AutoScaling

LoadBalancer **ASSIGNMENT**



Name – Maithely Sharma
College – University of Petroleum and Energy Studies

EmployeeID -4057

1.Differences between ELB, ALB, and NLB. Where will you use which one?

- Elastic Load Balancing Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and Lambda functions. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones. Elastic Load Balancing offers three types of load balancers that all feature the high availability, automatic scaling, and robust security necessary to make your applications fault tolerant.
- Application Load Balancer Application Load Balancer is best suited for load balancing of HTTP and HTTPS traffic and provides advanced request routing targeted at the delivery of modern application architectures, including microservices and containers. Operating at the individual request level (Layer 7), Application Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) based on the content of the request.
- Network Load Balancer Network Load Balancer is best suited for load balancing of Transmission Control Protocol (TCP), User Datagram Protocol (UDP) and Transport Layer Security (TLS) traffic where extreme performance is required. Operating at the connection level (Layer 4), Network Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) and is capable of handling millions of requests per second while maintaining ultra-low latencies. Network Load Balancer is also optimized to handle sudden and volatile traffic patterns.

2.Differences between step scaling and target scaling.

With **step** scaling and **simple** scaling, you choose scaling metrics and threshold values for the CloudWatch alarms that trigger the scaling process. You also define how your Auto Scaling group should be scaled when a threshold is in breach for a specified number of evaluation periods.

AWS recommends that you use a **target tracking scaling policy** to scale on a metric like **average CPU utilization** or the **RequestCountPerTarget** metric from the Application Load Balancer. Metrics that decrease when capacity increases and increase when capacity decreases can be used to proportionally scale out or in the number of instances using target tracking. This helps ensure that Amazon EC2 Auto Scaling follows the demand curve for your applications closely.

With target tracking scaling policies, you select a scaling metric and set a target value. Amazon EC2 Auto Scaling creates and manages the CloudWatch alarms that trigger the scaling policy and calculates the scaling adjustment based on the metric and the target value. The scaling policy adds or removes capacity as required to keep the metric at, or close to, the specified target value. In addition to keeping the metric close to the target value, a target tracking scaling policy also adjusts to the changes in the metric due to a changing load pattern.

3. Differences between Launch configuration and launch template.

Launch template is similar to launch configuration which usually Auto Scaling group uses to launch EC2 instances. However, defining a launch template instead of a launch configuration allows you to have multiple versions of a template.

AWS recommend that we should use launch templates instead of launch configurations to ensure that we can leverage the latest features of Amazon EC2, such as T2 Unlimited instances.

With launch templates, you can also provision capacity across multiple instance types using both On-Demand Instances and Spot Instances to achieve the desired scale, performance, and cost.

4.Differences between EC2 healthcheck and load balancer health check

Once your ELB has been associated with the auto scaling group, you can then instruct it to be used for the Health Check Type:



EC2 health check watches for instance availability from hypervisor and networking point of view. For example, in case of a hardware problem, the check will fail

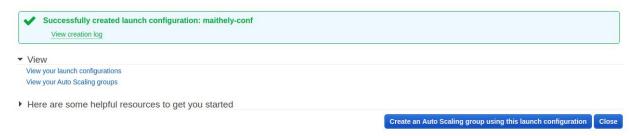
ELB health check verifies that a specified TCP port on an instance is accepting connections OR a specified web page returns 2xx code. Thus ELB health checks are a little bit smarter and verify that actual app works instead of verifying that just an instance works.

5. Create 2 auto-scaling groups with

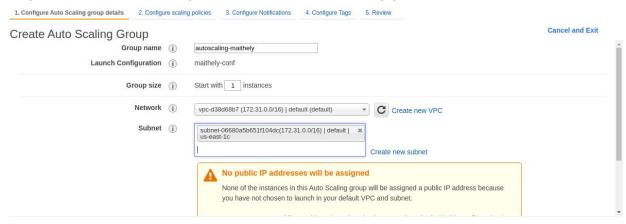
launch configuration and launch template

Create a launch configuratiom

Launch configuration creation status



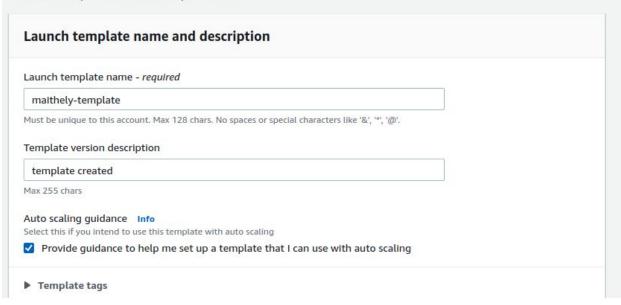
Using the above launch configuration, create an autoscaling group

