ECDSA key fingerprint is SHA256:Z/xeu6T+qSvKuwb5orQFic5UHNm+NTai/gWHPVo5JVQ.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'slave,172.31.42.235' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.4.0-1045-aws x86_64)
```

Edit core-site.xml, hdfs-site.xml and mapred-site.xml in master

```
ubuntu@master:~/hadoop$ sudo nano ~/hadoop/etc/hadoop/core-site.xml

ubuntu@master: ~/hadoop
GNU nano 4.8 /home/ubuntu/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://master:54310</value>
</property>
</configuration>

ubuntu@master:~/hadoop$ sudo nano ~/hadoop/etc/hadoop/hdfs-site.xml

ubuntu@master: ~/hadoop
GNU nano 4.8 /home/ubuntu/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>

ubuntu@master:~/hadoop$ sudo nano ~/hadoop/etc/hadoop/mapred-site.xml

ubuntu@master: ~/hadoop
GNU nano 4.8 /home/ubuntu/hadoop/etc/hadoop/mapred-site.xml
<?xml version="1.0"?>
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>hdfs://master:54311</value>
</property>
<property>
<name>mapred.job.tracker.http.address</name>
<value>0.0.0.0:50030</value>
</property>
</configuration>
```

Push config files to slave

scp /home/ubuntu/hadoop/etc/hadoop/* ubuntu@slave:/home/ubuntu/hadoop/etc/hadoop/

```
ubuntu@master:~$ scp /home/ubuntu/hadoop/etc/hadoop/* ubuntu@slave:/home/ubuntu/hadoop/etc/hadoop/
```

capacity-scheduler.xml	100%	4436	1.9MB/s	00:00
configuration.xsl	100%	1335	972.1KB/s	00:00
container-executor.cfg	100%	318	238.0KB/s	00:00
core-site.xml	100%	220	194.9KB/s	00:00
hadoop-env.cmd	100%	3589	2.5MB/s	00:00
hadoop-env.sh	100%	4246	7.3MB/s	00:00
hadoop-metrics.properties	100%	2490	1.3MB/s	00:00
hadoop-metrics2.properties	100%	2598	1.7MB/s	00:00
hadoop-policy.xml	100%	9683	5.2MB/s	00:00
hdfs-site.xml	100%	218	208.5KB/s	00:00
https-env.sh	100%	1449	1.0MB/s	00:00
https-log4j.properties	100%	1657	1.2MB/s	00:00
https-signature.secret	100%	21	15.7KB/s	00:00
https-site.xml	100%	620	204.2KB/s	00:00
kms-acls.xml	100%	3518	2.2MB/s	00:00
kms-env.sh	100%	1527	1.1MB/s	00:00
kms-log4j.properties	100%	1631	448.5KB/s	00:00
kms-site.xml	100%	5511	3.9MB/s	00:00
log4j.properties	100%	11KB	7.1MB/s	00:00
mapred-env.cmd	100%	931	652.2KB/s	00:00
mapred-env.sh	100%	1383	1.0MB/s	00:00
mapred-queues.xml.template	100%	4113	2.4MB/s	00:00
mapred-site.xml	100%	145	303.9KB/s	00:00
mapred-site.xml.template	100%	758	587.3KB/s	00:00
slaves	100%	10	8.3KB/s	00:00
ssl-client.xml.example	100%	2316	1.5MB/s	00:00
ssl-server.xml.example	100%	2268	1.5MB/s	00:00
yarn-env.cmd	100%	2191	1.2MB/s	00:00
yarn-env.sh	100%	4567	2.8MB/s	00:00
yarn-site.xml	100%	690	472.5KB/s	00:00

Edit masters and slaves configuration files on both instances. Note changed location in 2.7.3

```
ubuntu@master:~$ nano hadoop/etc/hadoop/masters
ubuntu@master:~$ nano hadoop/etc/hadoop/slaves

ubuntu@master: ~
GNU nano 4.8 hadoop/etc/hadoop/masters
172.31.41.162

ubuntu@master: ~
GNU nano 4.8 hadoop/etc/hadoop/slaves
localhost
172.31.42.235

ubuntu@slave:~$ nano hadoop/etc/hadoop/masters
ubuntu@slave:~$ nano hadoop/etc/hadoop/slaves

ubuntu@slave: ~
GNU nano 4.8 hadoop/etc/hadoop/slaves
localhost
172.31.42.235

Select ubuntu@slave: ~
GNU nano 4.8 hadoop/etc/hadoop/masters
172.31.41.162
```

Format namenode

