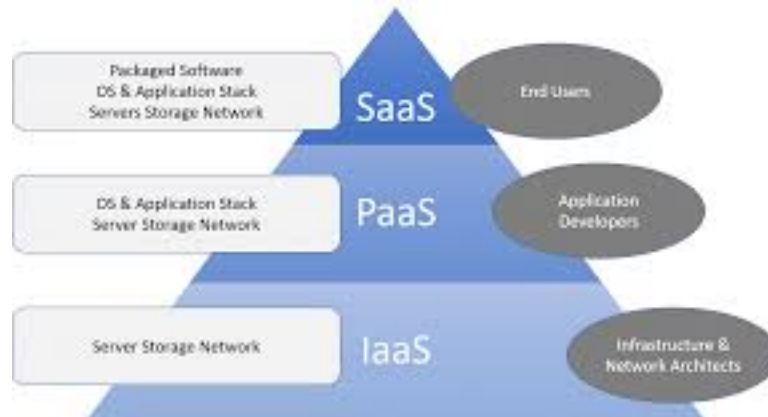




Cloud Service Models



CS352: Cloud Computing Jan-May 2019

SelfieLessActs App

A photo sharing web app being hosted on AWS Cloud.

CS352 | 30-04-2019

SNo	Name	USN	Class/Section
1	Rohan Sharma	01FB16ECS311	F
2	Royal Mehta	01FB16ECS315	F
3	Shashank Shekhar Pathak	01FB16ECS359	F
4	Shikhar Shrestha	01FB16ECS362	F

INTRODUCTION

SelfieLessActs app is a photo sharing web app being hosted on AWS Cloud. It is used to store information about anything that is good for the society that we observe.

Examples of such acts could be :

- Picking up a piece of garbage and dumping it in a garbage can
- Road getting laid in your area
- Someone helping a blind man cross the road.
- Helping your mother at home in the kitchen.

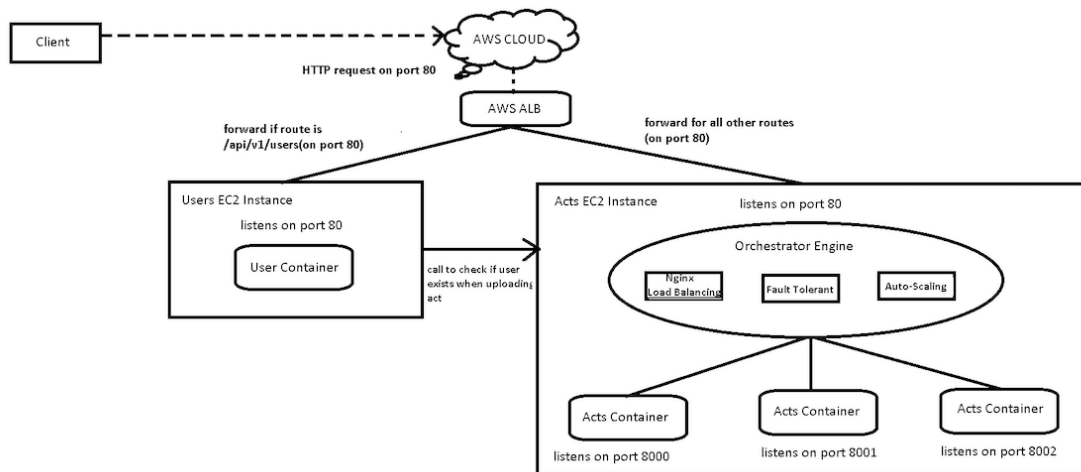
The SelfLessActs application allows the users of the application to upload image of the act with a small caption and categories. They can also view the acts which they have uploaded as well as on selecting a category, all the acts in that category can be viewed in reverse chronological order. Acts can also be liked by the users.