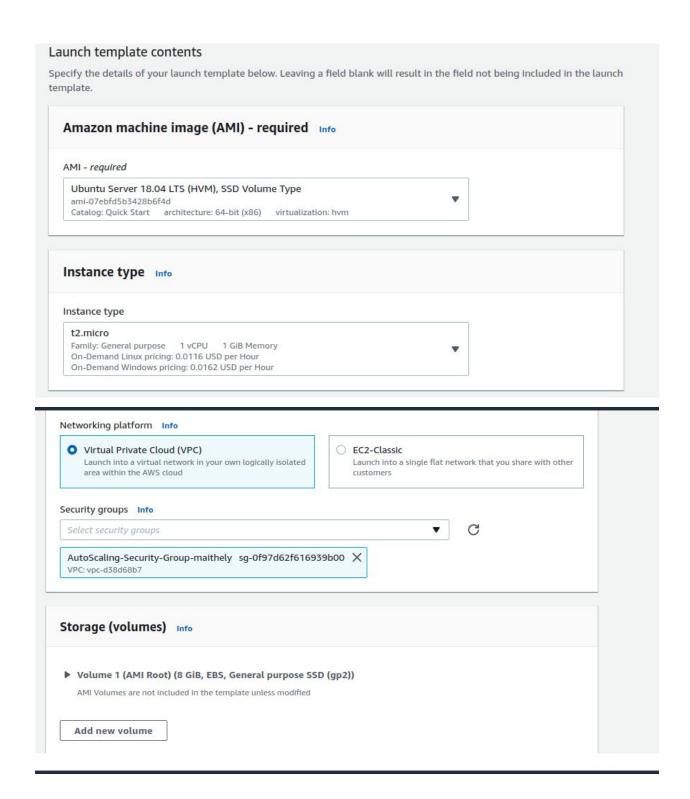


Now create a launch template

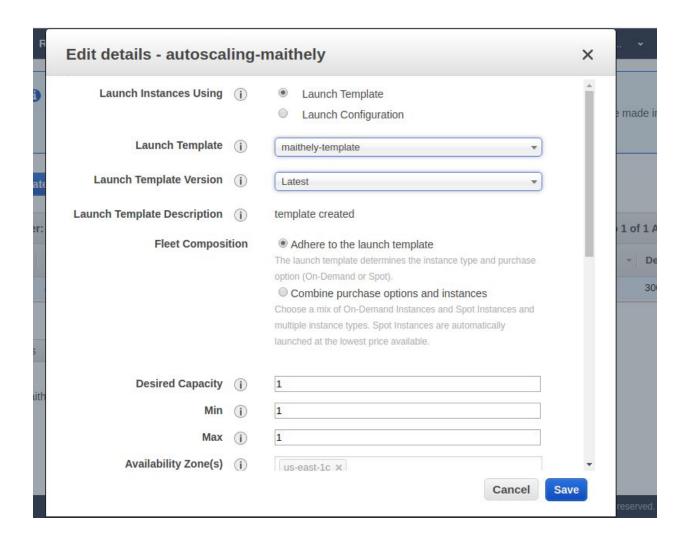
Create launch template

Creating a launch template allows you to create a saved instance configuration that can be reused, shared and launched at a later time. Templates can have multiple versions.



