



Andhra Pradesh State Skill Development Corporation



AWS CLOUD COMPUTING

CONFIGURE AMAZON AUTO SCALING WITH LOAD BALANCER AND SNAPSHOT



Andhra Pradesh State Skill Development Corporation (APSSDC)



Configure Amazon Auto Scaling with load balancer and snapshot

Amazon EC2 Auto Scaling

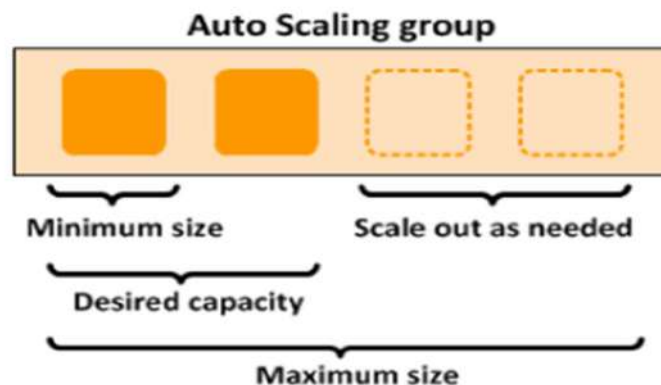




To configure Amazon Auto Scaling with load balancer and snapshot

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.

