Andhra Pradesh State Skill Development Corporation

Skill AP
APSSDC

eSkill AP
Learn Anytime Anywhere
Andhra Pradesh State Skill Development Corporation

AWS CLOUD COMPUTING

CONFIGURE AMAZON AUTO SCALING WITH LOAD BALANCER AND SNAPSHOT

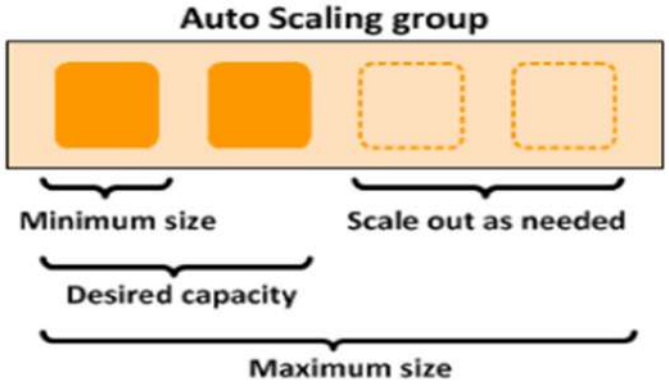
Configure Amazon Auto Scaling with load balancer and snapshot

Amazon EC2 Auto Scaling

To configure Amazon Auto Scaling with load balancer and snapshot

Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application. You create collections of EC2 instances, called *Auto Scaling groups*. You can specify the minimum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes below this size. You can specify the maximum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes above this size. If you specify the desired capacity, either when you create the group or at any time thereafter, Amazon EC2 Auto Scaling ensures that your group has this many instances. If you specify scaling policies, then Amazon EC2 Auto Scaling can launch or terminate instances as demand on your application increases or decreases.

For example, the following Auto Scaling group has a minimum size of 1 instance, a desired capacity of 2 instances, and a maximum size of 4 instances. The scaling policies that you define adjust the number of instances, within your minimum and maximum number of instances, based on the criteria that you specify.



Auto Scaling Component

The following table describes the key components of Amazon EC2 Auto Scaling.

Groups

Your EC2 instances are organized into *groups* so that they can be treated as a logical unit for the purposes of scaling and management. When you create a group, you can specify its minimum, maximum, and desired number of EC2 instances.

Launch configurations

Your group uses a *launch configuration* as a template for its EC2 instances. When you create a launch configuration, you can specify information such as the AMI ID, instance type, key pair, security groups, and block device mapping for your instances.

Scaling options

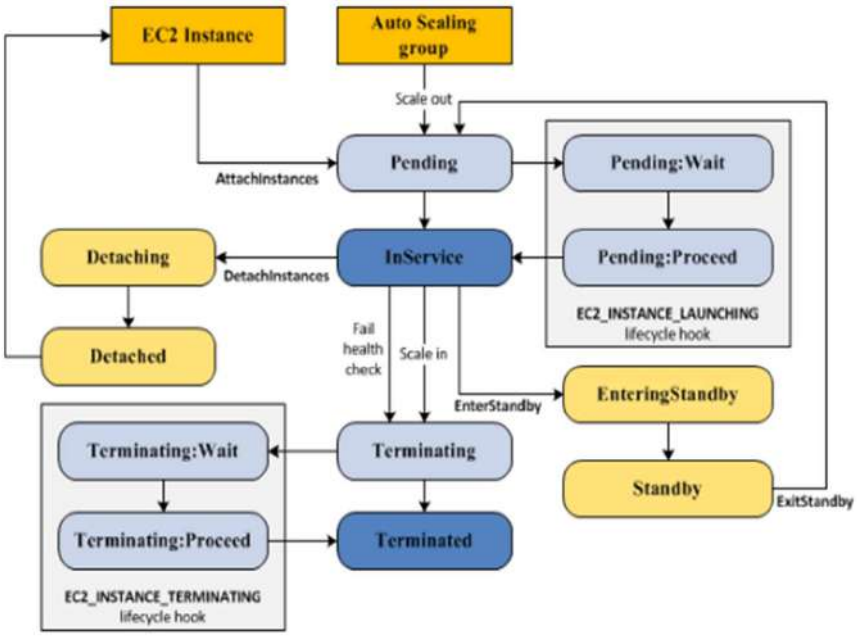
Amazon EC2 Auto Scaling provides several ways for you to scale your Auto Scaling groups. For example, you can configure a group to scale based on the occurrence of specified conditions (dynamic scaling) or on a schedule.

Auto Scaling Lifecycle

The EC2 instances in an Auto Scaling group have a path, or lifecycle, that differs from that of other EC2 instances. The lifecycle starts when the Auto Scaling group launches an instance and puts it into service. The lifecycle ends when you terminate the instance,or the Auto Scaling group takes the instance out of service and terminates it.

Note : You are billed for instances as soon as they are launched, including the time that they are not yet in service.

The following illustration shows the transitions between instance states in the Amazon EC2 Auto Scaling lifecycle.



PRE-REQUISITES

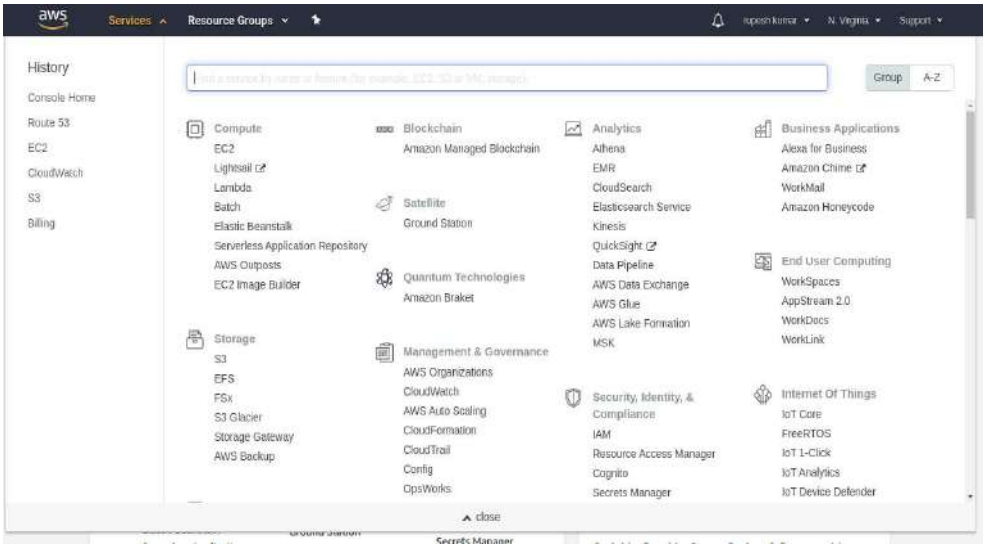
You should have AWS account, or IAM user with EC2 Full access

Task

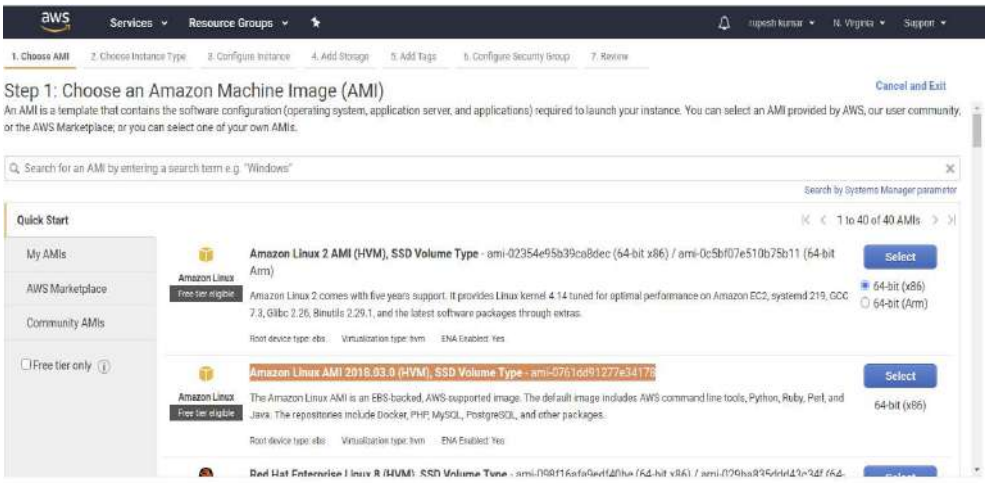
1. **Launch Amazon linux instance**
2. **Configure web server**
3. **Stop the instance**
4. **Create AMI image of above instance**
5. **Configure Auto Scaling launch configuration and Auto Scaling group**
6. **Configure load balancer auto scaling**

