Setup overview:

A screenshot of a computer

Description automatically generated with low confidence

In the load balancing measurement setup, we aim to simulate the situation of multiple concurrent users accessing the same network resource as a webpage to the exact location. Then we will install a load balance strategy such as HAproxy, EC2 load balancer, and Nginx onto the frontend server. Based on the technologies and methods that we choose; the frontend server will forward the request to one of the agent servers to distribute the user load. We expect that using this can achieve a higher performance than a single node server. In addition, we will compare the technologies we mentioned above and decide which is the best and justify the reason.

1. **Install a web server to accept user request in backend**

We can build a simple website with a web page that accepts user requests and returns the result using Spring.

Graphical user interface

Description automatically generated with medium confidence

Then we build the web container file

Graphical user interface, text, application

Description automatically generated

Create a t2.micro Instant on Amazon cloud, then upload the web container file just build to the t2.micro Instant directory.

Graphical user interface, text

Description automatically generated

To start up the server on one of the backend servers, we will require to install java JDK.

Type "sudo apt-get update" first if first time running the EC2 instant.

Then Type “sudo apt install openjdk-11-jdk”

Text

Description automatically generated

After that, we can start the web server by typing “sudo java -jar loadbalancetestweb-0.0.1-SNAPSHOT.war”

Text

Description automatically generated

Now the server should be up and running. We can access the testing website with the ec2 instant IP with port 80

\*Remember to update the inbound rules for accepting internet connection from outside

A screenshot of a computer

Description automatically generated

Perform this step to set up more instants of serval more time in each backend server will be used in load balancing.

1. **Setup frontend server to accept requests from the user and perform load balance strategy**

In this step, we will be required to set up the environment for running the load balancing solution. Each of them needs to configure differently. Therefore, the following step will be separated into three subsections.

2.1 HAproxy load balancer setup procedure.

Create a t2.micro Instant on Amazon cloud, but this time we will use it to install different types of load balancing strategies.

Graphical user interface, text, application, email

Description automatically generated

Remote into the EC2 instant and type "sudo apt-get update" if first startup

Then Type “sudo apt install haproxy” to install haproxy.

Text

Description automatically generated

Haproxy should run automatically, to check the service status, type "sudo systemctl status haproxy"

Text

Description automatically generated

After that, we will need to configure our loadbalancer. Location the directory location /etc/haproxy/ and find haproxy.cfg

Graphical user interface, application

Description automatically generated

Add the following line to setup the load balance strategy and backend server load we just set in step 1.

一張含有 文字 的圖片

自動產生的描述

Load balancing strategy

The EC2 backend instances running the web application

Example in text

frontend haproxy\_inbound

bind :80

default\_backend haproxy\_httpd

backend haproxy\_httpd

balance roundrobin

server webapp1 172.31.86.115:80/loadtest1 check

server webapp2 172.31.93.219:80/loadtest1 check

To make the new setting take effect, we will need to restart the HAproxy service by typeing "sudo systemctl stop haproxy" then "sudo systemctl start haproxy"

Text

Description automatically generated

Then the load balancer should be working correctly. Access the load balancer public IP address to test the result.

Graphical user interface, text, application, email

Description automatically generated

You can check the result of the webpage

A picture containing text, cellphone, phone, displayed

Description automatically generated

Since it is round robin, you can refresh the webpage and see each backend server taking turns serving the page.

Graphical user interface, application, website

Description automatically generated

2.2 Nginx load balancer setup procedure.

Again, we create a t2.micro instant on Amazon cloud to install our load balancer with Nginx.

Graphical user interface, text, application

Description automatically generated

Type "sudo apt-get update" if first startup

Then input “sudo apt install nginx” to install nginx.

Text

Description automatically generated

Check service running status with command “sudo systemctl status nginx”

A screenshot of a computer

Description automatically generated with medium confidence

Locate the directory location /etc/nginx/conf.d/

Graphical user interface, application

Description automatically generated

Upload the following configuration file with load balancing details.



Text

Description automatically generated

Restart the Nginx server to let the new setting take effect.

Type "sudo nginx -s reload" or "sudo systemctl stop nginx" follow with "sudo systemctl start nginx"

Graphical user interface, text, application

Description automatically generated

You can refresh the webpage and check if it is success.

2.3 AWS EC2 load balancer setup procedure.

Unlike Nginx and HAProxy, EC2 load balance in a build-in service in Amazon AWS. This means it does not require a dedicated server to hold the balancing solution.

Graphical user interface, application

Description automatically generated

We can locate the load balancing service by typing "EC2 Load Balancer" on AWS website.

Graphical user interface, text, application, chat or text message

Description automatically generated

Click Create Load Balancer

Diagram

Description automatically generated

Choose the type of load balance we want to create, this time we use application load balancer

The setup should be straightforward. You just need to type the name and select the parameter it requires.

Graphical user interface, text, application, email

Description automatically generated

One thing we need to watch out for is to add the target group, which assigns the instant as backend servers for load balancing.

Text

Description automatically generated with medium confidence

Assign EC2 instants into the target group. Then create the load balancer

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Copy the load balancer address and access the site.

A picture containing text, monitor

Description automatically generated

As you can see, it directed to our ec2 instant from the target group we assigned. This means the load balance was successfully created.

1. **Performance Load test with Jmeter**