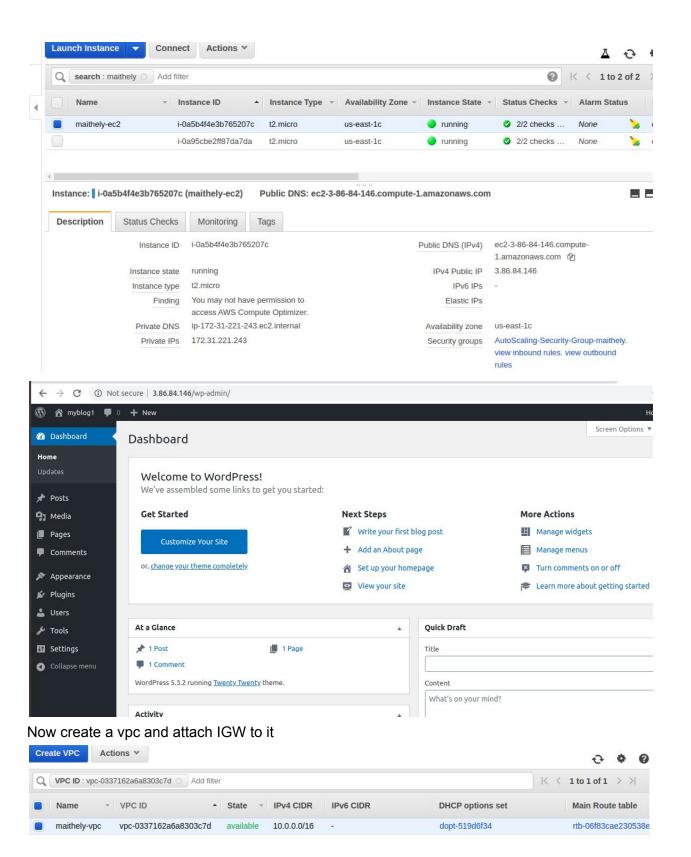


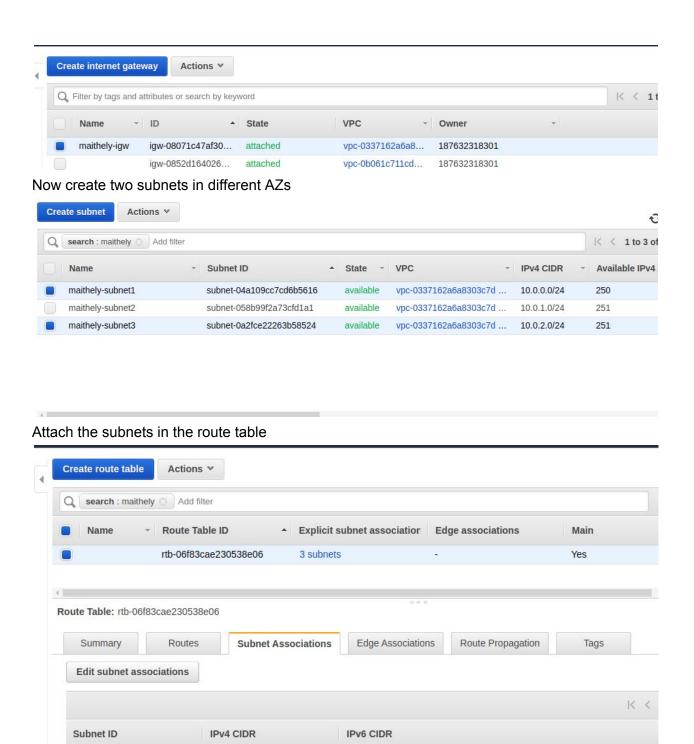
After creating the launch template, now attach it to the auto scaling group



6.Setup autoscaling Wordpress application with the Application load balancer. Auto-scaling should be triggered based on CPU usage of EC2 instances.

Create an instance and install install wordpress with nginx and mysql: https://www.journaldev.com/25670/install-wordpress-nginx-ubuntu





Now create an AMI of the instance created

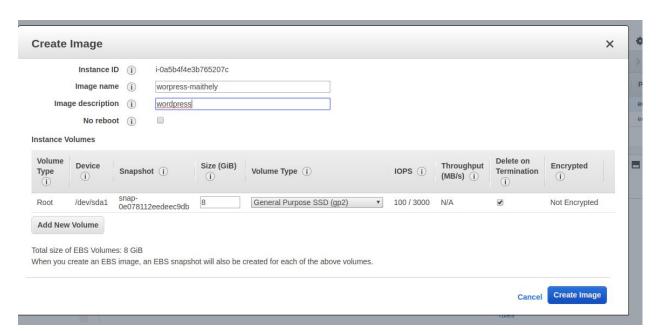
subnet-058b99f2a73cfd1a... 10.0.1.0/24

10.0.0.0/24

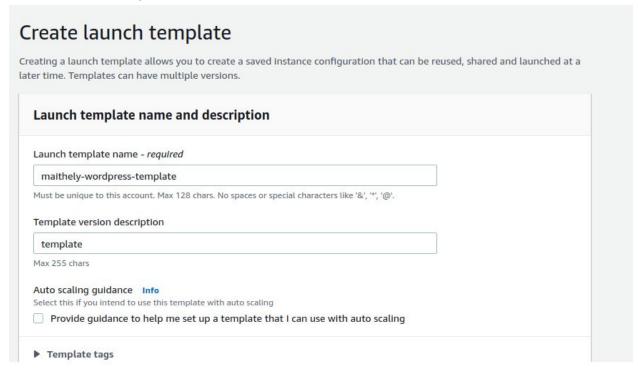
10.0.2.0/24

subnet-04a109cc7cd6b56...