```
ubuntu@master:~/hadoop$ bin/hdfs namenode -format
21/07/03 07:30:51 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = master/172.31.41.162
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.3
STARTUP_MSG: classpath = /home/ubuntu/hadoop/etc/hadoop:/home/ubuntu
/*****/
SHUTDOWN_MSG: Shutting down NameNode at master/172.31.41.162
*****/
```

start-dfs.sh

```
ubuntu@master:~/hadoop$ sbin/start-dfs.sh
Starting namenodes on [master]
master: starting namenode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-namenode-master.out
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:ddjHrERQEX6UPBvxegk1jZUKD9m7oZWpg5czbZKzNoc.
Are you sure you want to continue connecting (yes/no/[fingerprint])? 172.31.42.235: starting datanode, logging to /home/
ubuntu/hadoop/logs/hadoop-ubuntu-datanode-slave.out
yes
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
localhost: starting datanode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-datanode-master.out
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is SHA256:ddjHrERQEX6UPBvxegk1jZUKD9m7oZWpg5czbZKzNoc.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts.
0.0.0.0: starting secondarynamenode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-secondarynamenode-master.out
```

start-all.sh (deprecated, use start-yarn.sh)

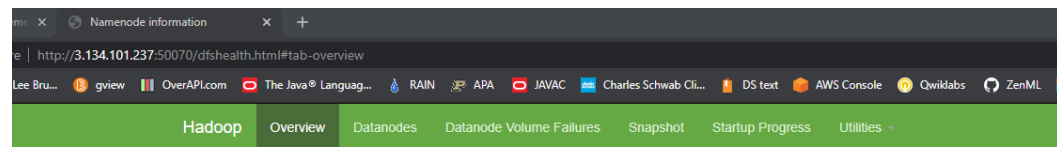
```
ubuntu@master:~/hadoop$ sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [master]
master: namenode running as process 17802. Stop it first.
localhost: datanode running as process 17984. Stop it first.
172.31.42.235: datanode running as process 18289. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 18192. Stop it first.
starting yarn daemons
starting resourcemanager, logging to /home/ubuntu/hadoop/logs/yarn-ubuntu-resourcemanager-master.out
172.31.42.235: starting nodemanager, logging to /home/ubuntu/hadoop/logs/yarn-ubuntu-nodemanager-slave.out
localhost: starting nodemanager, logging to /home/ubuntu/hadoop/logs/yarn-ubuntu-nodemanager-master.out
```

Verify namenode and secondary node are up

```
ubuntu@master:~/hadoop-2.7.3$ ps -aux | grep java | awk '{print $12}'
-Dproc_namenode
-Dproc_datanode
-Dproc_secondarynamenode
-Dproc_resourcemanager
-Dproc_nodemanager
```

