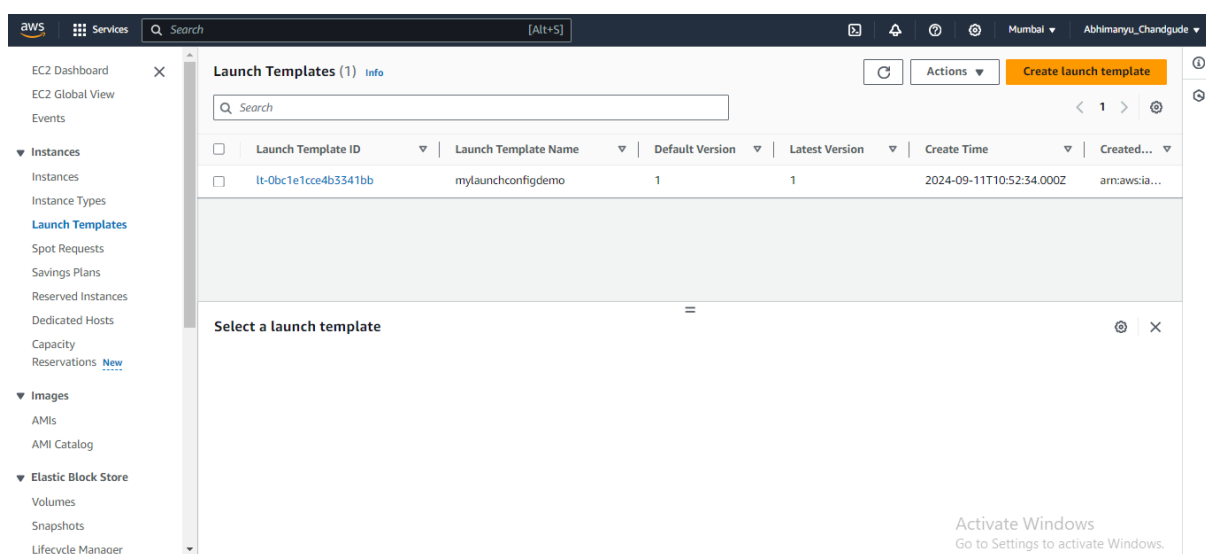


10th September 2024 – Autoscaling

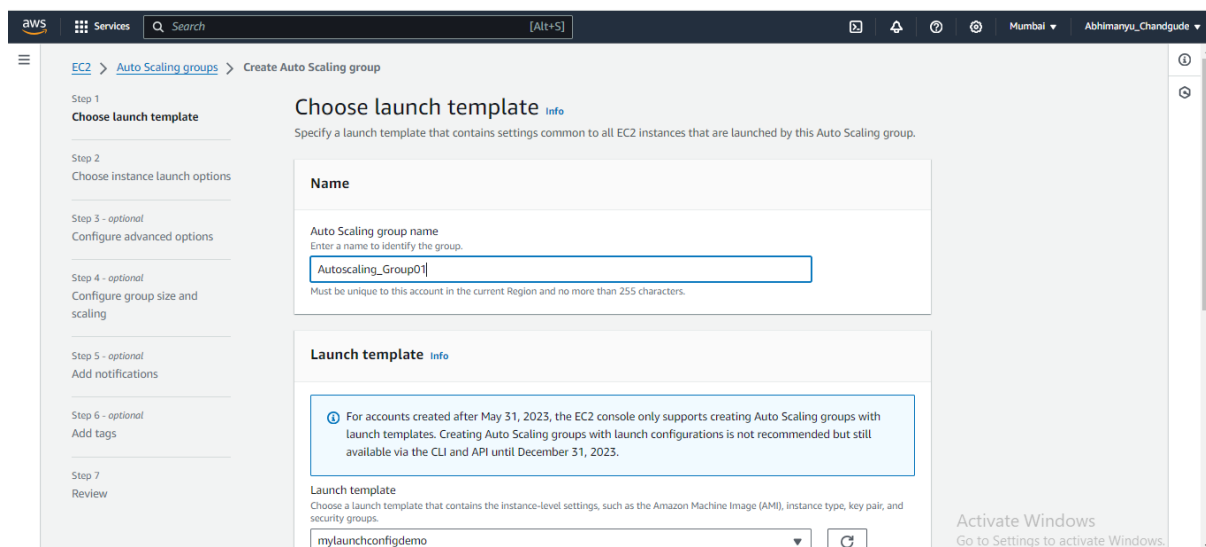
➤ Create template launch configuration

- Select AMI as Amazon Linux
- Select instance type is “t2.micro”
- Select key-pair
- Select security group
- Select role with s3FullAccess
- Advance setting >> Add bootstrap script
- Create launch template