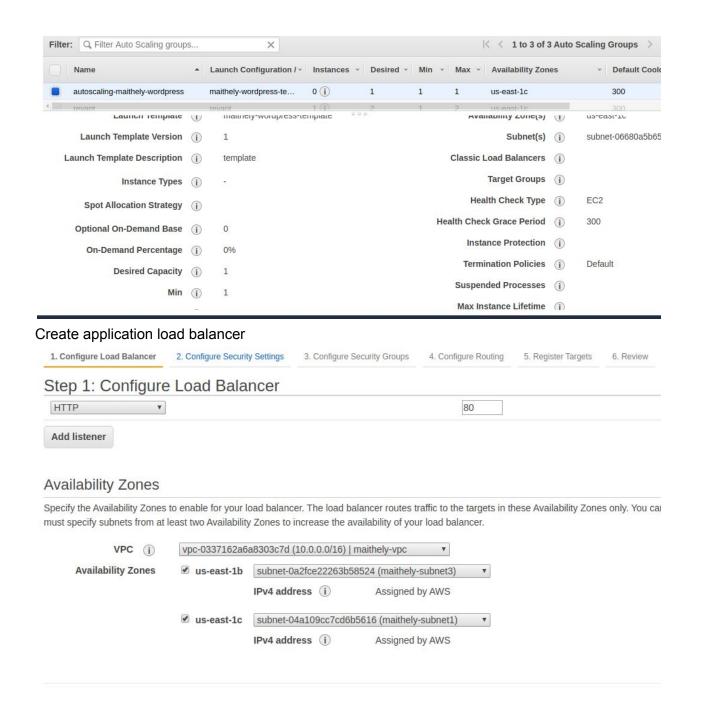
subnet-0a2fce22263b585...



Now create an configuration template and attach AMI to it

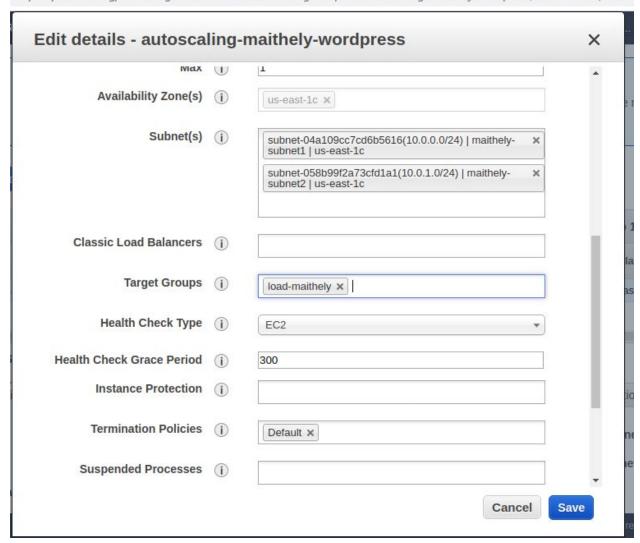


Now create auto scaling group

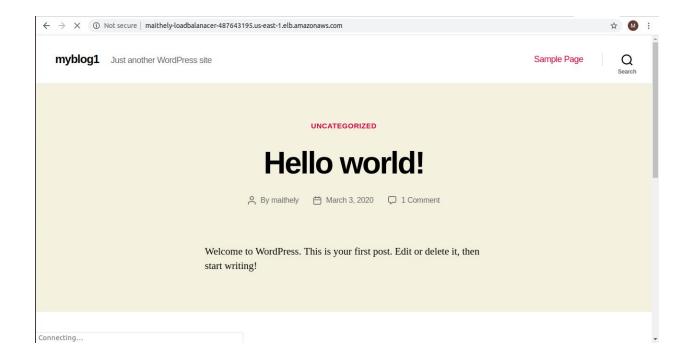


Now you have to edit the ASG and add target group

:om/ec2/autoscaling/home?region=us-east-1#AutoScalingGroups:id=autoscaling-maithely-wordpress;view=details;filter=