```
ubuntu@slave:~$ ps -aux | grep java | awk '{print $12}'
-Dproc_datanode
-Dproc_nodemanager
```

Open hadoop cluster status pages



Overview 'master:54310' (active)

Started:	Sat Jul 03 07:33:38 UTC 2021
Version:	2.7.3, rbaa91f7c6bc92be5982de4719c1c8af91ccff
Compiled:	2016-08-18T01:41Z by root from branch-2.7.3
Cluster ID:	CID-2bc3cebd-f00f-4278-91e2-d5946ab31a55
Block Pool ID:	BP-555548357-172.31.41.162-1625297453038

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 37.52 MB of 46.3 MB Heap Memory. Max Heap Memory is 966.69 MB.

Non Heap Memory used 42.83 MB of 43.69 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

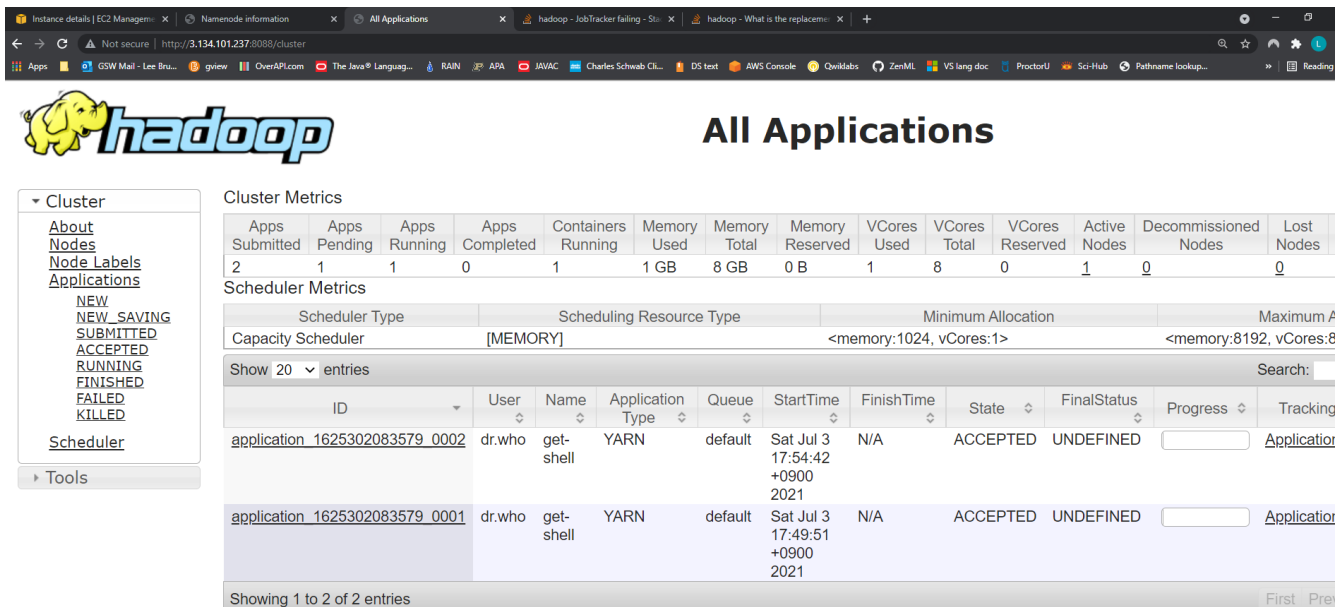
Configured Capacity:	15.38 GB
DFS Used:	52 KB (0%)
Non DFS Used:	4.73 GB
DFS Remaining:	10.65 GB (69.26%)
Block Pool Used:	52 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	2 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
slave:50010 (172.31.42.235:50010)	1	In Service	7.69 GB	24 KB	2.37 GB	5.33 GB	0	24 KB (0%)	0	2.7.3
master:50010 (172.31.41.162:50010)	1	In Service	7.69 GB	28 KB	2.36 GB	5.33 GB	0	28 KB (0%)	0	2.7.3

resourcemanager on :8088 replaced task tracker post Apache 2.3.0



hadoop

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCoers Used	VCoers Total	VCoers Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes
2	1	1	0	1	1 GB	8 GB	0 B	1	8	0	1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking
application_1625302083579_0002	dr.who	get-shell	YARN	default	Sat Jul 3 17:54:42 +0900 2021	N/A	ACCEPTED	UNDEFINED		Application
application_1625302083579_0001	dr.who	get-shell	YARN	default	Sat Jul 3 17:49:51 +0900 2021	N/A	ACCEPTED	UNDEFINED		Application

Showing 1 to 2 of 2 entries

Ch4Q2:

Note: Re-using instances from Q1.

```
ubuntu@master:~$ mkdir data
ubuntu@master:~$ ls
data  hadoop  hadoop-2.7.3.tar.gz
```

Copy text files to master 'data' folder

wget https://filesamples.com/samples/document/txt/sample3.txt