Steps :

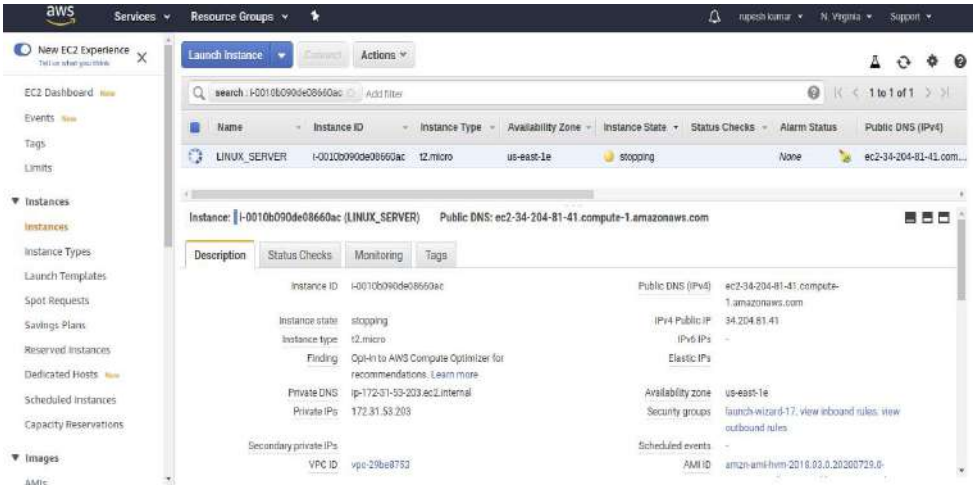
From the aws management console, click on services and select ec2 service from the compute section.



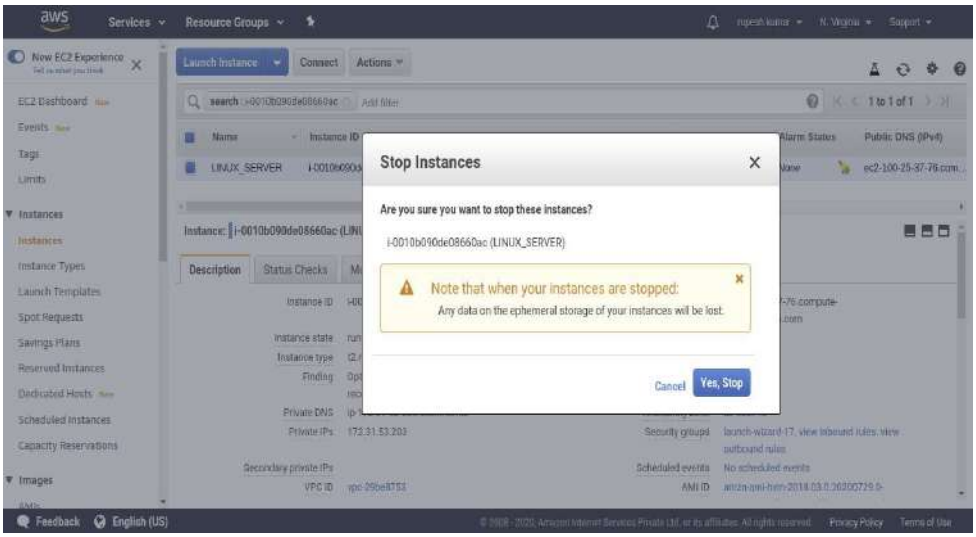
Launch an Amazon linux instance and Configure web server like NGINX



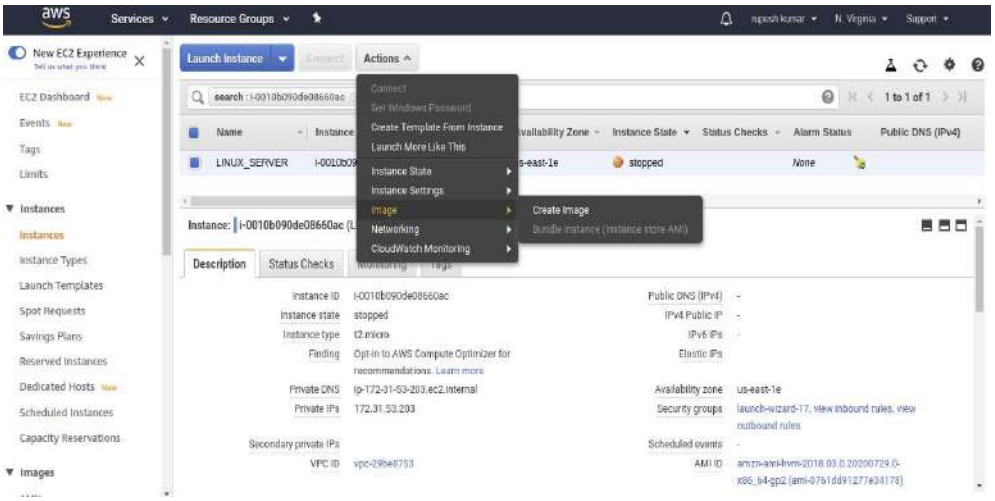
Stop the instance : To create AMI from this instance on ‘EC2 Dashboard Panel”. select the instance and click on action and select the instance state then stop.



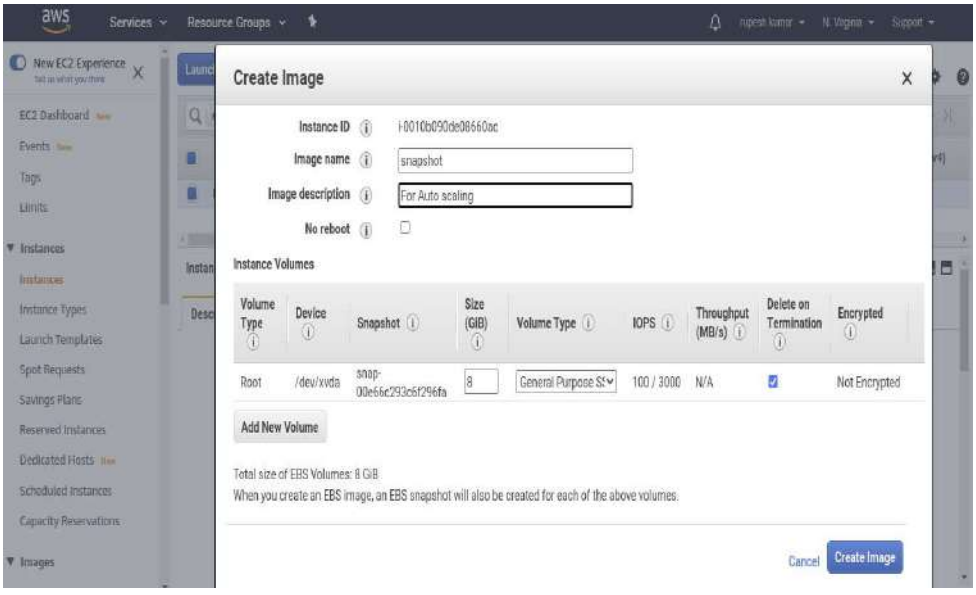
Click on Yes, Stop.