Browse the DNS of load balancer



7.Create another Wordpress website and use the ALB created above to send traffic to this website based on the hostname

Create 2 instances



Now ssh in both the instances and install nginx

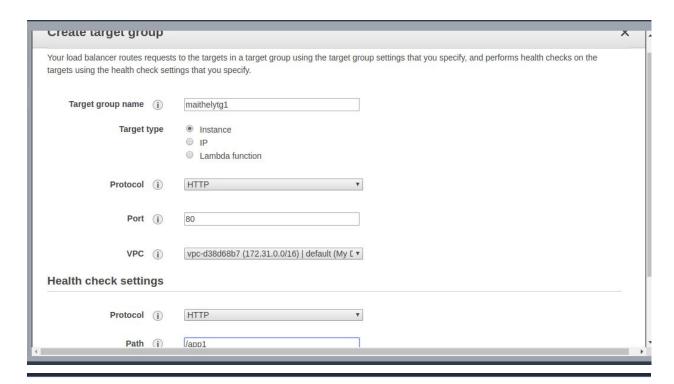
```
maithely@maithely:~/Downloads$ ssh -i "maithely.pem" ubuntu@54.205.14.30
The authenticity of host 54.205.14.30 54.205.14.30 can't be established.
ECDSA key fingerprint is SHA256:MLHRuMavfUM3/ZAlaoMRfV82IaFmUIX400toxQkKoU0.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '54.205.14.30' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 4.15.0-1057-aws x86_64)
 * Documentation: https://help.ubuntu.com
 * Management: https://landscape.canonical.com
 * Support:
                  https://ubuntu.com/advantage
  System information as of Wed Mar 4 10:04:55 UTC 2020
                                                      89
  System load: 0.04
                                 Processes:
  Usage of /: 13.6% of 7.69GB Users logged in:
                                                      0
  Memory usage: 15%
                                 IP address for eth0: 10.0.0.35
  Swap usage: 0%
0 packages can be updated.
O updates are security updates.
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
```

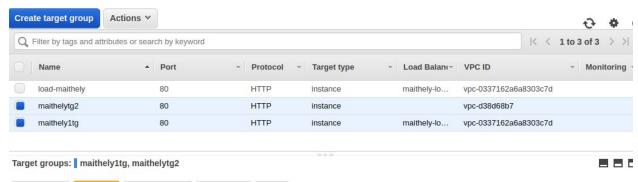
```
ubuntu@ip-10-0-0-35:~$ sudo apt install nginx
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  fontconfig-config fonts-dejavu-core libfontconfig1 libg
  libnginx-mod-http-image-filter libnginx-mod-http-xslt-1
  nginx-core
Suggested packages:
  libgd-tools fcgiwrap nginx-doc ssl-cert
```

And also create a directory app1 and a file index.html

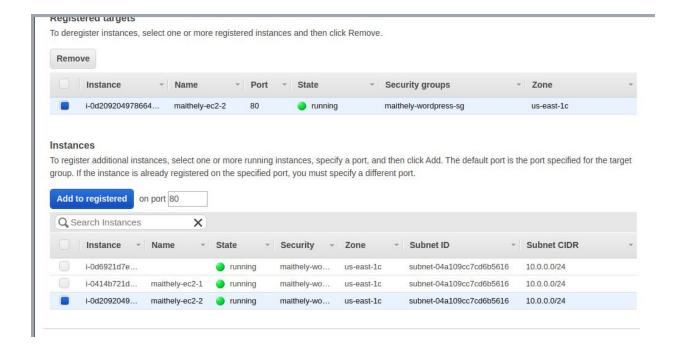
```
ubuntu@ip-10-0-0-35:~$ cd /var/www/html/app1/
ubuntu@ip-10-0-0-35:/var/www/html/app1$ sudo vim index.html
ubuntu@ip-10-0-0-35:/var/www/html/app1$ ls
index.html
```