```

ubuntu@master:~$ wget https://filesamples.com/samples/document/txt/sample3.txt -O /home/ubuntu/data
/home/ubuntu/data: Is a directory
ubuntu@master:~$ cd data
ubuntu@master:~/data$ wget https://filesamples.com/samples/document/txt/sample3.txt
--2021-07-04 10:43:02-- https://filesamples.com/samples/document/txt/sample3.txt
Resolving filesamples.com (filesamples.com)... 172.67.178.244, 104.21.17.252, 2606:4700:3035::6815:11fc, ...
Connecting to filesamples.com (filesamples.com)|172.67.178.244|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/plain]
Saving to: 'sample3.txt'

sample3.txt           [ <=> ] 3.46K --.-KB/s  in 0s

2021-07-04 10:43:02 (34.1 MB/s) - 'sample3.txt' saved [3541]

ubuntu@master:~/data$ wget https://filesamples.com/samples/document/txt/sample2.txt
--2021-07-04 10:43:15-- https://filesamples.com/samples/document/txt/sample2.txt
Resolving filesamples.com (filesamples.com)... 104.21.17.252, 172.67.178.244, 2606:4700:3035::ac43:b2f4, ...
Connecting to filesamples.com (filesamples.com)|104.21.17.252|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/plain]
Saving to: 'sample2.txt'

sample2.txt           [ <=> ] 2.79K --.-KB/s  in 0s

2021-07-04 10:43:16 (36.5 MB/s) - 'sample2.txt' saved [2859]

```

```

ubuntu@master:~/data$ ls
sample2.txt  sample3.txt

```

re-start dfs and yarn

```

ubuntu@master:~/hadoop$ sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [master]
master: starting namenode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-namenode-master.out
localhost: starting datanode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-datanode-master.out
172.31.42.235: starting datanode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-datanode-slave.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/ubuntu/hadoop/logs/hadoop-ubuntu-secondarynamenode-master.out
starting yarn daemons
starting resourcemanager, logging to /home/ubuntu/hadoop/logs/yarn-ubuntu-resourcemanager-master.out
172.31.42.235: starting nodemanager, logging to /home/ubuntu/hadoop/logs/yarn-ubuntu-nodemanager-slave.out
localhost: starting nodemanager, logging to /home/ubuntu/hadoop/logs/yarn-ubuntu-nodemanager-master.out

```

create HDFS directory on master

```

ubuntu@master:~/hadoop$ bin/hadoop fs -mkdir -p master/data

```

Copy the files from the Hadoop master node local filesystem to HDFS

bin/hadoop fs -copyFromLocal /home/ubuntu/data master/data

```

ubuntu@master:~/hadoop$ bin/hadoop fs -copyFromLocal /home/ubuntu/data master
21/07/05 13:03:11 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:609)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:370)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:546)

```

Add dfs.webhdfs, enabled = true to hdfs-site.xml

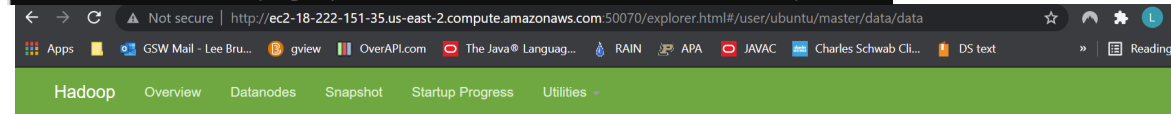
```
Select ubuntu@master: ~/hadoop/etc/hadoop
GNU nano 4.8 hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.webhdfs.enabled</name>
<value>true</value>
</property>
</configuration>
```

Verify files on webhdfs or using hadoop dfs

```
ubuntu@master:~/hadoop$ bin/hadoop dfs -ls master/data
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Found 2 items
-rw-r--r-- 1 ubuntu supergroup 2859 2021-07-05 13:03 master/data/sample2.txt
-rw-r--r-- 1 ubuntu supergroup 3541 2021-07-05 13:03 master/data/sample3.txt
```



Browse Directory

/user/ubuntu/master/data/data								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-rw-r--r--	ubuntu	supergroup	2.79 KB	7/5/2021, 7:22:56 PM	1	128 MB	sample2.txt	
-rw-r--r--	ubuntu	supergroup	3.46 KB	7/5/2021, 7:22:56 PM	1	128 MB	sample3.txt	

find / -name "hadoop-mapreduce-examples*.jar" -print

```
ubuntu@master:~/data$ sudo find / -name "hadoop-mapreduce-examples*.jar" -print
/home/ubuntu/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar
/home/ubuntu/hadoop/share/hadoop/mapreduce/sources/hadoop-mapreduce-examples-2.7.3-test-sources.jar
/home/ubuntu/hadoop/share/hadoop/mapreduce/sources/hadoop-mapreduce-examples-2.7.3-sources.jar
```

~/hadoop\$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount -D mapred.reduce.tasks=16 master/data output

*I edited text files to make them small after many run errors.