RELATED WORK

- **Flask** - To implement backend server for Acts/Users (<https://www.tutorialspoint.com/flask>)
- **SQLite3** - DataBase system to store information about Acts. (https://www.tutorialspoint.com/sqlite/sqlite_python.htm)
- **JavaScript** - To implement Client's Web App dynamically. (<https://www.w3schools.com/js/>)
- **XMLHttpRequest** - To make a request to server to get or post the data. (https://www.w3schools.com/xml/xml_http.asp)
- **Docker/Container** - To implement microservices. (<https://www.docker.com/resources/what-container>)
- **Nginx** - To implement load balancing among the containers. (<https://www.docker.com/resources/what-container>)

EXPERIMENTS/DESIGN

We have created two instances on AWS, one for users and one for acts. On each instance, we have created docker container on which acts and users api file runs. Both these instances are accessible from under the same public IP address and also the same port (80).



Img 1. Design for our Application

We have used AWS application load balancer for path based routing which will distribute incoming HTTP requests to one of the two EC2 instances based on the URL route of the request.

In the Acts instance, we have created an orchestration engine which performs load balancing, fault tolerance and auto-scaling. The orchestrator engine runs inside the Acts EC2 instance, listen on port 80 of the public IP, and load balance all HTTP incoming requests equally to every Acts container.

We have used Nginx server for load balancing which runs on port 80 of localhost. We have used round-robin algorithm in Nginx server for distributing the traffic among all the active acts container.

For fault tolerance and auto-scaling, we have written two separate python scripts. We are continuously monitoring the health of the active containers. Whenever we find an unhealthy container, we stop and remove that container and replace it with a new container which runs on the same port as the previous unhealthy container.

For auto-scaling, we use a threshold value to determine when to scale-up and when to scale-down the containers. For this we check the log file of the nginx server which has details of all the requests made. We have used a python timer library to create the timer for 2 minutes. Whenever the first request is made, the timer is started. At every 2 minute interval, we check the no. of requests made and the no. of containers which are active and the orchestrator must increase or decrease the number of Acts containers according to the following rules:

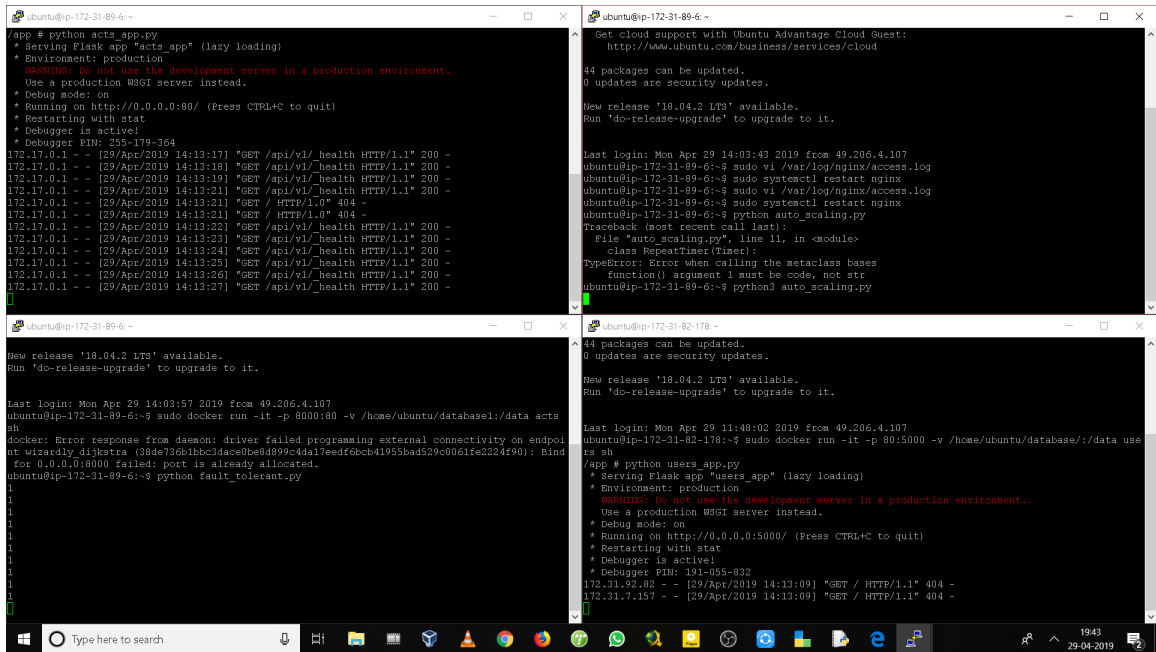
- If the number of requests is less than 20, then only 1 Acts container must be running.
- If the number is ≥ 20 and < 40 , then 2 Acts containers must be running.
- If the number is ≥ 40 and < 60 , then 3 Acts containers must be running.
- and so on...

