

Model Deployment Tutorial 4 - ML Model in AWS EC2 Instance Cloud(IAAS)

Deployment in AWS EC2 with NGINX, Guinicorn, Supervisor

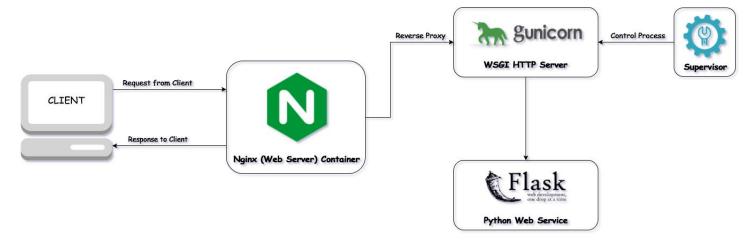
Architecture of Deployment

We will be using the following technologies:

→ Nginx: Reverse proxy, web server

→ Flask: Server backend→ Gunicorn: To run flask app

→ Supervisor: Monitor and control gunicorn process



Here is the architecture of this deployment, where client could be web browser or mobile device etc. NGINX as the web server and reverse proxy. This means that NGINX will sit between your Flask application and external clients and forward all client requests to your running Flask application. Gunicorn (Green Unicorn), is a Python web server gateway interface (WSGI) HTTP Server for UNIX. It will be used to forward requests from your NGINX web server to your Flask application and finally Supervisor is a client/server system that allows its users to monitor and control a number of processes on UNIX-like operating systems. Supervisor can handle auto-reloading Gunicorn if it crashes or if your server is rebooted unexpectedly

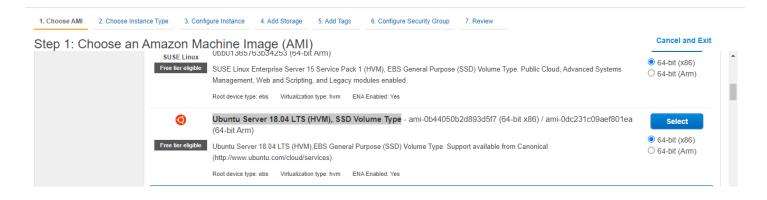
Starting up an EC2 instance

- Login to the AWS console here https://aws.amazon.com/console/
- → Select EC2 from AWS services.
- → Click on Launch Instance.
- → Choose Amazon Machine Image

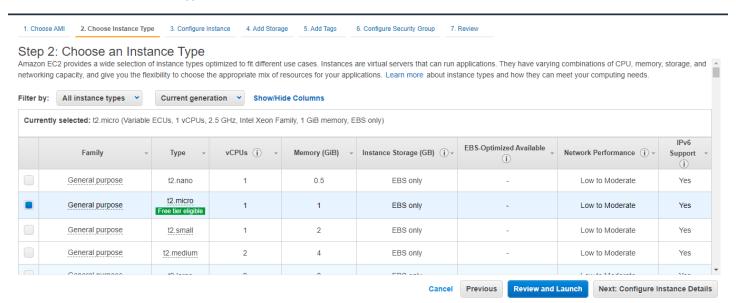
Select the Ubuntu Server 18.04 LTS (HVM), SSD Volume Type.

The exact versions may change with time.