```
ubuntu@master:~/hadoop$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount -D mapred.reduce
.tasks=16 master/data output
21/07/05 13:37:45 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
21/07/05 13:37:45 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
21/07/05 13:37:45 INFO input.FileInputFormat: Total input paths to process : 2
21/07/05 13:37:45 INFO mapreduce.JobSubmitter: number of splits:2
    File Input Format Counters
        Bytes Read=69
    File Output Format Counters
        Bytes Written=90
```

Check output directory and cat part-00000

```
ubuntu@master:~/hadoop$ bin/hadoop dfs -ls output
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Found 17 items
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/_SUCCESS
-rw-r--r--  1 ubuntu supergroup      7 2021-07-05 13:37 output/part-r-00000
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00001
-rw-r--r--  1 ubuntu supergroup    34 2021-07-05 13:37 output/part-r-00002
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00003
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00004
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00005
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00006
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00007
-rw-r--r--  1 ubuntu supergroup      9 2021-07-05 13:37 output/part-r-00008
-rw-r--r--  1 ubuntu supergroup      7 2021-07-05 13:37 output/part-r-00009
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00010
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00011
-rw-r--r--  1 ubuntu supergroup    13 2021-07-05 13:37 output/part-r-00012
-rw-r--r--  1 ubuntu supergroup      0 2021-07-05 13:37 output/part-r-00013
-rw-r--r--  1 ubuntu supergroup    14 2021-07-05 13:37 output/part-r-00014
-rw-r--r--  1 ubuntu supergroup      6 2021-07-05 13:37 output/part-r-00015
```

```
ubuntu@master:~/hadoop$ bin/hadoop dfs -cat output/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

AWS,      1
```

Ch5Q1: *Extra credit*

-Create a Django application that uses the Django authentication system.

Create python virtual environment

```
PS C:\python39> pip install virtualenv
```

```
PS C:\python39> py -m venv C:\Users\leebr\Desktop\HW3
```

install Django & verify version

```
PS C:\python39> pip install django
```

```
PS C:\python39> py -m django --version
3.2.5
```

Create new django project

```
PS C:\Users\leebr\Desktop\HW3CH5Q1> python -m django startproject IPA_proj
```

Check the project structure and start the server on localhost:8000 to verify

```
File Edit Selection View Go Run Terminal Help HW3CH5Q1 - Visual Studio Code
```

EXPLORER

- HW3CH5Q1
 - IPA_proj
 - IPA_proj
 - __pycache__
 - __init__.py
 - asgi.py
 - settings.py
 - urls.py
 - wsgi.py
 - db.sqlite3
 - manage.py

TERMINAL

```
PS C:\Users\leebr\Desktop\HW3CH5Q1> cd IPA_proj
PS C:\Users\leebr\Desktop\HW3CH5Q1\IPA_proj> python manage.py runserver
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).

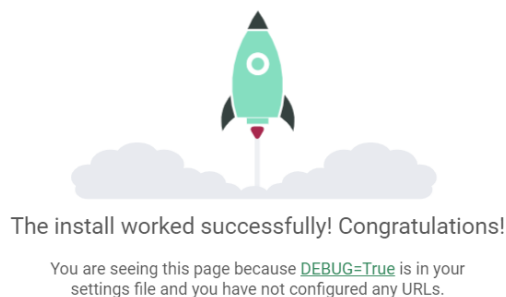
You have 18 unapplied migration(s). Your project may not work properly until you
apply the migrations for app(s): admin, auth, contenttypes, sessions.
Run 'python manage.py migrate' to apply them.
July 08, 2021 - 23:07:22
Django version 3.2.5, using settings 'IPA_proj.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

The install worked successfully! C X Yahoo | Mail, Weather, Search, P X +

localhost:8000 Not syncing

django

[View release notes](#) for Django 3.2



Create and application > python manage.py startapp appName

```
PS C:\Users\leebr\Desktop\HW3CH5Q1\IPA_proj> python manage.py startapp IPA
```

Add app urls to project urls path. **note:** can make the path an empty sting if home is desired
Add views.home to app path

```
views.py urls.py ..IPA_proj X urls.py ..IPA X
```

```
16 from django.contrib import admin
17 from django.urls import path, include
18
19 urlpatterns = [
20     path('admin/', admin.site.urls),
21     path('IPA/', include('IPA.urls')),
22 ]
```