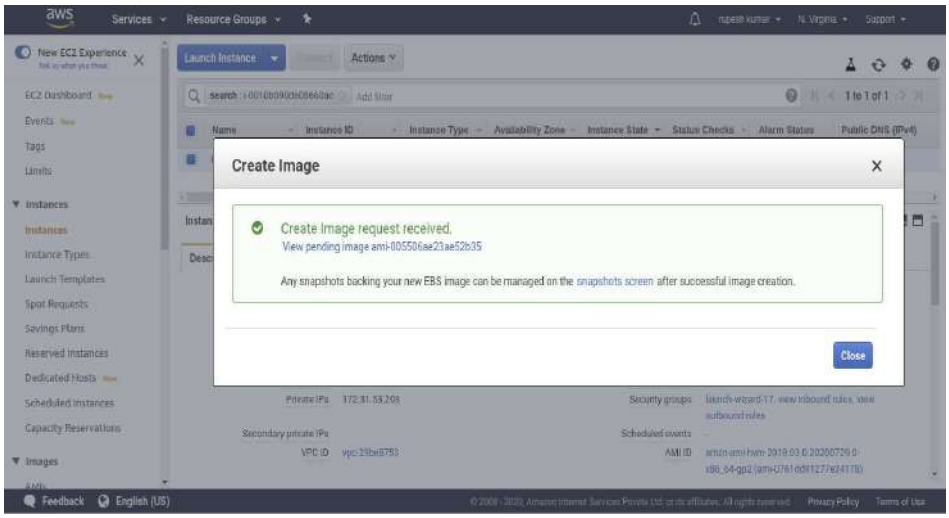
Select the stopped instance and click on action button
Select Image and then click on create image button



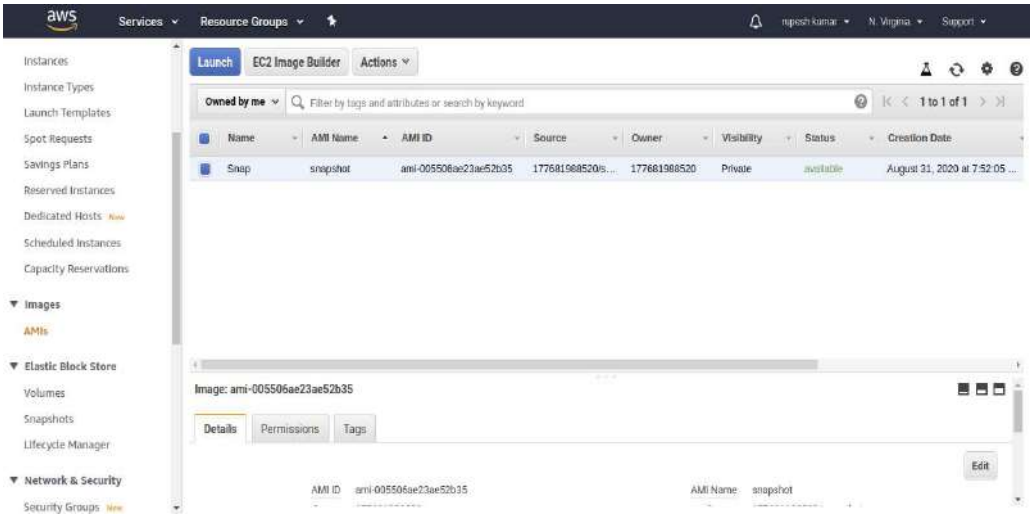
Click on create Image



Now, click on Close



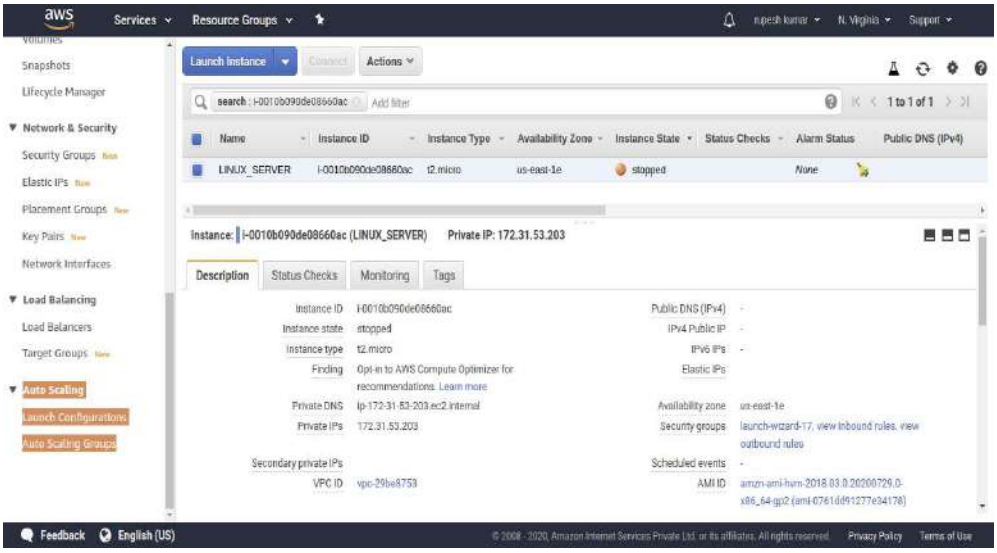
Verify AMI is created or not in the EC2 Dashboard panel. On the EC2 Dashboard panel, Select Images and Click on AMI's then check status is on Available state or Not.



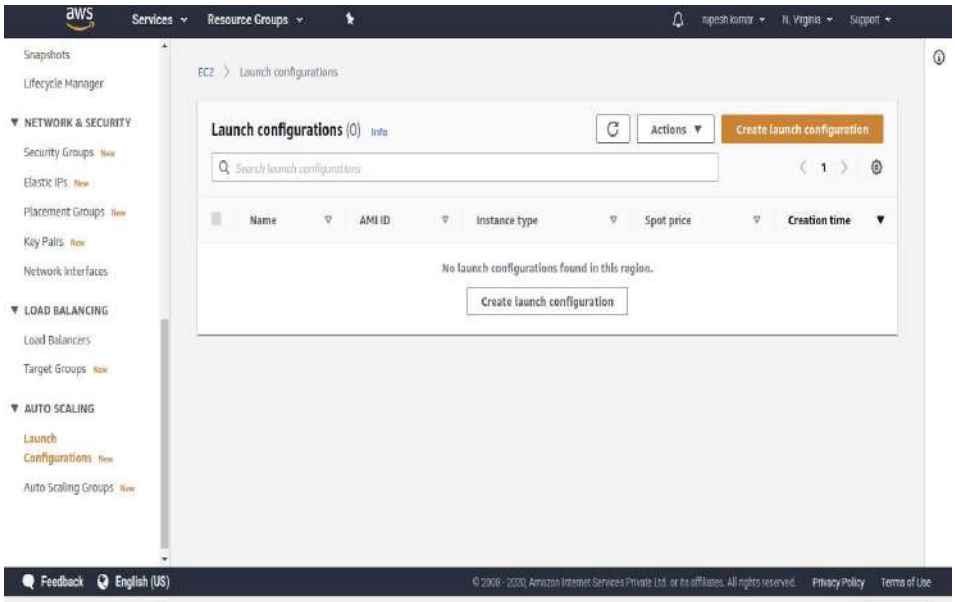
Configure Auto Scaling launch configuration and Auto Scaling group