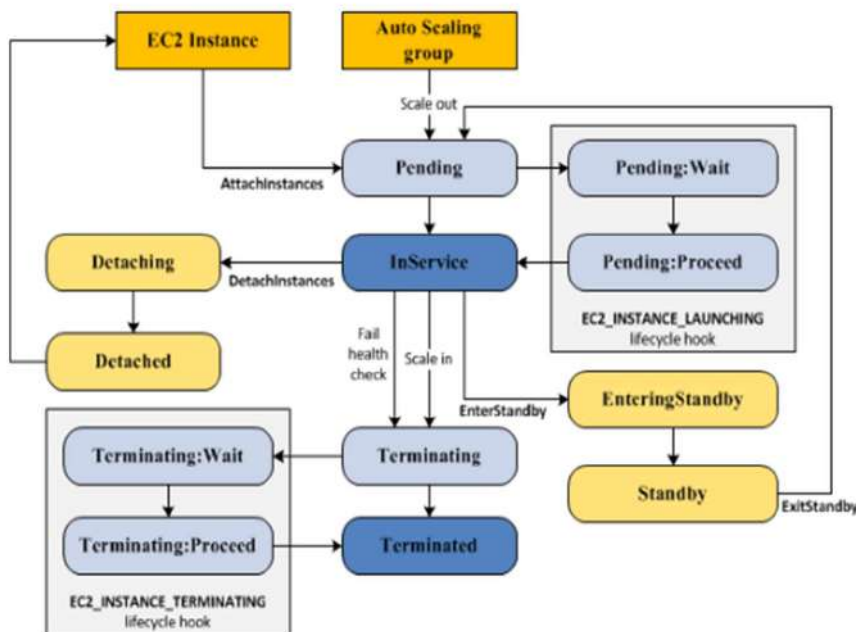
Auto Scaling Lifecycle

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

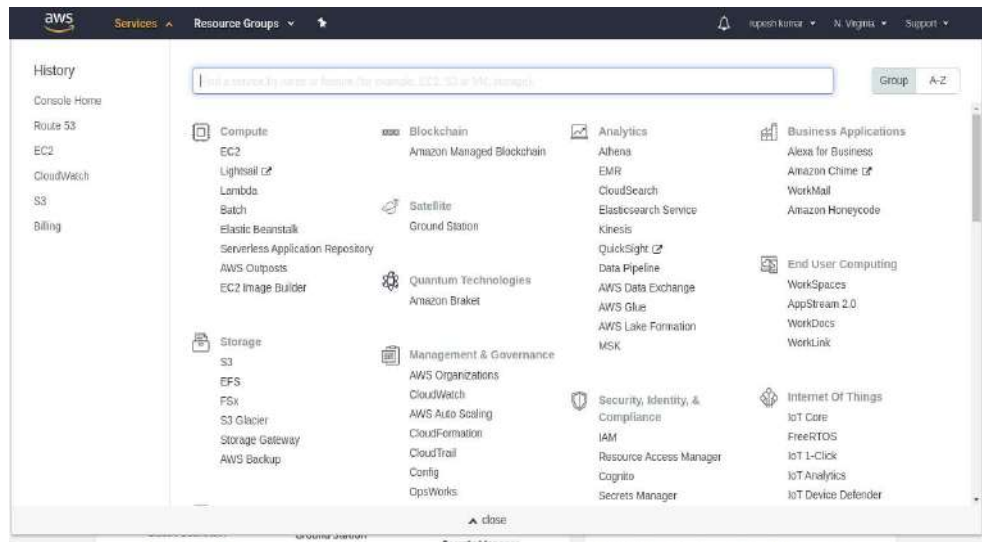
Task

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

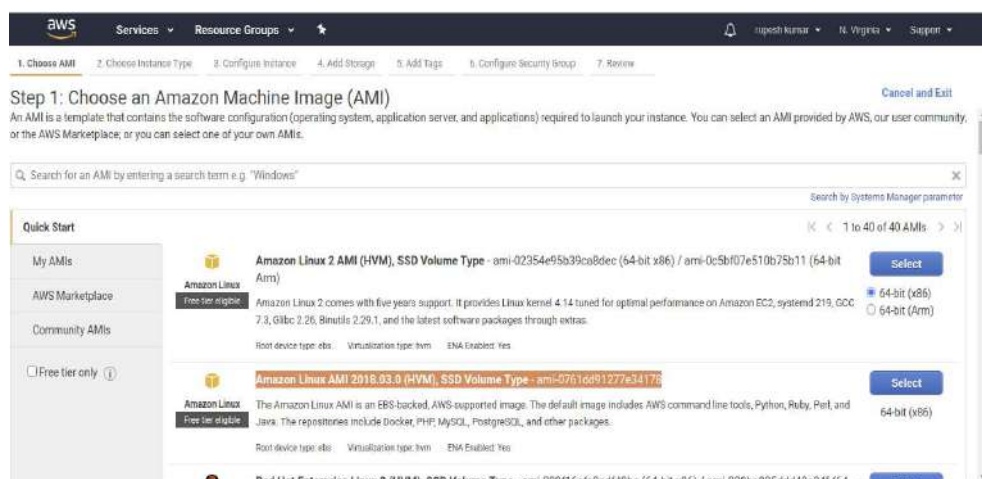


Steps :

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

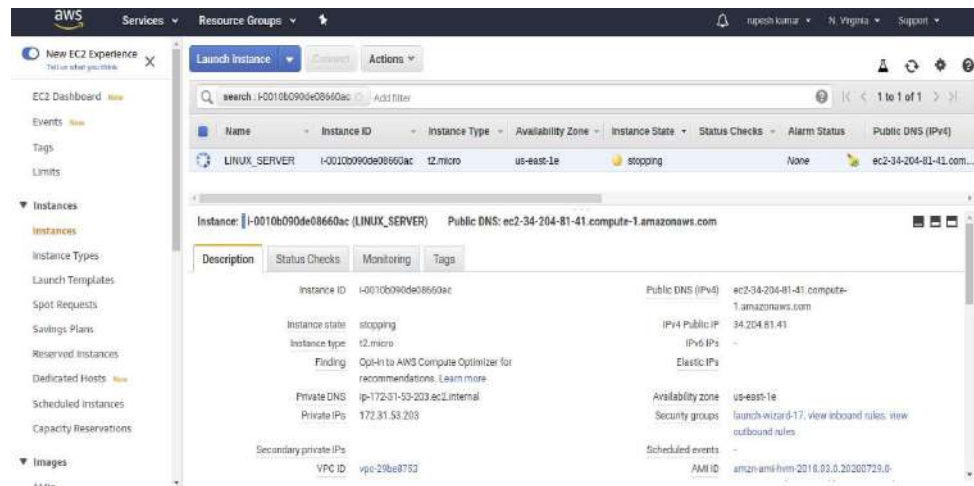


Launch an Amazon linux instance and Configure web server like NGINX

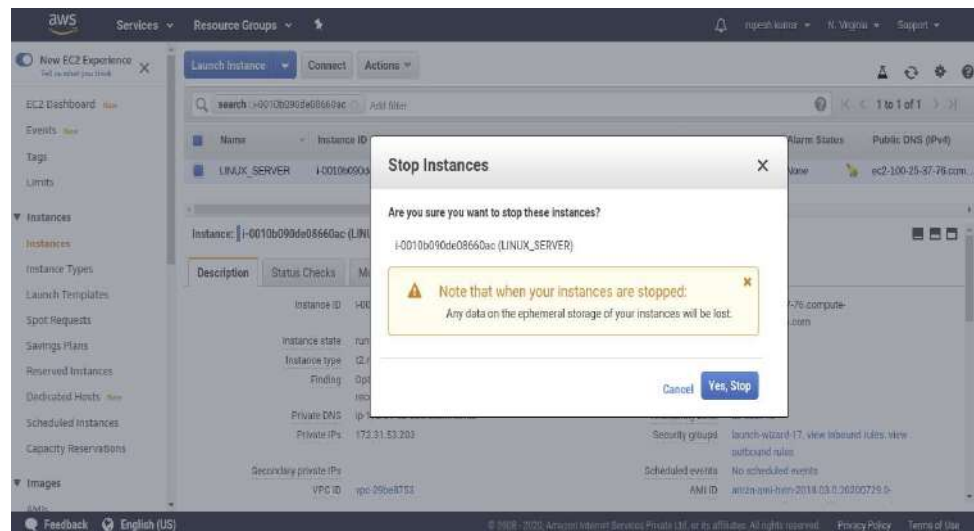




Stop the instance : To create AMI from this instance on ‘EC2 Dashboard Panel’”. select the instance and click on action and select the instance state then stop.