Using all this information, we update the `/etc/nginx/conf.d/default.conf` file with the port numbers for the new acts containers, whenever we increase the containers. And when we want to decrease the number of acts containers, we remove the required lines from the configuration file.

In this way, we have implemented load balancing, fault tolerance and auto-scaling

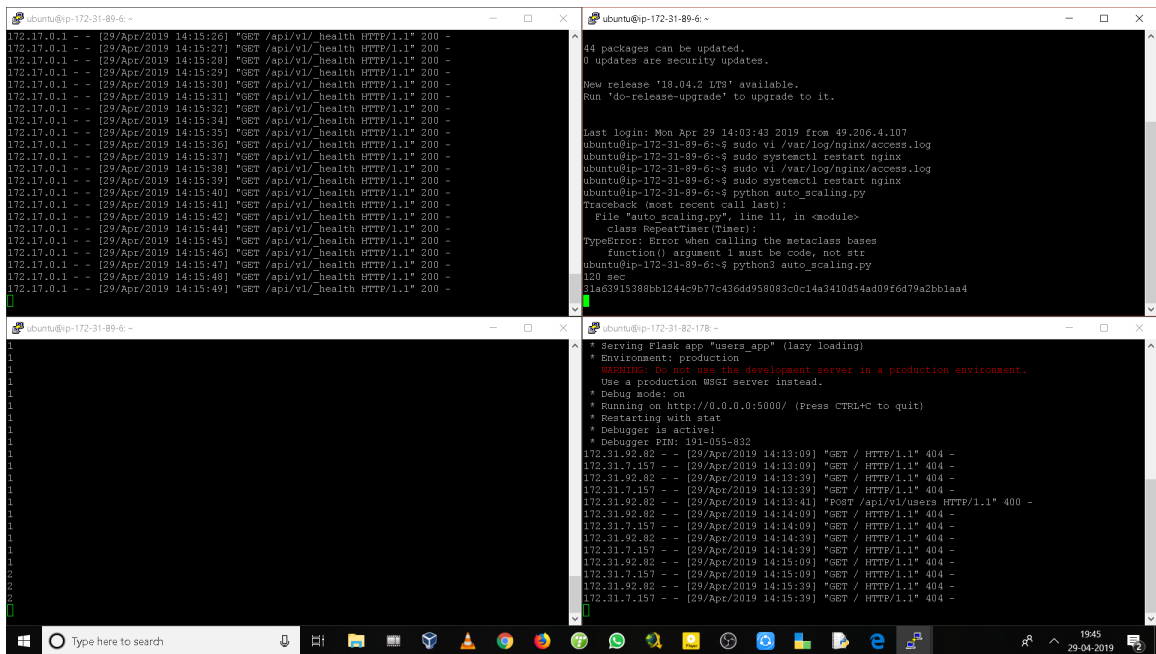
TESTING/RESULTS

We have tested our code for load balancing, fault tolerance and auto-scaling.