To configure Auto Scaling Group on the Ec2 Dashboard Panel

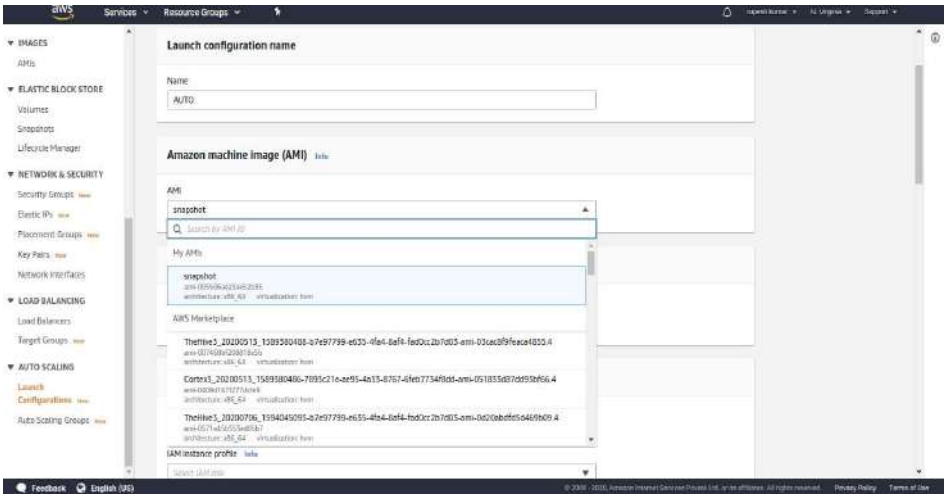
Select Auto Scaling



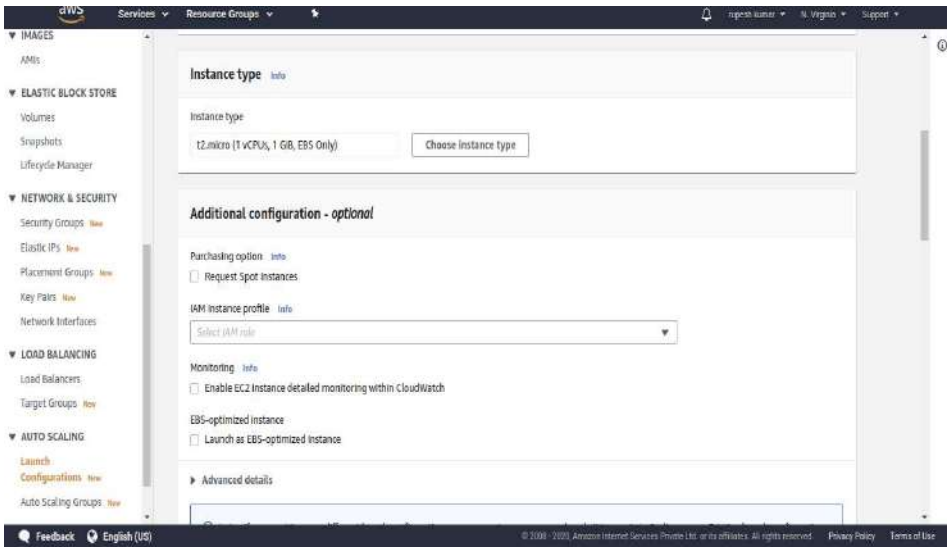
Click on Launch Configurations



Assign a name for the Launch Configurations and select the AMI that you want to configure

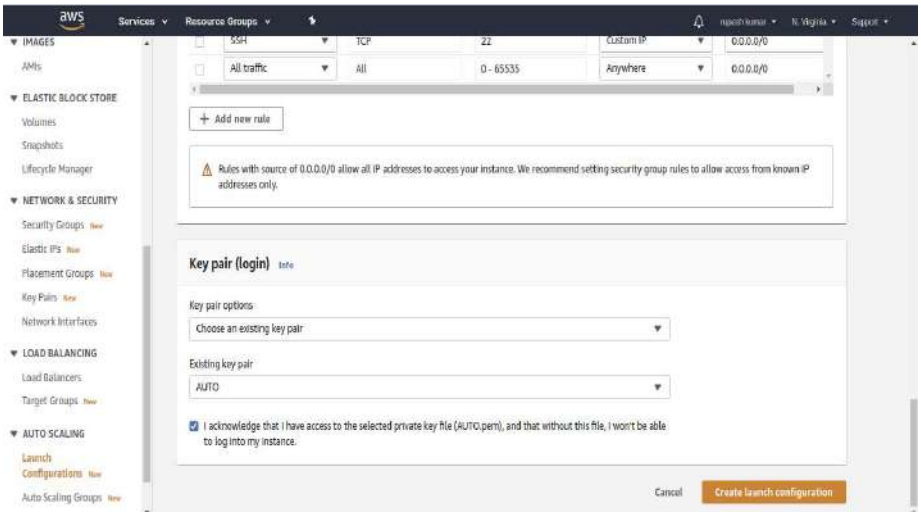


Select the instance type as t2.micro ,because the original instance has the same instance type .

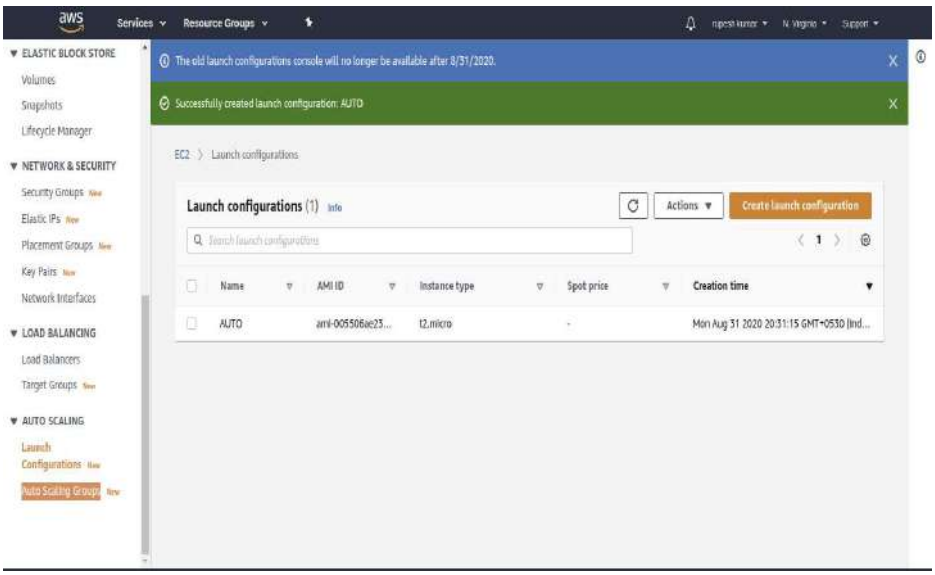


Under security groups , click on the Create a new security group .Scroll down a little bit, under the rules section ,click on the “add new rule” button and add a rule.(All traffic or HTTP and HTTPS)

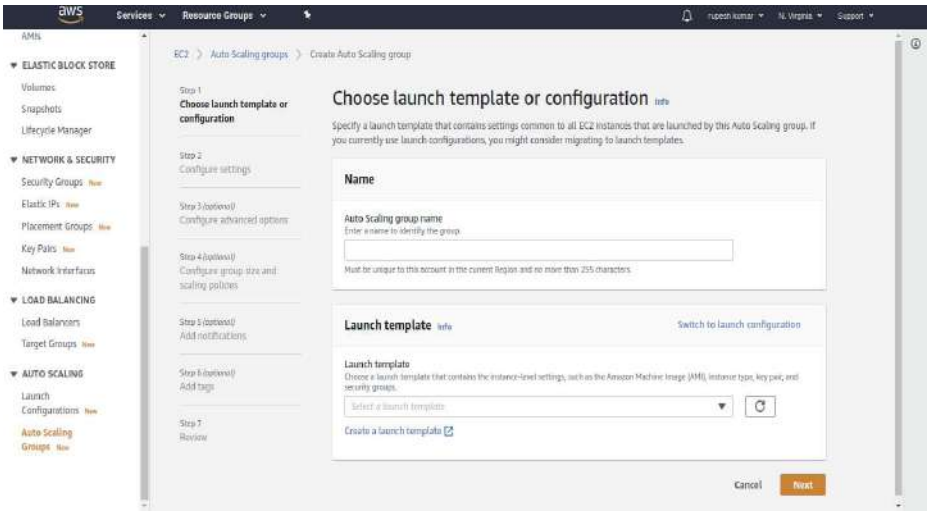
Select on the existing key pair, select your keypair(previously you're using in instance creation) click on i acknowledge checkbox and click on “create launch configuration” button.