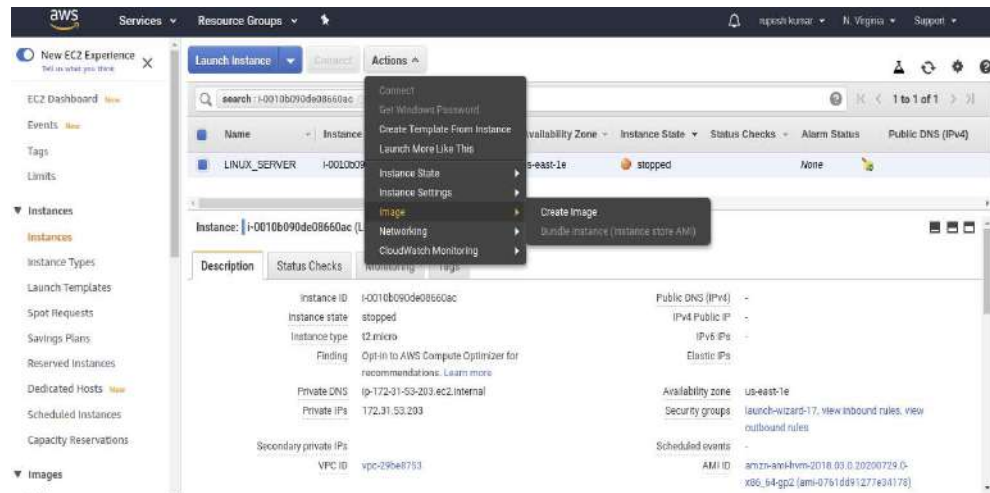


Click on Yes, Stop.

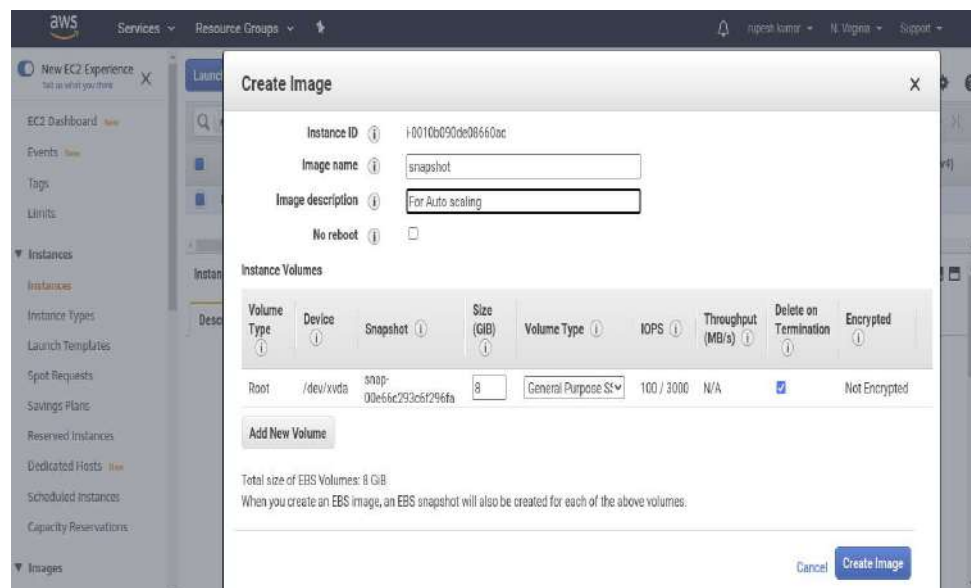




Select the stopped instance and click on action button
Select Image and then click on create image button