➤ Create Autoscaling Group

- Provide group name and select template which is created



b. Select Availability zones which you want to create instance

The screenshot shows the 'Network info' section of the AWS Management Console. It includes a sidebar with navigation links for Step 5 (optional), Step 6 (optional), and Step 7. The main content area is titled 'Network info' and contains the following sections:

- For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.**
- VPC:** Choose the VPC that defines the virtual network for your Auto Scaling group. A dropdown menu shows 'vpc-03ca67c5e94452d3b' with a refresh button.
- Availability Zones and subnets:** Define which Availability Zones and subnets your Auto Scaling group can use in the chosen VPC. A dropdown menu shows 'Select Availability Zones and subnets' with a refresh button.
- Selected subnets:** Two subnets are listed: 'ap-south-1a | subnet-00c35046390d8ad10' and 'ap-south-1b | subnet-092b7c23bcebd64ec'. Each has a refresh button.

At the bottom, there are buttons for 'Cancel', 'Skip to review', 'Previous', and 'Next'. An 'Activate Windows' watermark is visible in the bottom right corner.

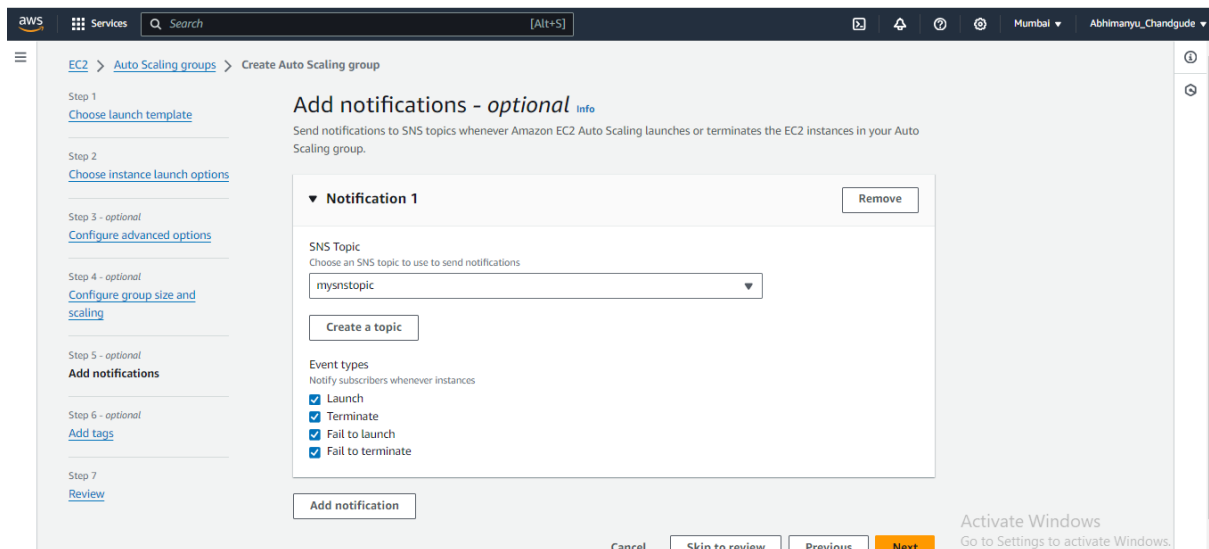
c. Configure group size and scaling

The screenshot shows the 'Configure group size and scaling' section of the AWS Management Console. It includes a sidebar with navigation links for Step 1, Step 2, Step 3 (optional), Step 4 (optional), Step 5 (optional), Step 6 (optional), and Step 7. The main content area is titled 'Configure group size and scaling - optional' and contains the following sections:

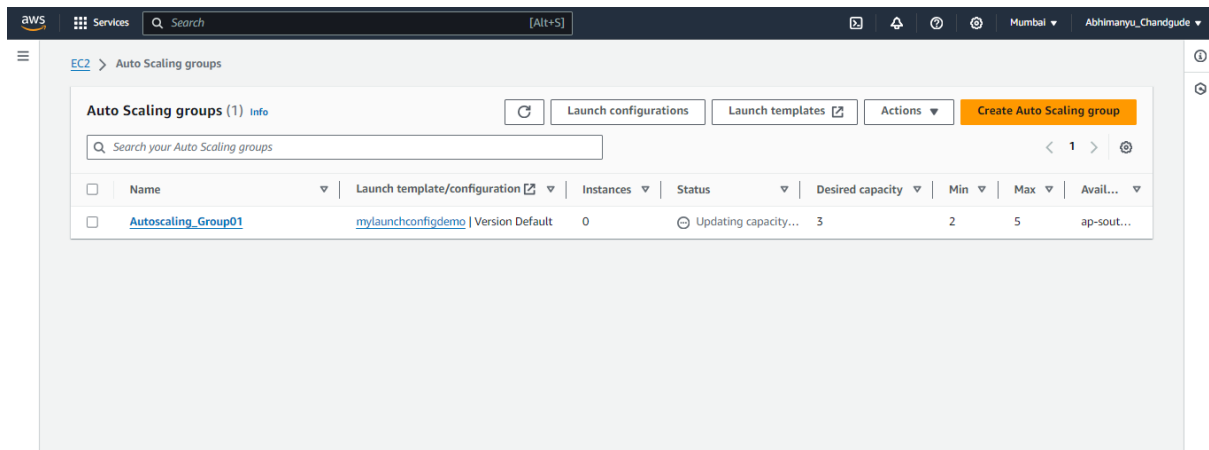
- Define your group's desired capacity and scaling limits. You can optionally add automatic scaling to adjust the size of your group.**
- Group size:** Set the initial size of the Auto Scaling group. After creating the group, you can change its size to meet demand, either manually or by using automatic scaling.
- Desired capacity type:** Choose the unit of measurement for the desired capacity value. vCPUs and Memory (GiB) are only supported for mixed instances groups configured with a set of instance attributes. A dropdown menu shows 'Units (number of instances)'.
- Desired capacity:** Specify your group size. A text input field contains the value '3'.
- Scaling:** You can resize your Auto Scaling group manually or automatically to meet changes in demand.
- Scaling limits:** Set limits on how much your desired capacity can be increased or decreased.
- Min desired capacity:** A text input field contains the value '1'.
- Max desired capacity:** A text input field contains the value '5'.

At the bottom, there are buttons for 'Equal or less than desired' and 'Equal or greater than desired'. An 'Activate Windows' watermark is visible in the bottom right corner.