Launch configuration was successfully created. Now scroll down the left side panel and click on auto scaling groups under the auto scaling section.

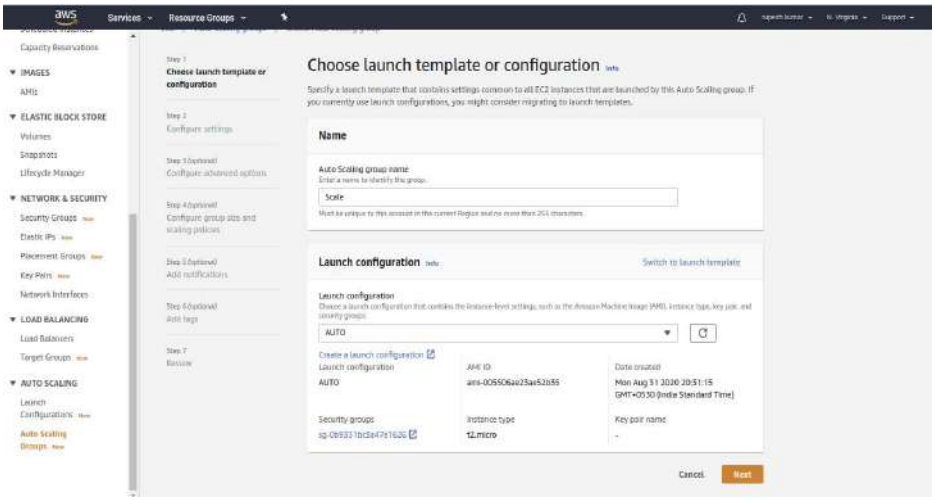


Click on Create Auto Scaling Group



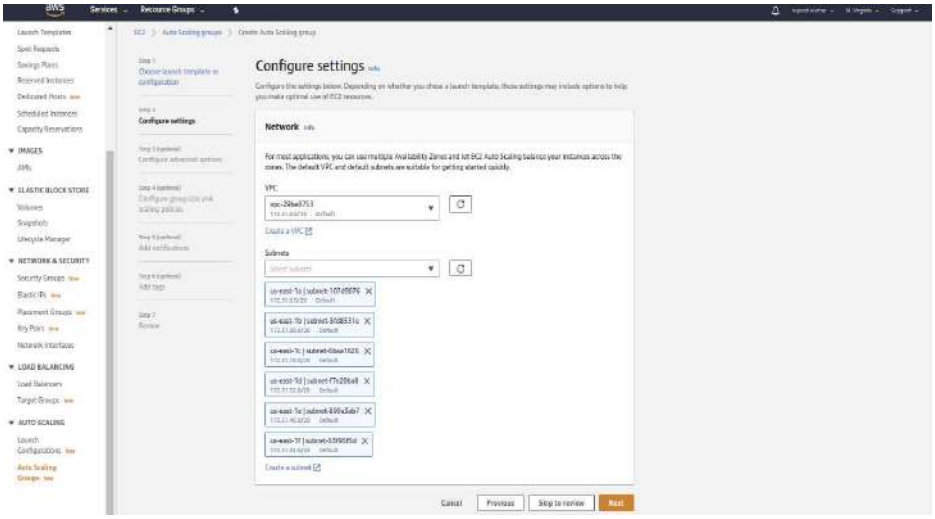
Choose launch template or configuration:

Enter Auto scaling group name and Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group. If you currently use launch configurations, you might consider migrating to launch templates.(*Here i am using Launch configuration option.Already i launched launch configuration with the name AUTO*).Click on Next option.



Configure settings:

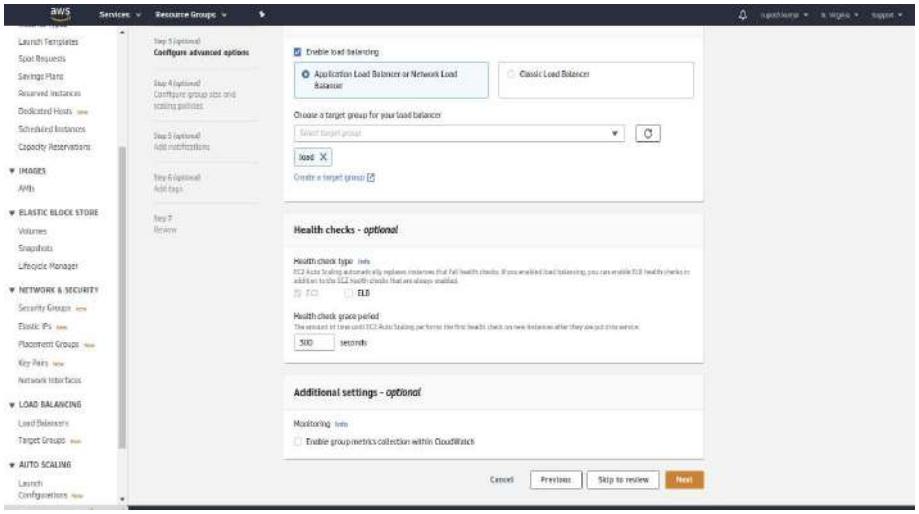
Configure the settings below. Depending on whether you chose a launch template, these settings may include options to help you make optimal use of EC2 resources.Select VPC and all subnets under that VPC.And click on Next.



Configure advanced options :

Choose a load balancer to distribute incoming traffic for your application across instances. You can also set options that give you more control over checking the health of instances.

Under Load balancing please enable that *Enable* ***load balancing*** check box.After that select the Target group of your load balancer.If you are not created previously please create a new one, by click on create new target group.I already created the Target group name as ***load.*** After that Health checks grace period make it 300 seconds.After that click on Next button.



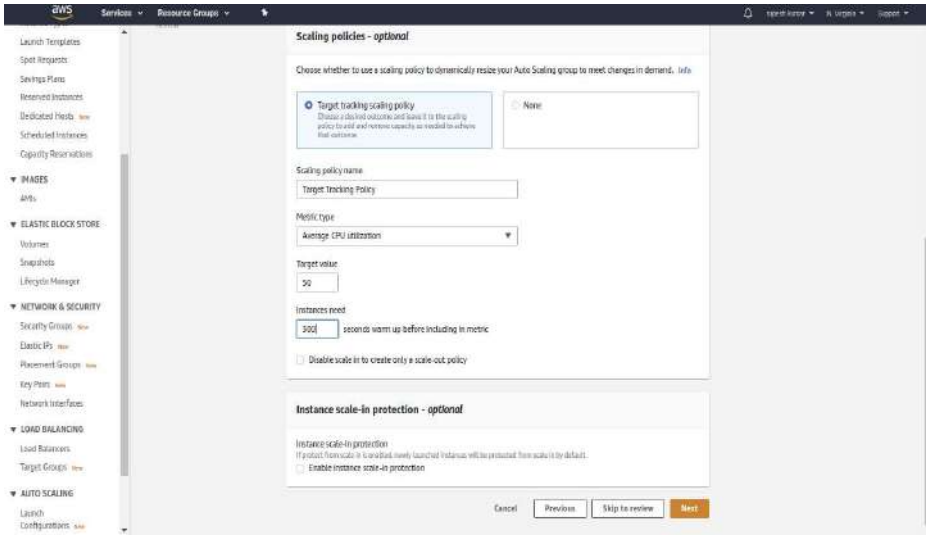
Configure group size and scaling policies:

Set the desired, minimum, and maximum capacity of your Auto Scaling group. You can optionally add a scaling policy to dynamically scale the number of instances in the group.