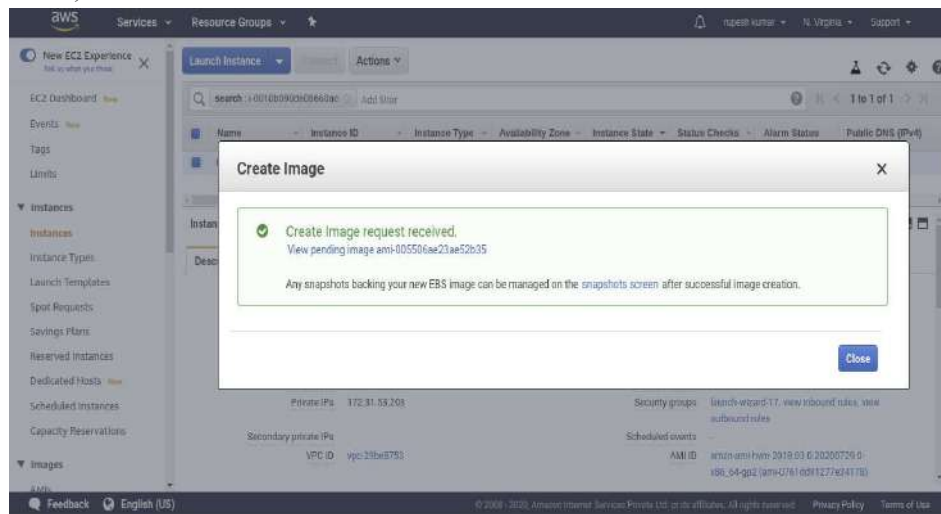


Click on create Image

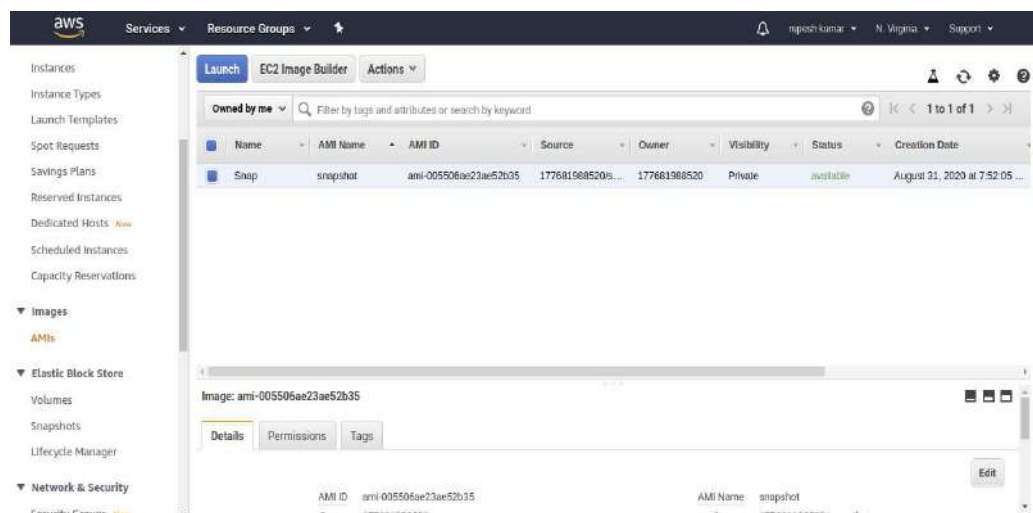




Now, click on Close



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

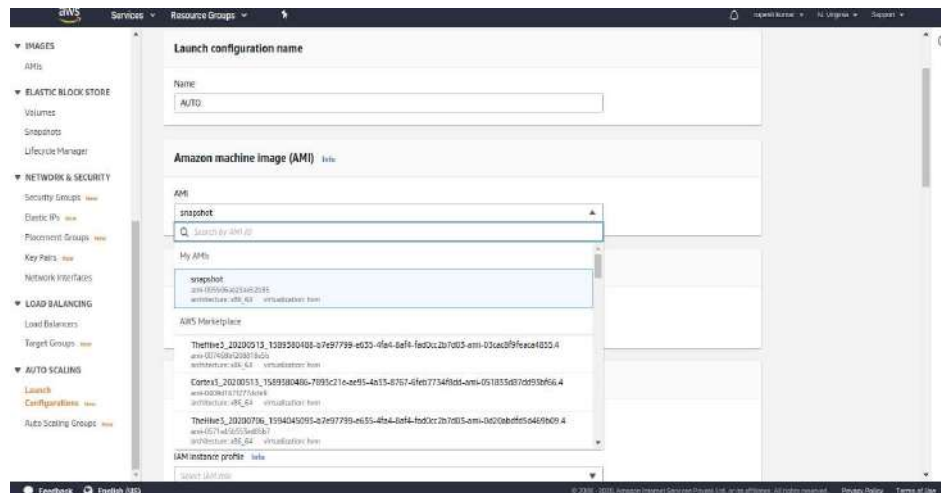
The screenshot shows the AWS Management Console interface. On the left sidebar, under 'Network & Security', the 'Auto Scaling' link is highlighted. The main content area displays the details for an EC2 instance named 'LINUX_SERVER' with Instance ID 'i-0010b090de08660ac'. The instance is in a 'stopped' state. Below the instance details, there are tabs for 'Description', 'Status Checks', 'Monitoring', and 'Tags'. The 'Description' tab is active, showing various instance attributes like Instance ID, Instance state, Instance type, Private DNS, Private IPs, Secondary private IPs, VPC ID, Public DNS (IPv4), IPv4 Public IP, IPv6 IPs, Elastic IPs, Availability zone, Security groups, and Scheduled events.

Click on Launch Configurations

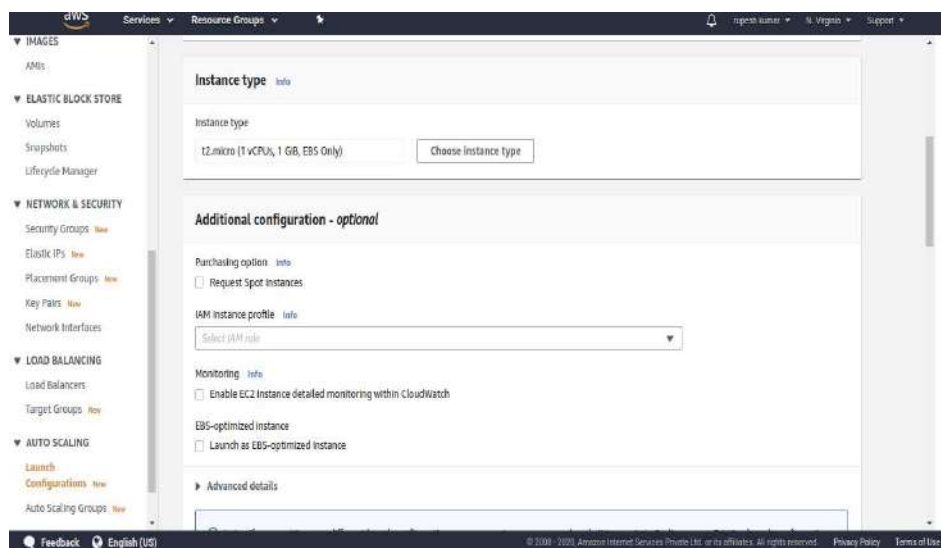
The screenshot shows the AWS Management Console interface for 'Launch configurations'. The left sidebar shows the 'Launch Configurations' link under the 'Auto Scaling' section. The main content area displays the 'Launch configurations (0)' page. It includes a search bar, a table with columns for Name, AMI ID, Instance type, Spot price, and Creation time. A message states 'No launch configurations found in this region.' with a 'Create launch configuration' button.