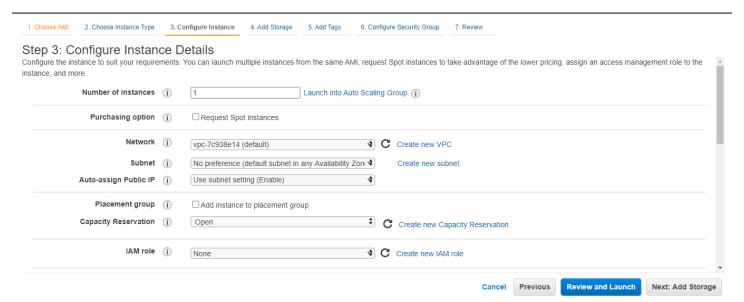




→ Choose the instance type

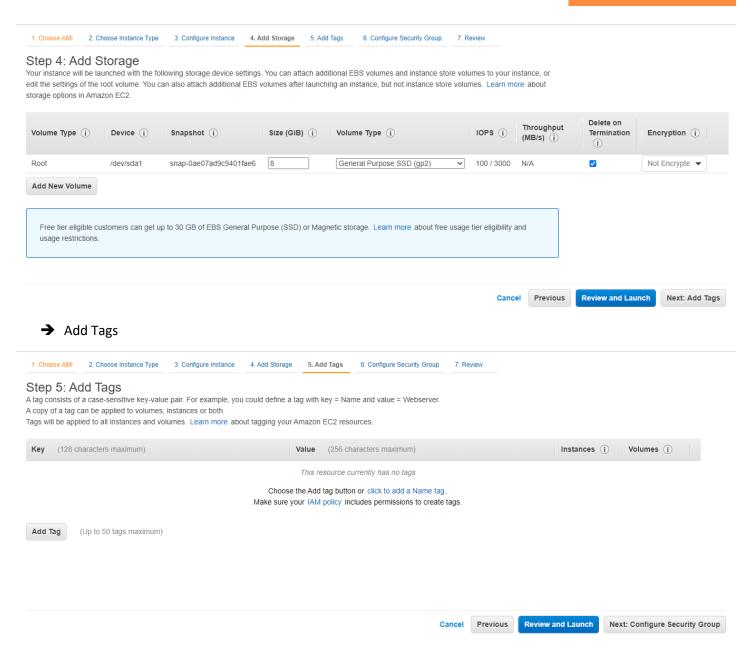


→ Configure the Instance



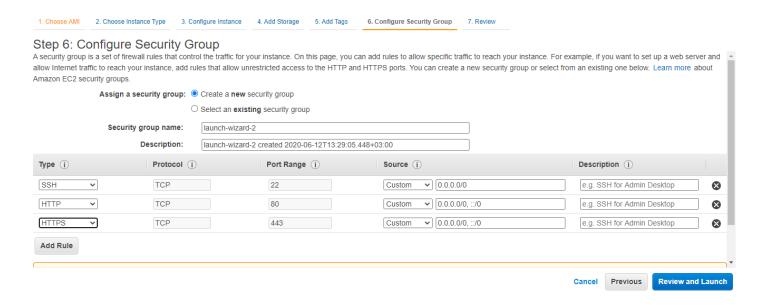
→ Add Storage



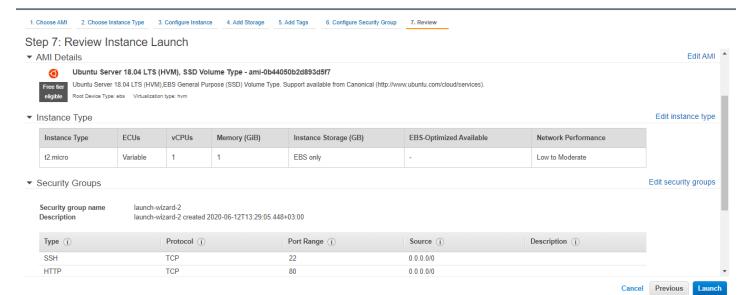


→ Configure the Security Group-> add HTTP and HTTPS





→ Review Instance Launch



→ Create new key pair

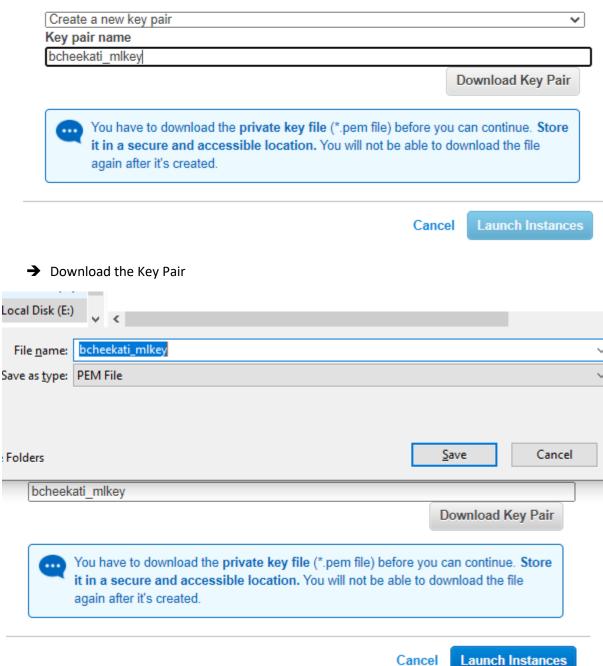


Select an existing key pair or create a new key pair

×

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about removing existing key pairs from a public AMI.



→ Finally click on Launch Instances button





→ Update the name of instance.



→ Note down the instance public ip, public DNS name and User