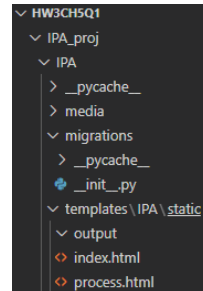
```
views.py urls.py ..IPA_proj X urls.py ..IPA X
```

```
1 from django.contrib import admin
2 from django.urls import path, include
3 from IPA import views
4
5 urlpatterns = [
6     path('', views.home, name='home'),
7 ]
```

Create templates/appName sub directory in application folder

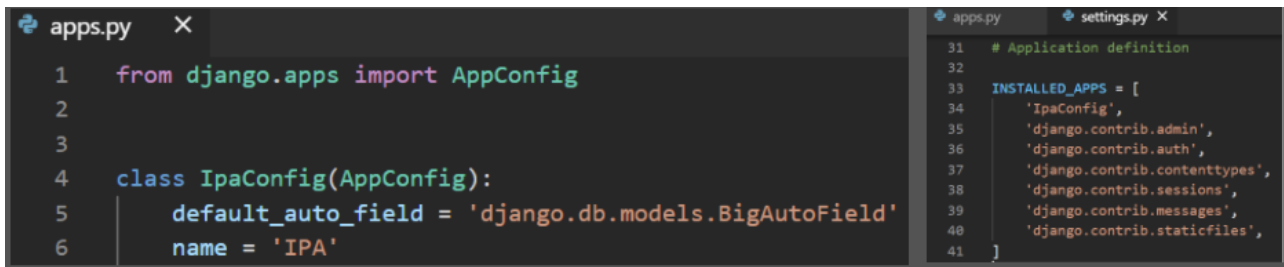
Add image submission(box 8.1) and results display (box 8.2) front-end html provided by text to respective templates

folders as index.html & process.html respectively

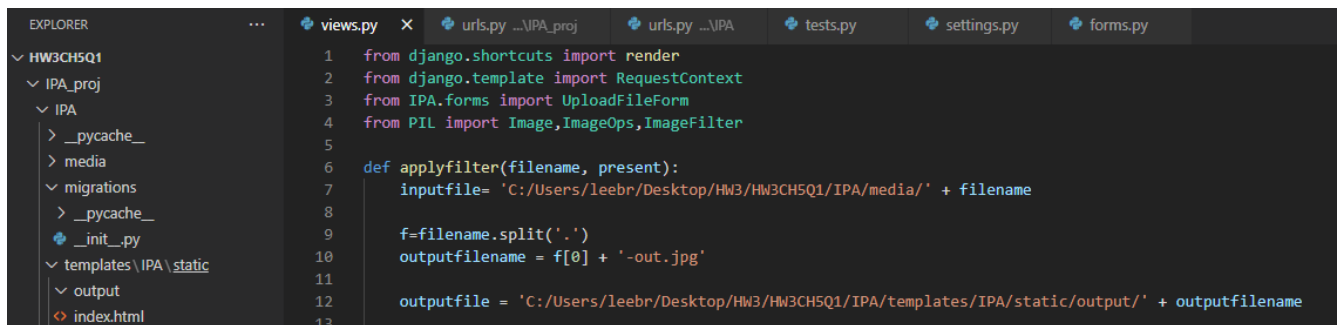
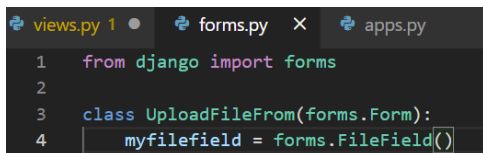


-Add application to project list of installed apps

From application apps.py, add configuration to project settings.py file



Add code given in text from Box 8.3 to Django forms.py & Box 8.4 to views.py



Note: render_to_render and context_instance are deprecated use render and RequestContext respectively in home function of views.py

```

52 def home(request):
53     if request.method == 'POST':
54         form = UploadFileForm(request.POST,request.FILES)
55         if form.is_valid():
56             present=request.POST['present']
57             outputfilename = handle_uploaded_file(request.FILES['myfilefield'],present)
58             return render(request, 'process.html',{'outputfilename': outputfilename})
59
60     else:
61         form = UploadFileForm()
62         return render(request, 'index.html',{'form': form})