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.

The screenshot shows the 'Key pair (login)' step in the AWS 'Create launch configuration' wizard. The 'Key pair options' dropdown is set to 'Choose an existing key pair'. The 'Existing key pair' dropdown is set to 'AUTO'. The 'I acknowledge' checkbox is checked. The 'Create launch configuration' button is visible at the bottom right.

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

The screenshot shows the 'Launch configurations' page in the AWS Management Console. A green notification bar at the top states 'Successfully created launch configuration: AUTO'. The 'Launch configurations (1)' table shows one configuration named 'AUTO' with AMI ID 'ami-005506ae23...', instance type 't2.micro', and creation time 'Mon Aug 31 2020 20:31:15 GMT+0530 (Ind...)'. The 'Auto Scaling Groups' link is highlighted in the left sidebar.



Click on Create Auto Scaling Group

The screenshot shows the AWS Management Console interface for creating an Auto Scaling group. The left sidebar contains navigation links for various AWS services. The main content area displays the 'Create Auto Scaling group' wizard, specifically Step 1: 'Choose launch template or configuration'. The 'Name' field is empty, and the 'Launch template' dropdown is set to 'Select a launch template'. The 'Next' button is visible at the bottom right.

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.

The screenshot shows the AWS Management Console interface for creating an Auto Scaling group. The left sidebar contains navigation links for various AWS services. The main content area displays the 'Create Auto Scaling group' wizard, specifically Step 1: 'Choose launch template or configuration'. The 'Name' field is filled with 'Scale'. The 'Launch configuration' dropdown is set to 'AUTO'. The 'Next' button is visible at the bottom right.



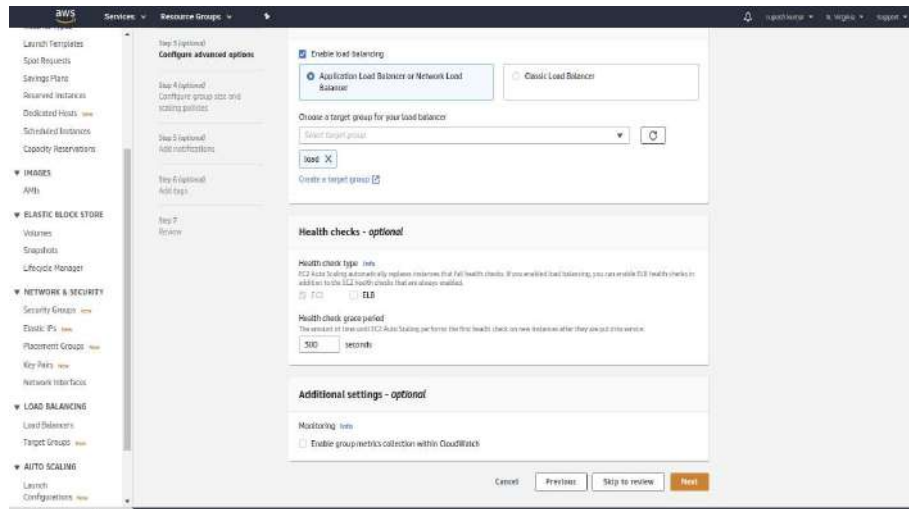
Configure settings:

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.



The screenshot shows the 'Scaling policies - optional' configuration page in the AWS Auto Scaling console. The left sidebar lists various AWS services. The main content area has a section titled 'Scaling policies - optional' with a sub-header 'Choose whether to use a scaling policy to dynamically resize your Auto Scaling group to meet changes in demand. Info'. There are two radio buttons: 'Target tracking scaling policy' (selected) and 'None'. Below this, there are input fields for 'Scaling policy name' (set to 'Target tracking Policy'), 'Metric type' (set to 'Average CPU utilization'), and 'Target value' (set to '50'). There is also a 'Instances need' field set to '300' seconds warm up before including in metric. A checkbox 'Disable scale-in to create only a scale-out policy' is present. Below this is the 'Instance scale-in protection - optional' section with a checkbox 'Enable instance scale-in protection'. At the bottom are 'Cancel', 'Previous', 'Skip to review', and 'Next' buttons.

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.

The screenshot shows the 'Add notifications' step in the AWS Auto Scaling console. The left sidebar lists various AWS services. The main content area has a section titled 'Add notifications' with a sub-header 'Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group. Info'. There is a 'Notification 1' section with a 'Remove' button. Below this, there are input fields for 'Send a notification to:' (set to 'gopikrishna') and 'With these recipients:' (set to 'gopikrishna@apssdc.in'). There is a checkbox 'Use existing topic'. Below this, there is a section 'Event types' with checkboxes for 'Launch', 'Terminate', 'Fail to launch', and 'Fail to terminate'. At the bottom are 'Add notification', 'Cancel', 'Previous', 'Skip to review', and 'Next' buttons.