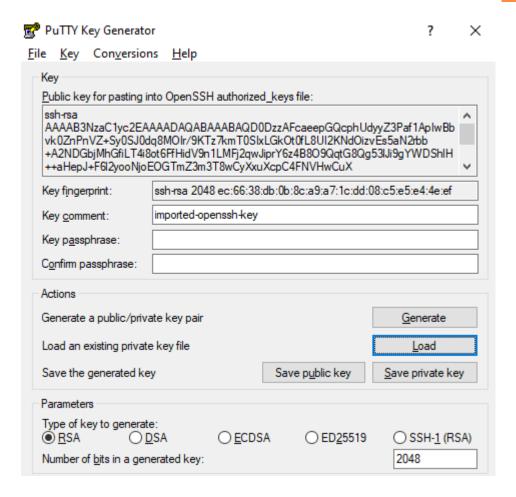
Public IP: 13.233.245.6

Public DNS name: ec2-13-233-245-6.ap-south-1.compute.amazonaws.com

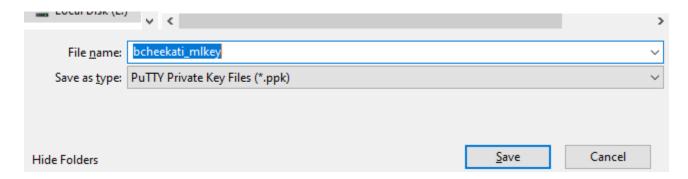
User: ubuntu

ssh -i "bcheekati_mlkey.pem" ubuntu@ec2-13-233-245-6.ap-south-1.compute.amazonaws.com

- → Download putty, puttygen and winscp
- → Load PEM private key file (bcheekati_mlkey.pem) using puttygen to generate private key in ppk format

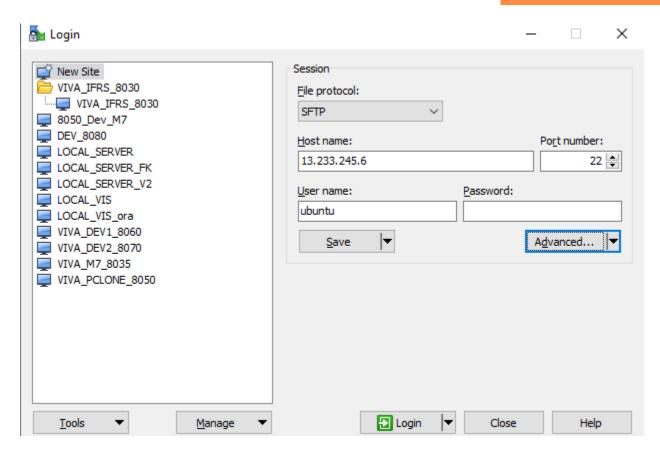


→ click on Save private key to save the private key file in ppk format. Private key will help to connect server from third party tools like putty and winscp without using password

