**Auto Scaling**

------------------------

Auto Scaling helps you ensure that you have the correct number of EC2 instances available to handle the load for your application.

You create collections of EC2 instances, called Auto Scaling groups.

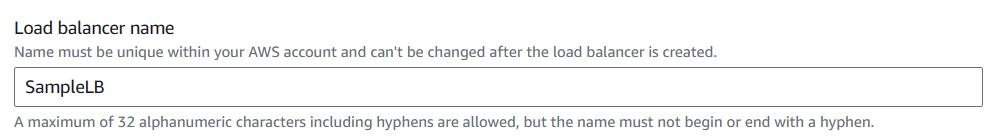
You can specify the minimum number of instances and the maximum number of instances in Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes above this size.

**Steps:**

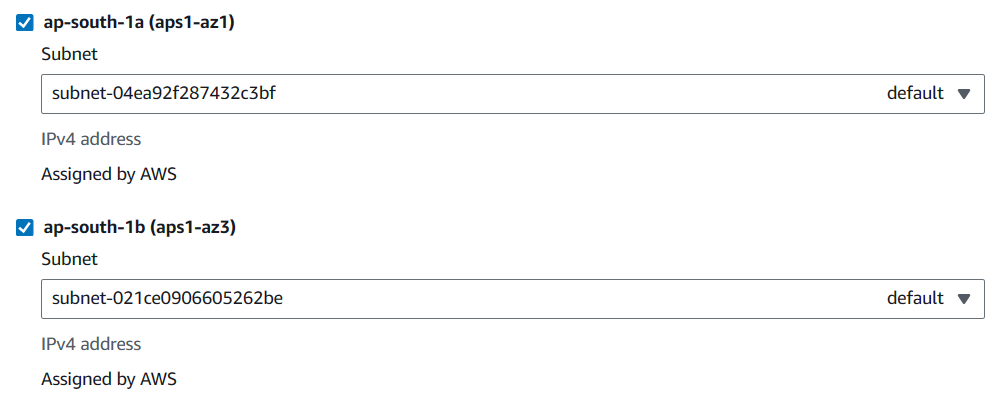
1. Create a load balancer.
2. Create a launch configuration.
3. Create a topic in SNS (Simple Notification Service)
4. Create an Auto Scaling group.
5. Create an alarm in CloudWatch.
6. Add a policy to the Auto Scaling group.

**Step 1: Creating a Load Balancer**

1. Go to the EC2 dashboard and select Load Balancers.
2. Click Create Load Balancer. (Classic LB)
3. Enter a name for your load balancer.