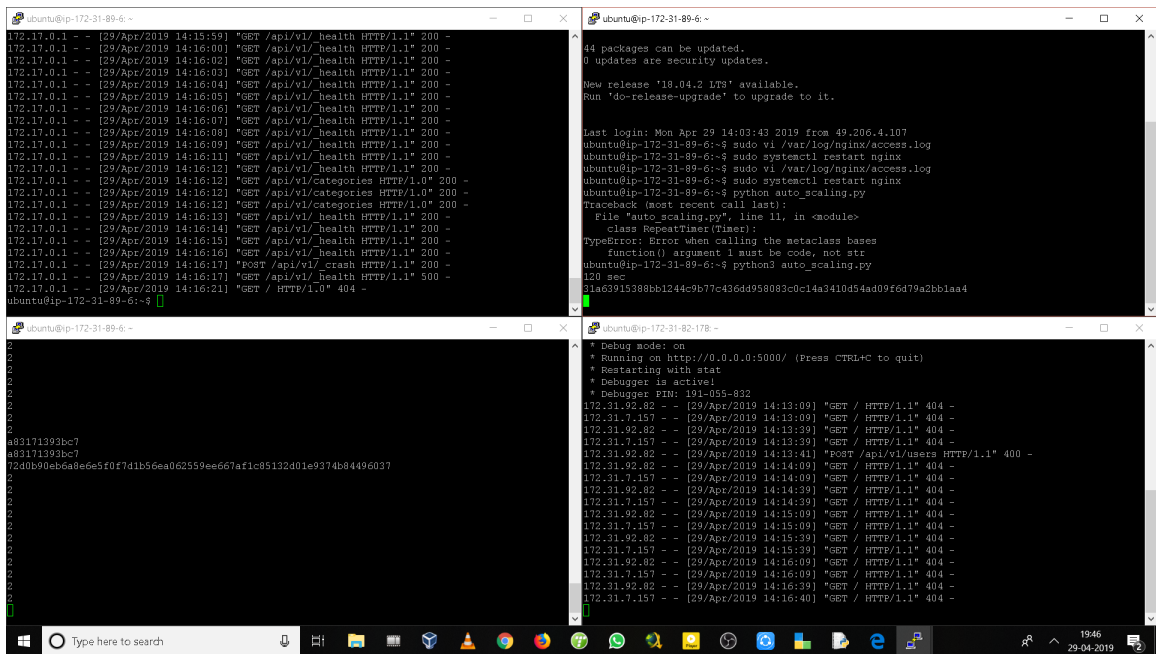
Img 2.

In the beginning, we had only one Acts container which was active.



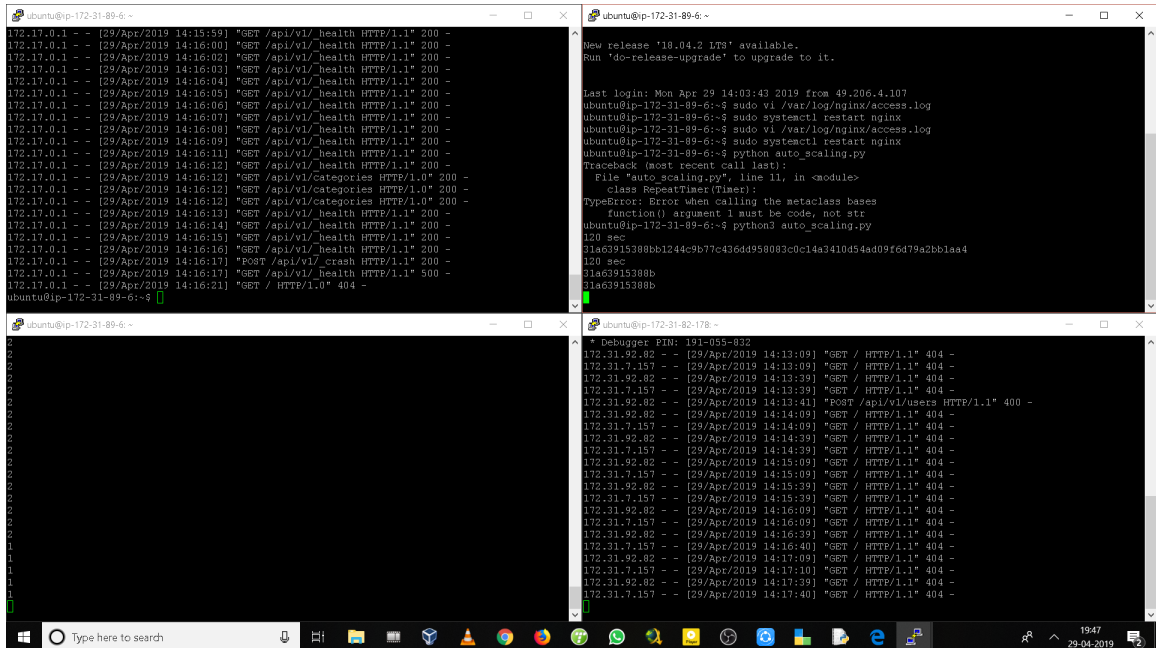
Img 3.