Similarly do the same for the second instance as well. Create 2 target groups

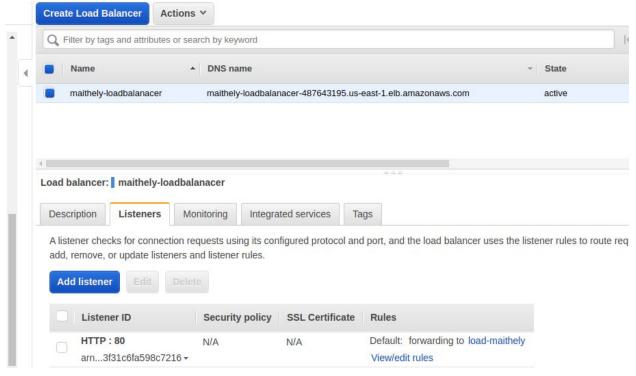




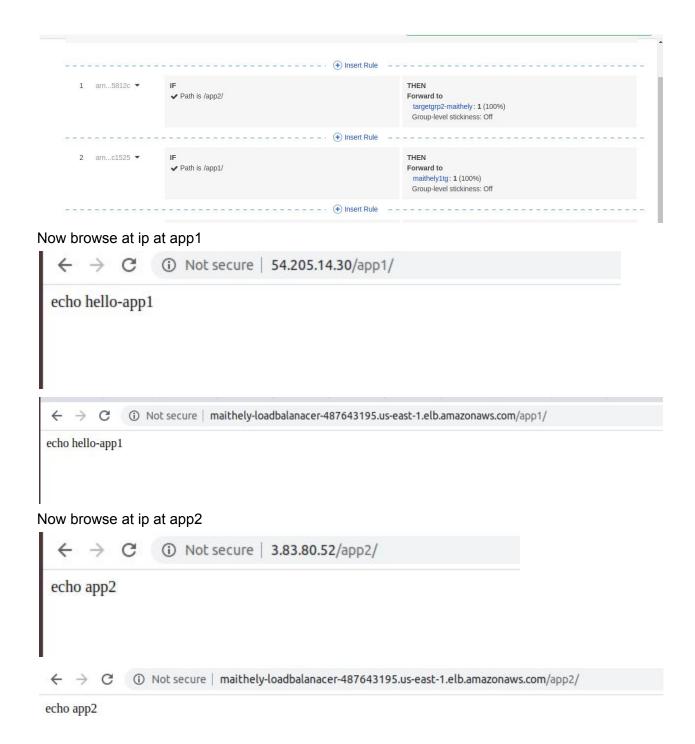
After creating the target groups edit to add instance



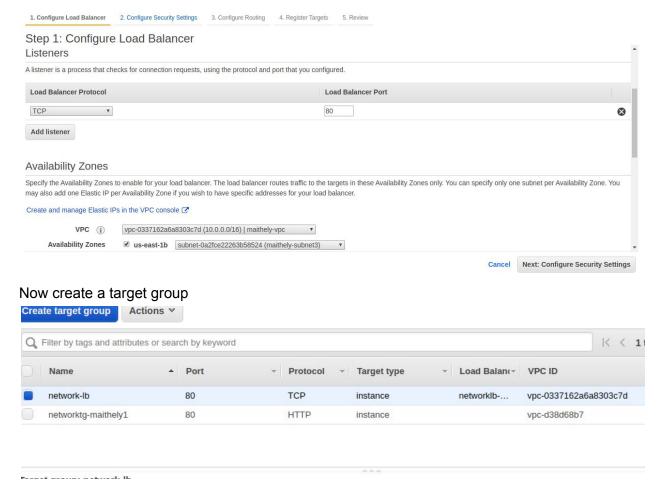
In the load balancer edit the rules in the listener