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.

Key	Value - optional	Tag new instances
Name	Auto Scaling Instance	<input checked="" type="checkbox"/>

40 remaining

Cancel Update

Review :

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

Instance scale-in protection: ☐ Exclude instances protection from scale in

Step 5: Add notifications Edit

Notifications

Notification 1

SNS Topic: Gopikrishna (gopikrishna.r@apssdc.in)

Event types

☒ Launch

☒ Terminate

☒ Fail to launch

☒ Fail to terminate

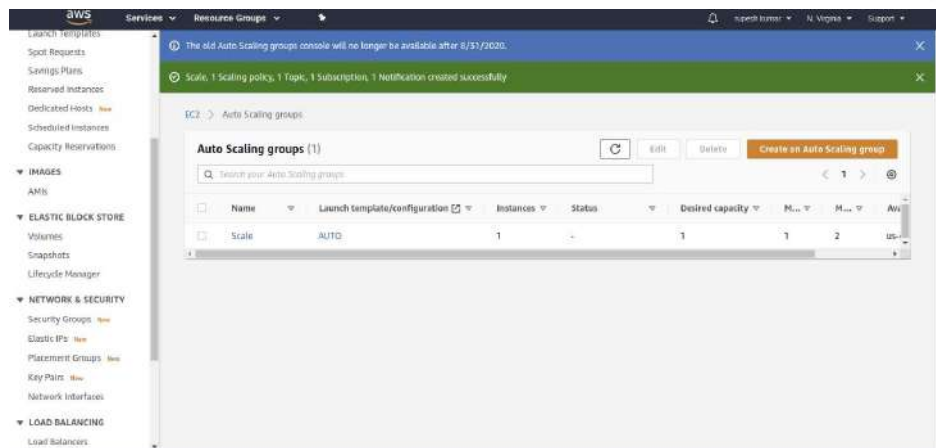
Step 6: Add tags Edit

Tags (1)

Key	Value	Tag new instances
Name	Auto Scaling Instance	Yes

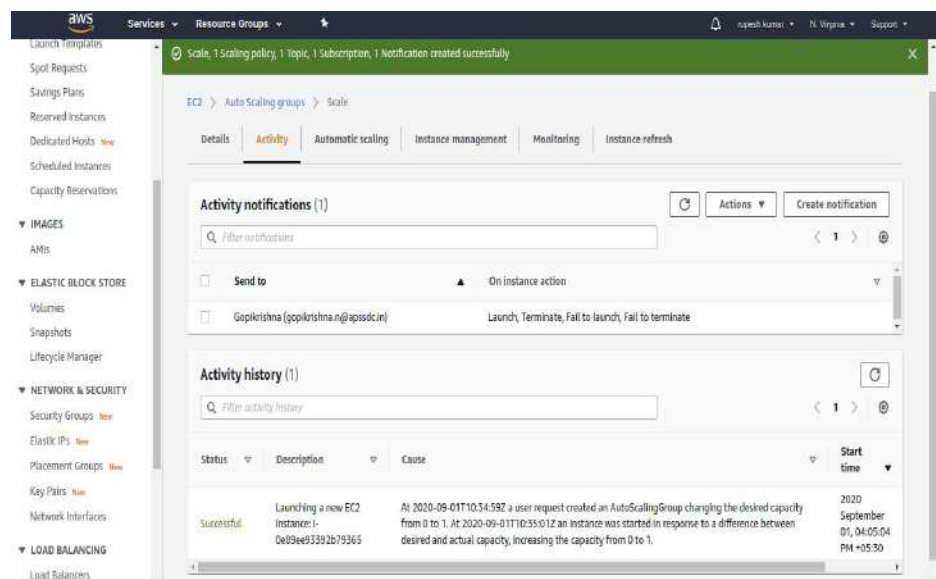
Cancel Create Auto Scaling group

Check the status of the Auto scaling group.

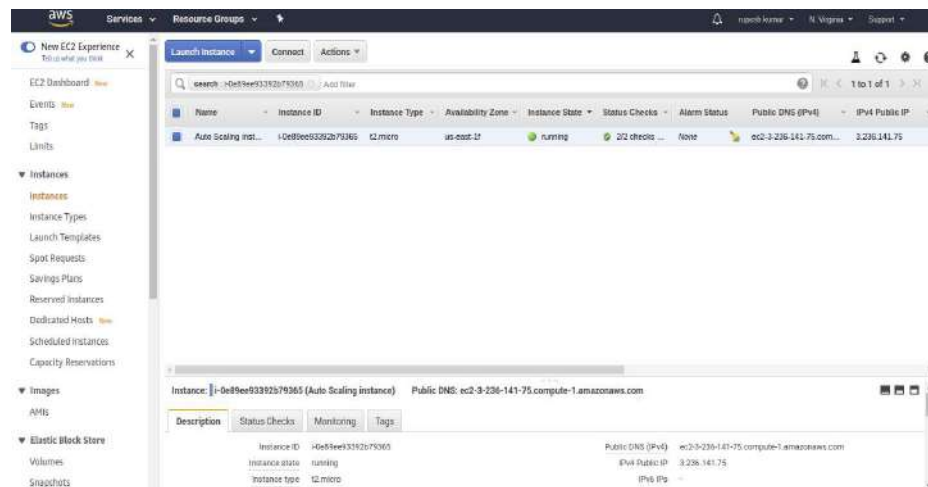


Click on the Auto scaling group name ,after that click on Activity and check the Activity History status.