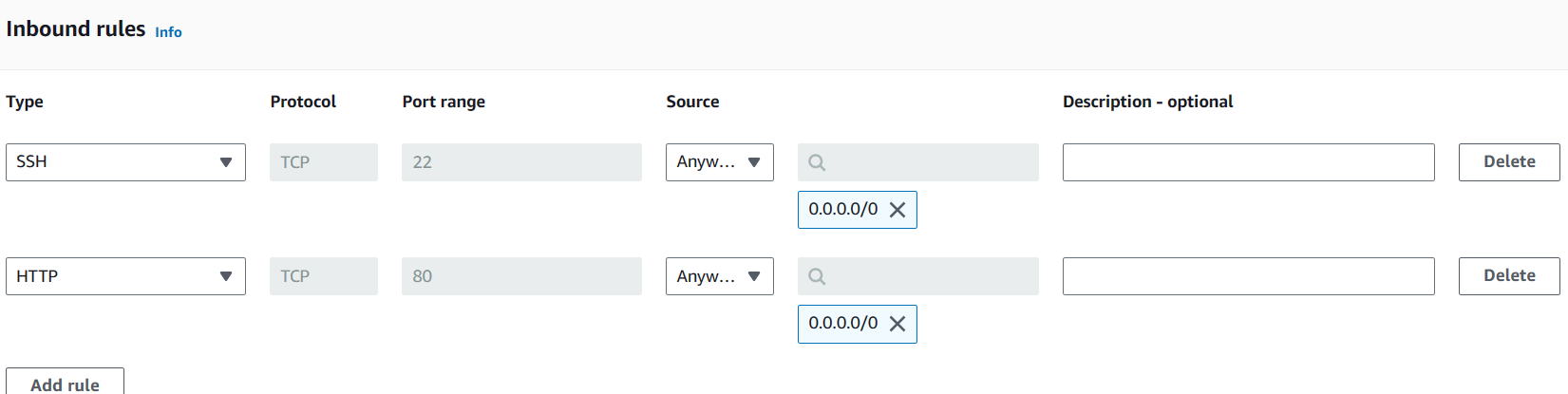


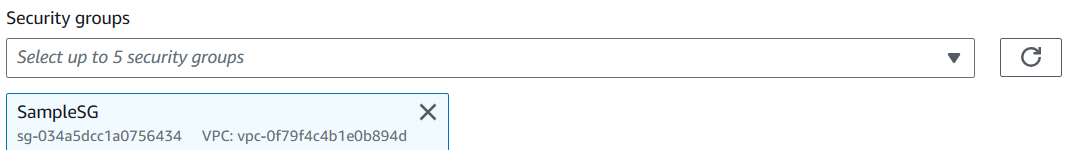
1. Select the location for your load balancer.



1. Create a new security group for your load balancer.

Open two ports SSH and HTTP





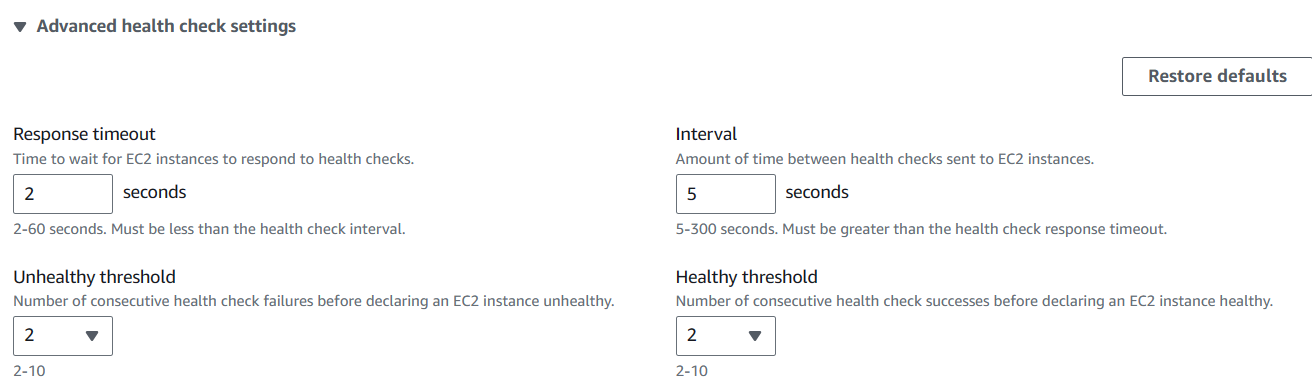
1. Configure a health check for your load balancer.

Response timeout -2

Interval -5

Unhealthy threshold - 2

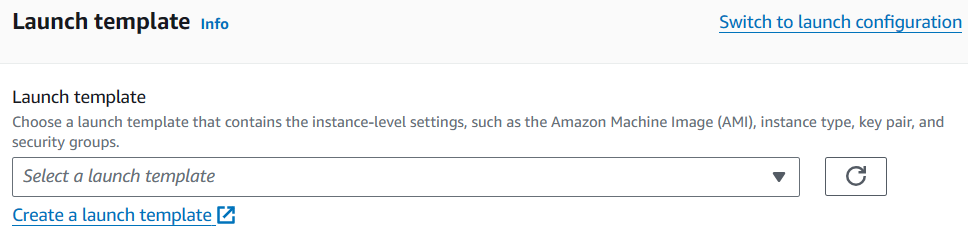
Healthy threshold -2



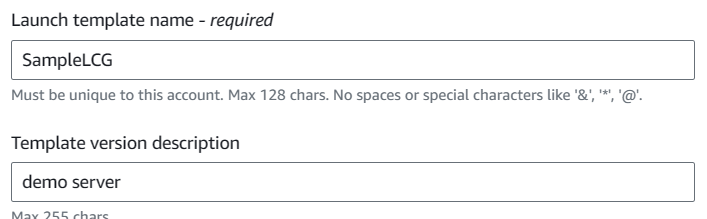
1. Review your settings and click Create.