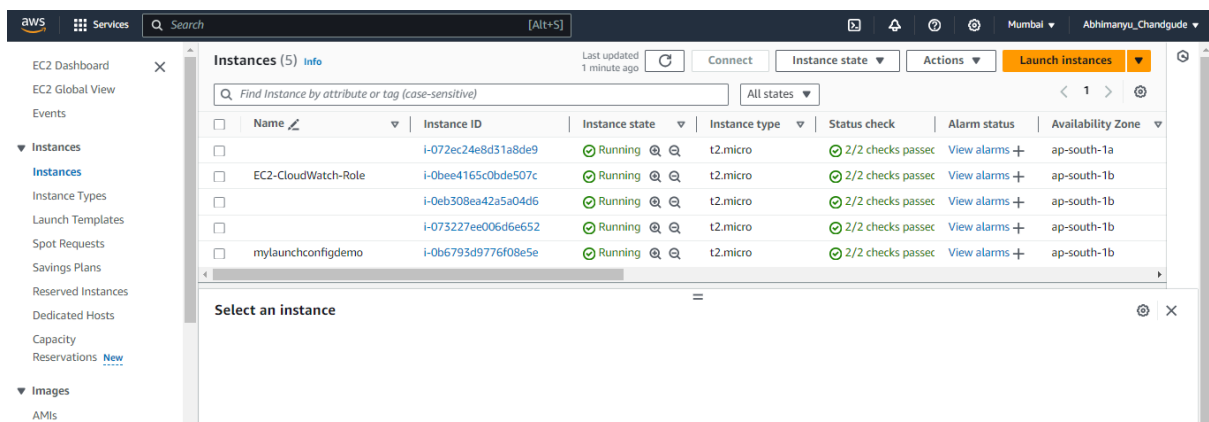
d. Add Notification



e. Autoscaling group have been created

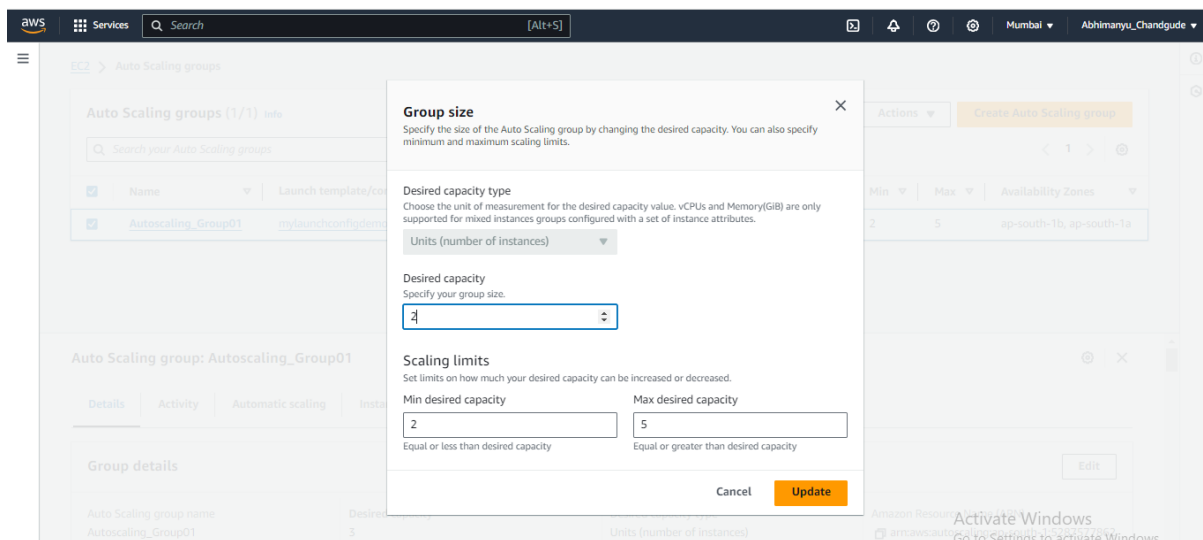


f. Autoscaling group have been created 3 instances as we mentioned in desire "3" count

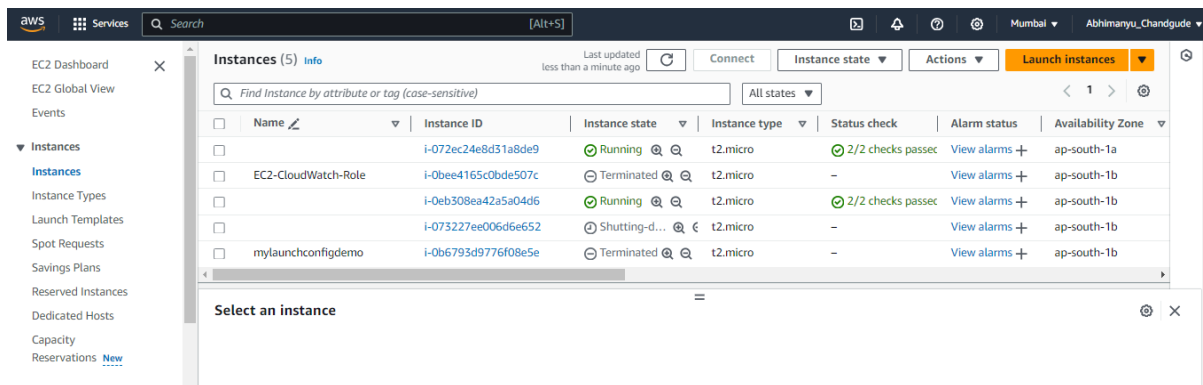


➤ Manual Scaling

If I change desired value manually as “2” so Autoscaling group should remove 1 instance