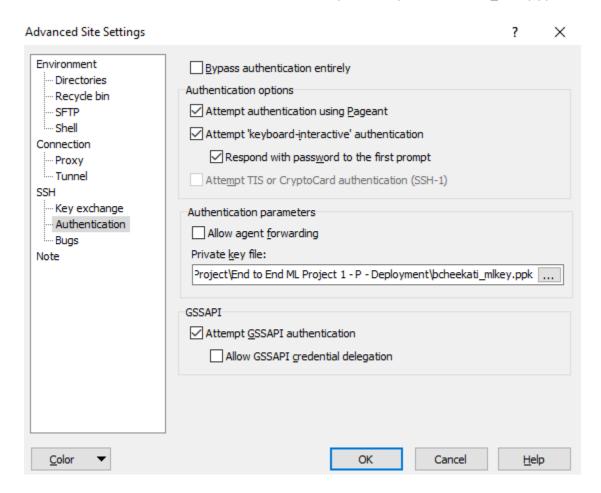


2. Connect WINSCP tool to upload the code files into server Enter host name and port and user name.



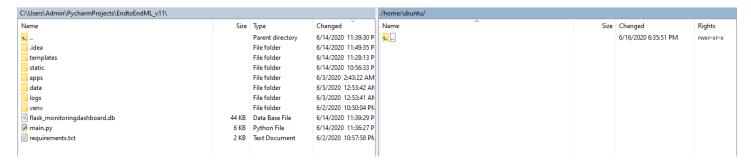


→ Click on Advanced -> SSH -> Authentication. and browse private key file (bcheekati_mlkey.ppk)

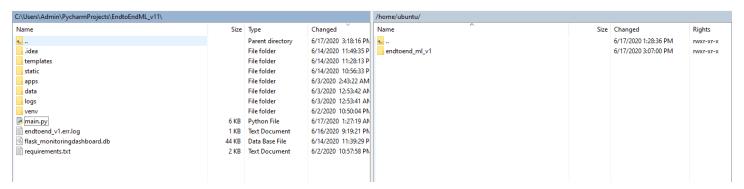




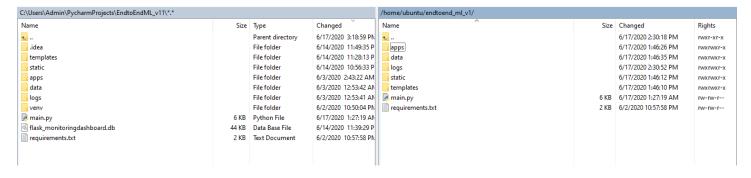
→ Finally click on Login Button to connect server



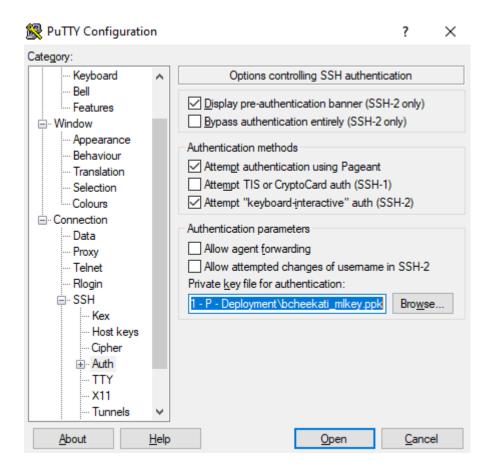
→ Create new folder in /home/ubunto/ directory to upload all code files



→ Upload required code files into server



- 3. Connect putty to install webserver called NGINX, Flask Web framework and all the required libraries
- → Upload private to connect server without password. Go to connection -> SSH -> Auth





- 4. Run below command to upgrade pip install utility in Ubuntu server before installing all required libraries
- → In your home directory, install Python 3: sudo apt install python3
- → Install pip3, the standard package manager for Python sudo apt-get update && sudo apt-get install python3-pip