**Step 2: Creating a Launch Configuration**

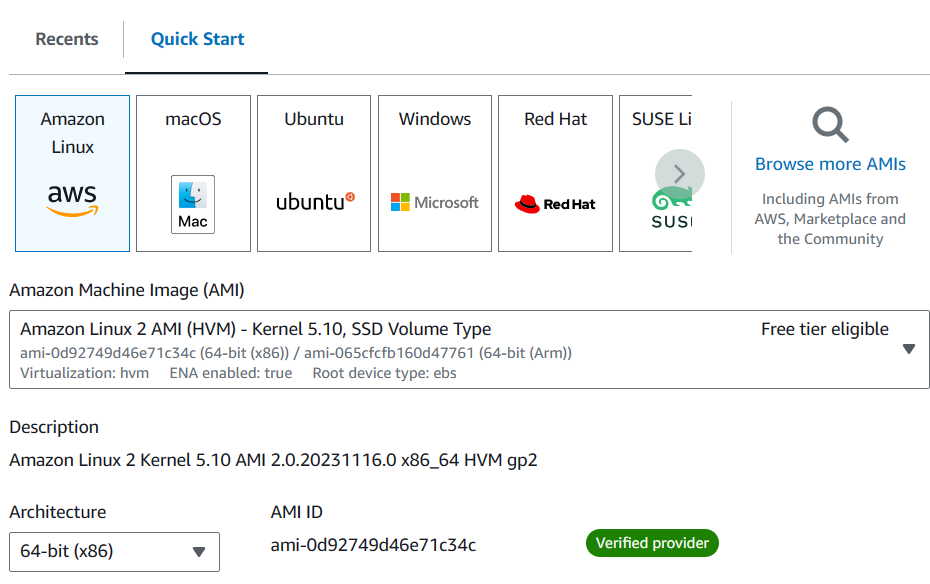
1. Go to the Auto Scaling dashboard and select Launch Configurations.



1. Click Create Launch Configuration.
2. Enter a name for your launch configuration, such as SampleLCG.

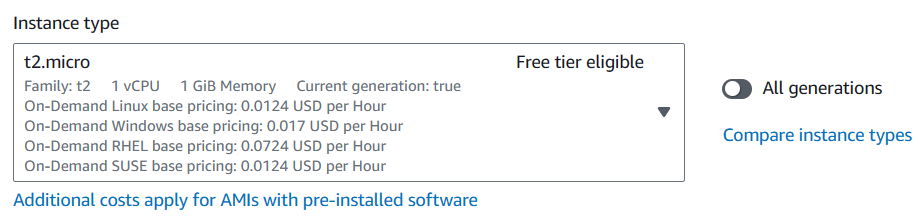


1. Select an AMI for your launch configuration.



1. Select an instance type for your launch configuration and choose keypair.

t2.micro



1. Select existing security group for your launch configuration.
2. Configure user data for your launch configuration.

#!/bin/bash

sudo su

yum update -y

yum install httpd -y

cd /var/www/html

echo "Welcome to AWS Learning" > index.html

service httpd start

chkconfig httpd on

1. Click Create Launch Configuration.

**Step 3: Creating a Topic in SNS**

1. Go to the SNS dashboard and select Topics.

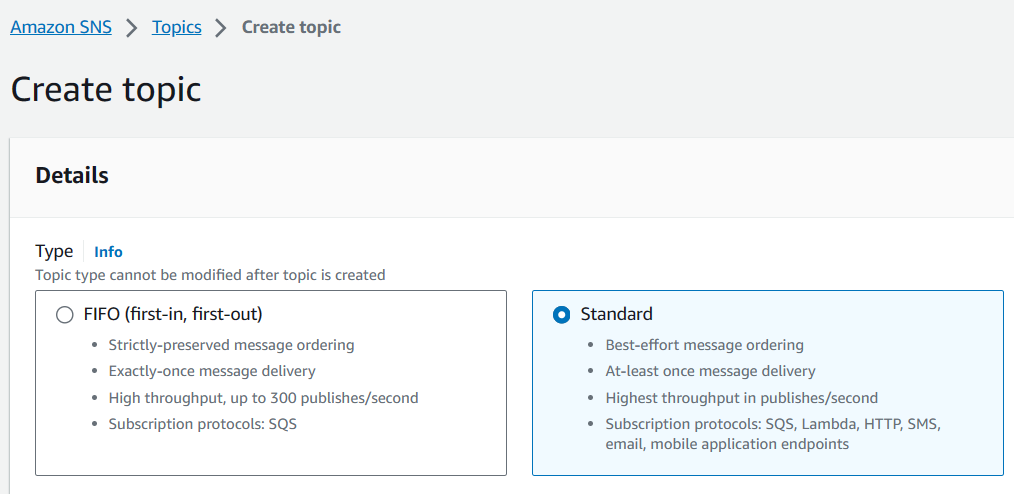
Services -- Application Integration -- Simple Notification Service