```

Add TEMPLATES DIRS paths

```

54
55 TEMPLATES = [
56     {
57         'BACKEND': 'django.template.backends.django.DjangoTemplates',
58         'DIRS': [os.path.join(BASE_DIR, 'IPA', 'templates', 'IPA', 'static')],

```

Add static root

```

STATIC_ROOT = ('C:/Users/leebr/Desktop/HW3CH5Q1/IPA_proj/IPA/templates/IPA/static')

```

When you run in deploy mode, static files are served from this directory.

```

PS C:\Users\leebr\Desktop\HW3CH5Q1\IPA_proj> python manage.py collectstatic

```

```

You have requested to collect static files at the destination
location as specified in your settings:

```

```

    C:\Users\leebr\Desktop\HW3CH5Q1\IPA_proj\IPA\templates\IPA\static

```

```

This will overwrite existing files!

```

```

Are you sure you want to do this?

```

```

Type 'yes' to continue, or 'no' to cancel: yes

```

```

128 static files copied to 'C:\Users\leebr\Desktop\HW3CH5Q1\IPA_proj\IPA\templates\IPA\static'.

```

or

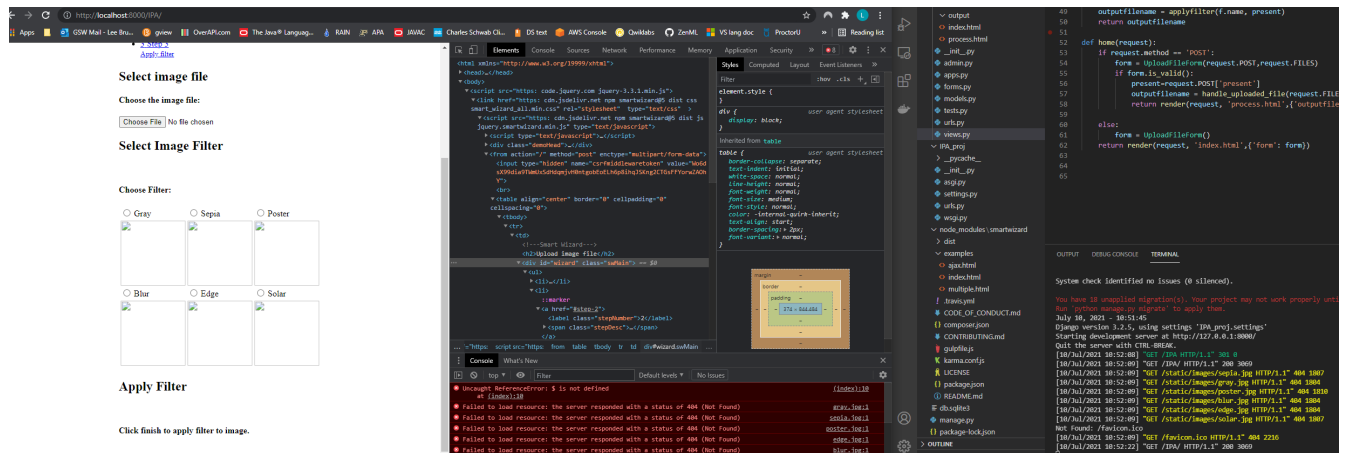
```

#STATICFILES_DIRS = [os.path.join(BASE_DIR, 'IPA', 'templates', 'IPA', 'static')]

```


Unfortunately, I have to stop here on account of time constraints and move on to begin the final project in order to meet the deadlines for the course (*and my other pending obligations*). I found this exercise one of the most interesting challenges in my GSW experience thus far, and wish I had more time to focus on it; however, I feel that I gave it an honest effort given the context. Definitely an area I would like to revisit in the future along with the AWS certs.

I still need to get the static images and CSS / JS bootstrapped as seen in the image 404 errors below.



Authentication steps after the smart_wizard issues are worked out

<https://docs.djangoproject.com/en/3.2/topics/auth/default/>

S3 and Cloud front-end steps (*this was accomplished in a more crude form in HW2*) for final steps.

<https://www.ordinarycoders.com/blog/article/serve-django-static-and-media-files-in-production>