In group size option select the Desired capacity as greater than or equal to Minimum capacity and less than or equal to Maximum capacity .

In Scaling policies , select Target tracking scaling policy then give the scaling policy name ,Metric type (select Metric as CPU utilization),and Target value 50 %.

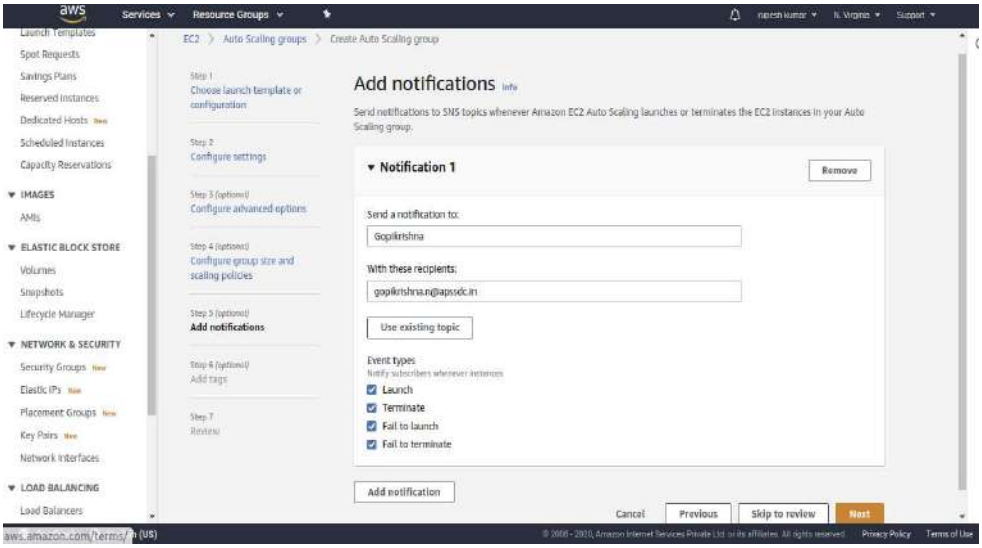
Instances need time that are 300 seconds warm up before including in the metric.After that click on the Next button.



Add notifications:

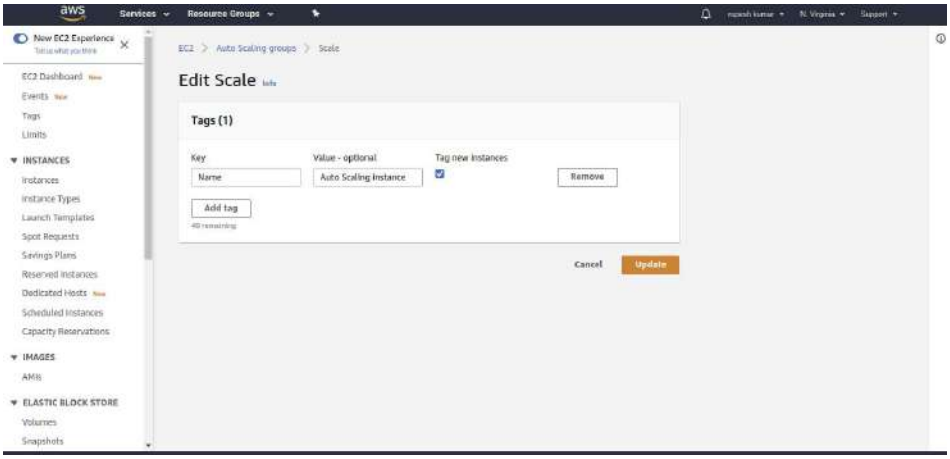
Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group.

Click on Add notification button and Enter the name of the person you want to send notification and mail id of the person.After that click on the next button.



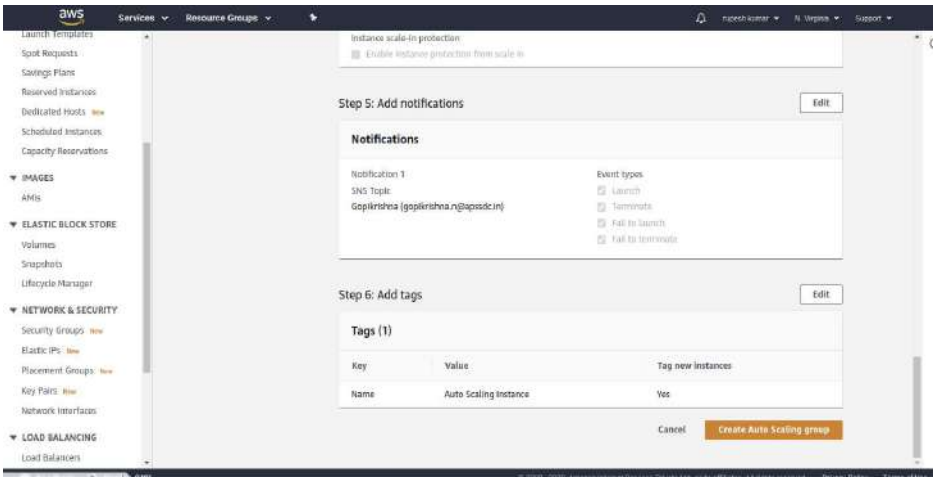
Add tags :

Add tags to help you search, filter, and track your Auto Scaling group across AWS. You can also choose to automatically add these tags to instances when they are launched.After that click on the Next button.

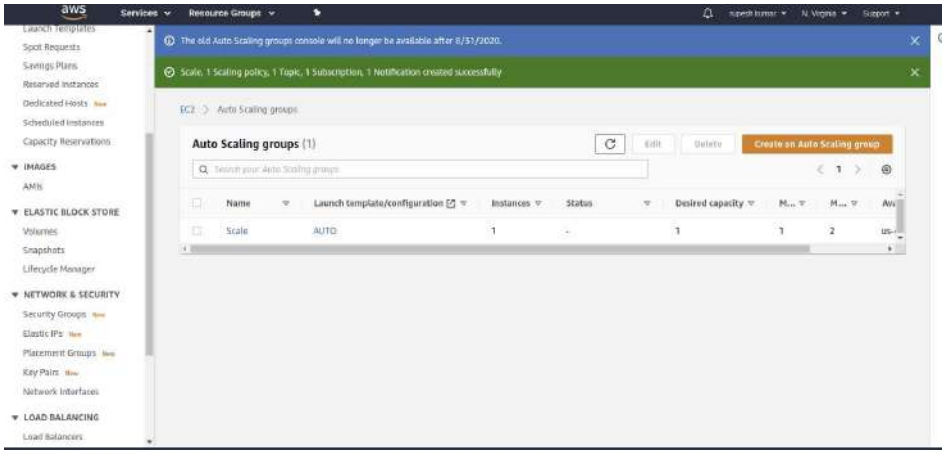


Review :

In the Review part once again select all and click on Create Auto scaling group.