1. Click Create Topic.
2. Enter a name for your topic, such as MyTopic1.

Type - Standard

Name - MyTopic1

Display Name - MyTopic1



1. Click Create Topic.

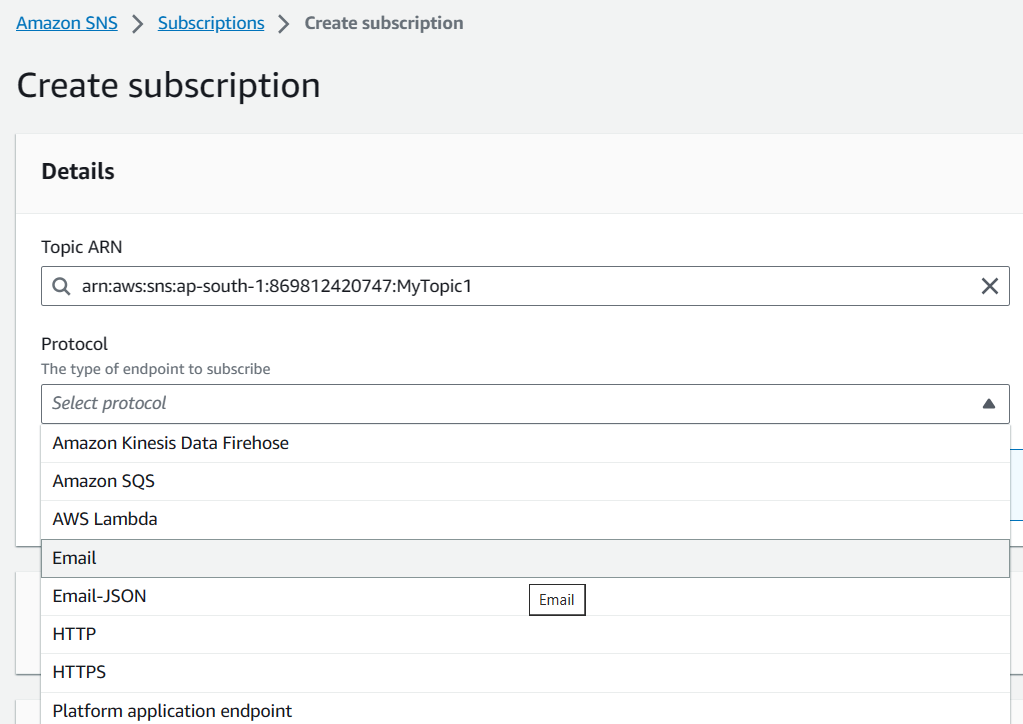
Add subscriptions to the topic

Create Subscription

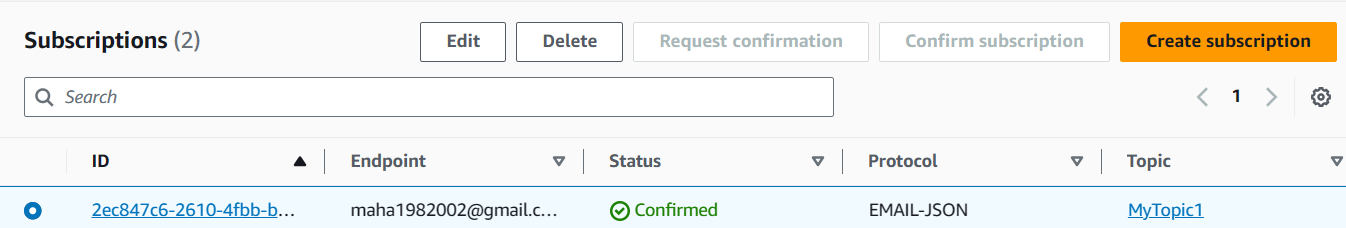
Protocol - Email

Endpoint –xyz@gmail.com

Create Subscription.