5. Install nginx web server by running below command

sudo apt-get install nginx

```
ubuntu@ip-172-31-40-150:-$ pwd
/home/ubuntu
ubuntu@ip-172-31-40-150:-$ sudo apt-get install nginx
Reading package lists... Done
Building dependency tree
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
    fontconfig-config fonts-dejavu-core libfontconfigl libgd3 libjbig0 libjpeg-turbo8 libjpeg8 libnginx-mod-http-geoip libnginx-mod-http-image-filter
    libnginx-mod-http-xslt-filter libnginx-mod-mail libnginx-mod-stream libtiff5 libwebp6 libxpm4 nginx-common nginx-core

Suggested packages:
    libgd-tools fogiwrap nginx-doc ssl-cert
The following NEW packages will be installed:
    fontconfig-config fonts-dejavu-core libfontconfigl libgd3 libjbig0 libjpeg-turbo8 libjpeg8 libnginx-mod-http-geoip libnginx-mod-http-image-filter
    libnginx-mod-http-sslt-filter libnginx-mod-mail libnginx-mod-stream libtiff5 libwebp6 libxpm4 nginx nginx-common nginx-core

0 upgraded, 18 newly installed, 0 to remove and 64 not upgraded.

Need to get 2462 kB of archives.

After this operation, 8210 kB of additional disk space will be used.

Do you want to continue? [Y/n] Y
Get:1 http://ap-south-l.ec2.archive.ubuntu.com/ubuntu bionic-main amd64 fonts-dejavu-core all 2.37-1 [1041 kB]
Get:2 http://ap-south-l.ec2.archive.ubuntu.com/ubuntu bionic/main amd64 fonts-dejavu-core all 2.13.6-0ubuntu2 [55.8 kB]

Get:3 http://ap-south-l.ec2.archive.ubuntu.com/ubuntu bionic/main amd64 libfonts-dejavu-core all 2.13.6-0ubuntu2 [55.8 kB]

Get:3 http://ap-south-l.ec2.archive.ubuntu.com/ubuntu bionic/main amd64 libfonts-dejavu-core all 2.13.6-0ubuntu2 [55.8 kB]
```

Above will install nginx as well as run it.

Check status of nginx using sudo service nginx status
Here are the commands to start/stop/restart nginx sudo service nginx start sudo service nginx stop sudo service nginx restart

→ Check the website by hitting public ip or with public DNS. It displays default page of nginx sever http://13.233.245.6/



→ remove the default page by deleting the default file to redirect our custom index.html page

sudo rm /etc/nginx/sites-enabled/default

→ create a new config file in the sites-available folder and create a symbolic link to it in the sites-enabled folder.

sudo vim /etc/nginx/sites-available/endtoend_v1.conf

```
server {

listen 80;
```

```
server_name 13.233.245.6;
  root /home/ubuntu/endtoend_v1;
  access_log/home/ubuntu/endtoend_v1/logs/nginx_log/access.log;
  error_log/home/ubuntu/endtoend_v1/logs/nginx_log/error.log;
  location / {
    proxy_set_header X-Forward-For $proxy_add_x_forwarded_for;
    proxy set header Host $http host;
    proxy_redirect off;
    if (!-f $request_filename) {
      proxy_pass http://127.0.0.1:8000;
      break;
    }
  }
 location /static {
    alias /home/ubuntu/endtoend_v1/static/;
    autoindex on;
}
:wq to save and exit
```

- → We have to create the directory for our nginx logs mkdir -p ~endtoend_v1/logs/nginx_log
- → Create symlink for this file in /etc/nginx/sites-enabled by running this command,

sudo In -s /etc/nginx/sites-available/endtoend_v1.conf /etc/nginx/sites-enabled/endtoend_v1.conf

- → Restart nginx server sudo service nginx restart
- 6. Create virtual environment