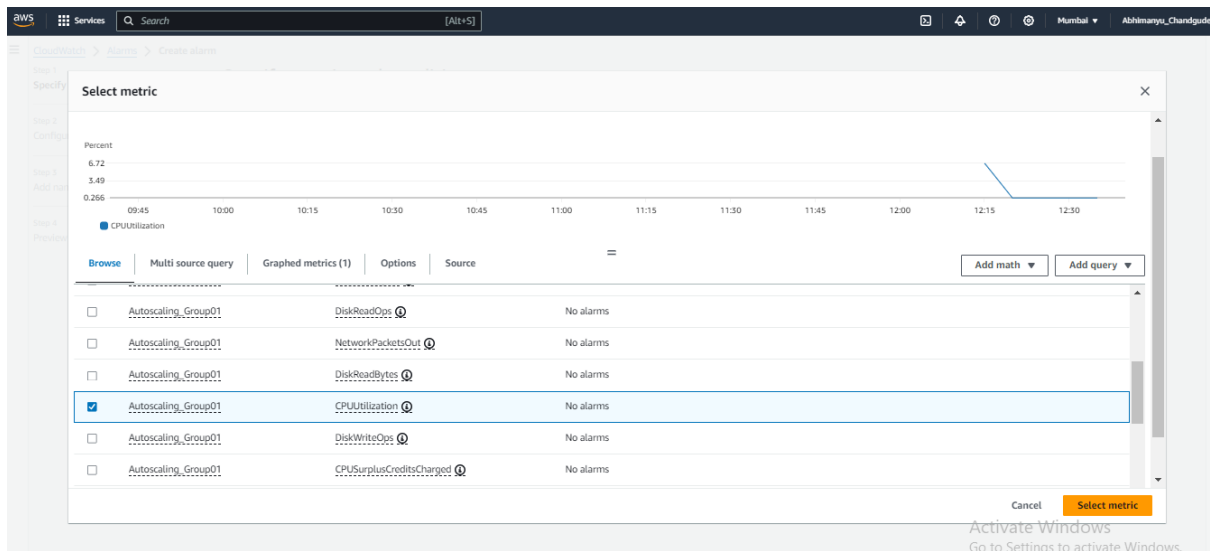
We can see one instance is going to terminate