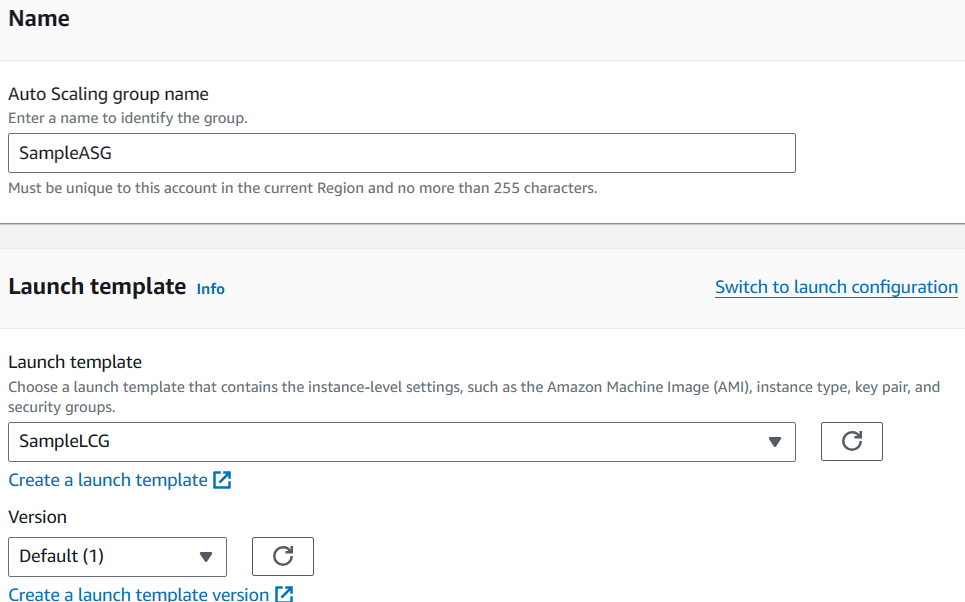


1. Confirm the subscription through the link received in email

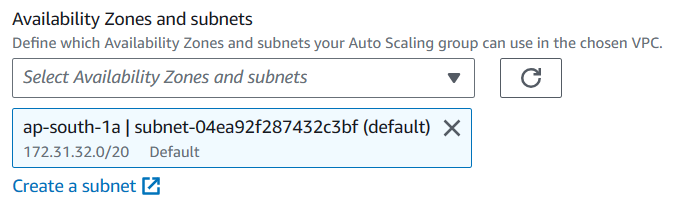


**Step 4: Creating an Auto Scaling Group**

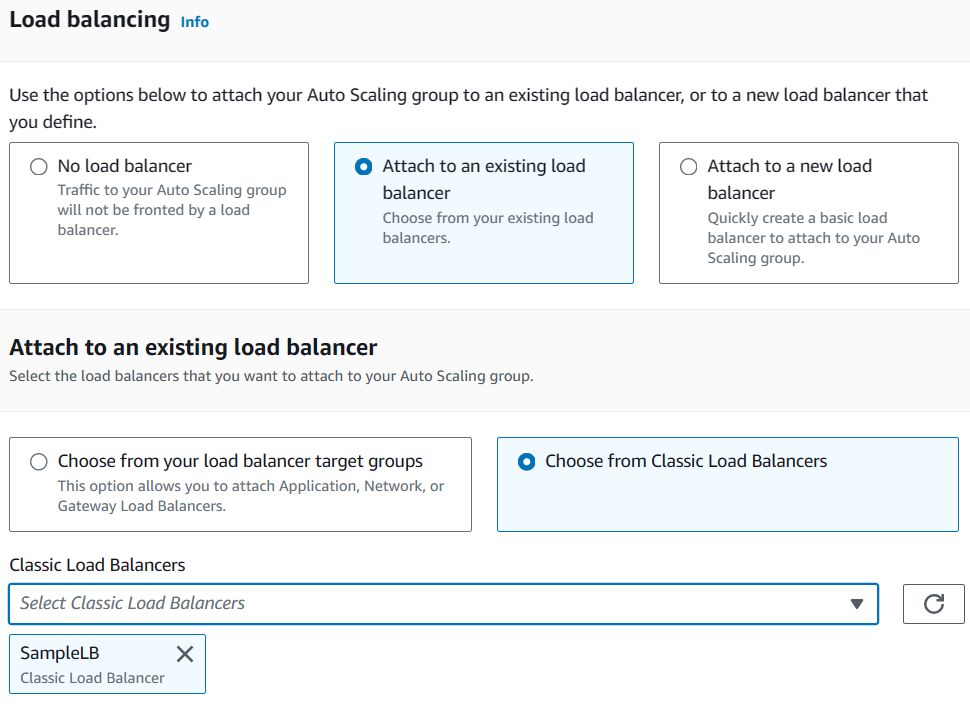
1. Go to the Auto Scaling dashboard and select Auto Scaling Groups.
2. Click Create Auto Scaling Group.
3. Enter a name for your Auto Scaling group, such as SampleASG.
4. Select a launch configuration for your Auto Scaling group.