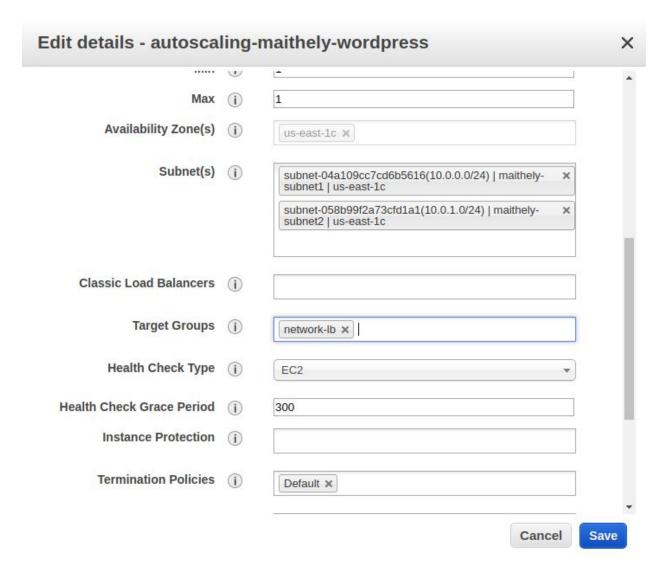
Where you add the actions

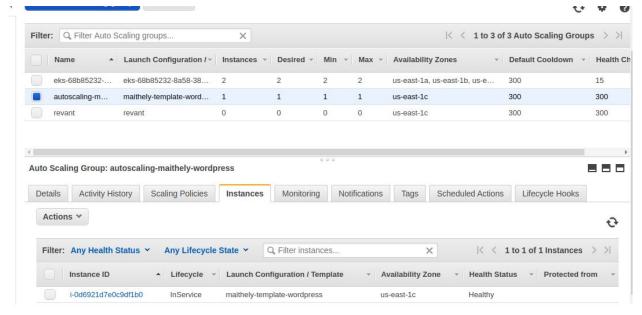


8.Use NLB that replaces the ALB in the above setup. Create a network load balancer



Now tadd target group to the asg





If we are hitting at the ip of the instance we get the app1

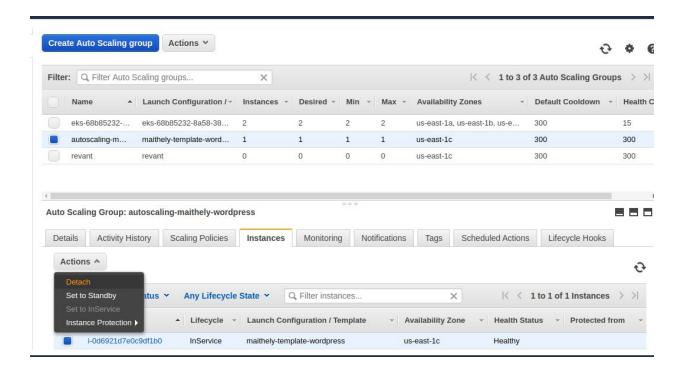


But not with DNS because path based at NLB cannot happen at layer 4 -transport layer tcp/ip

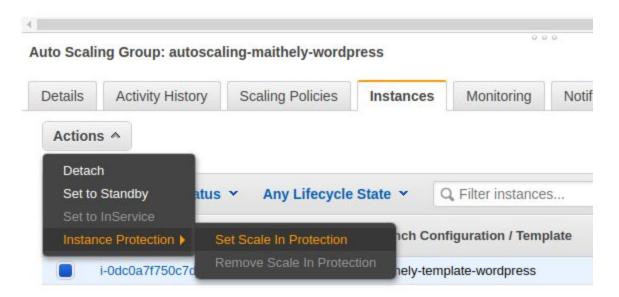


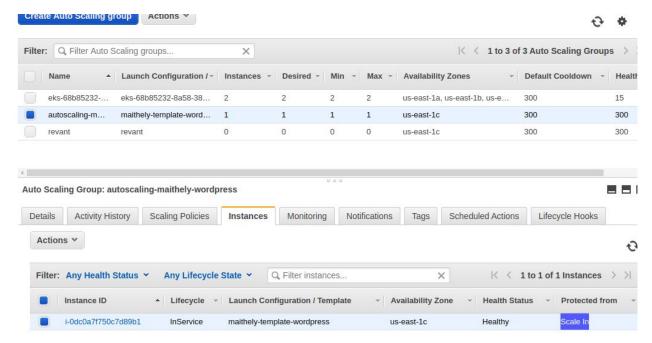
nginx/1.14.0 (Ubuntu)

9. Take an instance out of the ASG.



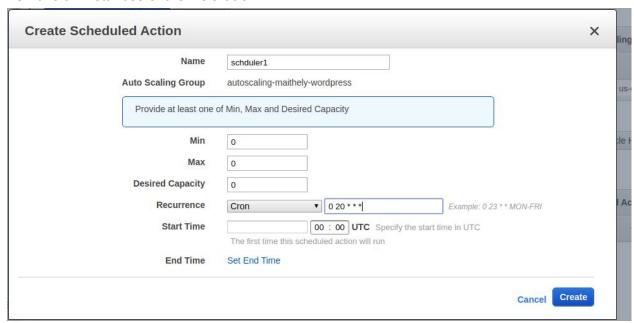
10. Put scale-in protection on an instance in the ASG.





11.Put Schedules in ASG to:

Remove all instances of the ASG at 8 PM



Launch a minimum of 2 instances at 10 AM