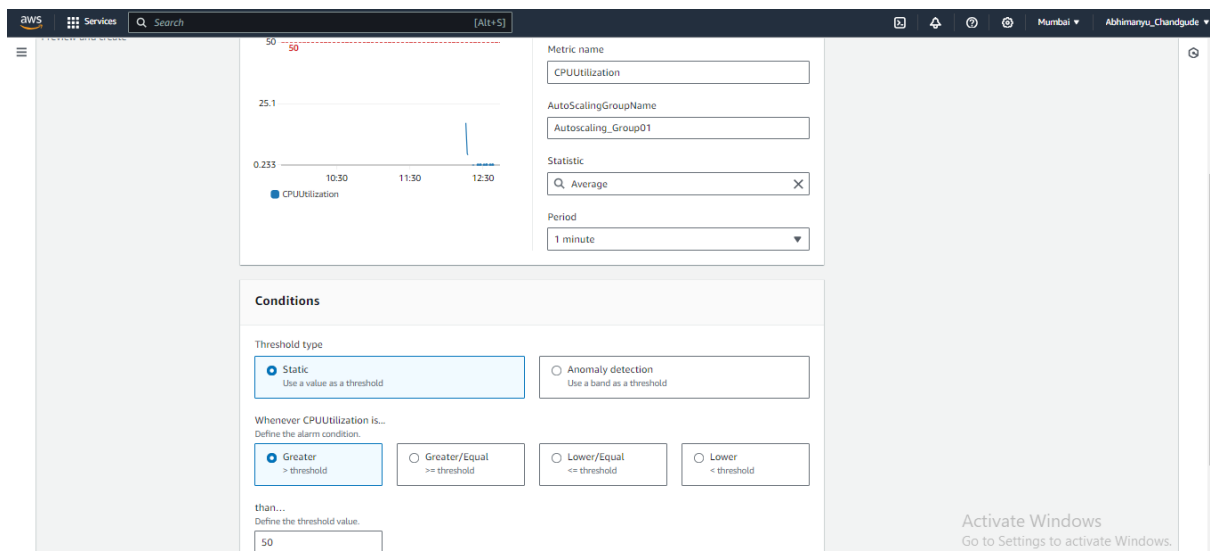
➤ Automatic Scaling

To create Automatic scaling, we required alarm to set Create dynamic scaling policies

1. Create alarm for Autoscaling group
 - a. Select metrics (CPU Utilization)



b. Select timeperiod “1 min” and threshold value (utilization>50)



c. Select in which state you want to trigger “In alarm” and add SNS topic

aws Services Search [Alt+S] Mumbai Abhimanyu_Chandgude

Step 1 Specify metric and conditions

Step 2 **Configure actions**

Step 3 Add name and description

Step 4 Preview and create

Configure actions

Notification

Alarm state trigger
Define the alarm state that will trigger this action.

☒ In alarm
The metric or expression is outside of the defined threshold.

☐ OK
The metric or expression is within the defined threshold.

☐ Insufficient data
The alarm has just started or not enough data is available.

Remove

Send a notification to the following SNS topic
Define the SNS (Simple Notification Service) topic that will receive the notification.

☒ Select an existing SNS topic

☐ Create new topic

mynstopic
mynstopic

Q mynstopic X

Only topics belonging to this account are listed here. All persons and applications subscribed to the selected topic will receive notifications.

Email (endpoints)
akc0107@gmail.com - View in SNS Console

Activate Windows
Go to Settings to activate Windows

d. Add name for Alarm

aws Services Search [Alt+S] Mumbai Abhimanyu_Chandgude

CloudWatch > Alarms > Create alarm

Step 1 Specify metric and conditions

Step 2 Configure actions

Step 3 **Add name and description**

Step 4 Preview and create

Add name and description

Name and description

Alarm name
increase_instance_alarm

Alarm description - optional View formatting guidelines

Edit Preview

Up to 1024 characters (0/1024)

Markdown formatting is only applied when viewing your alarm in the console. The description will remain in plain text in the alarm notifications.

Activate Windows
Go to Settings to activate Windows

e. Alarm have been created

aws Services Search [Alt+S] Mumbai Abhimanyu_Chandgude

CloudWatch

Successfully created alarm increase_instance_alarm. View alarm

CloudWatch > Alarms

Alarms (2) Hide Auto Scaling alarms Clear selection Create composite alarm Actions Create alarm

Search Alarm state: Any Alarm type: Any Actions status: Any

	Name	State	Last state update (UTC)	Conditions
<input type="checkbox"/>	increase_instance_alarm	Insufficient data	2024-09-11 12:51:56	CPUUtilization > 50 for 1 datapoints within 1 minute
<input type="checkbox"/>	CPU_utilization_j-0e7a239fbdde5cda9	Insufficient data	2024-09-11 06:59:57	CPUUtilization >= 50 for 1 datapoints within 2 minutes

- Create Dynamic Policy with alarm in take action added count as “1”. It means if utilization is goes above threshold i.e. 50% it should add one more instance.

The screenshot shows the 'Create dynamic scaling policy' form in the AWS Management Console. The form is for the 'Autoscaling_Group01' and includes the following fields:

- Policy type:** Simple scaling
- Scaling policy name:** increased_instance_policy
- CloudWatch alarm:** increased_instance_alarm
- Take the action:** Add 1 capacity units
- And then wait:** 300 seconds before allowing another scaling activity

Created Dynamic policy as “Increase_instance_alarm”

The screenshot shows the 'Auto Scaling groups' page in the AWS Management Console. A green notification bar at the top says "Dynamic scaling policy created or edited successfully." Below the table, the details for 'Autoscaling_Group01' are shown, including the policy 'increased_instance_policy'.