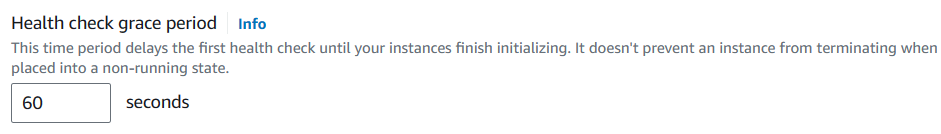
1. Select a subnet for your Auto Scaling group.



1. Attach a load balancer to your Auto Scaling group.

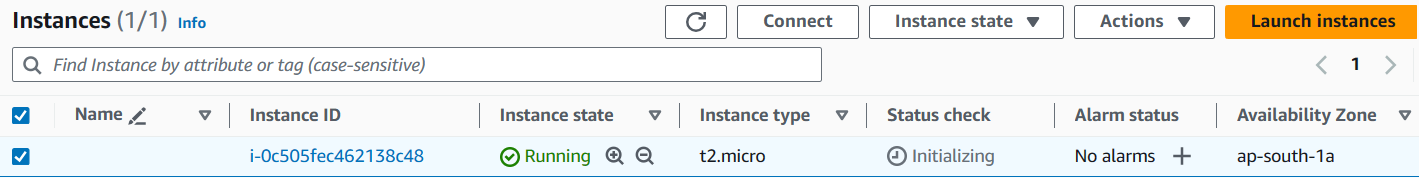


1. Configure the health check grace period for your Auto Scaling group.



1. Review your settings and click Create Auto Scaling Group.

As the desired capacity is 1, By this time one EC2 Machine could have been created.

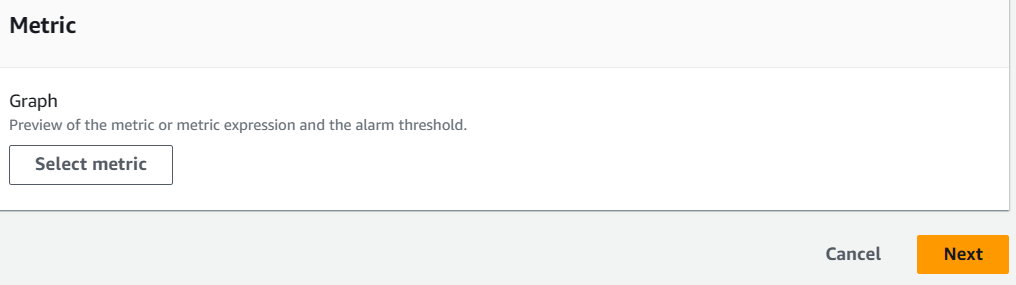


**Step 5: Creating an Alarm in CloudWatch**

1. Go to the CloudWatch dashboard and select Alarms.

Services ---- Management & Governance --- CloudWatch > Alarms > Inalarm

1. Click Create Alarm.



1. Select a metric for your alarm, such as CPUUtilization.

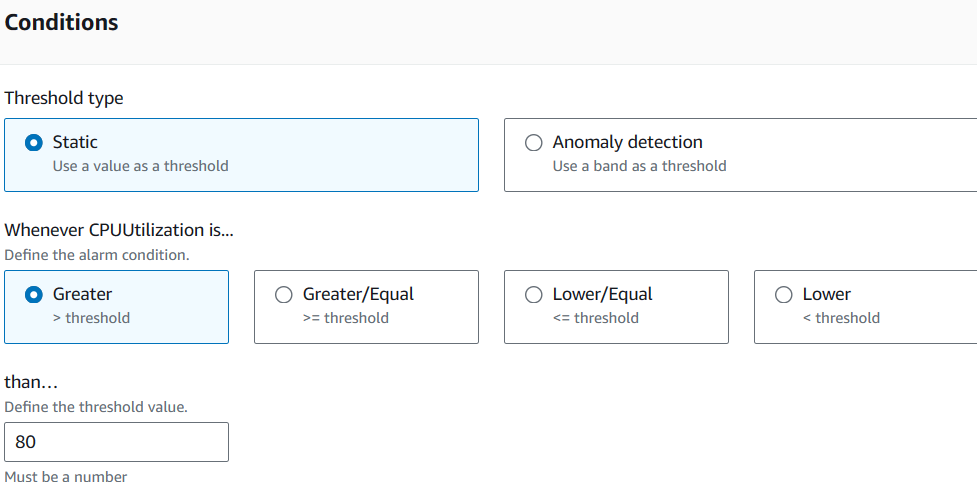
The metric being monitored, such as CPU utilization, memory usage, or network traffic.