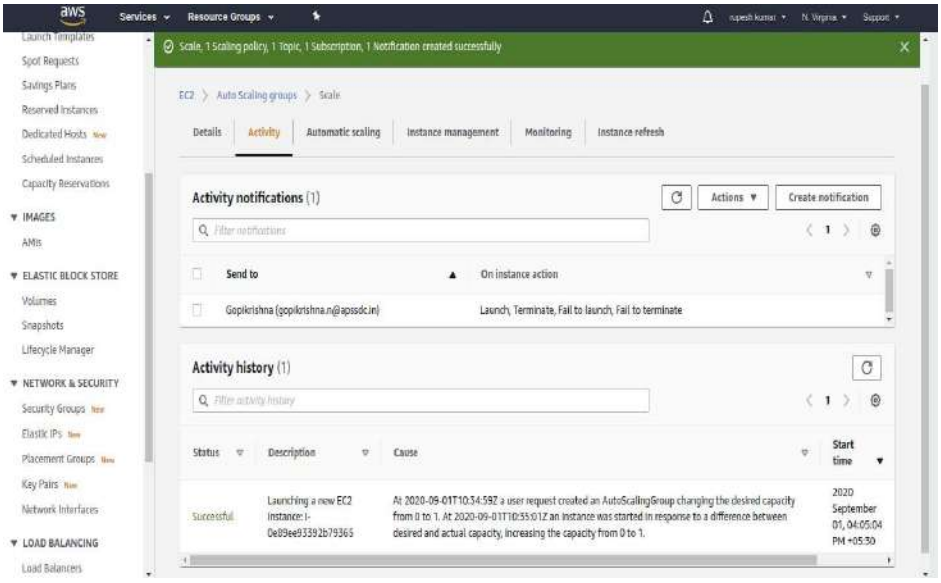


Check the status of the Auto scaling group.

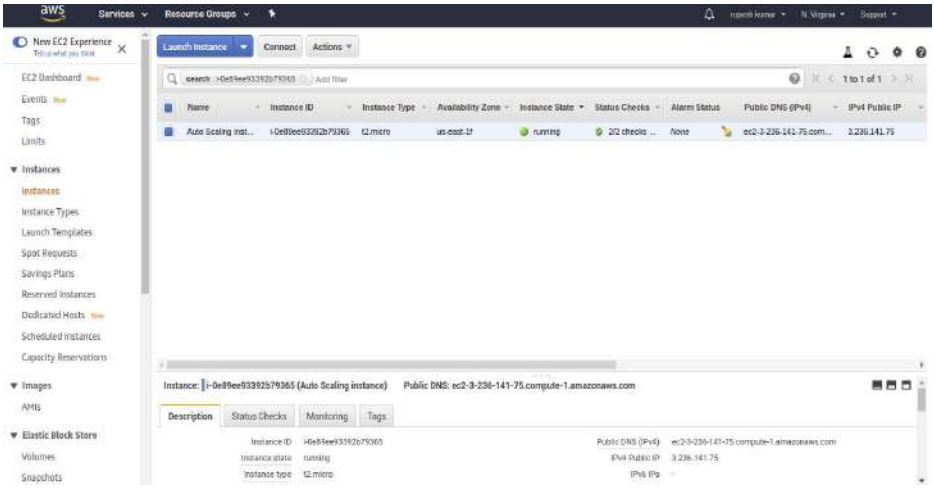


Click on the Auto scaling group name ,after that click on Activity and check the Activity History status.

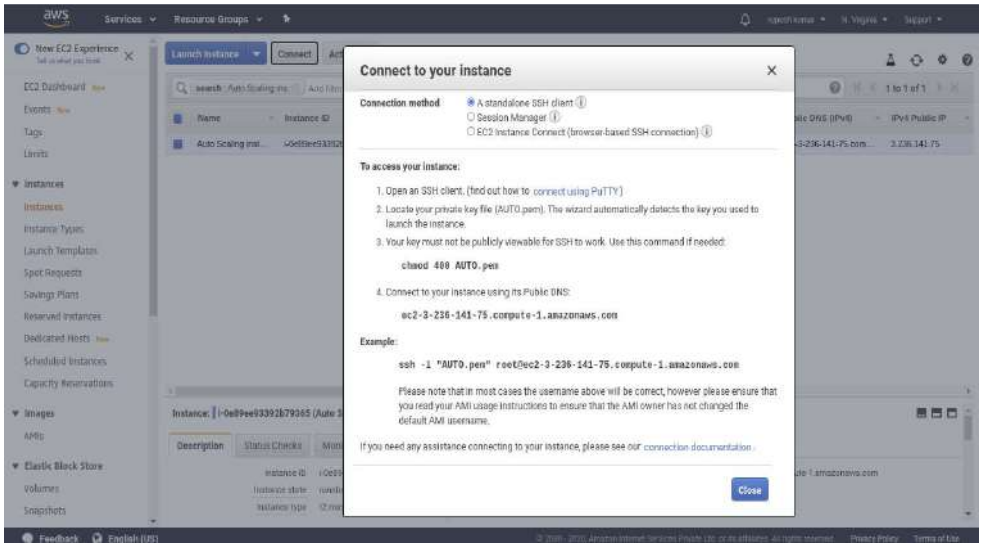
Launched on Instance successfully



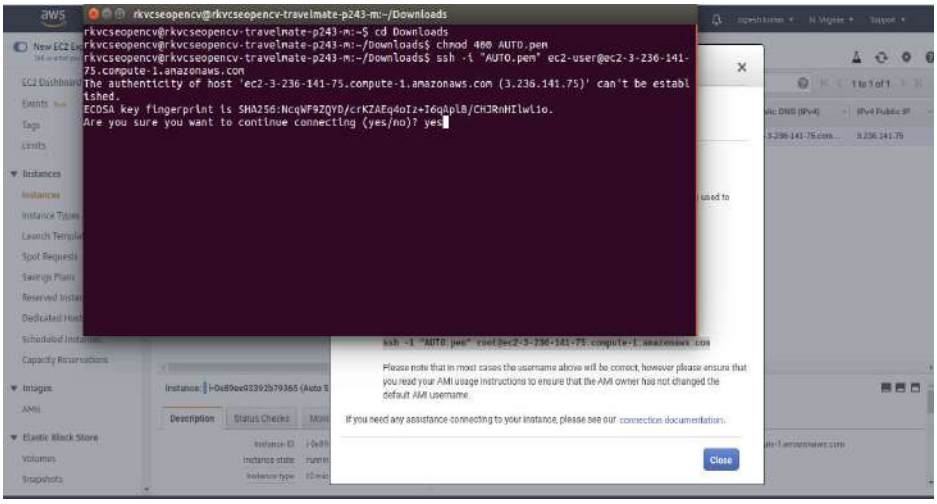
Open AWS Management console and click on running instances.You can find out the instance name as *Auto scaling instance*.



Now we can apply stress management on Auto scaling instance.Connecting the server through putty or Terminal or windows powershell.Now I am connecting through Terminal.



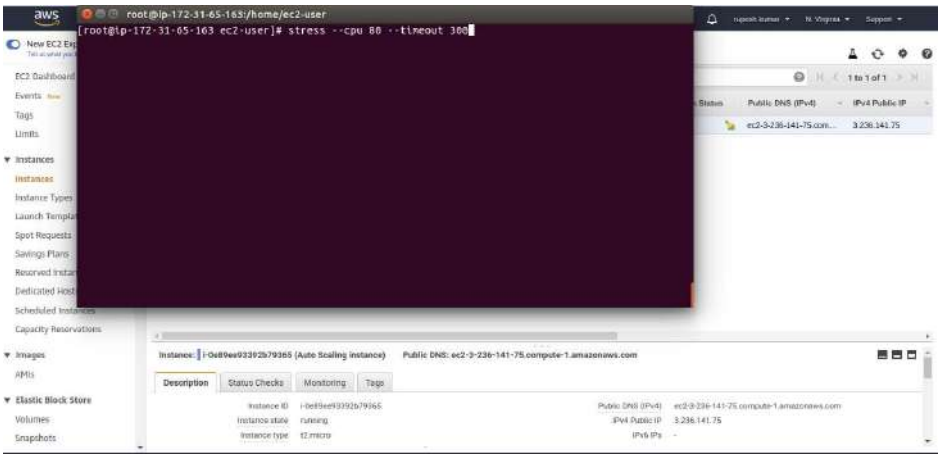
Now open Terminal and connect to your server /instance.