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Av...
Autoscaling_Group01	mylaunchconfigdemo Version Default	2	-	2	2	5	ap-so...

Auto Scaling group: Autoscaling_Group01

- increased_instance_policy**
- Policy type:** Simple scaling
- Enabled or disabled:** Enabled
- Execute policy when:** No alarm selected
- Take the action:** Add 1 capacity units

For one more dynamic policy we need to create one more alarm as “Decrease_instance_alarm”

1. Now condition is {if CPU utilization is below 20% it should trigger the alarm}

Step 4
Preview and create

Percent

20

10.1

0.253

10:30 11:30 12:30

CPUUtilization

Namespace
AWS/EC2

Metric name
CPUUtilization

AutoScalingGroupName
Autoscaling_Group01

Statistic
Average

Period
1 minute

Conditions

Threshold type

☒ Static
Use a value as a threshold

☐ Anomaly detection
Use a band as a threshold

Whenever CPUUtilization is...

Define the alarm condition.

☐ Greater
> threshold

☐ Greater/Equal
>= threshold

☐ Lower/Equal
<= threshold

☒ Lower
< threshold

than...

Define the threshold value.

20

Activate Windows
Go to Settings to activate Windows.

2. Alarm have been created

CloudWatch

Successfully created alarm decrease_instance_alarm.

View alarm

Alarms (2)

Hide Auto Scaling alarms

Clear selection

Create composite alarm

Actions

Create alarm

<input type="checkbox"/>	Name	State	Last state update (UTC)	Conditions	Actions
<input type="checkbox"/>	decrease_instance_alarm	Insufficient data	2024-09-11 13:07:12	CPUUtilization < 20 for 1 datapoints within 1 minute	Actions enabled
<input type="checkbox"/>	increase_instance_alarm	Insufficient data	2024-09-11 13:05:37	CPUUtilization > 50 for 1 datapoints within 1 minute	Actions enabled

- Create policy as “decrease_instance_policy” in take action select “Remove” and added count as “1”

It means if utilization is below its threshold i.e. 20% it should remove one instance.

EC2 > Auto Scaling groups > Autoscaling_Group01

Create dynamic scaling policy

Policy type
Simple scaling

Scaling policy name
decrease_instance_alarm

CloudWatch alarm
Choose an alarm that can scale capacity whenever:
decrease_instance_alarm

Create a CloudWatch alarm

breaches the alarm threshold: CPUUtilization < 20 for 1 consecutive periods of 60 seconds for the metric dimensions:

AutoScalingGroupName = Autoscaling_Group01

Take the action
Add 1 capacity units

And then wait
300 seconds before allowing another scaling activity

Cancel Create

Activate Windows
Go to Settings to activate Windows.

We have added two policies successfully and both are enabled.

Dynamic scaling policy created or edited successfully.

EC2 > Auto Scaling groups

Auto Scaling groups (1/1) Info

Search your Auto Scaling groups

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones
Autoscaling_Group01	mylaunchconfigdemo Version Default	2	-	2	2	5	ap-south-1b, ap-south-1a

Auto Scaling group: Autoscaling_Group01

Decrease_instance_alarm

Policy type
Simple scaling

Enabled or disabled
Enabled

Execute policy when
No alarm selected

increased_instance_policy

Policy type
Simple scaling

Enabled or disabled
Enabled

Execute policy when
increase_instance_alarm
breaches the alarm threshold: CPUUtilization > 50 for 1 consecutive periods of 60 seconds for the instance

➤ Output

We could see one alarm has been triggered “decrease_instance_alarm”. As checked average CPU utilization of instance is below 20% only.

CloudWatch > Alarms

Alarms (2)

Search

Alarm state: Any Alarm type: Any Actions status: Any

Name	State	Last state update (UTC)	Conditions	Actions
increase_instance_alarm	OK	2024-09-11 13:12:37	CPUUtilization > 50 for 1 datapoints within 1 minute	Actions enabled
decrease_instance_alarm	In alarm	2024-09-11 13:12:12	CPUUtilization < 20 for 1 datapoints within 1 minute	Actions enabled