Select Metric ---> EC2 --- By Auto Scaling Group ---- Select Auto Scaling Group Name, Metric Name – CPUUtilization

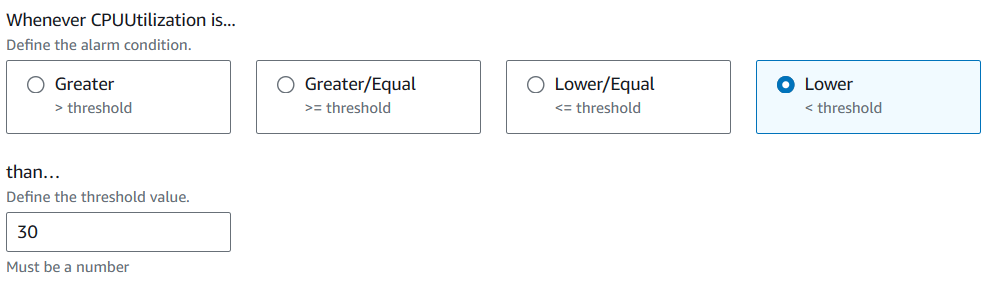
**Evaluation Periods:** The number of consecutive data points that need to meet the threshold condition before the alarm is triggered.

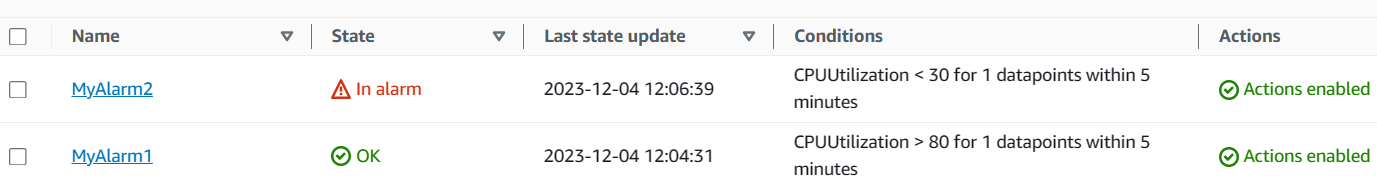


1. Set a threshold for your alarm, such as 80. **Threshold is t**he value that triggers the alarm.

****

1. Similarly Create another Alarm for CPU Utilization <30





**Alarm State:** The state of the alarm, which can be OK, ALARM, or INSUFFICIENT DATA.

**OK:** The metric being monitored is within the specified threshold and the alarm is not triggered.

**ALARM:** The metric being monitored has exceeded the specified threshold and the alarm is triggered.

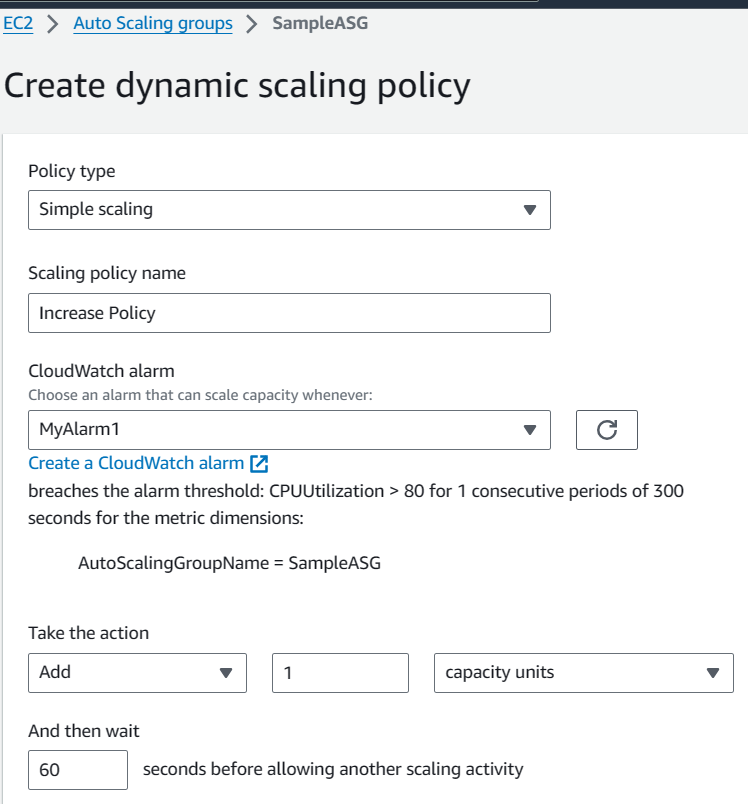
**INSUFFICIENT DATA:** There is not enough data available to determine the state of the alarm. This can occur when the metric is newly created or when there is a temporary interruption in data collection.