sudo apt install virtualenv

→ create a virtual environment and activate

cd /home/ubuntu/endtoend_v1 mkdir /home/ubuntu/endtoend_v1/.virtualenvs && cd /home/ubuntu/endtoend_v1/.virtualenvs virtualenv -p python3 endtoend_v1_venv



```
ubuntu@ip-172-31-40-150:-$ od /home/ubuntu/endtoend_vl
ubuntu@ip-172-31-40-150:-/endtoend_vl$ mkdir /home/ubuntu/endtoend_vl/.virtualenvs
ubuntu@ip-172-31-40-150:-/endtoend_vl$ ls
_pycache_ apps data flask_monitoringdashboard.db logs main.py requirements.txt static templates
ubuntu@ip-172-31-40-150:-/endtoend_vl$ ls -aa
... virtualenvs _pycache_ apps data flask_monitoringdashboard.db logs main.py requirements.txt static templates
ubuntu@ip-172-31-40-150:-/endtoend_vl$ cd /home/ubuntu/endtoend_vl/.virtualenvs
ubuntu@ip-172-31-40-150:-/endtoend_vl$ cd /home/ubuntu/endtoend_vl/.virtualenvs
ubuntu@ip-172-31-40-150:-/endtoend_vl/.virtualenvs$ virtualenv -p python3 endtoend_vl_venv
Already using interpreter /usr/bin/python3
Using base prefix '/usr'
New python executable in /home/ubuntu/endtoend_vl/.virtualenvs/endtoend_vl_venv/bin/python3
Also creating executable in /home/ubuntu/endtoend_vl/.virtualenvs/endtoend_vl_venv/bin/python
Installing setuptools, pkg_resources, pip, wheel...done.
```

→ Activate the virtual env

source /home/ubuntu/endtoend_v1/.virtualenvs/endtoend_v1_venv/bin/activate

```
ubuntu@ip-172-31-40-150:~/endtoend_v1/.virtualenvs$ source /home/ubuntu/endtoend_v1/.virtualenvs/endtoend_v1_venv/bin/activate
(endtoend_v1_venv) ubuntu@ip-172-31-40-150:~/endtoend_v1/.virtualenvs$
```

7. Install dependencies using requirement.txt. run below command

pip3 install -r /home/ubuntu/endtoend v1/requirements.txt

```
(endtoend_vl_venv) ubuntu@ip-172-31-40-150:~/endtoend_vl/.virtualenvs$ pwd
/home/ubuntu/endtoend_vl/.virtualenvs
(endtoend_vl_venv) ubuntu@ip-172-31-40-150:~/endtoend_vl/.virtualenvs$ pip3 install -r /home/ubuntu/endtoend_vl/requirements.txt

Collecting Flask==1.1.
    Using cached Flask=1.1.1-py2.py3-none-any.whl (94 kB)
Collecting Flask-Cors==3.0.8
    Using cached Flask Cors=3.0.8-py2.py3-none-any.whl (14 kB)
Collecting matplotlib==3.1.2
    Using cached matplotlib=3.1.2-cp36-cp36m-manylinuxl_x86_64.whl (13.1 MB)
Collecting numpy==1.18.1
    Using cached matplotlib=3.1.2-cp36-cp36m-manylinuxl_x86_64.whl (20.1 MB)
Collecting pandas==0.25.3
    Using cached upandas=0.25.3-cp36-cp36m-manylinuxl_x86_64.whl (10.4 MB)
Collecting soikti-learn==0.22.1
    Using cached scikit_learn=0.22.1-cp36-cp36m-manylinuxl_x86_64.whl (7.0 MB)
Collecting kneed=0.5.1
    Using cached kneed=0.5.1-py2.py3-none-any.whl (9.9 kB)
Collecting xpdoost=1.0.2
    Using cached kneed=0.5.1-py2.py3-none-manylinuxl_x86_64.whl (109.7 MB)
```

8. Install Gunicorn. It act as python WSGI HTTP server for Unix

pip3 install gunicorn

```
(endtoend vl venv) ubuntu@ip-172-31-40-150:~/endtoend_vl/.virtualenvs$ pwd
/home/ubuntu/endtoend_vl/.virtualenvs
(endtoend_vl_venv) ubuntu@ip-172-31-40-150:~/endtoend_vl/.virtualenvs$ pip3 install gunicorn
Collecting gunicorn
Using cached gunicorn-20.0.4-py2.py3-none-any.whl (77 kB)
Requirement already satisfied: setuptools>=3.0 in ./endtoend_vl_venv/lib/python3.6/site-packages (from gunicorn)
Installing collected packages: gunicorn
Successfully installed gunicorn-20.0.4
(endtoend_vl_venv) ubuntu@ip-172-31-40-150:~/endtoend_vl/.virtualenvs$
```