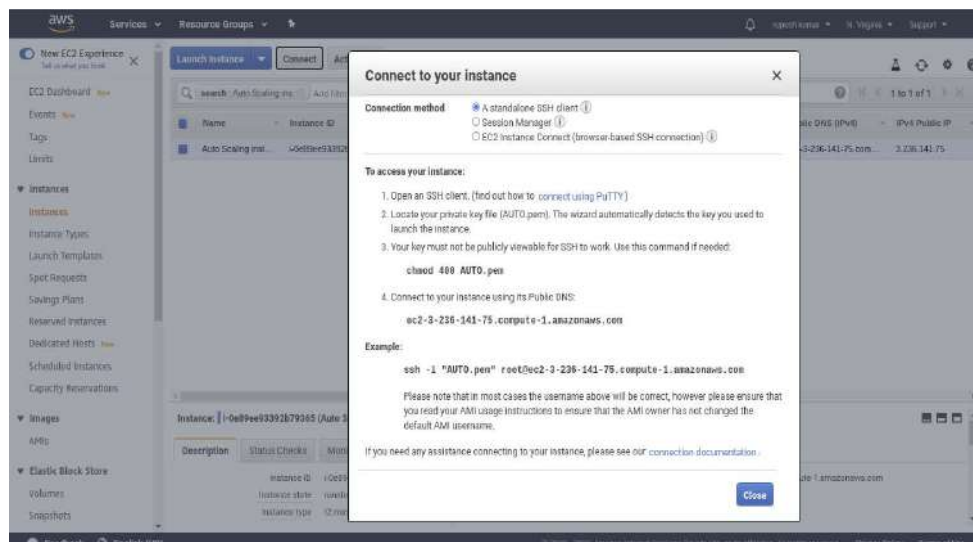
Launched on Instance successfully



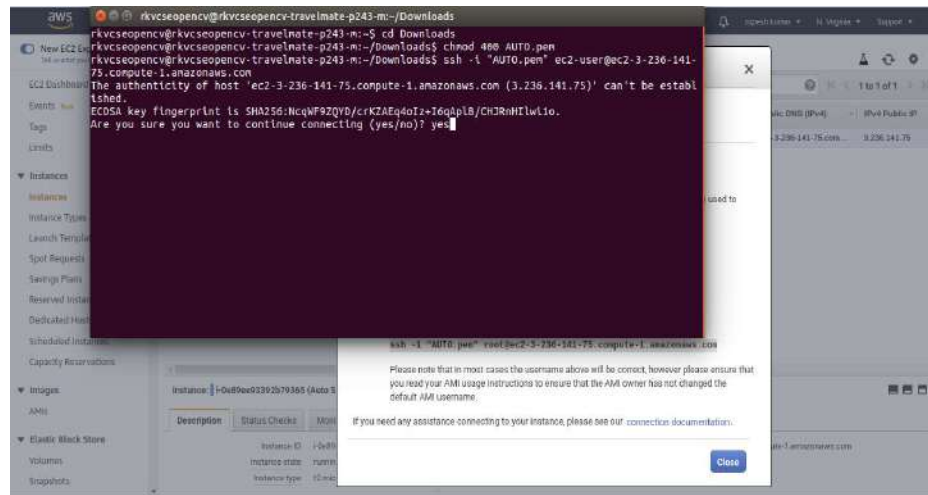
Open AWS Management console and click on running instances.You can find out the instance name as *Auto scaling instance*.



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

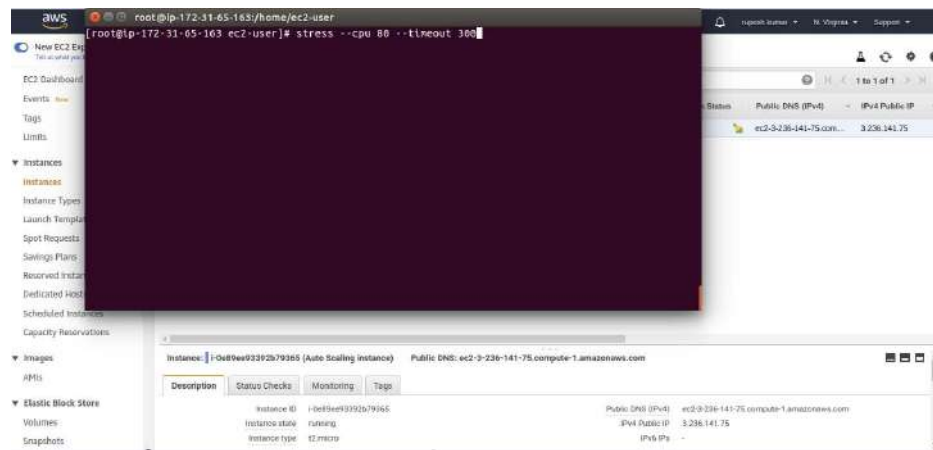
For Installation:

Sudo yum install stress (incase Linux server)

Sudo apt-get install stress (Inc case of Ubuntu server)

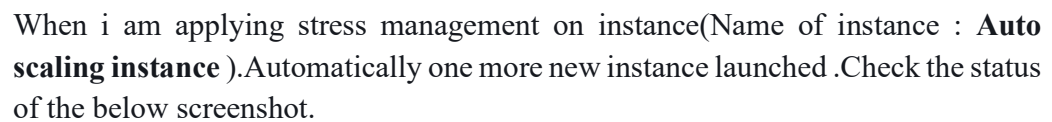
Apply stress by using :

stress --cpu 80 --timeout 300

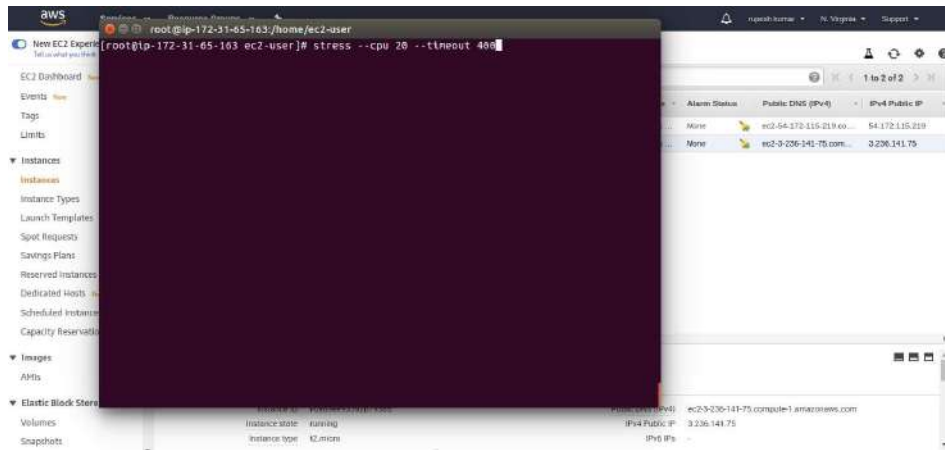


Now check the status of CPU utilization by using following command

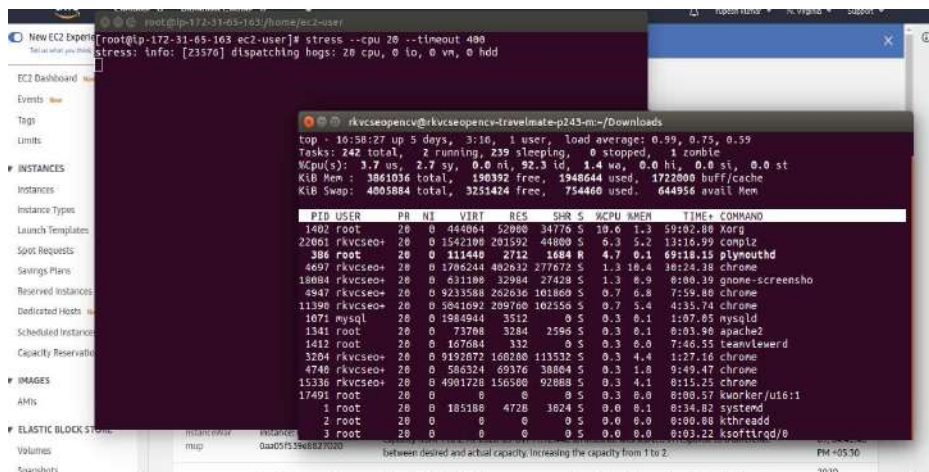
Open new Terminal and use this command : **top**



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.



Andhra Pradesh State Skill Development Corporation (APSSDC)



The screenshot shows the AWS Management Console interface. On the left, there is a navigation menu with options like 'New EC2 Experience', 'EC2 Dashboard', 'Events', 'Tags', 'Limits', 'Instances', 'Images', and 'Elastic Block Store'. The 'Instances' section is expanded, showing a list of instances. The main panel displays a table of instances managed by an Auto Scaling group. The table has columns for Name, Instance ID, Instance Type, Availability Zone, Instance State, Status Checks, Alarm Status, Public DNS (IPv4), and IPv4 Public IP. There are four instances listed: two are 'terminated' and two are 'running'.

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Public IP
Auto Scaling Inst...	i-0ba05f3b68827028	t2.micro	us-east-1b	terminated	None	None		
Auto Scaling Inst...	i-0ed3ae523320c79305	t2.micro	us-east-1f	terminated	None	None		
Auto Scaling Inst...	i-0123ae523320c79304	t2.micro	us-east-1c	running	2/2 checks ...	None	ec2-34-225-27-18.com...	34.225.27.18
Auto Scaling Inst...	i-00e3ae523320c79304	t2.micro	us-east-1e	running	2/2 checks ...	None	ec2-34-185-82.com...	34.185.82

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