**Step 6:** **Add Policy in Auto Scaling**

This step involves creating two scaling policies, one to increase the number of EC2 instances in the Auto Scaling group when CPU utilization exceeds 80%, and another to decrease the number of instances when CPU utilization falls below 30%.

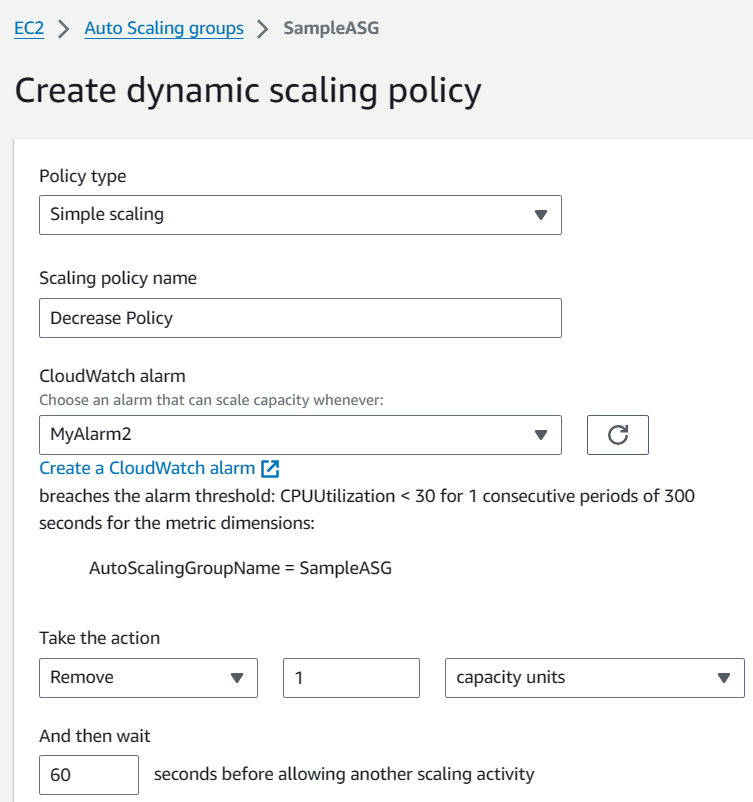
**Step 6.1: Create an Increase Policy**

1. Go to the Auto Scaling groups page and select the Auto Scaling group you created in Step 4.
2. Click the Automatic Scaling tab.
3. Click create dynamic scaling Add Policy.
4. Select Simple Scaling as the policy type.
5. Enter a name for the policy, such as Increase Policy.
6. Select My\_alarm1 as the CloudWatch alarm.
7. Under Take the action, select Add and then enter 1 Unit.
8. Click Create.



**Step 6.2: Create a Decrease Policy**

1. Repeat steps 1-4 from Step 6.1.
2. Enter a name for the policy, such as Decrease Policy.
3. Select My\_alarm2 as the CloudWatch alarm.
4. Under Take the action, select Remove and then enter 1 Unit.
5. Click Create.



**Deleting process**

-----------------

1) Delete Autoscaling

2) Delete launch templates(Instances will be terminated automatically)

3) Delete Load balancer

4) Delete Topic in notification service

5) Delete Alarm in cloud watch