After that 20 requests were made, as a result of which the containers were scaled up and now we have 2 active containers as is visible from Img 3, between which load balancing takes place.



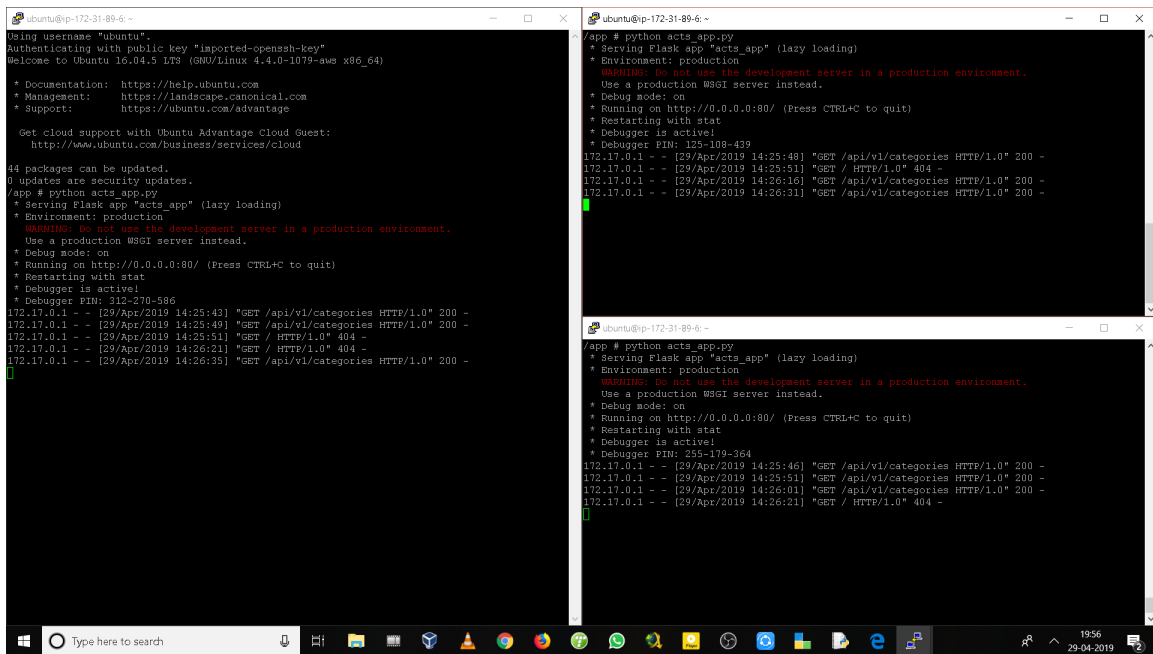
Img 4.

After this, the crash api request was sent as a result of which 2nd acts container running on port 8001 was crashed and it was detected by the fault tolerant script which stops and removes this container and in place of it starts a new container running on the same port and we can see the container id of with these containers from the Img 4.



Img 5.

Now since only 6 requests were made in this next two minutes, so the containers are scaled-down and the 2nd container is removed. As we can see from the Img 5. there is only one active container now.



Img 6.

When we send 9 requests, the requests are evenly distributed among the 3 active containers in round robin manner.

REFERENCES

- <https://www.tutorialspoint.com/flask>
- https://www.tutorialspoint.com/sqlite/sqlite_python.htm
- <https://www.w3schools.com/js/>
- https://www.w3schools.com/xml/xml_http.asp
- <https://www.docker.com/resources/what-container>
- <https://www.docker.com/resources/what-container>

EVALUATIONS (Leave this for the faculty)

Date	Evaluator	Comments	Score

CHECKLIST

SNo	Item	Status
1	Source code documented	
2	Source code uploaded to CCBD server	
3	Source code in GitLab. Please do not upload your source code to github where it can be seen by everyone.	