→ Let's start a Gunicorn process to serve your Flask app.

cd /home/ubuntu/endtoend_v1 gunicorn main:app -w 3

```
(endtoend vl venv) ubuntu@ip-172-31-40-150:~/endtoend_vl/.virtualenvs% cd /home/ubuntu/endtoend_vl
(endtoend_vl_venv) ubuntu@ip-172-31-40-150:~/endtoend_vl% gunicorn main:app -w 3
[2020-06-16 18:17:21 +0000] [9274] [INFO] Starting gunicorn 20.0.4
[2020-06-16 18:17:21 +0000] [9274] [INFO] Listening at: http://l27.0.0.1:8000 (9274)
[2020-06-16 18:17:21 +0000] [9274] [INFO] Booting worker: sync
[2020-06-16 18:17:21 +0000] [9277] [INFO] Booting worker with pid: 9277
[2020-06-16 18:17:21 +0000] [9278] [INFO] Booting worker with pid: 9278
[2020-06-16 18:17:21 +0000] [9279] [INFO] Booting worker with pid: 9279
Scheduler started
Scheduler started
```



This will set your Gunicorn process off running in the background, which will work fine for your purposes here. An improvement that can made here is to run Gunicorn via Supervisor.

9. Install supervisor lib which Supervisor allows to monitor and control a number of processes on UNIX-like operating systems. Supervisor will look after the Gunicorn process and make sure that they are restarted if anything goes wrong, or to ensure the processes are started at boot time.

sudo apt install supervisor

```
[2020-06-16 17:31:27 +0000] [8466] [INFO] Shutting down: Master

(endtoend vl_venv) ubuntu@ip-172-31-40-150:~/endtoend_vl$ sudo apt install supervisor

Reading package lists... Done

Building dependency tree

Reading state information... Done

The following additional packages will be installed:
python-meld3

Suggested packages:
supervisor-doc

The following NEW packages will be installed:
python-meld3 supervisor

0 upgraded, 2 newly installed, 0 to remove and 0 not upgraded.

Need to get 287 kB of archives.

After this operation, 1580 kB of additional disk space will be used.

Do you want to continue? [Y/n] Y

Get: Intp://ap-south-1.ec2.archive.ubuntu.com/ubuntu bionic/universe amd64 python-meld3 all 1.0.2-2 [30.9 kB]
```

→ Create a supervisor file in /etc/supervisor/conf.d/ and configure it according to your requirements.

sudo vim /etc/supervisor/conf.d/endtoend_v1.conf

```
[program:endtoend_v1]
directory=/home/ubuntu/endtoend_v1/
command=/home/ubuntu/endtoend_v1/.virtualenvs/endtoend_v1_venv/bin/gunicorn main:app
autostart=true
autorestart=true
stopasgroup=true
killasgroup=true
stderr_logfile=/home/ubuntu/endtoend_v1/logs/supervisor_log/endtoend_v1.err.log
stdout_logfile=/home/ubuntu/endtoend_v1/logs/supervisor_log/endtoend_v1.out.log
```

→ Create the log directories and files listed in the endtoend_v1.conf file. Make sure to replace endtoend_v1 if it was modified in the Supervisor script above:

```
sudo mkdir /home/ubuntu/endtoend_v1/logs/supervisor_log
sudo touch /home/ubuntu/endtoend_v1/logs/supervisor_log/endtoend_v1.err.log
sudo touch /home/ubuntu/endtoend v1/logs/supervisor log/endtoend v1.out.log
```

→ To enable the configuration, run the following commands:

```
sudo supervisorctl reread
sudo supervisorctl update
sudo supervisorctl reload
```

additional command sudo service supervisor restart



sudo service supervisor stop

→ This should start a new process. To check the status of all monitored apps, use the following command:

sudo supervisorctl status

10. run below URL into browser to see our custom index.html page

http://13.233.245.6/