Now install stress and apply stress Management on the server
Use this commands for installing and apply stress management

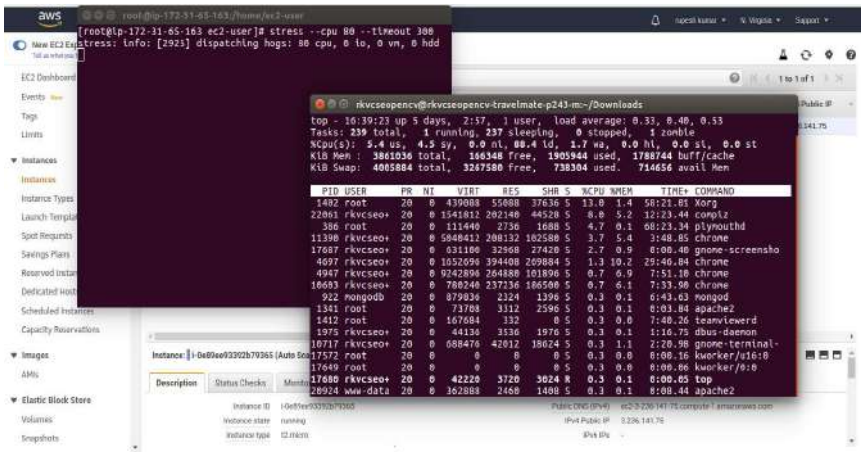
For Installation:
Sudo yum install stress (incase Linux server)
Sudo apt-get install stress (Incase of Ubuntu server)

Apply stress by using :
stress --cpu 80 --timeout 300

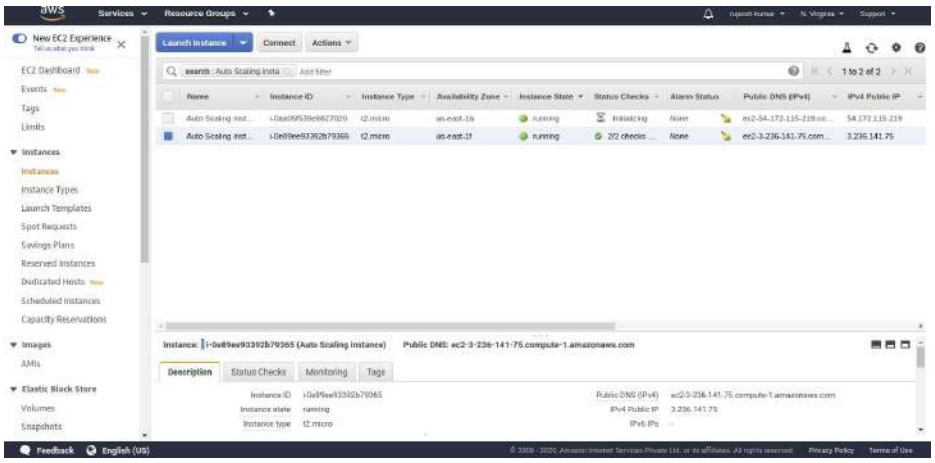


Now check the status of CPU utilization by using following command

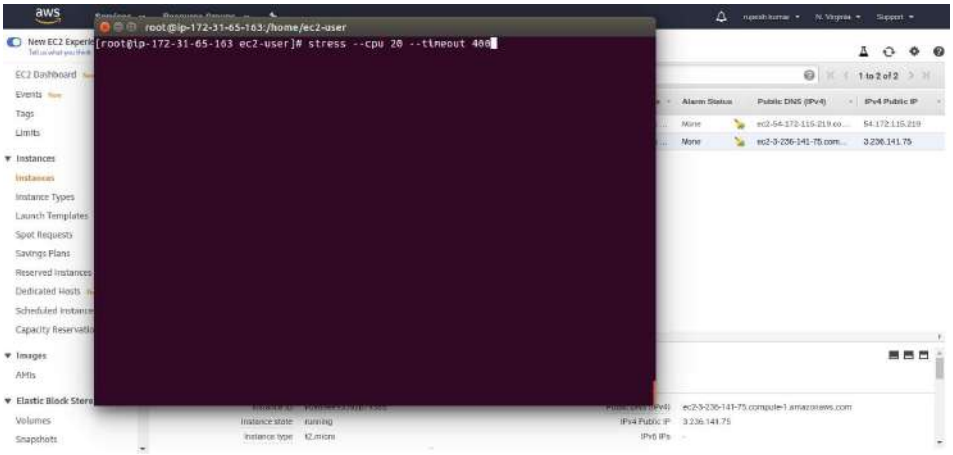
Open new Terminal and use this command **: top**



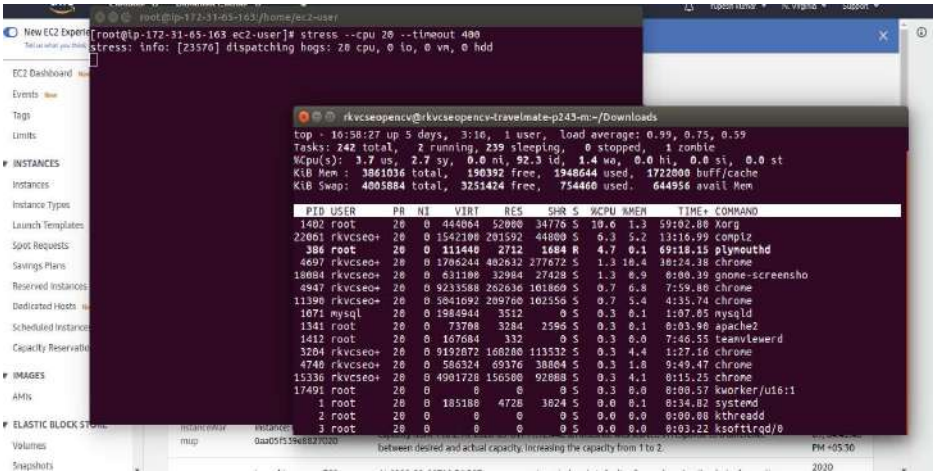
When i am applying stress management on instance(Name of instance : **Auto scaling instance**).Automatically one more new instance launched .Check the status of the below screenshot.



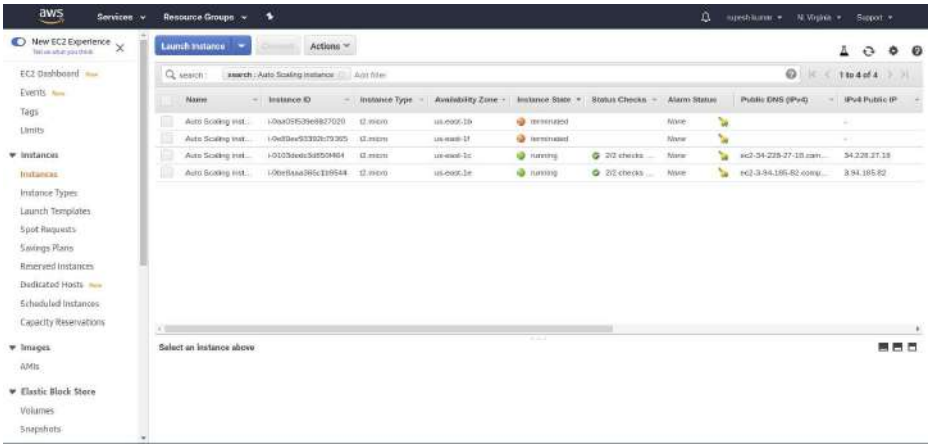
Now you can Decrease the Stress management for instance by using the following command.



Check the status of the CPU Utilization on the server in the below screenshot.



Now goto Ec2 dashboard and check the status of the instances.The CPU utilization goes down so that instances will be terminated which were launched when CPU utilization high.Look at this below screenshot.



The lab regarding Auto scaling with Load Balancer and snapshot was completed……...